Compendium of Plant Genomes
Series Editor: Chittaranjan Kole

Nikolai V. Ivanov
Nicolas Sierro
Manuel C. Peitsch *Editors*

The Tobacco Plant Genome

Compendium of Plant Genomes

Series Editor

Chittaranjan Kole, Raja Ramanna Fellow, Government of India,
ICAR-National Research Center on Plant Biotechnology, Pusa,
New Delhi, India

Whole-genome sequencing is at the cutting edge of life sciences in the new millennium. Since the first genome sequencing of the model plant Arabidopsis thaliana in 2000, whole genomes of about 100 plant species have been sequenced and genome sequences of several other plants are in the pipeline. Research publications on these genome initiatives are scattered on dedicated web sites and in journals with all too brief descriptions. The individual volumes elucidate the background history of the national and international genome initiatives; public and private partners involved; strategies and genomic resources and tools utilized; enumeration on the sequences and their assembly; repetitive sequences; gene annotation and genome duplication. In addition, synteny with other sequences, comparison of gene families and most importantly potential of the genome sequence information for gene pool characterization and genetic improvement of crop plants are described.

Interested in editing a volume on a crop or model plant? Please contact Prof. C. Kole, Series Editor, at ckoleorg@gmail.com

More information about this series at http://www.springer.com/series/11805

Nikolai V. Ivanov · Nicolas Sierro ·
Manuel C. Peitsch
Editors

The Tobacco Plant Genome

Editors
Nikolai V. Ivanov
PMI R&D
Philip Morris Products S.A.
Neuchâtel, Switzerland

Nicolas Sierro
PMI R&D
Philip Morris Products S.A.
Neuchâtel, Switzerland

Manuel C. Peitsch
PMI R&D
Philip Morris Products S.A.
Neuchâtel, Switzerland

ISSN 2199-4781 ISSN 2199-479X (electronic)
Compendium of Plant Genomes
ISBN 978-3-030-29492-2 ISBN 978-3-030-29493-9 (eBook)
https://doi.org/10.1007/978-3-030-29493-9

© Springer Nature Switzerland AG 2020
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

This book series is dedicated to my wife Phullara
and our children Sourav and Devleena

Chittaranjan Kole

Preface to the Series

Genome sequencing has emerged as the leading discipline in the plant sciences coinciding with the start of the new century. For much of the twentieth century, plant geneticists were only successful in delineating putative chromosomal location, function, and changes in genes indirectly through the use of a number of "markers" physically linked to them. These included visible or morphological, cytological, protein, and molecular or DNA markers. Among them, the first DNA marker, the RFLPs, introduced a revolutionary change in plant genetics and breeding in the mid-1980s, mainly because of their infinite number and thus potential to cover maximum chromosomal regions, phenotypic neutrality, absence of epistasis, and codominant nature. An array of other hybridization-based markers, PCR-based markers, and markers based on both facilitated construction of genetic linkage maps, mapping of genes controlling simply inherited traits, and even gene clusters (QTLs) controlling polygenic traits in a large number of model and crop plants. During this period, a number of new mapping populations beyond F_2 were utilized and a number of computer programs were developed for map construction, mapping of genes, and for mapping of polygenic clusters or QTLs. Molecular markers were also used in the studies of evolution and phylogenetic relationship, genetic diversity, DNA fingerprinting, and map-based cloning. Markers tightly linked to the genes were used in crop improvement employing the so-called marker-assisted selection. These strategies of molecular genetic mapping and molecular breeding made a spectacular impact during the last one and a half decades of the twentieth century. But still they remained "indirect" approaches for elucidation and utilization of plant genomes since much of the chromosomes remained unknown and the complete chemical depiction of them was yet to be unraveled.

Physical mapping of genomes was the obvious consequence that facilitated the development of the "genomic resources" including BAC and YAC libraries to develop physical maps in some plant genomes. Subsequently, integrated genetic–physical maps were also developed in many plants. This led to the concept of structural genomics. Later on, emphasis was laid on EST and transcriptome analysis to decipher the function of the active gene sequences leading to another concept defined as functional genomics. The advent of techniques of bacteriophage gene and DNA sequencing in the 1970s was extended to facilitate sequencing of these genomic resources in the last decade of the twentieth century.

As expected, sequencing of chromosomal regions would have led to too much data to store, characterize, and utilize with the-then available computer software could handle. But the development of information technology made the life of biologists easier by leading to a swift and sweet marriage of biology and informatics, and a new subject was born—bioinformatics.

Thus, the evolution of the concepts, strategies, and tools of sequencing and bioinformatics reinforced the subject of genomics—structural and functional. Today, genome sequencing has traveled much beyond biology and involves biophysics, biochemistry, and bioinformatics!

Thanks to the efforts of both public and private agencies, genome sequencing strategies are evolving very fast, leading to cheaper, quicker, and automated techniques right from clone-by-clone and whole-genome shotgun approaches to a succession of second-generation sequencing methods. The development of software of different generations facilitated this genome sequencing. At the same time, newer concepts and strategies were emerging to handle sequencing of the complex genomes, particularly the polyploids.

It became a reality to chemically—and so directly—define plant genomes, popularly called whole-genome sequencing or simply genome sequencing.

The history of plant genome sequencing will always cite the sequencing of the genome of the model plant Arabidopsis thaliana in 2000 that was followed by sequencing the genome of the crop and model plant rice in 2002. Since then, the number of sequenced genomes of higher plants has been increasing exponentially, mainly due to the development of cheaper and quicker genomic techniques and, most importantly, the development of collaborative platforms such as national and international consortia involving partners from public and/or private agencies.

As I write this preface for the first volume of the new series "Compendium of Plant Genomes," a net search tells me that complete or nearly complete whole-genome sequencing of 45 crop plants, eight crop and model plants, eight model plants, 15 crop progenitors and relatives, and 3 basal plants is accomplished, the majority of which are in the public domain. This means that we nowadays know many of our model and crop plants chemically, i.e., directly, and we may depict them and utilize them precisely better than ever. Genome sequencing has covered all groups of crop plants. Hence, information on the precise depiction of plant genomes and the scope of their utilization are growing rapidly every day. However, the information is scattered in research articles and review papers in journals and dedicated Web pages of the consortia and databases. There is no compilation of plant genomes and the opportunity of using the information in sequence-assisted breeding or further genomic studies. This is the underlying rationale for starting this book series, with each volume dedicated to a particular plant.

Plant genome science has emerged as an important subject in academia, and the present compendium of plant genomes will be highly useful both to students and teaching faculties. Most importantly, research scientists involved in genomics research will have access to systematic deliberations on the plant genomes of their interest. Elucidation of plant genomes is of interest not only for the geneticists and breeders, but also for practitioners of an array of plant science disciplines, such as taxonomy, evolution, cytology, physiology,

pathology, entomology, nematology, crop production, biochemistry, and obviously bioinformatics. It must be mentioned that information regarding each plant genome is ever-growing. The contents of the volumes of this compendium are, therefore, focusing on the basic aspects of the genomes and their utility. They include information on the academic and/or economic importance of the plants, description of their genomes from a molecular genetic and cytogenetic point of view, and the genomic resources developed. Detailed deliberations focus on the background history of the national and international genome initiatives, public and private partners involved, strategies and genomic resources and tools utilized, enumeration on the sequences and their assembly, repetitive sequences, gene annotation, and genome duplication. In addition, synteny with other sequences, comparison of gene families, and, most importantly, the potential of the genome sequence information for gene pool characterization through genotyping by sequencing (GBS) and genetic improvement of crop plants have been described. As expected, there is a lot of variation of these topics in the volumes based on the information available on the crop, model, or reference plants.

I must confess that as the series editor, it has been a daunting task for me to work on such a huge and broad knowledge base that spans so many diverse plant species. However, pioneering scientists with lifetime experience and expertise on the particular crops did excellent jobs editing the respective volumes. I myself have been a small science worker on plant genomes since the mid-1980s and that provided me the opportunity to personally know several stalwarts of plant genomics from all over the globe. Most, if not all, of the volume editors are my longtime friends and colleagues. It has been highly comfortable and enriching for me to work with them on this book series. To be honest, while working on this series I have been and will remain a student first, a science worker second, and a series editor last. And I must express my gratitude to the volume editors and the chapter authors for providing me the opportunity to work with them on this compendium.

I also wish to mention here my thanks and gratitude to the Springer staff, Dr. Christina Eckey and Dr. Jutta Lindenborn in particular, for all their constant and cordial support right from the inception of the idea.

I always had to set aside additional hours to edit books beside my professional and personal commitments—hours I could and should have given to my wife, Phullara, and our kids, Sourav, and Devleena. I must mention that they not only allowed me the freedom to take away those hours from them but also offered their support in the editing job itself. I am really not sure whether my dedication of this compendium to them will suffice to do justice to their sacrifices for the interest of science and the science community.

Kalyani, India Chittaranjan Kole

Preface

Tobacco (*Nicotiana tabacum*) is a crop from the Nicotiana genus that is cultivated worldwide, with the major producers located in China, Brazil, India, and the USA. Three main types of tobacco are cultivated for industrial use: Virginia, burley, and oriental. Virginia tobacco is typically flue-cured for a week in heated barns and has a light, bright aroma and taste. Burley tobacco is air-cured in barns for up to 2 months, thereby losing most of its natural sugars and developing a strong taste. The small-leafed aromatic oriental tobacco is sun-cured in the open air. Due to the ease of their genetic transformation and regeneration, *N. tabacum* and *N. benthamiana* have served as model plant systems in functional genomics, research on host–pathogen interactions, and plant-based protein expression in research settings.

Tobacco originates from middle and South America and was introduced in Europe in the early 1500s along with three other American plants: potato, tomato, and maize.

The Nicotiana genus, to which tobacco belongs, consists of about 78 species indigenous to North and South America, Southwest Africa, Australia, and the south Pacific. Among them, both diploid and tetraploid species are found with large genome sizes ranging from about 2 to 5 Gb. *Nicotiana tabacum* is a recent allotetraploid (2n = 4x = 48) species resulting from the hybridization of the diploid (2n = 12) ancestor species *Nicotiana sylvestris* and *Nicotiana tomentosiformis* about 200,000 years ago.

The first effort to obtain a draft tobacco genome was carried out within the Tobacco Genome Initiative, where methyl-filtered DNA was sequenced, and that allowed the development of simple sequence repeat genetic markers that subsequently permitted the construction of the first version of a high-resolution tobacco genetic map in 2007. With the availability of high-throughput sequencers in the early 2010s, whole-genome shotgun sequencing approaches became feasible, and draft genomes for *Nicotiana benthamiana* and modern descendants of the tobacco ancestors were first assembled, rapidly followed by draft genomes for a representative of each of the three main classes of cultivated tobacco. Since then, several additional Nicotiana species genomes have been sequenced, highlighting particularities of their respective species.

This book is intended as a resource for plant and genome biologists, geneticists, and biotechnologists. It starts by providing background and historical information about the tobacco plant and its genome in the context of other Solanaceae species.

The biodiversity, genetics, genome, and transcriptome of tobacco are then presented, including insights into the complexity of the genomes of tobacco and other Nicotiana species with regards to polyploidization, repeat expansion, and interspecies introgression; the Nicotiana resources available from the SOL Genomics Network, including metabolic databases, are described. Several chapters focus on key features of tobacco, such as tobacco leaf curing, nicotine biosynthesis, transport and regulation, or tobacco-specific nitrosamine formation, while others present the impact of genetics on tobacco flower colors or evidence of natural Agrobacterium-mediated transformation in the Nicotiana genus. Two important Nicotiana species beyond tobacco are also highlighted: *Nicotiana attenuata*, which can adapt to a wide range of pests, thanks to its phenotypic plasticity, and *Nicotiana benthamiana*, which is used as a model plant to study plant–pathogen interactions. The last chapter is dedicated to potential future use of tobacco and other Nicotiana species for the production of valuable compounds.

Neuchâtel, Switzerland

Nikolai V. Ivanov
Nicolas Sierro
Manuel C. Peitsch

Contents

1 Background and History of Tobacco Genome Resources 1
Nicolas Sierro and Nikolai V. Ivanov

2 Biodiversity of *Nicotiana* (Solanaceae) 21
Sandra Knapp

3 Genetic Markers in Tobacco, Usage for Map Development, Diversity Studies, and Quantitative Trait Loci Analysis 43
Nicolas Bakaher

4 Characterizing the Genome of *Nicotiana tabacum* 51
James N. D. Battey, Nicolas Sierro and Nikolai V. Ivanov

5 Tobacco Resources in the Sol Genomics Network and *Nicotiana* Metabolic Databases 59
Hartmut Foerster and Lukas A. Mueller

6 Leaf Curing Practices Alter Gene Expression and the Chemical Constituents of Tobacco Leaves 73
Lucien Bovet, Aurore Hilfiker, James N. D. Battey, Nicolas Sierro and Simon Goepfert

7 Repetitive DNA Dynamics and Polyploidization in the Genus *Nicotiana* (Solanaceae). 85
Steven Dodsworth, Ales Kovarik, Marie-Angèle Grandbastien, Ilia J. Leitch and Andrew R. Leitch

8 The Use of *Nicotiana* Species in Tobacco Improvement 101
Apoloniusz Berbeć and Teresa Doroszewska

9 Nicotine Biosynthesis, Transport, and Regulation in Tobacco: Insights into the Evolution of a Metabolic Pathway 147
Tsubasa Shoji

10 Impact of Genetics and Production Practices on Tobacco-Specific Nitrosamine Formation 157
Dongmei Xu, Marcos F. Lusso and James A. Strickland

11 The Genetic Basis of Flower Color Differences in *Nicotiana tabacum* 175
Elizabeth W. McCarthy, Jacob B. Landis, Amelda Kurti, Amber J. Lawhorn and Amy Litt

12 Natural *Agrobacterium*-Mediated Transformation in the Genus *Nicotiana* 195
Léon Otten

13 *Nicotiana attenuata* Genome Reveals Genes in the Molecular Machinery Behind Remarkable Adaptive Phenotypic Plasticity 211
Aura Navarro-Quezada, Klaus Gase, Ravi K. Singh, Shree P. Pandey and Ian T. Baldwin

14 *Nicotiana benthamiana*, A Popular Model for Genome Evolution and Plant–Pathogen Interactions 231
Marina A. Pombo, Hernan G. Rosli, Noe Fernandez-Pozo and Aureliano Bombarely

15 Production of Valuable Compounds in Tobacco 249
Kacper Piotr Kaminski, Simon Goepfert, Nikolai V. Ivanov and Manuel C. Peitsch

Check for updates

Background and History of Tobacco Genome Resources

1

Nicolas Sierro and Nikolai V. Ivanov

Abstract

Nicotiana species, which share significant similarities with tomato, potato, eggplant, and pepper, originated in South America and subsequently spread to North America, Africa, and Australia. They are known to produce and accumulate alkaloids as part of their defense mechanism against insects and herbivores. Since their description and classification by Goodspeed in 1947, research on *Nicotiana* species has been ongoing, including investigations on how the genus evolved and its relationship with other Solanaceae members. The presence of diploid and polyploid species in *Nicotiana* and their large genome sizes resulted in the generation of multiple resources aiming at better understanding of how polyploids formed and evolved over time. The use of *Nicotiana tabacum* and *Nicotiana benthamiana* as model organisms in plant research further helped increase the interest in understanding the molecular mechanisms in these species. Here, we review the literature and summarize what is currently known about the phylogeny, cytology, and genetics of *Nicotiana* species. We also present the current state of the *Nicotiana* genomes and introduce the functional genomics insights they have provided.

N. Sierro (✉) · N. V. Ivanov
PMI R&D, Philip Morris Products S.A., Quai Jeanrenaud 5, 2000 Neuchâtel, Switzerland
e-mail: nicolas.sierro@pmi.com

N. V. Ivanov
e-mail: nikolai.ivanov@pmi.com

1.1 Introduction

Plant species commonly referred to as "tobacco" belong to the genus *Nicotiana* and are members of the Solanaceae family, with at least 3,000 species described to date, the majority of which are assigned to the genus *Solanum* (Knapp et al. 2004a; Olmstead et al. 2008). Some Solanaceae species are important commercial crops or ornamental plants, such as potato (*S. tuberosum*), tomato (*S. lycopersicum*), tobacco (*N. tabacum*), eggplant (*S. melongena*), pepper (*Capsicum annuum*), and petunia (*Petunia* × *hybrida*). The evolutionary and geographical origin of Solanaceae is in South America; however, some species have been found as far away as Africa and Australia (Eich 2008). Owing to the close relationship with Solanaceae, coffee species (e.g., arabica coffee [*Coffea arabica*]), from the Rubiaceae family, are also often considered together with Solanaceae species, as these two families are both members of the Asterid I clade.

A large number of Solanaceae species belong to the so-called $x = 12$ clade (including *Solanum* and *Nicotiana* genera) (Sarkinen et al. 2013) and share the same diploid ($2n = 2x = 24$) or tetraploid ($2n = 4x = 48$) chromosome numbers and similar chromosome architecture. Similarly, coffee comprises species with either diploid

© Springer Nature Switzerland AG 2020

N. V. Ivanov et al. (eds.), *The Tobacco Plant Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-030-29493-9_1

($2n = 2x = 22$ in *C. canephora*) or allotetraploid ($2n = 4x = 44$ in *C. arabica*) architectures. Thus, the Solanaceae and related *Coffea* genus present an appealing system to study genome evolution and polyploidization (Wu et al. 2010).

The genomes of a number of important crop plants within Solanaceae and Rubiaceae have recently become available, including *S. tuberosum* (potato) (Xu et al. 2011), *S. lycopersicum* (tomato) (Sato et al. 2012), *S. pimpinellifolium* (currant tomato) (Sato et al. 2012), *S. pennellii* (wild tomato) (Bolger et al. 2014), *S. melongena* (eggplant) (Hirakawa et al. 2014), *C. annuum* (pepper) (Kim et al. 2014; Qin et al. 2014), *N. benthamiana* (Australian tobacco) (Bombarely et al. 2012; Naim et al. 2012), *N. sylvestris* and *N. tomentosiformis* (Sierro et al. 2013a), *N. tabacum* (Edwards et al. 2017; Sierro et al. 2014) and *N. otophora* (Sierro et al. 2014), *C. canephora* (robusta coffee) (Denoeud et al. 2014), *Petunia axillaris* and *P. inflata* (Bombarely et al. 2016), *N. attenuata* and *N. obtusifolia* (Xu et al. 2017), *N. glauca* (Khafizova et al. 2018), and *N. rustica*, *N. undulata*, *N. paniculata,* and *N. knightiana* (Sierro et al. 2018). Genome sequencing of other species, such as *Petunia x hybrida*, *C. arabica* (Tran et al. 2018), and numerous *Nicotiana* species, is currently ongoing. *N. attenuata* has been intensively studied and established as a model of a biotic or abiotic stress (Luu et al. 2015). Sequencing data and assembled genomes for the abovementioned Solanaceae species are available at the SOL Genomics Network (http://solgenomics.net) (Fernandez-Pozo et al. 2015).

The genus *Nicotiana* comprises about 78 naturally occurring species (Table 1.1), of which approximately three-quarters are found in the Americas and one-quarter are found in Australia and the South Pacific, with a single species indigenous to Africa (Chase et al. 2003). The species are semi-xerophytic in nature, prefer dry rather than humid heat, and are absent from tropical zones (Goodspeed 1947). The genus was last monographed by Goodspeed (Goodspeed 1954). The most recent sectional classification of the genus, which in part reflected recent molecular phylogenetic evidence, recognized 13 sections (Table 1.1). *N. tabacum* is placed in the monotypic section *Nicotiana* and its putative progenitor species are classified in section *Tomentosae* (*N. tomentosiformis*) and in the monotypic section *Sylvestres* (*N. sylvestris*). All of the Australian, South Pacific, and African species are classified in section *Suaveolentes*. Five sections are believed to be of allotetraploid origin (Knapp et al. 2004b); thus, hybridization is indicated to have made an important contribution to diversification within the genus.

Tobacco (*N. tabacum* L.) is an allotetraploid ($2n = 4x = 48$) that is considered to originate from a hybridization event between ancestors of *N. sylvestris* (S-genome; $2n = 2x = 24$) and *N. tomentosiformis* (T-genome; $2n = 2x = 24$) (Clarkson et al. 2005; Leitch et al. 2008; Lim et al. 2004; Sierro et al. 2013a, 2014). The hybridization and polyploidization event is estimated to have occurred within the last 200,000 years (Clarkson et al. 2004, 2005). Wild plants of *N. tabacum* are considered to be cultivated, and truly wild populations of the species cannot be unequivocally confirmed (Lewis and Nicholson 2007). Despite the scientific significance of *N. tabacum* as a model plant system and its importance as an agronomic crop, until recently, the genetic resources available for genome analysis, genetic mapping, and molecular marker-assisted selection were relatively limited, at least compared with other economically important crops (Davis and Nielsen 1999). Indeed, reference genome sequences for *Nicotiana* species have been published only recently (Bombarely et al. 2012; Edwards et al. 2017; Sierro et al. 2013a, 2014, 2018; Xu et al. 2017). In addition to *N. tabacum* and *N. benthamiana*, certain other *Nicotiana* species are grown as ornamentals (*N. sylvestris*, *N. alata*, and *N. langsdorffii*) or for industrial purposes (*N. rustica* and *N. glauca*). *N. benthamiana* (section *Suaveolentes*) is a model plant for studies of plant–microbe interactions on account of its well-characterized transient protein expression and its amenability to virus-induced gene silencing (VIGS) (Bombarely et al. 2012; Goodin et al. 2008).

Table 1.1 Sectional classification of *Nicotiana* proposed by Knapp et al. (2004b)

Section	Natural distribution	No. species	Haploid chromosome no.[a]
Alatae	N. & S. America	8	9, 10
Nicotiana	S. America	1	24
Noctiflorae	S. America	6	12
Paniculatae	S. America	7	12
Petunioides	N. & S. America	8	12
Polydicliae	N. America	2	24
Repandae	N. America	4	24
Rusticae	S. America	1	24
Suaveolentes	Australia, S. Pacific, Africa	28[b]	15, 16, 18, 19, 20, 21, 22, 23, 24, 32
Sylvestres	S. America	1	12
Tomentosae	S. America	5	12
Trigonophyllae	N. America	2	12
Undulatae	S. America	5	12, 24

[a]*Source* Lewis (2011) and Marks and Ladiges (2011)
[b]Includes the recently recognized *N. fatuhivensis* (Marks 2010) and *N. monoschizocarpa* (Symon and Lepschi 2007)

1.2 Phylogenetic Relationships Within *Nicotiana*

Phylogenetic relationships within *Nicotiana* have been explored using sequence data for four coding or noncoding chloroplast DNA (cpDNA) regions (Clarkson et al. 2004) and the internal transcribed spacer (ITS) region of nuclear ribosomal DNA (Chase et al. 2003). Relationships suggested from the two data sets for diploid taxa generally show good congruence; however, such analyses and their interpretation for elucidation of the evolution of allopolyploid taxa require extreme caution. The cpDNA trees only allow the line of maternal inheritance to be traced (Clarkson et al. 2004). The ITS region is inherited biparentally and is subject to recombination; therefore, it can provide strong evidence for the parentage of hybrid taxa, but conversely may be misleading for elucidating the evolutionary history of taxa of hybrid origin owing to concerted evolution. Interestingly, putative allopolyploid *Nicotiana* taxa show no evidence for “hybrid” ITS sequences, even among those taxa thought to be of recent origin, such as *N. tabacum*; this is believed to be a result of rapid gene conversion to one parental copy (Chase et al. 2003).

Previous phylogenetic reconstructions from cpDNA sequence data indicate that within the Solanaceae, *Nicotiana* is related to the Australian-endemic genus *Symonanthus* (tribe Anthocercideae) and diverged approximately 15 million years ago (Mya) (Clarkson et al. 2004). These authors hypothesized that *Nicotiana* originated in southern South America and subsequently dispersed to North America, Africa, and Australia. The divergence between *Nicotiana* and *Petunia* has been estimated at more than 23 Mya (Wikstrom et al. 2001). A recent Bayesian molecular dating analysis of the Solanaceae (Sarkinen et al. 2013) used sequence data for six plastid and nuclear DNA regions for 40% of the total number of species in the family with two fossil calibration points. The sister clades Solanoideae (containing *Solanum*, *Capsicum*, and *Physalis*) and Nicotianoideae (including *Nicotiana*, Anthocercideae, and *Symonanthus*) were estimated to have diverged about 24 Mya (95% highest posterior density 23–26 Mya). The two clades formed what was informally termed the “$x = 12$ clade” by Särkinen et al. (2013).

In both cpDNA and ITS molecular phylogenetic reconstructions, members of section *Tomentosae* (*N. kawakamii*, *N. otophora*, *N. tomentosa*, and *N. tomentosiformis*) were consistently members of the clade related to the remainder of the genus (Chase et al. 2003; Clarkson et al. 2004). In cpDNA phylogenies, the basal-most clades are consistently retrieved and generally well supported; thus, they are considered to be reliable groups. Although most of the major clades were well supported and generally corresponded with the sections of Knapp et al. (2004b), some internal nodes were only weakly supported, so intersectional relationships were not completely resolved.

Six independent polyploidy events and at least three homoploid hybrid events in *Nicotiana* are hypothesized (Kovarik et al. 2012). Polyploidization events have been estimated to occur between less than 0.2 Mya (*N. tabacum*, *N. rustica*, and *N. arentsii*) and more than 10 Mya (section *Suaveolentes*) according to the molecular clock analyses (Clarkson et al. 2004, 2005; Leitch et al. 2008). The sections containing allopolyploids are thought to originate from the ancestors of distantly related present taxa according to their placement in cpDNA trees (Knapp et al. 2004b). For example, the presumed progenitors of *N. tabacum* (*N. sylvestris* and *N. tomentosiformis*) are not phylogenetically close relatives. Sections *Tomentosae*, *Undulatae*, *Paniculatae*, *Trigonophyllae*, *Petunioides*, *Alatae*, *Repandae*, *Noctiflorae*, and *Suaveolentes* are all monophyletic, but in some cases with weak support, in the cpDNA phylogenies of Clarkson et al. (2004). The members of section *Polydicliae* are indicated to share the same maternal parent, but each existing allopolyploid taxon arose at different times; therefore, the section is not monophyletic in cpDNA phylogenies (Clarkson et al. 2004). Current knowledge of the origins of allopolyploid species is discussed by Kelly et al. (2013).

The monophyly of section *Suaveolentes* is only weakly or moderately supported, and the clade shows comparatively low genetic variation based on cpDNA sequence data compared with other clades (Clarkson et al. 2004). Evolutionary relationships in section *Suaveolentes* are complex, but recent studies provide more detailed insights into the genetic, morphological, and karyotypic evolution of the group (Kelly et al. 2013; Marks and Ladiges 2011; Marks et al. 2011). Nuclear and plastid DNA sequence data suggest that a species from either section *Noctiflorae* or section *Petunioides,* or their hybrid, was the maternal progenitor, while a species from the section *Sylvestres* was the paternal progenitor (Kelly et al. 2013).

Some of the diploid species (*N. glauca*, *N. spegazzinii*, and *N. linearis*) have been shown to be homoploid hybrids based on sequence data for low-copy nuclear genes (Kelly et al. 2010).

1.3 Cytology and Allopolyploidy in *Nicotiana*

The base haploid chromosome number in the *Nicotiana* genus is $n = 12$, which is the haploid chromosome number in seven of the 13 sections proposed by Knapp et al. (Knapp et al. 2004b), as shown in Table 1.1. Aneuploid species occur in section *Alatae* ($n = 9$ or 10) and section *Suaveolentes* ($n = 16$–22). Four sections (*Nicotiana*, *Rusticae*, *Polydicliae*, and *Repandae*) contain allopolyploid species with $n = 24$. These allopolyploids comprise >40% of the total number of *Nicotiana* species. Autoploidy is believed to have been unimportant, at least in the recent evolution of the genus, and aneuploid and/or dysploid reduction is thought to have contributed to the evolution of numerous species in sections *Alatae* and *Suaveolentes* (Reed 1991).

Cytogenetic evolution in *Nicotiana* is summarized by Reed (1991). The karyotype of *Nicotiana* species is useful to resolve species groupings. Goodspeed (1947) presented hypotheses for karyotypic evolution in the genus. A number of meiotic aberrations have been observed in some diploid and polyploid species, which may support the suggested hybrid origin of those taxa. Meiotic behavior in all of the monosomics (individuals in which one chromosome of the normal complement has been deleted) of *N. tabacum*, in all of the trisomics

(individuals that carry three copies of a specific chromosome) of *N. sylvestris*, and in >200 interspecific F_1 hybrids has been studied and provides evidence for chromosomal affinities among the species. *Nicotiana* exhibits a relatively high degree of interspecific cross-compatibility. In approximately 90% of the F_1 hybrids investigated, the amount of chromosomal pairing is consistent with the taxonomic affinity of the parental taxa (Goodspeed 1947). Allopolyploidy in *Nicotiana* has been associated with frequent chromosomal and cytogenetic alterations, such as intergenomic translocations (currently known only in *N. tabacum* (Leitch et al. 2008; Skalicka et al. 2005)), changes in copy number, distribution of both simple and complex repeats, rDNA homogenization, and loss of loci (e.g., Clarkson et al. 2004, 2005; Kenton et al. 1993; Lim et al. 2000, 2004, 2007; Petit et al. 2007). The extent of chromosomal evolution may be associated with the time of the polyploidy event (Lim et al. 2007).

Goodspeed (1954) used evidence from morphology, cytology, artificial hybridization experiments, and distribution to hypothesize on the progenitors of polyploid species; for example, Goodspeed suggested that a member of section *Alatae* was one parent of section *Suaveolentes,* that *N. tabacum* was derived from members of sections *Alatae* and *Tomentosae*, and that *N. rustica* was derived from the ancestors of *N. undulata* and *N. paniculata*. Meiotic behavior of interspecific F_1 hybrids may also shed light on the ancestry of allopolyploid taxa; for example, *N. sylvestris* and *N. otophora* (section *Tomentosae*) F_1 hybrids show almost no chromosome pairing, whereas F_1 progeny of *N. tabacum* with either *N. otophora* or *N. sylvestris* show a high frequency of multivalents (Goodspeed 1947).

Many crosses between members of section *Suaveolentes* and *N. tabacum*, which share a common progenitor (the ancestor of *N. sylvestris*), show hybrid lethality. Different mechanisms may be responsible in different crosses. Genes in both the S- and T-genomes of *N. tabacum* are indicated to be responsible for hybrid lethality in crosses with one member of section *Suaveolentes*, *N. occidentalis* (Tezuka and Marubashi 2012). Tezuka and Marubashi (2012) proposed an evolutionary model to explain the phenomenon of hybrid lethality in section *Suaveolentes*.

More than 300 synthetic interspecific hybrids in *Nicotiana* have been described (Lewis 2011). Goodspeed (1947) considered that hybridization was a major contributor to diversification and chromosomal evolution in *Nicotiana* and that interspecific relationships were complex and difficult to resolve, as illustrated in his intricate phyletic diagrams. Phylogenetic analyses of cpDNA sequence data have provided evidence for the maternal progenitors of the allopolyploid taxa (Clarkson et al. 2004; Knapp et al. 2004b). The ancestor of *N. sylvestris* is the putative maternal parent of *N. tabacum* and is hypothesized to have contributed to multiple allotetraploid events and the origin of three allotetraploid sections (Knapp et al. 2004b).

Genomic in situ hybridization (GISH) offered additional proof of the identity of the progenitors of *N. tabacum* (Kenton et al. 1993; Murad et al. 2002) to explore the origin of 15 allopolyploid *Nicotiana* species (Chase et al. 2003); to evaluate genome evolution in the allotetraploid species *N. arentsii*, *N. rustica*, and *N. tabacum* (Lim et al. 2004); and to examine the integration of geminiviral-related DNA into the *Nicotiana* genome (Lim et al. 2000). GISH was used to demonstrate that virus-derived sequences of *N. tabacum* are closer to those of *N. tomentosiformis* than *N. tomentosa* (Murad et al. 2002). On the basis of GISH with total genomic DNA of *N. sylvestris* and *N. tomentosiformis*, Lim et al. (2000) suggested that gene conversion overall had played a limited role in the evolution of *N. tabacum*.

Fluorescent in situ hybridization (FISH) has also been used to pinpoint the chromosomes of *N. tabacum* (Parokonny and Kenton 1995), for chromosomal localization of the 5S rDNA (Kitamura et al. 2005), and to examine the localization of the Tnt1 retrotransposon in 22 *Nicotiana* species in conjunction with GISH (Melayah et al. 2004).

1.4 Genetic Resources in *Nicotiana*

1.4.1 Wild and Cultivated Germplasm

Wild *Nicotiana* species show considerable variation in growth habit, inflorescence type, and ecophysiological adaptation. Flower variability is mainly manifested in the corolla form and color, stamen insertion, and aestivation (Goodspeed 1947). Wild *Nicotiana* species offer extensive phenotypic diversity, cross-compatibility, and amenity to ploidy manipulation and tissue culture (Lewis 2011) and thus represent a valuable resource for genetic improvement of cultivated tobacco.

Considerable phenotypic variation in agromorphological traits and disease resistance exists among cultivated *N. tabacum* (e.g., Darvishzadeh et al. 2013; Elliott et al. 2008), which demonstrates there is potential for genotypic selection to improve cultivars. Numerous studies have detected a higher level of genetic diversity among wild *Nicotiana* species or naturalized *N. tabacum* genotypes than among cultivated *N. tabacum* genotypes (e.g., Ren and Timko 2001; Yang et al. 2007).

Genetic variability within *N. tabacum* is believed to have been affected by several genetic bottlenecks during its evolution, reflecting the nature of its origin and breeding practices, but it remains unknown if more than one hybridization event occurred and whether subsequent introgression from either progenitor species occurred. Furthermore, it is speculated that only a minute portion of the genetic diversity of the diploid progenitor genomes contributed to that of *N. tabacum* (Lewis and Nicholson 2007).

Cultivars of some market classes of cultivated tobacco, such as cigar, dark fire-cured, and oriental, have remained largely the same. However, burley and flue-cured cultivars have been significantly improved through modern breeding practices, which considerably narrowed their germplasm diversity, resulting in limited number of available crosses between members of different classes (Lewis and Nicholson 2007).

In several studies, grouping of *N. tabacum* genotypes based on genetic distance is generally consistent with the main tobacco classes (e.g., Darvishzadeh et al. 2013; Davalieva et al. 2010; Fricano et al. 2012), which suggests that the gene pool of each class has largely remained isolated during subsequent selection. The degree of homogeneity of each class partially corresponds to differences in the stringency of the breeding strategy for each class. Fricano et al. (2012) resolved two heterogeneous clades that contained most wild accessions and a number of country-specific cultivars. These clades may be of particular interest for future breeding programs, as they represent the highest genetic diversity in the *Nicotiana* genus. Grouping of *N. tabacum* genotypes based on phenotypic traits or genetic distance also enables discrimination of groups of other genetically divergent genotypes.

As with many crops, breeding and selection of superior genotypes and intensification of cultivation have coincided with a decline in genetic diversity among cultivated *N. tabacum* genotypes. Thus, conservation of genetic resources and their use to improve cultivated genotypes is of increasing importance. For example, the U.S. Department of Agriculture *Nicotiana* Germplasm Collection (USDA 2016) maintains more than 1,700 tobacco accessions and seeds of more than 200 accessions representing about 60 wild *Nicotiana* species (Lewis and Nicholson 2007). The Tobacco Research Institute of the Chinese Academy of Agricultural Sciences (TRI 2015) maintains a collection of more than 5,300 tobacco accessions and claims to maintain the largest collection of tobacco varieties in the world.

1.4.2 Cytoplasmic Male Sterile Lines

Cytoplasmic male sterility (CMS) is frequently encountered in interspecific hybrid progeny between a wild *Nicotiana* species and *N. tabacum* (as the pollen donor) after several backcross generations. It is usually expressed as staminal sterility. Many modern *N. tabacum* genotypes in cultivation carry CMS (Lewis and Nicholson 2007). Lines with restored structural or functional fertility were identified within the CMS

lines (Reed 1991); however, tobacco leaves rather than seeds bring the highest production value.

1.4.3 Cytogenetic Stocks

Aneuploid lines of *Nicotiana* have been developed and are useful for breeding and genetic investigations, but they have not been used to the same extent as in other polyploid crops, such as *Triticum*. As allotetraploids, both *N. tabacum* and *N. rustica* seem to be tolerant to different forms of aneuploidy. A number of genes have been localized to individual chromosomes within the monosomic lines. Trisomic lines have been developed for *N. sylvestris* and *N. langsdorffii* (Lewis 2011). Nullisomics (in which a specific chromosome pair from both genomes is deleted) have been generated, mostly in *N. tabacum*. The fact that most nullisomics of *N. tabacum* are nonviable or sterile is thought to be because of the evolution of genomic interdependence subsequent to polyploidization (Reed 1991).

In attempts to transfer foreign genes (e.g., for tobacco mosaic virus [TMV] resistance) to *N. tabacum* from wild species and to localize genes or molecular markers to specific loci, chromosome addition lines can be produced in tobacco (Lewis 2011). The amenability of *N. tabacum* to transformation and tissue culture and the recent availability of genomic resources and genetic maps may stimulate interest in genetic engineering technologies for improvement of cultivated tobacco.

1.5 Reference Genomes in *Nicotiana*

A methyl filtration approach (Whitelaw et al. 2003) was used to address the combined challenges of the bulky genome size and complexity of the tobacco genome within the Tobacco Genome Initiative (TGI). The project aimed to sequence more than 90% of *N. tabacum* hypomethylated gene-rich genomic DNA sequences. In addition, several bacterial artificial chromosome (BAC) libraries were constructed to perform BAC-by-BAC sequencing. More than 1.3 million genome survey sequences and more than 50,000 expressed sequence tags (EST) were released to GenBank. The subsequent assembly produced a highly fragmented genome with the majority of contigs containing only partial gene structures. Nevertheless, the TGI data served as a source of genomic sequences for the design of a tobacco exon array (Martin et al. 2012) and a tobacco genetic map (Bindler et al. 2011).

High-quality draft genome sequences have been published for representative genotypes of the three main cultivar classes of *N. tabacum* (K326, flue-cured; TN90, burley; and Basma Xanthi, oriental) (Edwards et al. 2017; Sierro et al. 2014), its progenitor species *N. sylvestris* and *N. tomentosiformis* (Sierro et al. 2013a), *N. benthamiana* (Bombarely et al. 2012; Naim et al. 2012), *N. attenuata* (Xu et al. 2017), and *N. rustica* and its progenitor species *N. undulata*, *N. paniculate*, and *N. knightiana* (Sierro et al. 2018). In addition, Sierro et al. (2013b) constructed a physical map of *N. tabacum* Hicks Broadleaf using whole-genome profiling technology. The physical map, which was predicted to cover the whole tobacco genome, was subsequently used for super-scaffolding of the *N. tabacum* genome (Sierro et al. 2014).

Bombarely et al. (2012) mapped 56.5 and 58.5% of coding sequences from the tomato and potato genomes, respectively, to the *N. benthamiana* genome to demonstrate syntenic relationships. The genomic ancestry of *N. benthamiana* is uncertain, but the progenitors of the allopolyploid section *Suaveolentes* might belong to the lineages represented by section *Sylvestres* and either section *Noctiflorae* or section *Petunioides* (Kelly et al. 2013).

Multiple insertions of cellular transferred DNA sequences (cT-DNA), derived from *Agrobacterium* strains and integrated into the host nuclear genome, have occurred during the evolution of *Nicotiana*. cT-DNAs contain genes involved in plant growth and the synthesis of opines important for bacterial growth. A number of *Agrobacterium*-derived genes (including *mis*, *orf13*, and *rolC*) have been detected by

polymerase chain reaction amplification in the genomes of various *Nicotiana* species (see Chen et al. 2014). Four cT-DNA inserts in the genome of *N. tomentosiformis* have been characterized by deep sequencing. In addition, the cT-DNAs present in the genomes of *N. tabacum* Basma Xanthi and *N. otophora* have been identified, with a fifth cT-DNA detected in the latter species (Chen et al. 2014). The cT-DNA inserted exclusively in *N. otophora* contains intact open reading frames coding for three 6b genes derived from the original *Agrobacterium* 6b gene, which is known to modify plant growth. The phenotypes observed upon expression of these 6b genes in *N. tabacum* under the constitutive 2 × 35S promoter are however different from those previously described. Indeed, shorter plants exhibiting modified petiole wings and dark-green leaves with increased number of veins were observed upon expression of any of these three genes. Furthermore, the expression of one of them led to the additional development of outgrowths at the leaf margins and of modified flowers. In the capsules of these plants, germination of the embryos was observed at an early development stage in the capsules (Chen et al. 2018). Kovacova et al. (2014) showed that the agrobacterial *mis* gene has evolved independently in the *N. glauca*, *N. tabacum*, and *N. tomentosiformis* genomes.

1.5.1 Chloroplast Genomes

The complete chloroplast genome of *N. tabacum* was sequenced more than 30 years ago (Shinozaki et al. 1986) and comprises 155 kb, including two identical copies of a 25 kb inverted repeat separated by 86 kb and 18 kb unique regions.

The whole chloroplast genomes of *N. sylvestris* and *N. tomentosiformis* have also been described (Yukawa et al. 2006). The *N. sylvestris* chloroplast genome comprises 156 kb and, with the exception of one gene, is highly similar to the one of *N. tabacum.* Only seven sites were polymorphic between *N. tabacum* and *N. sylvestris*. The chloroplast genome of *N. tomentosiformis* is 156 kb long and also shows similar genome structure to that of *N. tabacum*, except for five regions with more than 1,000 single nucleotide polymorphism (SNP) differences. Overall, the chloroplast genomes of *N. sylvestris* and *N. tomentosiformis* were 99.99 and 98.54% identical to that of *N. tabacum.* The authors concluded that the detailed comparison clearly supported the previous finding that *N. sylvestris* was the source of the *N. tabacum* chloroplast genome (Yukawa et al. 2006).

1.5.2 Mitochondrial Genome

Bland et al. (1985) cloned fragments of known genes in the mitochondrial genome (mtDNA) of maize and unidentified genes from *N. sylvestris* mtDNA to use as labeled probes for Southern blot hybridization to explore *N. tabacum* mtDNA. The results supported the contention that *N. tabacum* mtDNA was inherited from *N. sylvestris* and not from *N. tomentosiformis*. Conservation in organization and sequence homology indicates that little evolutionary divergence in mtDNA between *N. tabacum* and *N. sylvestris* has occurred.

More recently, Sugiyama et al. (2005) sequenced the complete mtDNA of *N. tabacum*. The genome was assumed to conform to a conventional “master circle” model of mitochondrial genome structure (Sloan 2013) and comprised of 431 kb, including protein-coding, ribosomal RNA, and transfer RNA genes. The homology of repeated sequences, the gene organization, and the intergenic spacer regions differ markedly from those of other model plants (*Arabidopsis thaliana* or *Oryza sativa*), indicating that the genome structure has undergone multiple reorganizations during higher plant evolution.

1.6 Reference Genomes in Solanaceae and *Coffea*

The genomes of Solanaceae and *Coffea* species play an important role in comparative genome analysis of *Nicotiana* species. The knowledge

accumulated over the last decades of research can be leveraged at genetics, genomics, transcriptomics, and proteomics data analysis.

The genomes of tomato (*S. lycopersicum*) (Sato et al. 2012) and potato (*S. tuberosum*) (Xu et al. 2011) are recognized as the best-quality reference genomes for the Solanaceae family, mostly due to the significant decade-long efforts by the tomato and potato communities to generate a huge collection of genomics and genetics resources. Although tomato and potato are closely related, the cultivated tomato varieties are inbred self-pollinating diploids, whereas most potato cultivars are highly heterozygous self-incompatible autotetraploids.

Due to high heterozygosity and autotetraploidy of potato cultivars, a homozygous doubled-monoploid (DM) potato was constructed to facilitate the assembly of its high-quality genome and transcriptome (Xu et al. 2011). The 727 Mb assembly of the genome with estimated size of 884 Mb was split over 443 super-scaffolds and contained at least 62% of repeats. The majority of the assembly (86%) was anchored to the potato genetic and physical maps to construct 12 pseudomolecules. Large blocks of heterochromatin were attributed to pericentromeric regions, and gene-rich euchromatin was pushed toward the distal chromosome ends. The combination of ab initio and experimental evidence (proteins and ESTs) predicted 39,031 protein-coding genes in the DM potato. Approximately 10,000 genes clustered into 1,800 paired syntenic blocks, suggesting the occurrence of at least two whole-genome duplications in the potato. Comparison between the DM and heterozygous potato (RH line) uncovered high level of structural variations comparable to the one in maize; however, no specific genes or regions were identified to be responsible for the phenomenon of inbreeding depression in the potato.

The assembled tomato genome contained 760 Mb (84% of 900 Mb) split into only 91 scaffolds linked to the BAC-based physical, genetic, and introgression line maps (Sato et al. 2012). As in the case of the potato, the tomato euchromatic gene-rich regions (71 Mb) were situated at the distal parts of the chromosome arms, while the heterochromatin was found in the pericentromeric part. The tomato genome has an unusually low proportion of retrotransposons and other complex repeats. The annotation pipeline predicted 35,000 genes in the tomato and potato, 18,320 of which were orthologous between the two species. Comparison of the tomato and potato genomes revealed ~9% nucleotide divergence in euchromatin and ~30% in heterochromatin. More than 20% of genomic regions of the grape have three orthologous regions in the tomato genome, supporting the hypothesis that triplication in the tomato occurred ~71 Mya based on K_s of paralogous genes (i.e., a long time before the divergence of the tomato and potato [~7 Mya]).

One of the interesting features of the hot pepper genome (*C. annuum* cv. CM334) is its large diploid size of 3.48 Gb (Kim et al. 2014), compared to less than 1 Gb for tomato or potato genomes. Kim et al. assembled 88% (3.06 Gb) of the genome into 37,989 scaffolds (with N_{50} = 2.47 Mb) and anchored 86% of the latter (2.63 Gb) onto 4,562 markers of the combined *C. annuum* high-density genetic map (Kim et al. 2014; Yarnes et al. 2013). Transposable elements occupied 76% (2.34 Gb) of the *C. annuum* genome, with long-terminal repeats (LTR) being the most represented (~70%) with only slight distribution preference toward the centromeric regions. Substantial expansion of the Gypsy family retrotransposons (sevenfold higher than in tomato) was put forward as the main cause of the inflated size of *C. annuum*. Caulimoviridae retrotransposons was another unusually abundant repeat family present ninefold higher than in tomato. The number of protein-coding genes (34,903) was very close to other Solanaceae, 17,397 of which have an ortholog in the tomato gene set. However, the expression levels among different tissues varied significantly, affecting as many as 46% of orthologous genes. Hot pepper, tomato, and potato shared 2,139 gene families, and 756 were specific to *C. annuum*. Genomics analysis of two other varieties of *C. annuum* (Qin et al. 2014) largely confirmed the observations and provided further insights into its evolution,

such as evidence for the whole-genome triplication. Despite more than 600 chromosomal translocation events between the pepper and tomato genomes being uncovered, overall, the presence of a substantial number (>1,000) of large syntenic blocks indicated a close evolutionary relationship between the species.

The draft genome of *Solanum melongena* L. (eggplant) (Hirakawa et al. 2014) covered 833 Mb (74%) of its estimated 1.1 Gb size, with 586 Mb annotated as repetitive sequences (70% of 833 Mb). LTR of Copia and Gypsy types accounted for almost 30% of the repeats. Only 42,035 of 85,446 genes were considered protein-coding, whereas the rest were described as "transposable elements." A total of 4,018 genes were found to be exclusive to the eggplant. Finally, Hirawaka et al. (2014) used the obtained genomics sequences to identify SNP and simple sequence repeat (SSR) polymorphic markers in the interspecific F_2 population and constructed a genetic map containing on 574 SNPs and 221 SSRs.

Although the *Coffea* genus belongs to the Rubiaceae family, there are a few striking similarities to the *Nicotiana* genus, making the inclusion of *Coffea* species valuable. Some of the most popular cultivars of coffee belong to the allotetraploid *C. arabica* ($2n = 4x = 44$) species, which, similar to *N. tabacum*, emerged through an interspecific hybridization of two ancestral *Coffea* species still available today: *C. eugenioides* ($2n = 2x = 22$) and *C. canephora* ($2n = 2x = 22$). The latter is also known as *C. robusta* and is considered to be easier to produce yet more bitter coffee. While the genome of *C. arabica* is still in progress, the high-quality draft of *C. canephora* assembled at 80% of its estimated 710 Mb genome size was reported as harboring 25,574 protein-coding genes (Denoeud et al. 2014). A high-density genetic map covered 64% of the draft genome. About half of the genome was filled with repeats, 85% of which were LTR retrotransposons. Comparative genome analysis of *C. canephora* chromosomes provided no evidence of triplication since the origin of eudicots, thus displaying one-to-one mapping to the grape genome and very little syntenic divergence to other asterids. Denoeud et al. (2014) suggested using *C. canephora* as a reference species to study the evolution of asteroid angiosperms.

1.7 Genomic Evolution in *Nicotiana*

Goodspeed (1954) considered that genic rather than genomic evolution was of greater importance in the evolution of the allopolyploid lineages of *Nicotiana*. However, recent studies have provided detailed insights into the extent of genomic evolution that has occurred in *N. tabacum* and other allopolyploid *Nicotiana* species of different ages. Allopolyploidy is frequently linked to functional and structural genomic changes, and DNA loss or gain often occurs at the onset of polyploidization (Petit et al. 2007). Such genomic differentiation or "diploidization," which decreases pairing between homoeologous chromosomes, is documented in *Nicotiana* allopolyploid species of different ages and synthetic hybrids.

1.7.1 Chromosomal Rearrangements

The *N. tabacum* genome has experienced a variety of chromosomal rearrangements subsequent to polyploidization in either wild type or synthetic species compared with the progenitor diploid genomes. The rearrangements include up to nine intergenomic translocations, retrotransposon deletion/proliferation, and transposable element mobility and loss (Bindler et al. 2011; Kenton et al. 1993; Kovarik et al. 2012; Lim et al. 2007; Petit et al. 2010; Skalicka et al. 2005; Wu et al. 2010). The T-genome of *N. tabacum* shows reduced genetic stability compared with the S-genome (Kovarik et al. 2012). The T-genome of *N. tabacum* is shown to have experienced more rapid chromosomal evolution than the S-genome, and both genomes are indicated to have evolved more rapidly than the corresponding genomes of diploid relatives. The more rapid genomic changes in *N. tabacum* and other polyploid species might be attributable to

chromosome recombination, epigenetic variation, or higher transposable element activity (Wu et al. 2010). Matyasek et al. (2011) suggested that the lack of subtelomeric satellite repeats in the T-genome of *N. tabacum* may promote homoeologous pairing, which in turn might explain the frequency of intergenomic translocations in the *N. tabacum* genome. Wild *N. tabacum* plants contain reduced copy numbers of pararetroviral repeat sequences derived from *N. tomentosiformis* and of the Tnt1 retrotransposon compared with the number expected to be inherited from the progenitors (Gregor et al. 2004; Melayah et al. 2004). Given that most nullisomics of *N. tabacum* are nonviable or sterile, post-polyploidization evolution of interdependence of the progenitor genomes is also indicated (Lewis 2011). However, both dynamic and stable inheritance of different repeat sequences during *N. tabacum* evolution is implicated (Skalicka et al. 2005).

1.7.2 Post-polyploidization Changes

Rapid genetic changes in synthetic *N. sylvestris* × *N. tomentosiformis* hybrids have been documented, mostly targeted to the paternal T-genome donated by *N. tomentosiformis*. Lim et al. (2004) examined genomic rearrangements in three recent polyploid species (*N. arentsii*, *N. rustica*, and five cultivars and one wild genotype of *N. tabacum*; each allopolyploid originated <0.2 Mya) and three synthetic F_3 progeny derived from the cross *N. sylvestris* × *N. tomentosiformis*. Intergenomic translocations were observed in all natural genotypes of *N. tabacum* examined but not in *N. arentsii* or *N. rustica*. Skalická et al. (2005) reported that two repeats related to endogenous viruses and two classes of *N. tomentosiformis*-specific noncoding tandem repeats were deleted in S_4 plants generated from a synthetic allotetraploid S_0 plant. All of the sequences were derived from the T-genome of *N. tomentosiformis*. T-genome 35S rDNA genes underwent rapid homogenization and were replaced by novel gene variants in the same S_4 population (Skalicka et al. 2003, 2005). Some of the abovementioned changes are consistent with changes that are believed to have occurred during the evolution of *N. tabacum*. These studies also demonstrate that considerable genomic variation may arise rapidly from a single polyploid individual within only three or four generations.

1.7.3 Transposable Elements

Transposable elements are mobile DNA elements that are inserted in different positions in the genome and generate structural variations upon insertion, therefore having an important impact on genetic diversity, gene expression, and genome structure. Extensive and rapid turnover of retrotransposons and transcriptional activation and/or mobilization of transposable elements may follow interspecific hybridization and polyploidization (Petit et al. 2007, 2010). Sequence-specific amplification polymorphisms (S-SAP) have been used to examine retrotransposon polymorphism and genomic distribution in *Nicotiana* species (Melayah et al. 2004; Petit et al. 2007). Melayah et al. (2004) reported that Tnt1 copia-type retrotransposons are highly polymorphic and have evolved rapidly in *Nicotiana*, and the number of Tnt1 insertions differs widely among *Nicotiana* species. S-SAP profiles provide evidence for post-polyploidization changes in Tnt1 insertions in *N. tabacum*. Some 30% of *N. tabacum* S-SAP bands are shared with *N. tomentosiformis*, and 38% are shared with *N. sylvestris*. A notable proportion (28%) of Tnt1 insertions are unique to *N. tabacum*. A large proportion of S-SAP bands (52% for *N. sylvestris* and 47% for *N. tomentosiformis*) are present in the two progenitor species but not in *N. tabacum*. In *N. tomentosiformis*, Tnt1 insertions are distributed across all chromosomes but at different densities. *N. sylvestris* contains a higher number of Tnt1 insertions, which show a more homogeneous distribution in the genome. The number of insertions in *N. tabacum* is higher than the numbers in both progenitor species, although only 67% of the expected number based on the numbers detected in the progenitor

species. These insertions are more concentrated on S-genome chromosomes than on T-genome chromosomes. Low Tnt1 polymorphism among *N. tabacum* lines suggests that Tnt1 content has stabilized during *N. tabacum* evolution. Petit et al. (2007) showed that retrotransposon turnover occurs in *N. tabacum* with removals counterbalanced by new insertions.

Parisod et al. (2012) examined the insertion polymorphism of seven transposable elements in the four allopolyploid species of section *Repandae* that show opposing evolutionary trends in *N. tabacum* and its progenitors. Many novel S-SAP bands were observed for two transposable elements, but considerable loss of transposable elements was detected in the four allotetraploid species of section *Repandae* relative to the genomes of the diploid progenitor species. In the process of diploidization of the four allotetraploid species, the transposable element genome fractions have been significantly restructured.

1.7.4 Repetitive Elements

Nicotiana genomes are notably rich in simple repeats, three superfamilies of which have been characterized in *Nicotiana* species (Matyasek et al. 2011). Homologous subtelomeric repeats in the allotetraploid *N. arentsii*, descended from progenitors of *N. undulata* (maternal donor) and *N. wigandioides* (paternal donor; both members of section *Undulatae*) less than 0.2 Mya, were analyzed by Matyasek et al. (2011). Intergenomic homogenization of the two homologous satellites has not occurred in *N. arentsii*. The authors suggested that the dissimilarity in sequence and structure of the satellite repeats protected homoeologous chromosomes from genomic intergressions.

As reported by Koukalova et al. (2010), the *Nicotiana* genomes of more recent allotetraploid species formed ~1 Mya (section *Polydicliae*) contain rearranged yet intact progenitor repeat sequences, while in older allotetraploids (section *Repandae* ~5 Mya), different satellite repeats partially or completely replaced the ones of the corresponding progenitors. The authors proposed a mechanism that involves removal of progenitor heterochromatic repeat-containing blocks and rolling circle replication leading to formation of new uniform blocks of satellite repeats.

Gill (1991) proposed that during the allotetraploid formation, incompatibility between maternal and paternal cytoplasms may become a source of progenitor genome instability. Lim et al. (2004) concluded that nuclear–cytoplasmic interaction was indicated to have influenced genome evolution only in *N. tabacum* among the taxa studied. If this hypothesis is pertinent to *N. tabacum*, the *N. tomentosiformis*-derived repeat sequences in *N. tabacum* would continue to be eliminated well after the polyploidization event (Lim et al. 2007). Skalická et al. (2005) raised the possibility that the *N. tomentosiformis* genome may simply be inherently less stable than that of *N. sylvestris*.

1.7.5 Other Genetic Changes

Homogenization of rDNA is often noted among allopolyploid *Nicotiana* species, but the extent of interlocus homogenization differs among allopolyploid species and synthetic polyploid hybrids (Dadejova et al. 2007). Furthermore, 18S rDNA coding sequences show near-complete homogenization in the diploid species *N. sylvestris*, *N. tomentosiformis*, *N. otophora*, and *N. kawakamii*, whereas the ITS1 region shows greater heterogeneity. Dadejova et al. (2007) speculated that transcriptionally active rDNA genic regions are more susceptible to homologous recombination than those that are transcriptionally silenced and thus remain unchanged. Matyasek et al. (2012) proposed that the evolution of rDNA genes is dependent on the transcriptional activity and the number of copies. The detection of putative intragenic recombination in the several low-copy nuclear genes suggests that formation of chimeric regions arising from different alleles might have been a common event during *Nicotiana* evolution (Kelly et al. 2010).

1.7.6 Genomic Downsizing

"Genomic downsizing" has been a frequent phenomenon found in *Nicotiana* ancient and more recent polyploids (Leitch and Bennett 2004). Changes in genome size in allopolyploid *Nicotiana* species were examined by comparing the genome sizes of nine diploid and nine polyploid species, with the extant species believed to be descended from the putative progenitors (Leitch et al. 2008). Genome size was assessed using flow cytometry and Feulgen microdensitometry. Four polyploids displayed genome downsizing, whereas five polyploids showed genome size increases. *N. repanda* and *N. nudicaulis* share the same progenitor species (ancestors of *N. sylvestris* and *N. obtusifolia*) yet show contrasting trends in genome size changes (~29% increase and ~14% decrease, respectively). The amount of genome size divergence was enhanced with increasing estimated age of the polyploid.

The progenitor sequences were still detectable in the downsized polyploids, while the signal disappears in the older upsized ones (formed ~4.5 Mya). The authors concluded that genome upsizing is likely linked with replacement of the progenitor repeat regions with the new satellite repeats.

Lim et al. (2007) analyzed genome evolution in *Nicotiana* allopolyploid by comparing repeat sequence distribution in selected BAC clones, in conjunction with FISH and GISH, from four allopolyploids of contrasting ages: *N. nesophila* (formed ~4.5 Mya), *N. quadrivalvis* (~1 Mya), *N. tabacum* (<0.2 Mya), and synthetic F_1 *N. sylvestris* × *N. tomentosiformis* hybrids. The authors proposed the term "genome turnover" for the apparent genome-wide replacement of one sequence type by another, possibly through concerted evolution. The authors presented a hypothetical time-course after polyploidization of the genomic events occurring in *Nicotiana* allopolyploids from the initial "genomic shock," followed by sequence mixing, to swap of progenitor DNA with new sets of genomic DNA repeats and, ultimately, a reduction in chromosome number (Lim et al. 2007).

1.8 Functional Genomics and Gene Mining

1.8.1 Alkaloid Metabolism

The biosynthesis and transport of major tobacco alkaloids (e.g., nicotine, nornicotine, anabasine, and anatabine) have been extensively reviewed elsewhere (Dewey and Xie 2013; Eich 2008). Nicotine is synthesized in the roots of tobacco plants and is transported to the leaves. The majority of genes involved in the nicotine biosynthesis pathway are sequenced and well characterized. Nicotine is the principal alkaloid that accumulates in *N. sylvestris*, whereas nornicotine is predominant in *N. tomentosiformis,* typical for members of section *Tomentosae*. Unlike most *Nicotiana* species, *N. sylvestris* accumulates higher quantities of nicotine in the roots than in the leaves. Transcriptome data indicate that the marked differences in alkaloid metabolism in *N. sylvestris* and *N. tomentosiformis* are due to strong upregulation of a set of key enzymes in the nicotine biosynthesis pathway in the roots of *N. sylvestris* (e.g., aspartate oxidase, putrescine *N*-methyltransferase [PMT], quinolinate synthase, and quinolinic acid phosphoribosyl transferase) compared with *N. tomentosiformis* (Sierro et al. 2013a).

In the senescing leaf of *N. tabacum*, nicotine is accumulated as the main alkaloid, whereas in *N. sylvestris*, *N. tomentosiformis*, and synthetic *N. sylvestris* × *N. tomentosiformis* and *N. sylvestris* × *N. otophora* amphidiploids, nicotine is converted to nornicotine. This pathway is mediated by nicotine *N*-demethylase (*CYP82E*) genes, copies of which are present in the genomes of the progenitor species. Transcriptome data indicate that the higher accumulation of nornicotine in the leaf of *N. tomentosiformis* reflects expression of three *CYP82E* genes that enhance nornicotine production (Sierro et al. 2013a). Mutation of *CYP82E* in *N. tabacum* decreases conversion of nicotine to nornicotine compared with the wild type (Chakrabarti et al. 2007).

Gene expression analysis of the tobacco BY-2 cell cultures has identified a set of genes responsible for alkaloid biosynthesis (Goossens et al. 2003).

Transcription factors (TF) were screened using a transient expression assay, leading to identification of two AP-2 domain TFs that enhance the expression of PMT, a key enzyme in nicotine biosynthesis (De Sutter et al. 2005). Häkkinen et al. (2007) isolated full-length clones for several genes putatively involved in pyridine alkaloid metabolism. Upon their transformation into tobacco BY-2 cell and hairy root cultures, overexpression of a GH3-like enzyme (designated NtNEG1) was shown to increase nicotine accumulation.

In general, the number of isoforms of genes implicated in alkaloid biosynthesis present in the *N. tabacum* reference genomes corresponds roughly to the sum of genes inherited from the progenitor species, *N. tomentosiformis* and *N. sylvestris* (Sierro et al. 2014).

1.8.2 Cadmium Accumulation

An unusual property of *Nicotiana* species is the ability to accumulate relatively high concentrations of cadmium and other heavy metals compared with other plants. *N. sylvestris* and *N. tomentosiformis* genomic sequences and transcriptome data for the root, leaf, and flower have been analyzed to identify genes associated with heavy metal accumulation and transport (Sierro et al. 2013a). Certain differences in transcription were detected in the different organs and between the two *Nicotiana* species, and between *A. thaliana* and the *Nicotiana* species. Expression of osmotin- and thaumatin-like proteins in both leaf tissue and trichomes is induced by cadmium treatment (Harada et al. 2010).

1.8.3 Disease Resistance and Viral Infection

Many of the disease resistance genes currently present in *N. tabacum* germplasm were introgressed from wild *Nicotiana* species (Battey and Ivanov 2014). These genes often consist of a single locus (e.g., TMV resistance derived from *N. glutinosa* (Lewis et al. 2005) and tomato spotted wilt virus resistance inherited from *N. alata* (Moon and Nicholson 2007)), but often additional genetic material from the donor *Nicotiana* species is also introgressed with the resistance gene (Battey and Ivanov 2014).

TMV resistance was introduced into *N. tabacum* in the 1930s as a single dominant locus through interspecific hybridization with *N. glutinosa*. The locus harbors the *N* gene, which encodes a TIR-NBS-LRR protein, triggering a hypersensitive response involving formation of a necrotic lesion (Holmes 1938). It was cloned (Whitham et al. 1994) and subsequently successfully transformed (Lewis et al. 2007). Adoption of flue-cured cultivars carrying the *N* gene was low because of unfavorable “linkage drag effects” (Lewis et al. 2005). Among the three *N. tabacum* cultivars with genome sequences, only TN90 is TMV resistant and includes a putative *N. glutinosa*-derived genomic segment containing the *N* gene (Sierro et al. 2014).

N. tabacum are vulnerable to infections by potyviruses, such as potato virus Y (PVY), tobacco vein mottling virus, and tobacco etch virus. Unlike *N. tabacum* and *N. sylvestris*, *N. tomentosiformis* is resistant to PVY. Next-generation sequencing has been used to characterize the recessive PVY resistance gene in a recombinant inbred line population of *N. tabacum* (Julio et al. 2014). A correlation between susceptibility and expression of a eukaryotic translation initiation factor 4E (*eIF4E*) was confirmed. The *eIF4E* gene was mapped on linkage group 21 and was inherited from the progenitor of *N. sylvestris.* Missense mutations in the *eIF4E* gene conferred resistance of *N. tabacum* lines to PVY infection. PVY resistance is associated with deletion of a chromosomal segment containing the *eIF4E* gene, which is corroborated by the genome sequence of the PVY-resistant *N. tabacum* TN90 (Sierro et al. 2014), although identification of other PVY-resistant accessions with a functional *eIF4E* gene indicates that additional sources of resistance may exist (Julio et al. 2014).

N. benthamiana serves as a model system to study plant–pathogen interactions, in large part because of its susceptibility to a wide spectrum of bacteria and viruses (Goodin et al. 2008).

The unique virus susceptibility of *N. benthamiana* is at least partly attributed to an SNP in an RNA-dependent RNA polymerase gene (*NbRdRP1m*) (Yang et al. 2004). Senthil et al. (2005) used microarrays derived from *S. tuberosum* ESTs to monitor gene expression to shed light on the mechanism of *N. benthamiana* infection with a variety of viruses. In addition, progress is being made on the development of VIGS cDNA libraries for *N. benthamiana* (Goodin et al. 2008).

Bombarely et al. (2012) detected orthologs for more than 20 immunity-associated genes putatively transferred into the *N. benthamiana* genome from other species. Due to the allotetraploid nature of *N. benthamiana*, with a couple of exceptions, each gene was represented by at least two homoeologous sequences in its genome sequence.

1.9 Conclusion

Among *Nicotiana* species, *N. tabacum* and *N. benthamiana* have been extensively used as model systems for plant biology. The availability of a growing number of genomic resources for the genus contributed significantly to understanding the *Nicotiana* biology and evolution. Tobacco plants thus continue to play a significant role in plant research. Furthermore, *N. tabacum* remains an important agricultural crop for which these resources are opening new avenues (e.g., in numerous biotechnological applications for the production of valuable compounds for the pharma or cosmetic industries). With the release of additional resources and, consequently, the increased understanding of the specificities of each clade and species, other members of the *Nicotiana* genus will eventually also develop into efficient biological systems for the production of these compounds.

References

Battey JND, Ivanov NV (2014) Data mining of plant metabolic pathways. In: Hoffmann RD, Gohier A, Pospisil P (eds) Data mining in drug discovery. Wiley, pp 101–130

Bindler G, Plieske J, Bakaher N, Gunduz I, Ivanov N, Van der Hoeven R, Ganal M, Donini P (2011) A high density genetic map of tobacco (*Nicotiana tabacum* L.) obtained from large scale microsatellite marker development. Theor Appl Genet 123:219–230

Bland MM, Matzinger DF, Levings CS (1985) Comparison of the mitochondrial genome of *Nicotiana tabacum* with its progenitor species. Theor Appl Genet 69:535–541

Bolger A, Scossa F, Bolger ME, Lanz C, Maumus F, Tohge T, Quesneville H, Alseekh S, Sorensen I, Lichtenstein G et al (2014) The genome of the stress-tolerant wild tomato species *Solanum pennellii*. Nat Genet 46:1034–1038

Bombarely A, Rosli HG, Vrebalov J, Moffett P, Mueller L, Martin G (2012) A draft genome sequence of *Nicotiana benthamiana* to enhance molecular plant-microbe biology research. Mol. Plant Microbe Interact 25:1523–1530

Bombarely A, Moser M, Amrad A, Bao M, Bapaume L, Barry CS, Bliek M, Boersma MR, Borghi L, Bruggmann R et al (2016) Insight into the evolution of the Solanaceae from the parental genomes of *Petunia hybrida*. Nat Plants 2:16074

Chakrabarti M, Meekins KM, Gavilano LB, Siminszky B (2007) Inactivation of the cytochrome P450 gene CYP82E2 by degenerative mutations was a key event in the evolution of the alkaloid profile of modern tobacco. New Phytol 175:565–574

Chase MW, Knapp S, Cox AV, Clarkson JJ, Butsko Y, Joseph J, Savolainen V, Parokonny AS (2003) Molecular systematics, GISH and the origin of hybrid taxa in *Nicotiana* (Solanaceae). Ann Bot 92:107–127

Chen K, Dorlhac de Borne F, Szegedi E, Otten L (2014) Deep sequencing of the ancestral tobacco species *Nicotiana tomentosiformis* reveals multiple T-DNA inserts and a complex evolutionary history of natural transformation in the genus *Nicotiana*. Plant J 80:669–682

Chen K, Dorlhac de Borne F, Sierro N, Ivanov NV, Alouia M, Koechler S, Otten L (2018) Organization of the TC and TE cellular T-DNA regions in *Nicotiana otophora* and functional analysis of three diverged TE-6b genes. Plant J Cell Mol Biol 94:274–287

Clarkson JJ, Knapp S, Garcia VF, Olmstead RG, Leitch AR, Chase MW (2004) Phylogenetic

relationships in *Nicotiana* (Solanaceae) inferred from multiple plastid DNA regions. Mol Phylogenet Evol 33:75–90
Clarkson JJ, Lim KY, Kovarik A, Chase MW, Knapp S, Leitch AR (2005) Long-term genome diploidization in allopolyploid *Nicotiana* section Repandae (Solanaceae). New Phytol 168:241–252
Dadejova M, Lim KY, Souckova-Skalicka K, Matyasek R, Grandbastien MA, Leitch A, Kovarik A (2007) Transcription activity of rRNA genes correlates with a tendency towards intergenomic homogenization in *Nicotiana* allotetraploids. New Phytol 174:658–668
Darvishzadeh R, Mirzaei L, Maleki HH, Laurentin H, Alavi SR (2013) Genetic variation in oriental tobacco (*Nicotiana tabacum* L.) by agro-morphological traits and simple sequence repeat markers. Rev Ciênc Agronômica 44:347–355
Davalieva K, Maleva I, Filiposki K, Spiroski O, Efremov GD (2010) Genetic Variability of Macedonian tobacco varieties determined by microsatellite marker analysis. Diversity 2:439–449
Davis DL, Nielsen MT (1999) Tobacco : production, chemistry and technology. Blackwell Science Ltd
De Sutter V, Vanderhaeghen R, Tilleman S, Lammertyn F, Vanhoutte I, Karimi M, Inze D, Goossens A, Hilson P (2005) Exploration of jasmonate signalling via automated and standardized transient expression assays in tobacco cells. Plant J 44:1065–1076
Denoeud F, Carretero-Paulet L, Dereeper A, Droc G, Guyot R, Pietrella M, Zheng C, Alberti A, Anthony F, Aprea G et al (2014) The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. Science 345:1181–1184
Dewey RE, Xie J (2013) Molecular genetics of alkaloid biosynthesis in *Nicotiana tabacum*. Phytochemistry 94:10–27
Edwards KD, Fernandez-Pozo N, Drake-Stowe K, Humphry M, Evans AD, Bombarely A, Allen F, Hurst R, White B, Kernodle SP et al (2017) A reference genome for *Nicotiana tabacum* enables map-based cloning of homeologous loci implicated in nitrogen utilization efficiency. BMC Genom 18:448
Eich E (2008) Solanaceae and convolvulaceae: secondary metabolites: biosynthesis, chemotaxonomy, biological and economic significance (a handbook). Springer
Elliott PE, Lewis RS, Shew HD, Gutierrez WA, Nicholson JS (2008) Evaluation of tobacco germplasm for seedling resistance to stem rot and target spot caused by *Thanatephorus cucumeris*. Plant Dis 92:425–430
Fernandez-Pozo N, Menda N, Edwards JD, Saha S, Tecle IY, Strickler SR, Bombarely A, Fisher-York T, Pujar A, Foerster H et al (2015) The Sol Genomics Network (SGN)-from genotype to phenotype to breeding. Nucleic Acids Res 43:D1036–D1041
Fricano A, Bakaher N, Del Corvo M, Piffanelli P, Donini P, Stella A, Ivanov NV, Pozzi C (2012) Molecular diversity, population structure, and linkage disequilibrium in a worldwide collection of tobacco (*Nicotiana tabacum* L.) germplasm. BMC Genet 13:18
Gill BS (1991) Nucleocytoplasmic interaction (NCI) hypothesis of genome evolution and speciation in polyploid plants. In: Sasakuma T, Kinoshita T (eds) Proceedings of the Kihara Memorial International Symposium on Cytoplasmic Engineering in Wheat, pp 48–53
Goodin MM, Zaitlin D, Naidu RA, Lommel SA (2008) *Nicotiana benthamiana*: its history and future as a model for plant-pathogen interactions. Mol Plant Microbe Interact 21:1015–1026
Goodspeed TH (1947) On the evolution of the genus *Nicotiana*. Proc Natl Acad Sci USA 33:158–171
Goodspeed TH (1954) The genus *Nicotiana*. Chron Bot 16:1–536
Goossens A, Hakkinen ST, Laakso I, Seppanen-Laakso T, Biondi S, De Sutter V, Lammertyn F, Nuutila AM, Soderlund H, Zabeau M et al (2003) A functional genomics approach toward the understanding of secondary metabolism in plant cells. Proc Natl Acad Sci USA 100:8595–8600
Gregor W, Mette MF, Staginnus C, Matzke MA, Matzke AJM (2004) A distinct endogenous pararetrovirus family in *Nicotiana tomentosiformis*, a diploid progenitor of polyploid tobacco. Plant Physiol 134:1191–1199
Hakkinen ST, Tilleman S, Swiatek A, De Sutter V, Rischer H, Vanhoutte I, Van Onckelen H, Hilson P, Inze D, Oksman-Caldentey KM et al (2007) Functional characterisation of genes involved in pyridine alkaloid biosynthesis in tobacco. Phytochemistry 68:2773–2785
Harada E, Kim JA, Meyer AJ, Hell R, Clemens S, Choi YE (2010) Expression profiling of tobacco leaf trichomes identifies genes for biotic and abiotic stresses. Plant Cell Physiol 51:1627–1637
Hirakawa H, Shirasawa K, Miyatake K, Nunome T, Negoro S, Ohyama A, Yamaguchi H, Sato S, Isobe S, Tabata S et al (2014) Draft genome sequence of eggplant (*Solanum melongena* L.): the representative solanum species indigenous to the old world. DNA Res 21:649–660
Holmes FO (1938) Inheritance of resistance to tobacco-mosaic disease in tobacco. Phytopathology 28:553–561
Julio E, Cotucheau J, Decorps C, Volpatti R, Sentenac C, Candresse T, Dorlhac de Borne F (2014) A Eukaryotic translation initiation factor 4E (eIF4E) is responsible for the "va" tobacco recessive resistance to Potyviruses. Plant Mol Biol Report 1–15
Kelly LJ, Leitch AR, Clarkson JJ, Hunter RB, Knapp S, Chase MW (2010) Intragenic recombination events and evidence for hybrid speciation in *Nicotiana* (Solanaceae). Mol Biol Evol 27:781–799
Kelly LJ, Leitch AR, Clarkson JJ, Knapp S, Chase MW (2013) Reconstructing the complex evolutionary origin of wild allopolyploid tobaccos (*Nicotiana* section suaveolentes). Evolution 67:80–94
Kenton A, Parokonny AS, Gleba YY, Bennett MD (1993) Characterization of the *Nicotiana tabacum* L. genome by molecular cytogenetics. Mol Gen Genet 240: 159–169

Khafizova G, Dobrynin P, Polev D, Matveeva T (2018) *Nicotiana glauca* whole-genome investigation for cT-DNA study. BMC Res Notes 11:18

Kim S, Park M, Yeom SI, Kim YM, Lee JM, Lee HA, Seo E, Choi J, Cheong K, Kim KT et al (2014) Genome sequence of the hot pepper provides insights into the evolution of pungency in *Capsicum* species. Nat Genet 46:270–278

Kitamura S, Tanaka A, Inoue M (2005) Genomic relationships among *Nicotiana* species with different ploidy levels revealed by 5S rDNA spacer sequences and FISH/GISH. Genes Genet Syst 80:251–260

Knapp S, Bohs L, Nee M, Spooner DM (2004a) Solanaceae—a model for linking genomics with biodiversity. Comp Funct Genomics 5:285–291

Knapp S, Chase MW, Clarkson JJ (2004b) Nomenclatural changes and a new sectional classification in *Nicotiana* (Solanaceae). Taxon 53:73–82

Koukalova B, Moraes AP, Renny-Byfield S, Matyasek R, Leitch AR, Kovarik A (2010) Fall and rise of satellite repeats in allopolyploids of *Nicotiana* over c. 5 million years. New Phytol 186:148–160

Kovacova V, Zluvova J, Janousek B, Talianova M, Vyskot B (2014) The evolutionary fate of the horizontally transferred agrobacterial mikimopine synthase gene in the genera *Nicotiana* and *Linaria*. PLoS ONE 9:e113872

Kovarik A, Renny-Byfield S, Grandbastien M-A, Leitch A (2012) Evolutionary implications of genome and karyotype restructuring in *Nicotiana tabacum* L. In: Soltis PS, Soltis DE (eds) Polyploidy and genome evolution. Springer, Berlin Heidelberg, pp 209–224

Leitch IJ, Bennett MD (2004) Genome downsizing in polyploid plants. Biol J Linn Soc 82:651–663

Leitch IJ, Hanson L, Lim KY, Kovarik A, Chase MW, Clarkson JJ, Leitch AR (2008) The ups and downs of genome size evolution in polyploid species of *Nicotiana* (Solanaceae). Ann Bot 101:805–814

Lewis RS (2011) Nicotiana. Wild Crop Relat. Genomic Breed Resour 185–208

Lewis RS, Nicholson JS (2007) Aspects of the evolution of *Nicotiana tabacum* L. and the status of the United States *Nicotiana* germplasm collection. Genet Resour Crop Evol 54:727–740

Lewis RS, Milla SR, Levin JS (2005) Molecular and genetic characterization of *Nicotiana glutinosa* L. chromosome segments in tobacco mosaic virus-resistant tobacco accessions. Crop Sci 45:2355–2362

Lewis RS, Linger LR, Wolff MF, Wernsman EA (2007) The negative influence of N-mediated TMV resistance on yield in tobacco: linkage drag versus pleiotropy. Theor Appl Genet 115:169–178

Lim KY, Matyášek R, Lichtenstein CP, Leitch AR (2000) Molecular cytogenetic analyses and phylogenetic studies in the *Nicotiana* section Tomentosae. Chromosoma 109:245–258

Lim KY, Matyasek R, Kovarik A, Leitch AR (2004) Genome evolution in allotetraploid *Nicotiana*. Biol J Linn Soc 82:599–606

Lim KY, Kovarik A, Matyasek R, Chase MW, Clarkson JJ, Grandbastien MA, Leitch AR (2007) Sequence of events leading to near-complete genome turnover in allopolyploid *Nicotiana* within five million years. New Phytol 175:756–763

Luu VT, Schuck S, Kim SG, Weinhold A, Baldwin IT (2015) Jasmonic acid signalling mediates resistance of the wild tobacco *Nicotiana attenuata* to its native Fusarium, but not Alternaria, fungal pathogens. Plant Cell Env. 38:572–584

Marks CE (2010) Definition of South Pacific taxa of *Nicotiana* section Suaveolentes. Muelleria 28:74–84

Marks CE, Ladiges PY (2011) Comparative morphology and phylogeny of *Nicotiana* section Suaveolentes (Solanaceae) in Australia and the South Pacific. Aust Syst Bot 24:61–86

Marks CE, Ladiges PY, Newbigin E (2011) Karyotypic variation in *Nicotiana* section Suaveolentes. Genet Resour Crop Evol 58:797–803

Martin F, Bovet L, Cordier A, Stanke M, Gunduz I, Peitsch MC, Ivanov NV (2012) Design of a Tobacco Exon Array with application to investigate the differential cadmium accumulation property in two tobacco varieties. BMC Genom 13:674

Matyasek R, Fulnecek J, Leitch AR, Kovarik A (2011) Analysis of two abundant, highly related satellites in the allotetraploid *Nicotiana arentsii* using double-strand conformation polymorphism analysis and sequencing. New Phytol 192:747–759

Matyasek R, Renny-Byfield S, Fulneček J, Macas J, Grandbastien M-A, Nichols R, Leitch A, Kovařík A (2012) Next generation sequencing analysis reveals a relationship between rDNA unit diversity and locus number in *Nicotiana* diploids. BMC Genom 13:722

Melayah, D, Lim, KY, Bonnivard, E, Chalhoub, B, Dorlhac De Borne, F, Mhiri, C, Leitch, AR, Grandbastien, MA (2004). Distribution of the Tnt1 retrotransposon family in the amphidiploid tobacco (*Nicotiana tabacum*) and its wild Nicotiana relatives. Biol. J. Linn. Soc. *82*, 639–649

Moon H, Nicholson JS (2007) AFLP and SCAR markers linked to tomato spotted wilt virus resistance in tobacco. Crop Sci 47:1887–1894

Murad L, Lim KY, Christopodulou V, Matyasek R, Lichtenstein CP, Kovarik A, Leitch AR (2002) The origin of tobacco's T genome is traced to a particular lineage within *Nicotiana tomentosiformis* (Solanaceae). Am J Bot 89:921–928

Naim F, Nakasugi K, Crowhurst RN, Hilario E, Zwart AB, Hellens RP, Taylor JM, Waterhouse PM, Wood CC (2012) Advanced engineering of lipid metabolism in *Nicotiana benthamiana* using a draft genome and the V2 viral silencing-suppressor protein. PLoS ONE 7:e52717

Olmstead RG, Bohs L, Migid HA, Santiago-Valentin E, Garcia VF, Collier SM (2008) A molecular phylogeny of the Solanaceae. Taxon 57:1159–1181

Parisod C, Mhiri C, Lim KY, Clarkson JJ, Chase MW, Leitch AR, Grandbastien M-A (2012) Differential

dynamics of transposable elements during long-term diploidization of *Nicotiana* section Repandae (Solanaceae) allopolyploid genomes. PLoS ONE 7:e50352

Parokonny, A.S., and Kenton, A.Y. (1995). Comparative physical mapping and evolution of the *Nicotiana tabacum* L. karyotype. pp. 301–320

Petit M, Lim KY, Julio E, Poncet C, Dorlhac de Borne F, Kovarik A, Leitch AR, Grandbastien MA, Mhiri C (2007) Differential impact of retrotransposon populations on the genome of allotetraploid tobacco (*Nicotiana tabacum*). Mol Genet Genomics 278:1–15

Petit M, Guidat C, Daniel J, Denis E, Montoriol E, Bui QT, Lim KY, Kovarik A, Leitch AR, Grandbastien M-A (2010) Mobilization of retrotransposons in synthetic allotetraploid tobacco. New Phytol 186:135–147

Qin C, Yu C, Shen Y, Fang X, Chen L, Min J, Cheng J, Zhao S, Xu M, Luo Y et al (2014) Whole-genome sequencing of cultivated and wild peppers provides insights into *Capsicum* domestication and specialization. Proc Natl Acad Sci USA 111:5135–5140

Reed, S.M. (1991). Cytogenetic evolution and aneuploidy in *Nicotiana*. Chromosome Eng Plants Genet Breed Evol Part B Elsevier Dordr Neth 483–505

Ren N, Timko MP (2001) AFLP analysis of genetic polymorphism and evolutionary relationships among cultivated and wild *Nicotiana* species. Genome 44:559–571

Sarkinen T, Bohs L, Olmstead RG, Knapp S (2013) A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): a dated 1000-tip tree. BMC Evol Biol 13:214

Sato S, Tabata S, Hirakawa H, Asamizu E, Shirasawa K, Isobe S, Kaneko T, Nakamura Y, Shibata D, Egholm M (2012) The tomato genome sequence provides insights into fleshy fruit evolution. Nature 485:635–641

Senthil G, Liu H, Puram VG, Clark A, Stromberg A, Goodin MM (2005) Specific and common changes in *Nicotiana benthamiana* gene expression in response to infection by enveloped viruses. J Gen Virol 86:2615–2625

Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T, Zaita N, Chunwongse J, Obokata J, Yamaguchi-Shinozaki K (1986) The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. EMBO J 5:2043

Sierro N, Battey JND, Ouadi S, Bovet L, Goepfert S, Bakaher N, Peitsch MC, Ivanov NV (2013a) Reference genomes and transcriptomes of *Nicotiana sylvestris* and *Nicotiana tomentosiformis*. Genome Biol 14:R60

Sierro N, van Oeveren J, van Eijk MJ, Martin F, Stormo KE, Peitsch MC, Ivanov NV (2013b) Whole genome profiling physical map and ancestral annotation of tobacco hicks broadleaf. Plant J 75:880–889

Sierro N, Battey JND, Ouadi S, Bakaher N, Bovet L, Willig A, Goepfert S, Peitsch MC, Ivanov NV (2014) The tobacco genome sequence and its comparison with those of tomato and potato. Nat Commun 5:3833

Sierro N, Battey JND, Bovet L, Liedschulte V, Ouadi S, Thomas J, Broye H, Laparra H, Vuarnoz A, Lang G et al (2018) The impact of genome evolution on the allotetraploid *Nicotiana rustica*—an intriguing story of enhanced alkaloid production. BMC Genom 19:855

Skalicka K, Lim KY, Matyášek R, Koukalová B, Leitch AR, Kovařík A (2003) Rapid evolution of parental rDNA in a synthetic tobacco allotetraploid line. Am J Bot 90:988–996

Skalicka K, Lim KY, Matyasek R, Matzke M, Leitch AR, Kovarik A (2005) Preferential elimination of repeated DNA sequences from the paternal, *Nicotiana tomentosiformis* genome donor of a synthetic, allotetraploid tobacco. New Phytol 166:291–303

Sloan DB (2013) One ring to rule them all? Genome sequencing provides new insights into the 'master circle' model of plant mitochondrial DNA structure. New Phytol 200:978–985

Sugiyama Y, Watase Y, Nagase M, Makita N, Yagura S, Hirai A, Sugiura M (2005) The complete nucleotide sequence and multipartite organization of the tobacco mitochondrial genome: comparative analysis of mitochondrial genomes in higher plants. Mol Genet Genomics 272:603–615

Symon DE, Lepschi BJ (2007) A new status in *Nicotiana* (Solanaceae): *N. monoschizocarpa* (P. Horton) Symon & Lepschi. J. Adelaide Bot. Gard. 21:92

Tezuka T, Marubashi W (2012) Genes in S and T subgenomes are responsible for hybrid lethality in interspecific hybrids between *Nicotiana tabacum* and *Nicotiana occidentalis*. PLoS ONE 7:e36204

Tran HTM, Ramaraj T, Furtado A, Lee LS, Henry RJ (2018) Use of a draft genome of coffee (*Coffea arabica*) to identify SNPs associated with caffeine content. Plant Biotechnol J 16:1756–1766

TRI (2015). Tobacco Research Institute of the Chinese Academy of Agricultural Sciences

USDA (2016). National Plant Germplasm System

Whitelaw CA, Barbazuk WB, Pertea G, Chan AP, Cheung F, Lee Y, Zheng L, van Heeringen S, Karamycheva S, Bennetzen JL et al (2003) Enrichment of gene-coding sequences in maize by genome filtration. Science 302:2118–2120

Whitham S, Dinesh-Kumar SP, Choi D, Hehl R, Corr C, Baker B (1994) The product of the tobacco mosaic virus resistance gene N: Similarity to toll and the interleukin-1 receptor. Cell 78:1101–1115

Wikstrom N, Savolainen V, Chase MW (2001) Evolution of the angiosperms: calibrating the family tree. Proc Biol Sci 268:2211–2220

Wu F, Eannetta NT, Xu Y, Plieske J, Ganal M, Pozzi C, Bakaher N, Tanksley SD (2010) COSII genetic maps of two diploid *Nicotiana* species provide a detailed picture of synteny with tomato and insights into chromosome evolution in tetraploid *N. tabacum*. Theor Appl Genet 120:809–827

Xu X, Pan S, Cheng S, Zhang B, Mu D, Ni P, Zhang G, Yang S, Li R, Wang J (2011) Genome sequence and analysis of the tuber crop potato. Nature 475:189

Xu S, Brockmoller T, Navarro-Quezada A, Kuhl H, Gase K, Ling Z, Zhou W, Kreitzer C, Stanke M, Tang H et al (2017) Wild tobacco genomes reveal the evolution of nicotine biosynthesis. Proc Natl Acad Sci USA 114:6133–6138

Yang S-J, Carter SA, Cole AB, Cheng N-H, Nelson RS (2004) A natural variant of a host RNA-dependent RNA polymerase is associated with increased susceptibility to viruses by *Nicotiana benthamiana*. Proc Natl Acad Sci USA 101:6297–6302

Yang BC, Xiao BG, Chen XJ, Shi CH (2007) Assessing the genetic diversity of tobacco germplasm using intersimple sequence repeat and inter-retrotransposon amplification polymorphism markers. Ann Appl Biol 150:393–401

Yarnes SC, Ashrafi H, Reyes-Chin-Wo S, Hill TA, Stoffel KM, Van Deynze A (2013) Identification of QTLs for capsaicinoids, fruit quality, and plant architecture-related traits in an interspecific *Capsicum* RIL population. Genome 56:61–74

Yukawa M, Tsudzuki T, Sugiura M (2006) The chloroplast genome of *Nicotiana sylvestris* and *Nicotiana tomentosiformis*: complete sequencing confirms that the *Nicotiana sylvestris* progenitor is the maternal genome donor of *Nicotiana tabacum*. Mol Genet Genomics 275:367–373

2

Biodiversity of *Nicotiana* (Solanaceae)

Sandra Knapp

Abstract

This brief synopsis outlines current knowledge of the biodiversity of the genus *Nicotiana* L., the fifth largest genus in the family Solanaceae. Of the 82 species of *Nicotiana* currently recognized, most are from the New World (North, Central, and South America), but a significant radiation has occurred in arid-zone Australia (ca. 35 species). The genus is divided into 13 sections, the largest of which is section *Suaveolentes*, which comprises species from Africa, Australia, and the Pacific. The last complete monograph of the genus *Nicotiana* was published in the 1950s, but considerable work in the field has identified new species, particularly in Australia. The biology and phenotype of species of *Nicotiana* are reviewed, as are phylogeny and biogeography. Advances in knowledge of species-level diversity in *Nicotiana* are summarized and some personal priorities for future research in *Nicotiana* biodiversity are suggested. A table of all 82 *Nicotiana* species with distribution, chromosome number, and reference to a botanical description is provided, along with photographs of representative habits and flowers.

S. Knapp (✉)
Department of Life Sciences, Natural History Museum, Cromwell Road, London SW7 5BD, UK
e-mail: s.knapp@nhm.ac.uk

2.1 Introduction

Nicotiana L. is a member of the medium-sized flowering plant family Solanaceae that is replete with species used for food (*Solanum tuberosum* L., potato; *S. lycopersicum* L., tomato; *Capsicum* spp., peppers), ornament (*Petunia* L. spp., petunias; *Salpiglossis sinuata* Ruiz and Pav., velvet tongue), and drugs and medicines (*Brugmansia* Pers. spp, huanduj; *Hyoscyamus* L. spp., henbane; *Atropa belladonna* L., deadly nightshade). Several *Nicotiana* species fall into this last category (e.g., Symon 2005); one of them, *N. tabacum* L., is among the most widely used drug plants in many cultures. Solanaceae occur in a wide variety of habitats, from deserts to tropical rainforests to montane areas above treeline, and range in life form from tiny annual herbs to large canopy trees (Särkinen et al. 2013). Both generic and species diversities in the Solanaceae are concentrated in the New World tropics, and are highest in South America, with 83 of the 103 currently accepted genera of the family occurring there, and with 79 of these known only from the New World. Significant generic diversity is also found in Australia, where the seven endemic genera of the tribe Anthocercidae occur (see Dupin et al. 2016 and the section “Biogeography of *Nicotiana*” in this chapter for details). *Solanum* L. is the largest genus in the family, with 1,200–1,400 species. It accounts for approximately half of the species-level diversity in the family and is distributed globally. Other genera

© Springer Nature Switzerland AG 2020
N. V. Ivanov et al. (eds.), *The Tobacco Plant Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-030-29493-9_2

Table 2.1 The naturally occurring species of *Nicotiana* L. Infraspecific taxa recognized by various authors are not included, nor are scientifically named artificial hybrids produced in the laboratory. For botanical illustrations of most of these species, see Goodspeed (1954), Cocucci (2013), and the series of papers accompanying Chase et al. 2018. Online descriptions of Australian and Brazilian species can be found in floristic websites (e.g., http://www.flora.sa.gov.au; https://florabase.dpaw.wa.gov.au; http://floradobrasil.jbrj.gov.br)

Species	Section (Knapp et al. 2004)	Chromosome number (n)	Distribution	Reference to most recent description, (chromosome number notes)
Nicotiana azambujae L.B. Sm. and Downs	*Alatae* Goodsp.	Not known	Brazil	Smith and Downs (1964) (known only from the type specimen)
Nicotiana alata Link and Otto	*Alatae* Goodsp.	9	Argentina, Brazil, Paraguay, Uruguay	Cocucci (2013)
Nicotiana bonariensis Lehm.	*Alatae* Goodsp.	9	Argentina, Brazil, Uruguay	Cocucci (2013)
Nicotiana forgetiana Hemsl.	*Alatae* Goodsp.	9	Brazil	Cocucci (2013)
Nicotiana langsdorffii Weinm.	*Alatae* Goodsp.	9	Argentina, Brazil, Paraguay	Cocucci (2013)
Nicotiana longiflora Cav.	*Alatae* Goodsp.	10	Argentina, Bolivia, Brazil, Paraguay, Uruguay (adventive elsewhere, e.g., USA)	Cocucci (2013)
Nicotiana mutabilis Stehmann and Samir	*Alatae* Goodsp.	9	Brazil	Stehmann et al. (2002)
Nicotiana plumbaginifolia Viv.	*Alatae* Goodsp.	10	Argentina, Bolivia, Paraguay (adventive elsewhere as a weed, e.g., Mexico, Caribbean, India)	Knapp (in press)
Nicotiana tabacum L.	*Nicotiana*	24	Worldwide (cultivated)	Knapp (in press)
Nicotiana acaulis Speg.	*Noctiflorae* Goodsp.	12	Argentina	Cocucci (2013)
Nicotiana ameghinoi Speg.	*Noctiflorae* Goodsp.	12	Argentina	Cocucci (2013)
Nicotiana glauca Graham	*Noctiflorae* Goodsp.	12	Argentina, Bolivia (worldwide as an invasive weed)	Cocucci (2013)
Nicotiana noctiflora Hook.	*Noctiflorae* Goodsp.	12	Argentina	Cocucci (2013)
Nicotiana paa Mart.Crov.	*Noctiflorae* Goodsp.	Not known	Argentina, Chile	Cocucci (2013) (likely to be $n = 12$)
Nicotiana petunioides (Griseb.) Millán	*Noctiflorae* Goodsp.	12	Argentina, Chile	Cocucci (2013)
Nicotiana benavidesii Goodsp.	*Paniculatae* Goodsp.	12	Peru	Goodspeed (1954)
Nicotiana cordifolia Phil.	*Paniculatae* Goodsp.	12	Chile	Goodspeed (1954)
Nicotiana cutleri D'Arcy	*Paniculatae* Goodsp.	Not known	Bolivia	D'Arcy (1977) (probably $n = 12$)
Nicotiana knightiana Goodsp.	*Paniculatae* Goodsp.	12	Peru	Goodspeed (1954)
Nicotiana paniculata L.	*Paniculatae* Goodsp.	12	Peru	Goodspeed (1954)
Nicotiana raimondii J.F. Macbr.	*Paniculatae* Goodsp.	12	Peru	Goodspeed (1954)

(continued)

Table 2.1 (continued)

Species	Section (Knapp et al. 2004)	Chromosome number (n)	Distribution	Reference to most recent description, (chromosome number notes)
Nicotiana solanifolia Walp.	*Paniculatae* Goodsp.	12	Chile (Juan Fernández Islands)	Goodspeed (1954)
Nicotiana acuminata (Graham) Hook.	*Petunioides* G. Don	12	Argentina, Chile	Cocucci (2013)
Nicotiana attenuata S. Wats.	*Petunioides* G. Don	12	United States of America	Knapp (in press)
Nicotiana corymbosa J. Rémy	*Petunioides* G. Don	12	Argentina, Chile	Cocucci (2013)
Nicotiana linearis Phil.	*Petunioides* G. Don	12	Argentina, Chile	Cocucci (2013)
Nicotiana longibracteata Phil.	*Petunioides* G. Don	12	Argentina, Chile	Cocucci (2013)
Nicotiana miersii J. Rémy	*Petunioides* G. Don	12	Argentina	Cocucci (2013)
Nicotiana pauciflora J. Rémy	*Petunioides* G. Don	12	Chile	Goodspeed (1954)
Nicotiana spegazzinii Millán	*Petunioides* G. Don	12	Argentina	Cocucci (2013)
Nicotiana clevelandii A. Gray	*Polydicliae* G. Don	24	United States of America, Mexico	Knapp (in press)
Nicotiana quadrivalvis Pursh	*Polydicliae* G. Don	24	Canada, United States of America, Mexico (cultivated)	Knapp (in press)
Nicotiana nesophila I.M. Johnst.	*Repandae* Goodsp.	24	Mexico (Revillagigedo Islands)	Goodspeed (1954)
Nicotiana nudicaulis S. Wats.	*Repandae* Goodsp.	24	Mexico	Goodspeed (1954)
Nicotiana repanda Willd.	*Repandae* Goodsp.	24	Mexico, Cuba	Goodspeed (1954)
Nicotiana stocktonii Brandegee	*Repandae* Goodsp.	24	Mexico (Revillagigedo Islands)	Goodspeed (1954)
Nicotiana rustica L.	*Rusticae* G.Don	24	Worldwide (cultivated)	Knapp (in press)
Nicotiana africana Merxm.	*Suaveolentes* Goodsp.	23	Namibia	Merxmüller and Butler (1975)
Nicotiana amplexicaulis N.T. Burb.	*Suaveolentes* Goodsp.	18	Australia	Purdie et al. (1982)
Nicotiana benthamiana Domin.	*Suaveolentes* Goodsp.	19	Australia	Chase and Christenhusz (2018f)
Nicotiana burbidgeae Symon	*Suaveolentes* Goodsp.	21	Australia	Chase et al. (2018d)
Nicotiana cavicola N.T. Burb.	*Suaveolentes* Goodsp.	23; 20	Australia	Burbidge (1960); Williams (1975); Purdie et al. (1982)
Nicotiana excelsior (J.M. Black) J.M. Black	*Suaveolentes* Goodsp.	19	Australia	Chase and Christenhusz (2018c)
Nicotiana exigua H.-M. Wheeler	*Suaveolentes* Goodsp.	Not known	Australia	Purdie et al. (1982)

(continued)

Table 2.1 (continued)

Species	Section (Knapp et al. 2004)	Chromosome number (n)	Distribution	Reference to most recent description, (chromosome number notes)
Nicotiana fatuhivensis F. Br.	*Suaveolentes* Goodsp.	Not known	Marquesas Islands	Wagner and Lorence (2002) (presumed to be n = 24 by Marks et al. 2011)
Nicotiana faucicola M.W. Chase and Christenh.	*Suaveolentes* Goodsp.	Not known	Australia, New Caledonia	Chase et al. (2018c)
Nicotiana forsteri Roem. and Schult.	*Suaveolentes* Goodsp.	24	Australia	Purdie et al. (1982) (as *N. debneyi* Domin, now recognized as a synonym of *N. forsteri*)
Nicotiana fragrans Hook.	*Suaveolentes* Goodsp.	24	New Caledonia, Tongatapu	Goodspeed (1954)
Nicotiana gascoynica M.W. Chase and Christenh.	*Suaveolentes* Goodsp.	20	Australia	Chase and Christenhusz (2018b)
Nicotiana goodspeedii H.-M. Wheeler	*Suaveolentes* Goodsp.	16	Australia	Purdie et al. (1982)
Nicotiana gossei Domin	*Suaveolentes* Goodsp.	18	Australia	Chase and Christenhusz (2018d)
Nicotiana hesperis N.T. Burb.	*Suaveolentes* Goodsp.	21	Australia	Purdie et al. (1982)(treated as infraspecies of *N. occidentalis* by Chase et al. 2018a)
Nicotiana heterantha Kenneally and Symon	*Suaveolentes* Goodsp.	Not known	Australia	Kenneally and Symon (19941994)
Nicotiana ingulba J.M. Black	*Suaveolentes* Goodsp.	20	Australia	Purdie et al. (1982) (sometimes treated as infraspecies of *N. rosulata*)
Nicotiana karijini M.W. Chase and Christenh.	*Suaveolentes* Goodsp.	Not known	Australia	Chase and Christenhusz (2018e)
Nicotiana maritima H.-M. Wheeler	*Suaveolentes* Goodsp.	16	Australia	Chase et al. (2018e)
Nicotiana megalosiphon Van Huerck and Müll.-Arg.	*Suaveolentes* Goodsp.	20	Australia	Purdie et al. (1982)
Nicotiana monoschizocarpa (P.Horton) Symon and Lepschi	*Suaveolentes* Goodsp.	24	Australia	Symon and Lepschi (2007)
Nicotiana occidentalis H.-M. Wheeler	*Suaveolentes* Goodsp.	21	Australia	Purdie et al. (1982); Chase and Christenhusz (2018g) (subsp. *obliqua* only)
Nicotiana rosulata (S.Moore) Domin	*Suaveolentes* Goodsp.	20	Australia	Purdie et al. (1982)
Nicotiana rotundifolia Lindl.	*Suaveolentes* Goodsp.	22	Australia	Purdie et al. (1982)
Nicotiana simulans N.T. Burb.	*Suaveolentes* Goodsp.	20	Australia	Purdie et al. (1982)
Nicotiana stenocarpa H.-M. Wheeler	*Suaveolentes* Goodsp.	20	Australia	Chase and Christenhusz (2018h)
Nicotiana truncata Symon	*Suaveolentes* Goodsp.	Not known	Australia	Symon (1984)

(continued)

Table 2.1 (continued)

Species	Section (Knapp et al. 2004)	Chromosome number (n)	Distribution	Reference to most recent description, (chromosome number notes)
Nicotiana suaveolens Lehm.	*Suaveolentes* Goodsp.	15	Australia	Chase and Christenhusz (2018h)
Nicotiana umbratica N.T. Burb.	*Suaveolentes* Goodsp.	23	Australia	Purdie et al. (1982)
Nicotiana velutina H.-M. Wheeler	*Suaveolentes* Goodsp.	16	Australia	Purdie et al. (1982)
Nicotiana wuttkei Clarkson and Symon	*Suaveolentes* Goodsp.	16	Australia	Clarkson and Symon (1991)
Nicotiana yandinga M.W. Chase and Christenh.	*Suaveolentes* Goodsp.	21	Australia	Chase et al. (2018b)
Nicotiana sylvestris Speg.	*Sylvestres* S. Knapp	12	Argentina, Bolivia	Cocucci (2013)
Nicotiana kawakamii Y. Ohashi	*Tomentosae* Goodsp.	12	Bolivia	Ohashi (1985)
Nicotiana leguiana J.F. Macbr.	*Tomentosae* Goodsp.	12	Bolivia	Macbride (1930) (synonym of *N. tomentosa* in Goodspeed 1954)
Nicotiana otophora Griseb.	*Tomentosae* Goodsp.	12	Argentina, Bolivia	Cocucci (2013)
Nicotiana setchellii Goodsp.	*Tomentosae* Goodsp.	12	Peru	Goodspeed (1954)
Nicotiana tomentosa Ruiz and Pav.	*Tomentosae* Goodsp.	12	Bolivia, Ecuador, Peru	Goodspeed (1954)
Nicotiana tomentosiformis Goodsp.	*Tomentosae* Goodsp.	12	Bolivia	Goodspeed (1954)
Nicotiana obtusifolia M. Martens and Galeottii	*Trigonophyllae* Goodsp.	12	Mexico, United States of America	Knapp (in press)
Nicotiana arentsii Goodsp.	*Undulatae* Goodsp.	24	Bolivia	Goodspeed (1954)
Nicotiana glutinosa L.	*Undulatae* Goodsp.	12	Bolivia, Ecuador, Peru	Goodspeed (1954)
Nicotiana thrysiflora Goodsp.	*Undulatae* Goodsp.	12	Peru	Goodspeed (1954)
Nicotiana undulata Ruiz and Pav.	*Undulatae* Goodsp.	12	Argentina, Bolivia, Peru	Goodspeed (1954)
Nicotiana wigandioides Koch and Fintelm.	*Undulatae* Goodsp.	12	Bolivia	Goodspeed (1954)

in the Solanaceae such as *Lycium* L., *Lycianthes* Hassl., and *Nicotiana*, also have species outside as well as within the New World.

Nicotiana, with 82 currently recognized species (but several still awaiting description, see section "Species diversity of *Nicotiana* and future prospects" in this chapter for details), is the fifth largest genus in the family Solanaceae, and *Nicotiana* species are distributed in the Americas and Australia, with a single taxon occurring in Africa and another three taxa in the Pacific (see Table 2.1). The two primary cultivated species, *N. tabacum* and *N. rustica* L., have a worldwide distribution in cultivation, but both are of Andean origin. *Nicotiana* was the subject of a worldwide monograph by Goodspeed

(1954), and relationships among its species and sections have been studied more recently using a combination of molecular and morphological methods (e.g., Aoki and Ito 2000; Chase et al. 2003; Knapp et al. 2004; Clarkson et al. 2005). Goodspeed (1954) divided the genus into a number of sections whose limits and composition were tested, and the subsequently revised classification of Knapp et al. (2004) is followed here. Polyploidization and chromosome number reduction have been common genomic events in *Nicotiana* (e.g., Kenton et al. 1993; Lim et al. 2006), and the genus has been used as a model for studying genomic rearrangements after polyploidy events (Leitch et al. 2008). The availability of whole genome sequences of several *Nicotiana* species (e.g., Bombarely et al. 2012; Sierro et al. 2013, 2014; Edwards et al. 2017; Fernie and Usadel 2020) will allow chromosome number and polyploid age (Chase et al. 2003; Clarkson et al. 2005) to be investigated more fully, and new insights into the evolution of the great diversity in both morphology and molecular patterns in these plants are now within reach.

2.2 Biology and Phenotype of *Nicotiana*

The 82 currently recognized (see Table 2.1) species of *Nicotiana* are extremely varied in life form, leaf shape and size, and flower morphology. Species of *Nicotiana* range from annual herbs (e.g., *N. linearis* Phil., *N. plumbaginifolia* Viv.) to small soft-wooded trees (e.g., *N. tomentosa* Ruiz and Pav., *N. wigandioides* Koch and Fintelm.) Most taxa grow in wet or seasonally wet environments (Fig. 2.1), and are often ephemeral (e.g., *N. attenuata* S.Wats. in the Great Basin, Knapp in press; or the species of the Australian deserts in section *Suaveolentes*, Chase et al. 2018a). The leaves are typically large, membranous, and covered with sticky glandular hairs; these features give the plants their distinctive odors and viscid feel. Some species, however, have very small leaves (e.g., *N. linearis*). *Nicotiana glauca* Graham and some of its relatives (sect. *Noctiflorae*, see Table 2.1) lack copious glands and have rubbery and somewhat waxy or glaucous leaves. Specialist and generalist lepidopteran and other insect herbivores feed on the leaves of *Nicotiana*, despite their high alkaloid content and toxicity (e.g., Baldwin 2001; Kessler and Baldwin 2001; Wu et al. 2008; Adam et al. 2018). Coyote tobacco, *N. attenuata*, has become a model system for studying the evolution of plant defense responses in natural populations through the investigations of Baldwin and colleagues of this plant in the deserts of the Great Basin in Utah (Kessler et al. 2008, 2015; Diezel et al. 2011; Quezada et al. (2020), Chap. 13).

Nicotiana flowers are tubular and produce copious nectar at the base as a reward for pollinators (Goodspeed 1954). In general, flowers open either diurnally and are pollinated by day-flying moths, butterflies, hummingbirds, and/or bees, or at night (crepuscular) and are pollinated by night-flying moths (Aigner and Scott 2002; Kaczorowski et al. 2005; Kessler and Baldwin 2007; Nattero and Cocucci 2007). Some species have flowers that open at night and then remain open during the day for at least some time, allowing a mixed pollination system (e.g., *N. attenuata*, Kessler et al. 2010). Flower odor can be quite strong, especially in night-blooming species, but the diurnal Australian *N. umbratica* N.T. Burb. has flowers with a strong odor of cloves (Chase and Christenhusz 2018e). Flower odor is not routinely characterized in taxonomic treatments, but could prove evolutionarily interesting (e.g., Haverkamp et al. 2018). Despite all *Nicotiana* flowers being tubular to some degree, they all have a limb (spreading petals of the corolla) that is more or less pentagonal, but because of differences in the length and width of the corolla tube they have a relatively wide variety of overall shapes. Many night-flowering species (e.g., *N. longiflora* Cav., *N. repanda* Willd., *N. sylvestris* Speg.) have very long, narrow tubes (ca. 1 mm in diameter in some species), while others (e.g., *N. obtusifolia* M. Martens and Galeotti, *N. rustica*, *N. wigandioides*) have shorter, broader tubes (McCarthy et al. 2016). Allopolyploid species (see the next

Fig. 2.1 Representative habits and habitats of *Nicotiana* species. **A**. *Nicotiana longiflora* in the swampy regions of the Paraguayan humid Chaco (Peña-Chocarro et al. 1488); **B**. *Nicotiana glauca* in stream bed in northern Argentina (Knapp et al. IM-10191); **C**. *Nicotiana raimondii* in the dry valleys of southern Peru (Knapp et al. 10361); **D**. *Nicotiana noctiflora* in open rocky plains in the high Andes of western central Argentina (Knapp et al. 10469); **E**. *Nicotiana tomentosa* in clouds of the southern Peruvian Andes (Knapp et al. 10330); **F**. *Nicotiana linearis* in Andean foothills of central Argentina (Barboza et al. 3781); **G**. *Nicotiana ameghinoi* in rocky plains of Patagonia (Barboza et al. 3707); **H**. *Nicotiana gossei* growing in rock fissures, central Australia (Dodsworth s. n.). Photographs A-G, S. Knapp; H S. Dodsworth

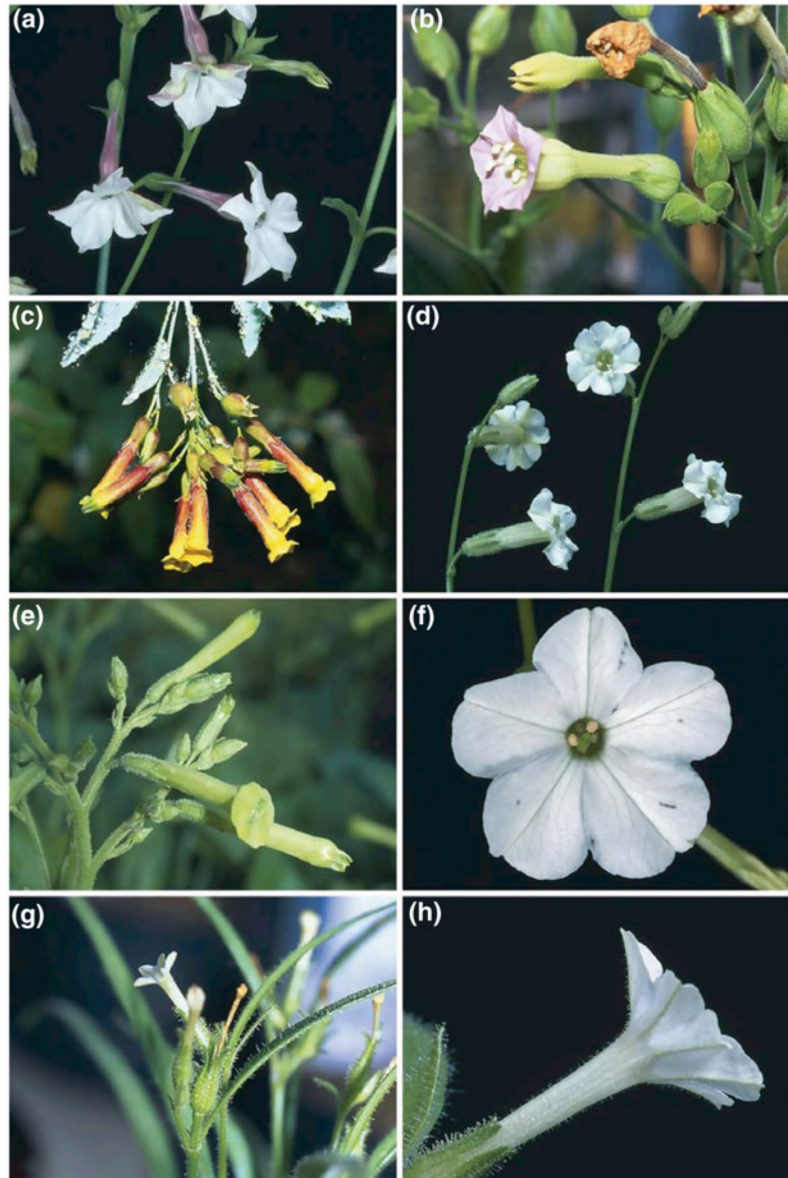

Fig. 2.2 Floral diversity in *Nicotiana*. **A**. Section *Alatae*, *Nicotiana bonariensis* (cult. RBG Kew); **B**. Section *Nicotiana*, *Nicotiana tabacum* (cult. Radboud University, acc. # A04750053); **C**. Section *Noctiflorae*, *Nicotiana glauca* (Bolivia; Nee et al. 51725); **D**. Section *Noctiflorae*, *Nicotiana noctiflora* (cult. RBG Kew); **E**. Section *Paniculatae*, *Nicotiana paniculata* (cult. RBG Kew); **F**. Section *Petunioides*, *Nicotiana attenuata* (cult. RBG Kew); **G**. Section *Petunioides*, *Nicotiana linearis* (cult. Radboud University, acc. # 96450099); **H**. Section *Polydicliae*, *Nicotiana clevelandii* (cult. RBG Kew). All voucher details can be found in Chase et al. 2003 and Clarkson et al. 2005 (reproduced with permission of the International Association for Plant Taxonomy from Knapp et al. 2004)

paragraph for details) tend to have flowers with shorter, broader tubes than their progenitors, suggesting that they are transgressively evolving to a more generalist pollination syndrome (McCarthy et al. 2016).

Human-perceived flower color in *Nicotiana* varies from clear white to cream and varying shades of pink or red, and some species have green or yellow flowers (Goodspeed 1954; McCarthy et al. 2015; Table 2.1 and see Figs. 2.2 and 2.3). Many species are tinged with grayish purple or green owing to the presence of chlorophyll in petal cells. Pollinators, of course, see flower color differently, and McCarthy et al. (2015) characterized *Nicotiana* flowers using spectral reflectance, and then used the obtained values to examine flower color in bee and hummingbird color space, which is more realistic than using human-perceived colors. They found that evolution of floral color (including spectral

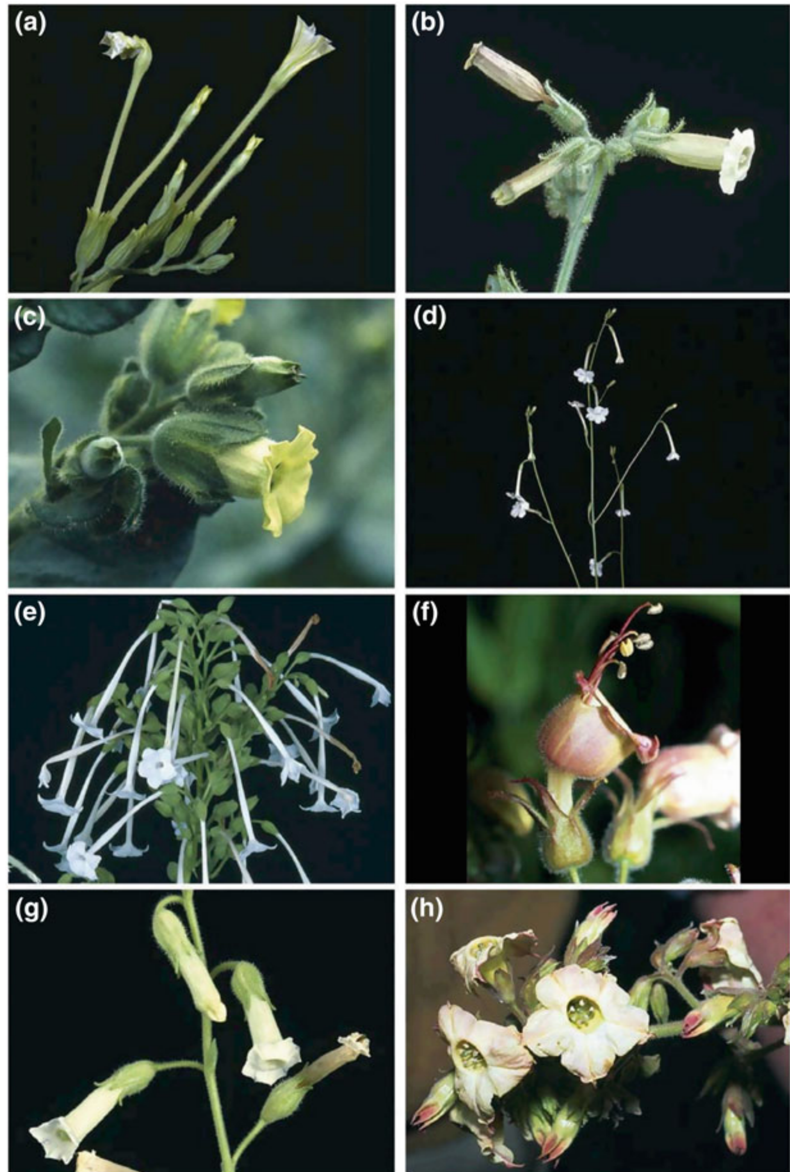

Fig. 2.3 Floral diversity in *Nicotiana*. **A**. Section *Repandae*, *Nicotiana nesophila* (cult. RBG Kew); **B**. Section *Repandae*, *Nicotiana nudicaulis* (cult. RBG Kew); **C**. Section *Rusticae*, *Nicotiana rustica* (cult. RBG Kew); **D**. Section *Suaveolentes*, *Nicotiana suaveolens* (cult. RBG Kew); **E**. Section *Sylvestres*, *Nicotiana sylvestris* (cult. RBG Kew); **F**. Section *Tomentosae*, *Nicotiana otophora* (Bolivia, Nee et al. 51727); **G**. Section *Trigonophyllae*, *Nicotiana obtusifolia* (cult. RBG Kew); **H**. Section *Undulatae*, *Nicotiana wigandioides* (Bolivia, Nee et al. 51764). All voucher details can be found in Chase et al. 2003 and Clarkson et al. 2005 (reproduced with permission of the International Association for Plant Taxonomy from Knapp et al. 2004)

reflectance) in *Nicotiana* is not entirely phylogenetically constrained, and that multiple independent origins of different pigment combinations occurred, especially in young allopolyploids such as *N. tabacum* where unexpected colors occur (McCarthy et al. 2016). Flower color in *Nicotiana* is due largely to flavonol and anthocyanin pigments (McCarthy et al. 2016), and the genetics of these pigments in *N. tabacum* is reviewed by McCarthy et al. (2020), Chap. 11). Chlorophyll is also present and contributes to color and hue in many species (McCarthy et al. 2015). Ancestral state reconstruction of parental flower color of polyploids (McCarthy et al. 2019) has shown that patterns of divergence are heavily influenced by early events in polyploidization (see the next paragraph for details). Pollinator selection in sympatry has

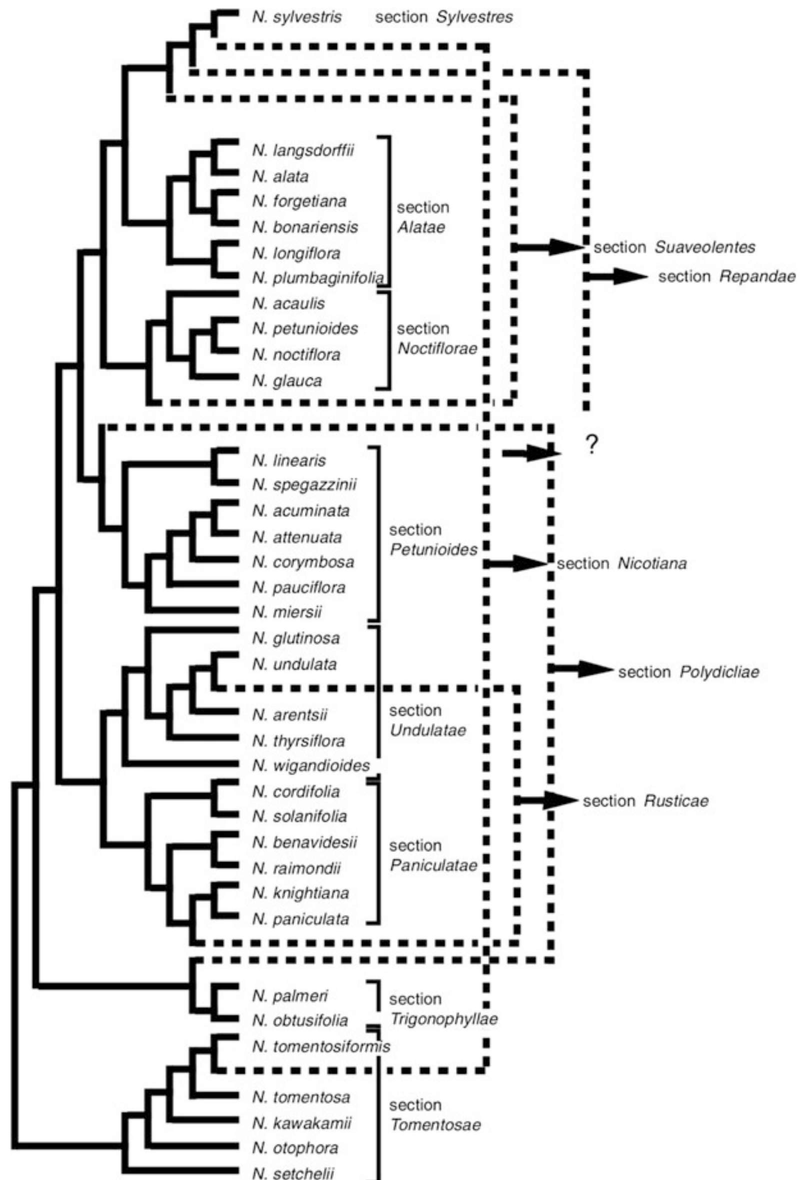

Fig. 2.4 Heuristic cladogram summary of evolutionary relationships in *Nicotiana*. The question mark (?) representing the unknown origin of the members of section *Repandae* has been resolved (Clarkson et al. 2005; Table 2.2). The progenitor lineages of section *Repandae* are those that currently comprise section *Sylvestres* and section *Trigonophyllae*. For information on data sources and methods see references cited in the text (reproduced with permission of the International Association for Plant Taxonomy from Knapp et al. 2004)

almost certainly influenced flower color in *Nicotiana*, but field experiments on most species in their native environments remain to be conducted.

Nicotiana is one of several lineages in the Solanaceae that has a large number of allopolyploid species (see Chase et al. 2003; other solanaceous polyploid hotspots are in *Solanum* [potatoes, Spooner et al. 2014; the black nightshades, Särkinen et al. 2018] and the Australian tribe Anthocercidae [Hunziker 2001]). Within the genus, allopolyploidy has occurred several times independently, with subsequent evolution at the polyploid level in several lineages (e.g., sections *Polydicliae*, *Repandae*, *Suaveolentes*; see Fig. 2.4). Allopolyploids of varying ages occur in the genus (Chase et al. 2003; Clarkson et al. 2004; Clarkson et al. 2017), from young (e.g. *N. tabacum* <600 Kya) to intermediate (e.g., *N. quadrivalvis* Pursh, ca. 1.2 Mya) to relatively

ancient (the mainly Australian section *Suaveolentes*, ca. 6 Mya). Rounds of polyploidization in angiosperms are generally followed by genome downsizing and diploidization (Leitch and Bennett 2004; Leitch et al. 2008; Dodsworth et al. 2016), and *Nicotiana* is no exception, with genomic in situ hybridization able to distinguish parental genomes in young allopolyploids such as *N. tabacum*, *N. rustica*, and *N. arentsii* Goodsp. (Kenton et al. 1993; Lim et al. 2004) but unable to distinguish progenitor genomic contributions in older polyploid lineages (Clarkson et al. 2005). Lim et al. (2007) suggested that the inability of genomic in situ hybridization (GISH) to distinguish progenitor genomes in lineages >5 Mya was due to near-complete genomic turnover within a conserved karyotype structure involving the replacement of nongenic sequences with new or previously rare sequence types. In naturally occurring young allopolyploids (*N. tabacum* and *N. rustica*), rDNA repeat change is thought to be the result of concerted evolution and the elimination of either maternal or paternal sequences (Kovarik et al. 2004; Skalicka et al. 2005). Clarkson et al. (2017) used time-calibrated phylogenetic trees to further examine the effect of polyploidy age on diversification, and showed that the oldest of the allopolyploid lineages in the genus, the mainly Australian section *Suaveolentes* (see Table 2.1), showed a significant lag between putative initial polyploidization and diversification. Diversification in this lineage has involved significant genome rearrangement and diploidization through chromosome fusion, perhaps involving strong selective sweeps (Clarkson et al. 2017; Chase et al. 2018a). Diploid species of section *Alatae* (one of the putative progenitors of the Australian taxa, see Goodspeed 1954; Clarkson et al. 2004; Clarkson et al. 2017) have also undergone chromosome number reduction (Lim et al. 2006), suggesting that these types of genomic rearrangements are not confined to polyploid species. New genomic tools, such as those outlined in this volume, will allow a better understanding of these patterns and their origins.

2.3 Phylogeny and Relationships of *Nicotiana*

Goodspeed (1954) suggested that the genus *Nicotiana* was related to the genera *Cestrum* and *Petunia*, and that an "ancestral reservoir" of diversity led to "primitive" pre-*Cestrum*, a more "advanced" pre-*Petunia*, and an "intermediate" pre-*Nicotiana* that combined characters of the other two genera. He postulated that the base chromosome number of the group was $n = 6$, and that *Nicotiana* was the result of an allopolyploid event between pre-*Cestrum* and pre-*Petunia*. His concept hypothesized the existence of two nuclei of species—"Petunioid" (such as *N. attenuata*, *N. noctiflora* Hook., *N. sylvestris*) and "Cestroid" (such as *N. paniculata* L., *N. glutinosa* L., *N. tomentosa* Ruiz and Pav.) (see Fig. 2.1 in Chase et al. 2003). He divided the genus into sections based mostly on floral morphology and chromosome number. The use of DNA sequence data in phylogenetic reconstruction of Solanaceae relationships, however, showed that *Nicotiana* was not closely related to either of these genera, but instead was part of a previously unidentified and strongly supported group with a synapomorphy of a base chromosome number of $n = 12$, characterized as the "$X = 12$" clade (Olmstead and Palmer 1992; Olmstead and Sweere 1994). Olmstead et al. (1999) recovered the endemic Australian genera of the tribe Anthocercidae (sensu Hunziker 2001) as the sister group of *Nicotiana*, and together these two lineages formed the provisional subfamily Nicotianioidae. Subsequent phylogenetic analyses of the family using more taxa and more markers confirmed these results (Olmstead et al. 2008; Särkinen et al. 2013).

The first subgeneric classification of *Nicotiana* was proposed by George Don (1838) who recognized four "sections": *Tabacum*, *Rustica*, *Petunioides*, and *Polydiclia*. Goodspeed (1954) elevated all but *Polydiclia* to subgeneric status and divided the genus into 14 sections, some of which were exclusively polyploid. Previously he had included polyploid species in otherwise

diploid sectional groupings (Goodspeed 1945). Knapp et al. (2004) used the results of molecular phylogenetic studies (Chase et al. 2003) and the rules of botanical nomenclature to revise Goodspeed's 1954 classification, and recognized 13 sections (see Table 2.1). Most of these correspond to Goodspeed's groupings, but four of the diploid species were found to belong to sections other than those into which Goodspeed had placed them. *Nicotiana glutinosa* was treated by Goodspeed as a member of his section *Tomentosae* and *N. thyrsiflora* Goodsp. was the sole member of his section *Thyrsiflorae*; both these taxa are members of the monophyletic group section *Undulatae* (Knapp et al. 2004). Because of its yellow flowers, Goodspeed (1954) included *N. glauca* in section *Paniculatae* (although with reservation), but Knapp et al. (2004) included it in section *Noctiflorae*, with which it shares many vegetative characteristics. *Nicotiana sylvestris* was placed in section *Alatae* (Goodspeed 1954), but molecular phylogenetic results showed that it was not well supported as a member of that or any other diploid group, but was instead sister to section *Alatae* and was an isolated lineage (Chase et al. 2003; Clarkson et al. 2017; McCarthy et al. 2019) that has been involved in many of the allopolyploid events in *Nicotiana* (see Table 2.2).

Internal relationships within *Nicotiana* are characterized by reticulate evolution through hybridization, through both allopolyploidy (e.g., Aoki and Ito 2000; Chase et al. 2003) and homoploid hybridization (Kelly et al. 2010). Chase et al. (2003); and subsequent authors treated groups of diploids and polyploids separately, acknowledging the reticulate origin of the latter. Relationships between and within the diploid sections have not been explicitly examined, but most phylogenetic reconstructions suggest that section *Tomentosae* (Clarkson et al. 2005) or sections *Tomentosae* plus *Trigonophyllae* is sister to the rest of the genus (e.g., Clarkson et al. 2017; McCarthy et al. 2019). Phylogenetic work in *Nicotiana* has focused on higher level rather than species-level relationships.

Progenitors for the various allopolyploid lineages in *Nicotiana* have been elucidated using a variety of cytogenetic methods, including genomic in situ hybridization (GISH) and molecular sequence data (e.g., Kenton et al. 1993; Clarkson et al. 2005; Kelly et al. 2013; Dodsworth et al. 2017; see Table 2.2). As reviewed in the section "Biology and phenotype of *Nicotiana*" in this

Table 2.2 Genomic origins of allopolyploid lineages in *Nicotiana*

Polyploid lineage/species	Maternal genome	Paternal genome	Progenitors now sympatric?	Reference
Nicotiana arentsii	*N. undulata*	*N. wigandioides*	Yes	Lim et al. (2004)
Nicotiana rustica (section *Rusticae*)	*N. paniculata*	*N. undulata*	No	Lim et al. (2004)
Nicotiana tabacum (section *Nicotiana*)	*N. sylvestris*	*N. tomentosiformis*	No	Kenton et al. (1993)
Nicotiana section *Repandae*	*N. sylvestris*	*N. obtusifolia*	No	Clarkson et al. (2005)
Nicotiana section *Polydicliae*	*N. obtusifolia*	*N. attenuata*	Yes	Anssour et al. (2009)
Nicotiana section *Suaveolentes*	Section *Noctiflorae*/*Petunioides*, probably of hybrid origin	*N. sylvestris*	Uncertain (see text)	Kelly et al. (2013); Clarkson et al. (2017)

Note although species names are used here as progenitors, it is clear that these extant taxa are not the same as those involved in the original polyploidy events; see Chase et al. 2003. For sectional membership, see Table 2.1

chapter, the ability to distinguish progenitor genomes becomes more difficult with event age, and the identification of the closest extant relatives of the members of section *Suaveolentes*, the oldest of the *Nicotiana* polyploid lineages, is not yet completely clear (Clarkson et al. 2017; Chase et al. 2018a).

Diversification at the diploid level has occurred at a relatively low level in *Nicotiana*. Sections *Alatae*, *Paniculatae*, and *Petunioides* have eight species each, a far cry from the approximately 35 species of the monophyletic allopolyploid section *Suaveolentes* that occur in Australia. Goodspeed (1954) suggested that the large number of species in Australia was the result of several different polyploidy events, but all phylogenetic results to date suggest that a single event gave rise to the group, and extensive speciation at the polyploid level followed (Clarkson et al. 2017).

2.4 Biogeography of *Nicotiana*

Nicotiana, like the other genera of the family Solanaceae, is a Gondwanan lineage, with its diversity and origins centered on the Southern Hemisphere (see Crisp et al. 2009). The timing of the split between the first branching lineages of Solanaceae and the monophyletic "$X = 12$" clade to which *Nicotiana* belongs is approximately 24 Mya (28.4–23 Mya depending on the model used; Särkinen et al. 2013). Beautifully preserved Eocene (ca. 52 Mya) fossils of solanaceous inflated calyces (Wilf et al. 2017) found recently in Patagonia (Argentina) suggest an earlier date for this split, and analysis of the position of these "lantern fruit" fossils from Argentina is currently a priority for dating the Solanaceae family tree (Deanna et al. 2019; R. Deanna personal communication). The position of the continents at the Eocene or Miocene means that vicariance is less likely to account for the distribution of clades in the "$X = 12$" clade and that at least two distinct long-distance dispersal events gave rise to the genera of the tribe Anthocercidae (sister group to *Nicotiana*) and the diverse Australian *Nicotiana* section *Suaveolentes* (Olmstead et al. 2008).

Within the New World, the sections of *Nicotiana* have distinct non-overlapping centers of species diversity (see Table 2.1). Species of section *Alatae* occur in the grassland habitats of the Southern Cone, other groups are concentrated in the Andes (e.g., sections *Tomentosae*, *Paniculatae*, and *Undulatae*), and others occur in the arid zones of North America (e.g., sections *Repandae* and *Polydicliae*). Members of section *Noctiflorae* occur in Argentina and adjacent Chile often at high elevations on either side of the Andean cordillera, and members of section *Petunioides* have disjunct distributions in South and North America.

The split between the Southern Cone and Andean lineages and subsequent diversification of the Andean lineages is suggested to have been driven by the uplift and expansion of the Andean mountain range (Chase et al. 2018a); this is a common pattern in Andean-centered groups (e.g., Hughes and Eastwood 2006; Lagomarsino et al. 2016). The species of these Andean groups often occur in isolated dry valleys, suggesting that evolution in these groups may have occurred in isolated dry forest "islands" (e.g., Särkinen et al. 2012). This isolation may explain the strong geographic differentiation in flower morphology in *N. glutinosa* along its extensive range from southern Ecuador to northern Bolivia (Goodspeed 1954). No explicit biogeographic hypotheses have been tested in these groups of *Nicotiana*. It is interesting to note that the extant species that have descended from the putative progenitors of allopolyploid lineages in the Andes are no longer sympatric, suggesting that range contraction of lineages has occurred.

Members of section *Noctiflorae* are all more or less sympatric across Patagonia and the southern Andean foothills, with *N. paa* Mart. Crov. and *N. petunioides* also occurring in adjacent Andean Chile. The other largely southern South American lineage, section *Petunioides*, is more broadly distributed in Patagonia and coastal Chile and has a single member in North America (*N. attenuata*) that is sister to *N. acuminata* (Graham) Hook. from coastal Chile (Clarkson et al. 2017). *Nicotiana acuminata* is thus a member of a monophyletic group that is

otherwise confined to southern South America; the section has an amphitropical distribution pattern (Raven 1963; Simpson et al. 2017). This suggests a long-distance dispersal from south to north, in contrast to the prevailing pattern of north-to-south dispersal seen in other flowering plant groups (Wen and Ickert-Bond 2009). The Chilean species *N. acuminata* has dispersed and become established in Mediterranean climates on the west coast of the United States of America (Goodspeed 1954; Knapp in press).

The genome of the extant North American diploid *N. obtusifolia* is present in both of the allopolyploid lineages in North America (see Table 2.2). The two species that comprise section *Polydicliae* (*N. clevelandii* A.Gray and *N. quadrivalvis*) are sympatric in western North America, but the native distribution of *N. quadrivalvis* is hard to discern because of its use and spread by American First Peoples prior to European colonization. Section *Repandae* contains four taxa, two of which (*N. repanda* Willd. and *N. nudicaulis* S.Wats.) are found in northeastern Mexico and two (*N. nesophila* I.M. Johnst. and *N. stocktonii* I.M.Johnst.) that are endemic to the recently formed (<2 Mya) volcanic Revillagigedo Islands off the western coast of Mexico, where they certainly arrived via a long-distance dispersal event (Clarkson et al. 2005). *Nicotiana nudicaulis* is sister to the other three species but differs from them in floral morphology (McCarthy et al. 2016). The molecular sequences of *N. repanda* and the two Revillagigedo endemic species are so similar that timings for this dispersal event have not yet been determined (Clarkson et al. 2005). Mummenhoff and Franzke (2007) suggested that the small seeds of *Nicotiana* predispose them to long-distance dispersal by birds, and Chase et al. (2018a) suggested that the same might be true for seeds that could adhere to bats.

The only group in *Nicotiana* for which formal biogeographic analysis has been undertaken is the Australian section *Suaveolentes* (Ladiges et al. 2011). Ladiges et al. (2011) suggested that diversification of *Nicotiana* occurred in the central arid zone of the continent (the Eremean zone) via in situ evolution and vicariance through repeated cycles of drying throughout the Miocene, and migration and exchange between wetter and drier regions (Crisp et al. 1999). The putative progenitors of this group (see Table 2.1 and Kelly et al. 2013) are from dry areas in South America and, given that strong biome conservatism exists on a global scale (Crisp et al. 2009), it is perhaps not surprising that arid zones feature heavily in the diversification of section *Suaveolentes* after arrival in Australia. Aridification of the central part of Australia began in the Middle Miocene (10–7 Mya) and continued through the climatic oscillations of the Pleistocene (ca. 2.5 Mya to 11,700 years ago) (Byrne et al. 2008). The species in lineages along the earliest branches in section *Suaveolentes* occur in wetter habitats at the periphery of the continent, and reduction in chromosome numbers occurred as species evolved to fill niches in the drying central Eremean zone. Clarkson et al. (2017) concur with this general hypothesis, but suggest that the timing of the arrival of the group in Australia is much more recent (10–7 Mya) than the early Miocene age (ca. 20 Mya) supported by Ladiges et al. (2011). Species-level diversification then lagged and occurred mainly in the later cycles of drying (after a wetter period in the Pliocene (ca. 4 Mya); Clarkson et al. 2017). The differences in these age estimates for the origins and diversification of section *Suaveolentes* are inconsistent because of differing calibration methods (see Clarkson et al. 2017 for a discussion).

The sister taxon to the entire Australian clade (section *Suaveolentes*) is *N. africana* Merxm. and Butler (Chase et al. 2003), a narrowly endemic species found on rocky outcrops in the Central Plateau of Namibia (Merxmüller and Butler 1975). This implies two potential routes for long-distance dispersal to Australia, one across the Atlantic to Africa and then to Australia, or island-hopping across the Pacific with subsequent extinction (Aoki and Ito 2000); however, the precise origins of section *Suaveolentes* may be difficult to discern because of the long timelines involved, and local extinction and population expansion within arid-zone Australia (Chase et al. 2018a).

Section *Suaveolentes* represents an unusual case in Australia of a group diversifying in response to extreme aridification, although studies of the California flora suggest that a preponderance of short branches occurring in arid regions indicates that we may have to rethink ideas about aridity as an evolutionary stimulus (Thornhill et al. 2017). Recent work in other groups has shown that geographic diversification is still ongoing in other dry areas of Australia, such as the monsoon tropics of the north (Edwards et al. 2018), and that clades are in the early stages of lineage sorting. and that previously recognized barriers are not impervious. Patterns in Australian *Solanum* mirror those found in *Nicotiana*, with explosive recent diversification occurring in dry habitats (Echeverría-Londoño et al. 2018).

2.5 Species Diversity of *Nicotiana* and Future Prospects

After Goodspeed (1954) published his comprehensive and detailed monograph, research in *Nicotiana* focused mainly on cytogenetics, phylogenetic relationships, or regional taxonomic treatments in floristic works (e.g., Purdie et al. 1982; Wagner and Lorence 2002; Marks 2010; Cocucci 2013; Knapp in press). Recent interest in the species-level diversity of *Nicotiana* has led to the discovery of a number of new taxa in Australia (Symon 1998; Chase et al. 2018a, b, c; Chase and Christenhusz 2018a, b) and South America (J.R. Stehmann, personal communication), several of which are still to be formally described and given scientific names. For example, the laboratory strain of *N. benthamiana* that has been the subject of much research in plant pathology (e.g., Kamoun et al. 1998) is likely to be an undescribed taxon (Chase and Christenhusz 2018d). With wild species of crops attracting ever greater interest for plant breeding and crop improvement (Dempewolf et al. 2014) and for understanding basic biochemical and genetic processes (e.g., Hübner et al. 2019), a detailed and field-based understanding of the diversity both within and between species in *Nicotiana* will yield many benefits. The increasing number of collections held in herbaria and museums worldwide will greatly aid in the tasks, especially as they are imaged and their data digitized and made available for widespread use (e.g., Funk 2018). New field work such as that being undertaken in Australia (e.g., Chase et al. 2018a) is also critical to improving our understanding of *Nicotiana* species, both newly discovered and those already described.

Although Goodspeed (1954: 285) placed much emphasis on the role of interspecific hybridization in the evolutionary history of *Nicotiana*, the complex genetic processes involved in diploid *Nicotiana* evolution have also only had their surface scratched when compared with studies of the genetic processes involved in allopolyploid taxa. Clarkson et al. (2009) identified putative homoploid hybrid origins for *N. glutinosa* and *N. linearis* by analyzing incongruence between nuclear and plastid gene trees and morphology. Kelly et al. (2010) inferred homoploid (interspecific) hybrid origin for three diploid species: *N. glauca*, *N. linearis*, and *N. spegazzini* Millán based on a more comprehensive analysis of gene-tree incongruence using five low-copy nuclear genes, and suggested that intergenic recombination was relatively common in the genus (as it is in other plant genera, e.g., *Helianthus* (sunflowers) as reviewed by Rieseberg 2006). It is difficult to accurately identify homoploid hybridization and it is probably underestimated in nature (e.g., Mallet 2005), and *Nicotiana* may be no exception. Population-level field studies are needed in conjunction with incongruence studies, now possible using genomic data (e.g., Pease et al. 2016; Wu et al. 2019), to fully analyze these patterns and ascertain their importance in the evolution of *Nicotiana*.

Some surprising relationships have emerged from phylogenetic studies involving more extensive genomic regions, for example, Dodsworth (2015) suggested that the close relationship of the Pacific species *N. fragrans* and *N. fatuhivensis* F.Br. (see Table 2.1 for distributions) is not supported by multiple gene regions. Instead of being sister species in an early lineage within section *Suaveolentes*, as hypothesized by

Marks et al. (2011), or two subspecies of a single species, based on morphology (Goodspeed 1954), *N. fragrans* is more closely related to the central South Australian species *N. truncata* Symon, a long-distance disjunction of more than 3,000 km (Dodsworth 2015; Chase et al. 2018a).

While some species of *Nicotiana* are narrowly endemic with restricted ranges and very specific substrate requirements (e.g., *N. africana* and *N. ameghinoi* Speg.; also see species described in the papers in Chase et al. 2018), two species have shown somewhat invasive tendencies and have become well established outside their putatively native ranges. *Nicotiana plumbaginifolia* whose native distribution is not well known, most likely originated from the Southern Cone of South America, as have all other species of section *Alatae*, is now a well-established weed throughout Central America and the Caribbean (e.g., Knapp 2012) as well as the southern regions of the United States of America (Knapp in press), and was introduced early into India (Goodspeed 1954). *Nicotiana plumbaginifolia* is self-compatible and often autogamous, which may account for its spread, perhaps as a ballast weed (Goodspeed 1954). *Nicotiana glauca* has been spread through the agency of humans far from its original native range in Argentina and Bolivia (Goodspeed 1954) and is established as a naturalized species in Mediterranean climates worldwide. In the Americas, *N. glauca* is hummingbird-pollinated, while in its introduced range in the Old World, the limited ability of local flower visitors and potential pollinators to deal with the alkaloid content of the nectar may be driving local evolution in situ (e.g., Tadmor-Melamed et al. 2004). In Australia, *N. glauca* readily hybridizes with native species and is considered an invasive weed, especially in disturbed areas (e.g., Florentine and Westbrooke 2005), with leachates from leaves inhibiting seed germination of other species. In Mexico, *N. glauca* is also considered an introduced weed and is widely distributed in arid habitats throughout the country (Hernández 1981); the species was thought to have arrived in California via Mexico and was first recorded there in 1879 (Sauer 1988). In California, the mainland populations of *N. glauca* are self-incompatible and exhibit considerable herkogamy (differential position of stigma and anthers), while those established on California's Channel Islands are self-compatible (Schueller 2004); both are pollinated by hummingbirds. Selfing-related traits such as distance between stigma and anthers are most developed in populations that are recently colonized and appear to be unrelated to pollinator service (Schueller 2004). This is in stark contrast to the situation in the native range of *N. glauca*, where hummingbird pollinator community composition is highly correlated with flower differences (Nattero and Cocucci 2007). Given the probable hybrid origin of *N. glauca* (see the second paragraph in this section), further investigation of geographic variation in ecological factors associated with both morphological and genomic variation will undoubtedly reveal more about the ongoing evolutionary trajectories within this widely distributed species.

Aridification and expansion of dryland environments has been a significant part of Earth's history (e.g., Byrne et al. 2008; Thornhill et al. 2017) and is predicted to rapidly accelerate under more drastic scenarios of climate change (e.g., Sherwood and Fu 2014; Huang et al. 2016). Expansion of ranges of dry- and disturbance-adapted species of *Nicotiana* coupled with increased rates of environmental change, including disturbance from human activity, may be significant drivers of evolution in the genus in the future, driving interspecific hybridization and potentially new allopolyploid formation.

2.6 Some Priorities for the Study of *Nicotiana* Biodiversity at the Species Level

I list here a few personal priorities for increasing our understanding of the natural history of the wild tobaccos. These are by no means the only gaps in our knowledge, and other scientists will have other priorities, but working on these in collaborations between the biodiversity and genomics communities will certainly advance both fields.

1. Verifying chromosome numbers in *Nicotiana* species where this information is lacking or conflicting (see Table 2.1).
2. Better documentation of the ranges and distributions of *Nicotiana* species in both space and time. This will require documentation of specimens held in herbaria and museums as well as targeted in-depth field work.
3. Better understanding of the role of homoploid hybridization in diploid species of *Nicotiana*; although Kelly et al. (2010) have made initial steps, the degree to which homoploid hybridization has driven evolution at the diploid level will require population sampling and genomic tools.
4. Description of new species known to occur in nature is often not a priority; the work being undertaken in Australia by Chase et al. (2018) is a model to follow for other parts of the *Nicotiana* range.
5. Population-level sampling for both morphology and molecular markers across species' ranges will certainly lead to new insights, and potentially more cryptic species.

New genomic tools will improve our ability to discover new molecular markers and ways of looking at relationships, but as ongoing taxonomic work has shown, improving taxon sampling both in terms of species and populations is equally important. The combination of the two will yield much in the years to come.

Acknowledgements I thank Nikolai Ivanov for inviting me to write this chapter; understanding of *Nicotiana* taxonomy and biodiversity is based on the foundational work done by Thomas Harper Goodspeed in the 1950s, and his is still the last comprehensive monograph of the genus; Andrew Leitch, Jim Clarkson, Mark Chase, Laura Kelly, Elizabeth McCarthy, Ales Kovarik, the late Anne Kenton, and the late Yoong Lim have all been instrumental in increasing my knowledge and understanding of the genus; funding for work on *Nicotiana* came from multiple sources—field work in South America was undertaken with permission from the governments of Argentina, Bolivia, Brazil, Paraguay, and Peru; travel for herbarium visits was financed by the SYNTHESYS Project (http://www.synthesys.info/) financed by European Community Research Infrastructure Actions under the FP6 and FP7 "Structuring the European Research Area" Programme, and the Special Funds of the Natural History Museum; field work was financed by the National Geographic Society and the Natural History Museum Enhancement Grant.

References

Adam N, Kellenbach M, Meldau S, Veit D, van Dam NM, Baldwin IT, Schuman MC (2018) Functional variation in a key defense gene structures herbivore communities and alters plant performance. PLoS ONE 13(6):e0197221. https://doi.org/10.1371/journal.pone.0197221

Aigner PA, Scott PF (2002) Use and pollination of a hawkmoth plant, *Nicotiana attenuata*, by migrant hummingbirds. Southwestern Nat 47:1–11

Anssour S, Krügel T, Sharbel TF, Slauz HP, Bonaventure G, Baldwin IT (2009) Phenotypic, genetic and genomic consequences of natural and synthetic polyploidization of *Nicotiana attenuata* and *Nicotiana obtusifolia*. Ann Bot 103:1207–1217

Aoki S, Ito M (2000) Molecular phylogeny of *Nicotiana* (Solanaceae) based on nucleotide sequence of the *matK* gene. Plant Biol 2:316–324

Baldwin IT (2001) An ecologically motivated analysis of plant-herbivore interactions in native tobacco. Plant Physiol 127:1449–1458

Baldwin IT, Pandey S, Gase K, Navarro A (2019) *Nicotiana attenuata*, an ecological model. In: Ivanov NV et al (eds), The tobacco genome. Compendium of Plant Genomes, vol 20. Springer, Berlin, pp 30

Bombarely A, Rosli AG, Vrebelov J, Moffett P, Mueller LA, Martin GB (2012) Draft genome sequence of *Nicotiana benthamiana* to enhance molecular plant-microbe biology research. Mol Plant Microbe Interact 25:1523–1530

Burbidge NT (1960) The phytogeogeography of the Australian region. Aust J Bot 8:75–209

Byrne M, Yeates DK, Joseph L, Kearney M, Bowler J, Williams MAJ, Cooper S, Donnellan SC, Keogh JS, Leys R, Melville J, Murphy DJ, Proch N, Wyrwoll K-H (2008) Birth of a biome: insights into the assembly and maintenance of the Australian arid zone biota. Mol Ecol 17:4398–4417

Chase MW, Christenhusz MJM (2018a) 883. *Nicotiana karijini*. Curtis' Bot Mag 35:228–236

Chase MW, Christenhusz MJM (2018b) 885. *Nicotiana gascoynica*. Curtis' Bot Mag 35:245–252

Chase MW, Christenhusz MJM (2018c) 887. *Nicotiana excelsior*. Curtis' Bot Mag 35:261–268

Chase MW, Christenhusz MJM (2018d) 888. *Nicotiana gossei*. Curtis' Bot Mag 35:269–277

Chase MW, Christenhusz MJM (2018e) 889. *Nicotiana umbratica*. Curtis' Bot Mag 35:278–294

Chase MW, Christenhusz MJM (2018f) 890. *Nicotiana benthamiana*. Curtis' Bot Mag 35:286–285

Chase MW, Christenhusz MJM (2018g) 891. *Nicotiana occidentalis* subsp. *obliqua*. Curtis' Bot Mag 35:295–303

Chase MW, Christenhusz MJM (2018h) 894. *Nicotiana stenocarpa*. Curtis' Bot Mag 35:319–327

Chase MW, Knapp S, Cox AV, Clarkson J, Butsko Y, Joseph J, Savolainen V, Parokonny AS (2003) Molecular systematics, GISH and the origin of hybrid taxa in Nicotiana (Solanaceae). Annals of Botany 92: 107–127.

Chase MW, Christenhusz MJM, Conran JG, Dodsworth S, Medeiros Nollet, de Assis F, Felix LP, Fay MF (2018a) Unexpected diversity of Australian tobacco species (*Nicotiana* section *Suaveolentes*, Solanaceae). Curtis' Bot Mag 35:212–227

Chase MW, Conran JG, Christenhusz MJM (2018b) 884. *Nicotiana yandinga*. Curtis' Bot Mag 35:237–244

Chase MW, Conran JG, Christenhusz MJM (2018c) 886. *Nicotiana faucicola*. Curtis' Bot Mag 35:253–260

Chase MW, Conran JG, Christenhusz MJM (2018d) 892. *Nicotiana burbidgeae*. Curtis' Bot Mag 35:304–311

Chase MW, Conran JG, Christenhusz MJM (2018e) 893. *Nicotiana maritima*. Curtis' Bot Mag 35:312–318

Clarkson JR, Symon DE (1991) Nicotiana wuttkei (Solanaceae), a new species from north-eastern Queensland with an unusual chromosome number. Austrobaileya 3:389–392

Clarkson JJ, Knapp S, Garcia VF, Olmstead RG, Leitch AR, Chase MW (2004) Phylogenetic relationships in *Nicotiana* (Solanaceae) inferred from multiple plastid DNA regions. Mol Phylogenet Evol 33:75–90

Clarkson JJ, Lim KY, Kovarik A, Chase MW, Knapp S, Leitch AR (2005) Long-term genome diploidization in allopolyploid *Nicotiana* section *Repandae* (Solanaceae). New Phytol 168:241–252

Clarkson JJ, Kelly LJ, Leitch AR, Knapp S, Chase MW (2009) Nuclear glutamine synthetase evolution in *Nicotiana*: phylogenetics and the origins of allotetraploid and homoploid(diploid) hybrids. Mol Phylogenet Evol 55:99–112

Clarkson JJ, Dodsworth S, Chase MW (2017) Time-calibrated phylogenetic trees establish a lag between polyploidisation and diversification in *Nicotiana* (Solanaceae). Plant Syst Evol 303:1001–1012

Cocucci AA (2013) *Nicotiana*. In: Anton AM, Zuloaga FO (eds), Barboza GE (coord.) Flora Argentina, Solanaceae, vol 13. IOBDA- IMBIV, CONICET: Buenos Aires & Córdoba, Argentina, pp 75–89

Crisp MD, West JG, Weston PH (1999) Biogeography of the terrestrial flora. In: Orchard AE, Thompson HS (eds) Introduction, Flora of Australia, edn 2, vol 1. ABRS/CSIRO, Melbourne, pp 321–368

Crisp MD, Arroyo MTK, Cook LG, Gandolfo MA, Jordan GJ, McGlone MS, Weston PH, Westoby M, Wilf P, Linder HP (2009) Phylogenetic biome conservatism on a global scale. Nature 458:754–756

D'Arcy WG (1977) [1976]) New names and taxa in the Solanaceae. Ann Mo Bot Gard 63:363–369

Deanna R, Larter MD, Barboza GE, Smith SD (2019) Repeated evolution of a morphological novelty: a phylogenetic analysis of the inflated fruiting calyx in the Physalideae tribe (Solanaceae). Am J Bot 106 (2):270–279

Dempewolf H, Eastwood RJ, Guarino L, Khoury CK, Müller JV, Toll J (2014) Adapting agriculture to climate change: a global initiative to collect, conserve, and use crop wild relatives. Agroecol Sustain Food Syst 38:369–377

Diezel C, Kessler D, Baldwin IT (2011) Pithy protection: *Nicotiana attenuata*'s jasmonic acid-mediated defenses are required to resist stem-boring weevil larvae. Plant Physiol 155:1936–1946

Dodsworth S (2015) Genome skimming for phylogenomics. PhD thesis, Queen Mary University of London, London

Dodsworth S, Chase MW, Leitch AR (2016) Is post-polyploidization diploidization the key to the evolutionary success of angiosperms? Bot J Linn Soc 180:1–5

Dodsworth S, Jang T-S, Chase Struebig M, Weiss-Scheeweiss MW, Leitch AR (2017) Genome wide repeat dynamics reflect phylogenetic distance in closely related allotetraploid *Nicotiana* (Solanaceae). Plant Syst Evol 303:1013–1020. https://doi.org/10.1007/s00606-016-1356-9

Echeverría-Londoño S, Särkinen T, Fenton IS, Knapp S, Purvis A (2018) Dynamism and context dependence in the diversification of the megadiverse genus *Solanum* L. (Solanaceae). https://www.biorxiv.org/content/early/2018/11/02/348961. Posted 2 Nov 2018

Edwards KD, Fernandez-Pozo N, Drake-Stowe K, Humphry M, Evans AD, Bombarely A, Allen F, Hurst R, White B, Kernodle SP, Bromley JR, Sanchez-Tamburrino JP, Lewis RS, Mueller LA (2017) A reference genome for *Nicotiana tabacum* enables map-based cloning of homeologus loci implicated in nitrogen utilization efficiency. BMC Genom 18:448. https://doi.org/10.1186/s12684-017-3791-6

Edwards RD, Crisp MD, Cook LG (2018) Species limits and cryptic biogeographic structure in a widespread complex of Australian monsoon tropics trees (broad-leaf paperbarks: *Melaleuca*, Myrtaceae). Aust Syst Bot 31:495–503

Fernie A, Usadel B (2020) The *Nicotiana glauca* genome. In: Ivanov NV et al (eds), The tobacco genome. Compendium of plant genomes, vol 20. Springer, Berlin, p 30

Florentine SK, Westbrooke ME (2005) Invasion of the noxious weed *Nicotiana glauca* R. Graham after an episodic flooding event in the arid zone of Australia. J Arid Environ 60:531–545

Funk VA (2018) Collections-based science in the 21st century. J Syst Evol 56:175–193. https://doi.org/10.1111/jse.12315

Goodspeed TH (1945) Studies in *Nicotiana*: III. A taxonomic organization of the genus. Univ Calif Publ Bot 18:335–344

Goodspeed TH (1954) The genus *Nicotiana*. Chron Bot 16:1–536
Haverkamp A, Hansson BS, Baldwin IT, Knaden M, Yon F (2018) Floral trait variations among wild tobacco populations influence the foraging behaviour of hawkmoth pollinators. Front Ecol Evol 6:19. https://doi.org/10.3389/fevo.2018.00019
Hernández H (1981) Sobre la ecología reproductive de *Nicotiana glauca* Grah.: una maleza de distribución cosmopolita. Boletin de la Sociedad Botánica de Mexico 41:47–73
Huang J, Yu H, Guan X, Wang G, Guo R (2016) Accelerated dryland expansion under climate change. Nat Clim Change 6:166–171
Hübner S, Bercovich N, Todesco M, Mandel JR, Odenheimer J, Zregler E, Lee JS, Baute GJ, Owens GL, Grassa CJ, Ebert DP, Ostevik KL, Moyers BT, Yakimoski S, Masalia RR, Gao L, Ćalić I, Bowers JE, Kane NC, Swanevelder DZH, Muños S, Langlade NB, Burke JM, Rieseberg LH (2019) Sunflower pangenome analysis shows that hybridization altered gene content and disease resistance. Nat Plants 5:64–72. https://doi.org/10.1038/s41477-018-0329-0
Hughes CE, Eastwood R (2006) Island radiation on a continental scale: exceptional rates of plant diversification after uplift of the Andes. Proc Nat Acad Sci USA 103:10334–10339. https://doi.org/10.1073/pnas.0601928103
Hunziker AT (2001) Genera Solanacearum. ARG Gantner Verlag, Rügen
Kaczorowski RL, Gardener MC, Holtsford TP (2005) Nectar traits in *Nicotiana* sect. *Alatae* (Solanaceae) in relation to floral traits, pollinators, and mating system. Am J Bot 92:1270–1283
Kamoun S, van West P, Vleeshouwers VGAA, de Groot KE, Govers F (1998) Resistance of *Nicotiana benthamiana* to *Phytophthora infestans* is mediated by the recognition of the elicitor protein INF1. Plant Cell 10:1413–1425
Kelly LJ, Leitch AR, Clarkson JJ, Hunter RB, Knapp S, Chase MW (2010) Intragenic recombination events and evidence for hybrid speciation in *Nicotiana* (Solanaceae). Mol Biol Evol 27:781–799
Kelly LJ, Leitch AR, Clarkson JJ, Knapp S, Chase MW (2013) Reconstructing the complex evolutionary origin of wild allopolyploid tobaccos (*Nicotiana* section *Suaveolentes*). Evolution 76:80–94
Kenton A, Parokonny AS, Gleba YY, Bennett MD (1993) Characterization of the *Nicotiana tabacum* L. genome by molecular cytogenetics. Mol Gen Genet 240:159–169
Kessler A, Baldwin IT (2001) Defensive function of herbivore-induced plant volatile emissions in nature. Science 291:2141–2144
Kessler D, Baldwin IT (2007) Making sense of nectar scents: the effects of nectar secondary metabolites on floral visitors of Nicotiana attenuata. Plant J 49(5):840–854
Kessler D, Gase K, Baldwin IT (2008) Field experiments with transformed plants reveal the sense of floral scents. Science 321:1200–1202
Kessler D, Diezel C, Baldwin IT (2010) Changing pollinators as a means of escaping herbivores. Curr Biol 20:237–242
Kessler D, Kallenbach M, Diezel C, Rothe E, Mardock M, Baldwin IT (2015) How scent and nectar influence flora antagonists and mutalists. eLife 4:e07641. http://doi.org/10:7554/eLife.07641
Knapp S (2012) Solanaceae. In: Acevedo-Rodriguez P, Strong MT (eds) Catalogue of seed plants of the west Indies, vol 98. Smithsonian Contributions to Botany, pp 898–913. https://botany.si.edu/Antilles/WestIndies/
Knapp S, Chase MW, Clarkson JJ (2004) Nomenclatural changes and a new sectional classification in *Nicotiana* (Solanaceae). Taxon 53:73–82
Knapp S (in press) *Nicotiana*. In: Sullivan J (ed) Solanaceae, in Flora of North America vol 14. Missouri Botanical Garden Press, St Louis, p 30
Kovarik A, Matyasek R, Lim KY, Skalická K, Koukalová B, Knapp S, Chase M, Leitch AR (2004) Concerted evolution of 18-5.8-26S rDNA repeats in Nicotiana allotetraploids. Biol J Linn Soc 82(4):615–625
Ladiges PY, Marks CE, Nelson G (2011) Biogeography of *Nicotiana* section *Suaveolentes* (Solanaceae) reveals geographical tracks in arid Australia. J Biogeogr 38:2066–2077
Lagomarsino L, Condamine FL, Antonelli A, Mulch A, Davis CC (2016) The abiotic and biotic drivers of rapid diversification in Andean bellflowers. New Phytol 210:1430–1442. https://doi.org/10.1111/nph.13920
Leitch IJ, Bennett MD (2004) Genome downsizing in polyploid plants. Biol J Linn Soc 82(4):651–663
Leitch IL, Hanson L, Lim YK, Kovarik A, Clarkson JJ, Chase MW, Leitch AR (2008) The ups and downs of genome size evolution in polyploid species of *Nicotiana* (Solanaceae). Ann Bot 101:805–814
Lim KY, Matyasek R, Kovarik A, Leitch AR (2004) Genome evolution in allotetraploid *Nicotiana*. Biol J Lin Soc 82:599–606
Lim KY, Kovarik A, Matyasek R, Chase MW, Knapp S, McCarthy E, Clarkson JJ, Leitch AR (2006) Comparative genomics and repetitive sequence divergence in the species of diploid *Nicotiana* section *Alatae*. Plant J 48:907–919
Lim KY, Kovarik A, Matyasek R, Chase MW, Clarkson JJ, Grandbastien MA, Leitch AR (2007) Sequence of events leading to near-complete genome turnover in allopolyploid *Nicotiana* within five million years. New Phytol 175:756–763
Macbride JF (1930) Spermatophytes, mostly Peruvian. 3. Peruvian Solanaceae. Publ Field Mus Nat Hist Bot Ser 8(2):105–112
Mallet J (2005) Hybridization as an invasion of the genome. Trends Ecol Evol 20:229–237
Marks CE (2010) Definition of South Pacific taxa of Nicotiana section Suaveolentes (Solanaceae). Muelleria 28:74–84

Marks CE, Newbigin E, Ladiges PY (2011) Comparative morphology and phylogeny of *Nicotiana* section *Suaveolentes* (Solanaceae) in Australia and the South Pacific. Aust Syst Bot 24:61–86
McCarthy EW, Arnold SEJ, Chittka L, LeComber SC, Verity R, Dodsworth S, Knapp S, Lj Kelly, Chase MW, Baldwin IT, Kovarik A, Mhiri C, Taylor L, Leitch AR (2015) The effect of polyploidy and hybridization on the evolution of flower colour in *Nicotiana* (Solanaceae). Ann Bot 115:1117–1131
McCarthy EW, Chase MW, Knapp S, Litt A, Leitch AR, LeComber SC (2016) Transgressive phenotypes and generalist pollination in the floral evolution of *Nicotiana* polyploids. Nat Plants 2016:119. https://doi.org/10.1038/nplants.2016.119
McCarthy EW, Landis JB, Kurti A, Lawhorn AJ, Chase MW, Knapp S, Le Comber SC, Leitch AR, Litt A (2019) Early consequences of allopolyploidy alter floral evolution in *Nicotiana* (Solanaceae). BMC Plant Biol
McCarthy EW, Landis JB, Kurti A, Lawhorn AJ, Litt A (2020) The Genetic Basis of Flower Color Differences in Nicotiana tabacum. In: N. V. Ivanov et al. (eds) The Tobacco Plant Genome. Springer Nature Switzerland AG, Switzerland, pp 175–192
Merxmüller H, Butler KP (1975) *Nicotiana* in der Afrikanischen Namib—ein Pflanzengeographisches und Phylogenetisches Ratsel. Mitteilungen aus der Botanischen Staatssammlung München 12:91–104
Nattero J, Cocucci AA (2007) Geographical variation in floral traits of the tree tobacco in relation to its hummingbird pollinator fauna. Biol J Lin Soc 90:657–667
Navarro-Quezada A, Gase K,Singh RK, Pandey SP, Baldwin IT (2020) Nicotiana attenuata Genome RevealsGenes in the Molecular MachineryBehind Remarkable AdaptivePhenotypic Plasticity. In: Ivanov et al. (eds) The Tobacco Plant Genome. Springer Nature, Switzerland, pp 209–227
Ohashi Y (1985) Thremmatological studies of wild species related to *Nicotiana tabacum*, with special reference to disease resistance. Iwata Tabako Shikenjo Hokoku (Bull Iwata Tob Exp Station) 17:1–67
Olmstead RG, Palmer JD (1992) A chloroplast DNA phylogeny of the Solanaceae: subfamilial relationships and character evolution. Ann Mo Bot Gard 79:249–265
Olmstead RG, Sweere JA (1994) Combining data in phylogenetic systematics: an empirical approach using three molecular data sets in Solanaceae. Syst Biol 43:467–481
Olmstead RG, Sweere JA, Spangler RE, Bohs L, Palmer JD (1999) Phylogeny and provisional classification of the Solanaceae based on chloroplast DNA data. In: Nee M, Symon DE, Lester RN, Jessop JP (eds), Solanaceae IV: advances in biology and utilization. Royal Botanic Gardens, Kew, pp 111–137
Olmstead RG, Bohs L, Migid HA, Santiago-Valentín E, Garcia VF, Collier SM (2008) A molecular phylogeny of the Solanaceae. Taxon 57:1159–1181
Pease JB, Haak DC, Hahn MW, Moyle LC (2016) Phylogenomics reveals three sources of adaptive variation during a rapid radiation. PLoS Biol 14: e1002379. https://doi.org/10.1371/journal.pbio.1002379
Purdie RW, Symon DE, Haegi L (1982) *Nicotiana*. In: George AS (ed) Flora of Australia, vol 29. Australian Government Publishing Service, Canberra, pp 38–57
Raven PH (1963) Amphitropical relationships in the floras of North and South America. Q Rev Biol 38:151–177
Rieseberg LH (2006) Hybrid speciation in sunflowers. Ann Mo Bot Gard 93:34–48
Särkinen T, Bohs L, Olmstead RG, Knapp S (2013) A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): a dated 1000-tip tree. BMC Evol Biol 13:214
Särkinen T, Pennington RT, Lavin M, Simon MC, Hughes CE (2012) Evolutionary islands in the Andes: persistence and isolation explain high endemism in Andean dry tropical forests. J Biogeogr 39:884–900
Särkinen T, Poczai P, Barboza GE, van der Weerden GM, Baden M, Knapp S (2018) A revision of the old world black nightshades (Morelloid clade of *Solanum* L., Solanaceae). PhytoKeys 106:1–223. https://doi.org/10.3897/phytokeys.106.21991
Sauer JD (1988) Plant migration: the dynamics of geographic patterning in seed plant species. University of California Press, Los Angeles
Schueller SK (2004) Self-pollination in island and mainland populations of the introduced hummingbird-pollinated plant, *Nicotiana glauca* (Solanaceae). Am J Bot 91:672–681
Sherwood S, Fu Q (2014) A drier future? Science 343:737–739
Sierro N, Battey JND, Ouadi S, Bovet L, Goepfert S, Bakaher N, Peitsch MC, Ivanov NV (2013) Reference genomes and transcriptomes of *Nicotiana sylvestris* and *Nicotiana tomentosiformis*. Genome Biol 14:R60
Sierro N, Battey JND, Ouadi S, Bakaher N, Bovet L, Willig A, Goepfert S, Peitsch MC, Ivanov NV (2014) The tobacco genome sequence and its comparison to those of tomato and potato. Nat Commun 5:3822
Simpson MG, Johnson LA, Villaverde T, Guilliams CM (2017) American amphitropical disjuncts: perspectives from vascular plant analyses and prospects for future research. Am J Bot 104:1600–1650
Skalicka K, Lim Y, Matyasek R, Matzka M, Leitch A, Kovarik A (2005) Preferential elimination of repeated DNA sequences from the paternal *N. tomentosiformis* genome donor of a synthetic allotetraploid tobacco. New Phytol 166:291–303
Smith LB, Downs RJ (1964) Notes on the Solanaceae of southern Brazil. Phytologia 10:422–453
Spooner DM, Ghislain M, Simon R, Jansky SH, Gavrilenko T (2014) Systematics, diversity, genetics, and evolution of wild and cultivated potatoes. Bot Rev 80:283–383
Stehmann JR, Semir J, Ippolito A (2002) *Nicotiana mutabilis* (Solanaceae), a new species from southern Brazil. Kew Bull 57:639–646

Symon DE (1984) A new species of *Nicotiana* (Solanaceae) from Dalhousie Springs (South Australia). J Adelaide Bot Gard 7:117–121

Symon DE (1998) A new *Nicotiana* (Solanaceae) from near Coober Pedy, South Australia. J Adelaide Bot Gard 18:1–4

Symon DE (2005) Native tobaccos (Solanaceae: *Nicotiana* spp.) in Australia and their use by Aboriginal peoples. Beagle 21:1–10

Symon DE, Kenneally KF (1994) A new species of *Nicotiana* (Solanaceae) from near Broome, Western Australia. Nuytsia 9:421–425

Symon DE, Lepschi BJ (2007) A new status in *Nicotiana* (Solanaceae): *N. monoschizocarpa* (P.Horton) Symon & Lepschi. J Adelaide Bot Gard 21:92

Tadmor-Malamed H, Markman S, Arieli A, Distl M, Wink M, Izhaki I (2004) Limited ability of Palestine Sunbirds *Nectarinia osea* to cope with pyridine alkaloids in nectar of Tree Tobacco *Nicotiana glauca*. Funct Ecol 18:844–850

Thornhill AH, Baldwin BG, Freyman WA, Nosratinia S, Kling MM, Morueta-Holme N, Madsen TP, Ackerly DD, Mischler BD (2017) Spatial phylogenetics of the California flora. BMC Biol 15:96. https://doi.org/10.1186/s12915-017-0435-x

Wagner WL, Lorence DH (2002) Flora of the Marquesas Islands website. http://botany.si.edu/pacificislandbiodiversity/marquesasflora/index.htm. Accessed Dec 2018

Wen J, Ickert-Bond SM (2009) Evolution of the Madrean-Tethyan disjunctions and the North and South American amphitropical disjunctions in plants. J Syst Evol 47:331–348

Wilf P, Carvalho MR, Gandolfo MA, Cúneo NR (2017) Eocene lantern fruits from Gondwanan Patagonia and the early origins of Solanaceae. Science 355:71–75

Williams E (1975) A new chromosome number in the Australian species *Nicotiana cavicola* L. (Burbidge). N Z J Bot 13:811–812

Wu J, Hettenhausen C, Schuman MC, Baldwin IT (2008) A comparison of two *Nicotiana attenuata* accessions reveals large differences in signaling induced by oral secretions of the specialist herbivore *Manduca sexta*. Plant Physiol 146:927–939

Wu M, Kostyun JL, Moyle LC (2019) Genome sequence of *Jaltomata* addresses rapid reproductive trait evolution and enhances comparative genomics in the hyperdiverse Solanaceae. Genome Biol Evol. https://doi.org/10.1093/gbe/evy274

Check for updates

3 Genetic Markers in Tobacco, Usage for Map Development, Diversity Studies, and Quantitative Trait Loci Analysis

Nicolas Bakaher

Abstract

Tobacco (*Nicotiana tabacum* L.) is an allotetraploid species in the Solanaceae family and has a complex genome. The first genetic maps of tobacco have been developed using microsatellite markers. Microsatellite markers were also used to create links between genetic maps of various Nicotiana species, providing a platform for synteny analysis. Markers are also an efficient tool to decipher the population genetics of tobacco. The three main tobacco types used in smoking products are flue-cured or Virginia, Burley, and Oriental. The commercial type is translated in genetic divergence resulting from the selection of adapted cultivars associated with agronomical and curing crop practices. Genetic markers are needed for trait discovery via quantitative trait locus mapping. They were used to identify loci involved in leaf surface components (cis-abienol and sucrose esters) and contributing to flavor and aroma characteristics. Another QTL mapping study targeting resistance to soilborne diseases such as black shank and bacterial wilt led to the identification of several loci that provide a better understanding of polygenic disease resistance in tobacco. Today with the availability of the genome sequence, single-nucleotide polymorphisms are being developed in tobacco. They are a fast, reproducible, and cost-efficient screening method for any genetic application from basic research to commercial breeding that may accelerate the pace of discoveries in tobacco.

N. Bakaher (✉)
PMI R&D, Philip Morris Products S.A., Quai Jeanrenaud 5, Neuchâtel 2000, Switzerland
e-mail: nicolas.bakaher@pmi.com

3.1 Genetic Maps

Tobacco (*Nicotiana tabacum* L.) is an allotetraploid species and has a genome size among the largest in the Solanaceae family. As a consequence of this complexity, the first version of a high-resolution genetic map of the species was published only in 2007 (Bindler et al. 2007). This map was constructed with microsatellite markers identified from the genomic libraries of TN90 Burley tobacco and sequence data from the Tobacco Genome Initiative (Gadani et al. 2003). This map contains fewer than 300 microsatellite markers mapped onto the 24 linkage groups.

A second version of the microsatellite genetic map (Bindler et al. 2011), containing 2,317 microsatellite markers spanning the 24 linkage groups, was published in 2011 using the same mapping population (Fig. 3.1). The microsatellite markers were assigned to their respective genome of origin, either *N. sylvestris* (S) or

© Springer Nature Switzerland AG 2020
N. V. Ivanov et al. (eds.), *The Tobacco Plant Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-030-29493-9_3

N. tomentosiformis (T). In most linkage groups, the markers could be attributed predominantly to one of the two genomes: 11 linkage groups could be clearly assigned to the S-genome, and nine linkage groups could be clearly assigned to the T-genome. Four linkage groups, mapped with both S- and T-genome-specific markers, were grouped mainly in different parts of the respective linkage group. At that time, the genetic map of tobacco was comparable in marker density and resolution to those of the tomato (*Solanum lycopersicum*) and potato (*S. tuberosum*) (Bindler et al. 2011). Since then, several other genetic maps have been developed using microsatellite markers for the flue-cured tobacco type (Tong et al. 2012 and 2016) or linked to quantitative trait locus (QTL) detection studies (Vontimitta et al. 2010; Drake-Stowe et al. 2017). These published maps refer to the backbone of the Bindler genetic map for the linkage group names (Bindler et al. 2011).

Microsatellite markers, together with single-copy conserved ortholog set II, have also been used to develop two genetic maps for diploid *Nicotiana* species, *N. tomentosiformis* and *N. acuminata* (Wu et al. 2010). *N. acuminata* is phylogenetically close to *N. sylvestris* and is selected because of the lack of polymorphisms within the *N. sylvestris* genome. These markers create links among the genetic maps of the cultivated *N. tabacum*, the wild diploid *Nicotiana* species, and tomato markers, thus providing a platform for synteny analysis among these and other solanaceous species, such as potato, eggplant, and pepper. The recent release of several *Nicotiana* genome sequences in the public domain allowed researchers to mine for novel genetic markers sequences inside and across species. Wang et al. (2018) have characterized a large set of so-called NIX (*Nicotiana* multiple (X) genome) microsatellite markers from the diploids *N. otophora*, *N. sylvestris* and

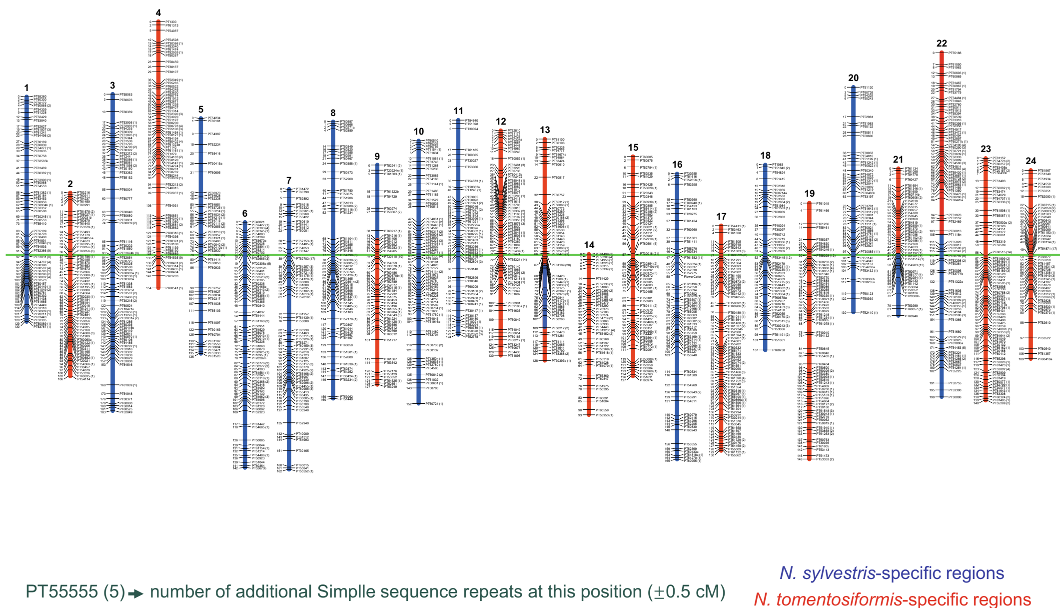

Fig. 3.1 High-density genetic map of the tobacco genome. Adapted from Bindler et al. 2011. Red indicates chromosomal segments assigned to the T-genome of *Nicotiana tomentosiformis*. Blue indicates chromosomal segments assigned to the S-genome of *N. sylvestris*. Numbers on the left of the chromosomes are centiMorgans (cM) counted from the top of the chromosome. Numbers in brackets after the marker name on the right of the chromosomes are the numbers of additional cosegregating markers (±0.5 cM). The green line represents the possible positions of the centromeres

N. tomentosiformis and from four allotetraploids, namely, *N. benthamiana* and three major commercial varieties of *N. tabacum*. They are available in the online database TMPD (http://biodb.sdau.edu.cn/tmpd/index.html). A fraction of them was experimentally validated in the five abovementioned species and demonstrated that they can be a valuable resource to screen multiple *Nicotiana* species or varieties.

Single-nucleotide polymorphisms (SNPs) are the most abundant type of genomic DNA variations currently used as genetic markers. The development and affordability of next-generation DNA sequencing represent a key process in the discovery of such variations in tobacco. The first SNP genetic map of tobacco was published in 2015 (Xiao et al. 2015). This map contains 4,138 SNP markers mapped onto 24 linkage groups. The availability of reference genome sequences (Sierro et al. 2014; Edwards et al. 2017) enables the discovery of more SNPs. As an additional tool in improving the mapping of tobacco sequences, BioNano Genomics optical mapping can anchor genomic sequences into pseudomolecules based on their locations on a high-density consensus genetic map containing 9,688 SNP markers and is available on the Sol Genomics Network (https://solgenomics.net/cview/map.pl?map_version_id=178). This map was generated using a 30k Infinium HD array (unpublished work).

With the growing availability of genome assemblies, more SNP arrays are expected to be developed, as has occurred for other commercial crops (tomato in Sim et al. 2012; potato in Vos et al. 2015). The number of markers addressed on each array may vary depending on the application. Nonetheless, SNP arrays will greatly improve the evaluation of tobacco genetic diversity and QTL or association mapping analysis, and enable the mapping of more genomic regions, eventually identifying candidate genes associated with these regions, such as the *Yellow Burley* (*YB*) loci involved in the Burley phenotype reported by Edwards et al. (2017).

SNP arrays are also fast, reproducible, and cost-efficient screening methods for molecular breeding, unlike the screening of microsatellite markers. They are nowadays a common tool used for commercial variety development in many crops.

3.2 Genetic Diversity

Tobacco is an economically important nonfood crop that has been cultivated for several centuries. European countries established tobacco state monopolies as early as the seventeenth century (Goodman 2005). Paper-rolled cigarettes were invented around 1832 (Goodman 2005) and their mass production commenced at the beginning of the twentieth century. Previously, tobacco was smoked in pipes or in a cigar-like configuration containing wrapper, filler, and binder leaves. To accommodate the environmental conditions and leaf usage in smoking or chewing products, several specific cultivars or types of tobacco were grown (Garner et al. 1937).

Conservative breeding methods used in the development of tobacco cultivars during the twentieth century have resulted in only a small portion of the germplasm pool being considered as acceptable material for new variety development (Clayton 1958). The genetic divergence between the various tobacco types marketed results from the selection of adapted cultivars associated with the agronomical and curing crop practices. Availability of microsatellite markers allowed better characterization of the tobacco germplasm. In Binder et al. (2007), microsatellite markers were used to select the best parents for the mapping populations by searching for the highest number of polymorphisms.

Using a limited number of microsatellite markers, the genetic diversity of tobacco can be assessed (Fig. 3.2). In their extensive study of the US Department of Agriculture (USDA) germplasm collection, Moon et al. (2009a) found that the tobacco type names referenced in the collection were given according to agricultural practice. The USDA classifications were therefore not always directly linked to the genetic pools. Nonetheless, these classifications are used when describing the three main tobacco types

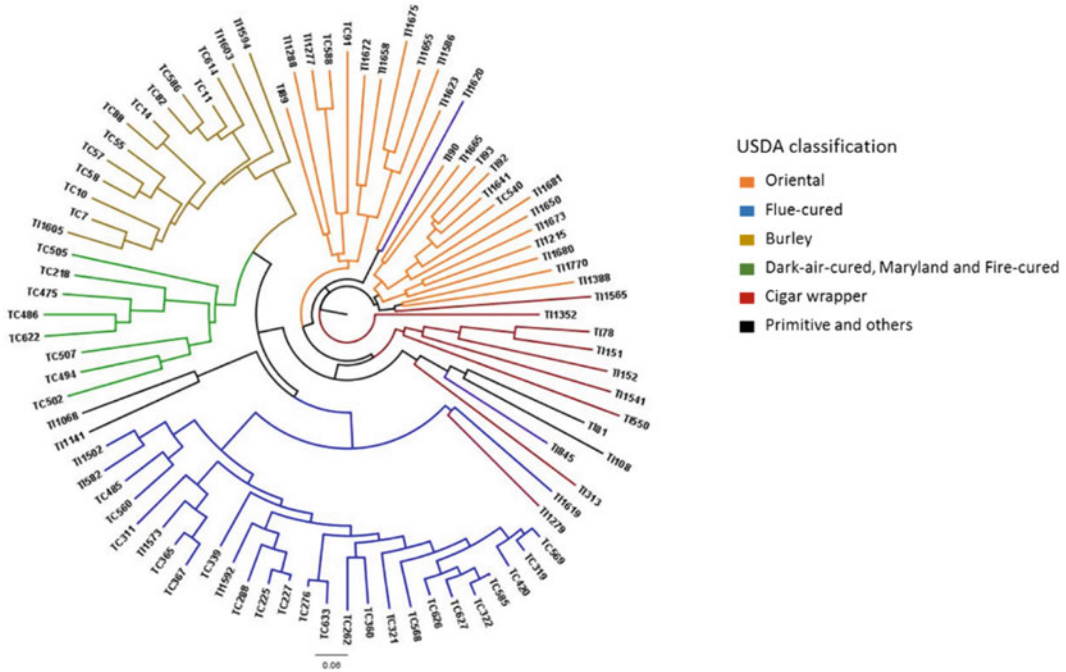

Fig. 3.2 Unrooted phylogenetic tree of 90 tobacco accessions. Labels correspond to the TI/TC code of the US Nicotiana Germplasm Collection. The tree was constructed with 27 microsatellite loci using Nei's genetic distance and the neighbor-joining method in the software package PowerMarker (Liu and Muse 2005). USDA, United States Department of Agriculture

used in smoking products: flue-cured or Virginia, Burley, and Oriental.

Currently, the highest volume of tobacco grown and exchanged worldwide is the flue-cured or Virginia type. Tobacco was previously only air-cured until the discovery of flue-curing in 1839 (Goodman 2005). Flue-cured tobacco is dried in a closed barn with oven-driven heat directed from pipes. The temperature of the curing barn is raised gradually, and the humidity is controlled until the leaves and stems are completely dried. Flue-curing takes approximately 1 week and fixes the natural sugar of the leaf, which has high sugar and medium nicotine contents. The initial gene pool used to develop the modern flue-cured cultivars is thought to be narrow (Murphy et al. 1987). In addition, application of scientific breeding methods commenced at the beginning of the twentieth century and contributed to further reductions in the pool of alleles detected in this tobacco type (Moon et al. 2009b). The elite flue-cured varieties available today are high yielding and produce leaves of good quality, but are genetically poorer.

Burley tobacco is the second most popular type for use in cigarettes and is thought to originate from a mutant plant of the Red Burley variety, known as the White Burley (Killebrew 1884). This variety quickly became popular in the US at the end of the nineteenth century because of its good quality and curability. Burley leaves are air-cured in barns for 4–8 weeks. Detached leaves can be hung on strings, but more commonly, the full stalk with leaves is hung in the barn. Burley tobacco has very low sugar content and medium-to-high nicotine content. Commercial Burley varieties are mostly available today as hybrids. This type is genetically close to the flue-cured type (Fricano et al. 2012).

Oriental tobacco is typically considered the aromatic component in smoking products and is originally produced in the Mediterranean (Greece, Macedonia, and Turkey). Oriental

tobacco is sun-cured to remove most of the moisture and is characterized by high aroma and low sugar and nicotine content. Oriental tobaccos are considered the most divergent market type because they have undergone more than 300 years of evolution and selection (Wolf and Wolf 1948; Moon et al. 2009a). Isolation, stressful environments, and specific agricultural conditions have selected for a genetic pool different from that of flue-cured and Burley tobaccos (Davalieva et al. 2010; Fricano et al. 2012).

Other major tobacco market types include cigar wrapper and filler, dark air-cured, fire-cured, and Maryland tobaccos. Some varieties of dark air-cured tobacco can also be fire-cured. Several curing types can be identified in the same genetic pool (Fig. 3.2); however, the genetic structure of the tobacco population does not always match agronomical and curing practices.

3.3 QTL Mapping Studies

There is a limited number of QTL mapping studies in tobacco, probably because of the unavailability of a high number of genetic markers and the absence of a core genetic map until recently. The first published QTL study in tobacco targeted agronomic traits, leaf quality, chemical composition, and smoke properties (Julio et al. 2006). The availability of the microsatellite map allowed for improved accuracy in QTL studies. Two loci involved in leaf surface components (cis-abienol and sucrose esters) were identified by Vontimitta et al. (2010). These compounds contribute to flavor and aroma characteristics of certain tobacco types, such as Oriental. The mapping of the *NtCPS2* gene at the cis-abienol locus was confirmed by Sallaud et al. (2012). *NtCPS2* is a major gene controlling cis-abienol production in tobacco. The survey of tobacco germplasm also showed that polymorphism in this gene accounts for the presence or absence of cis-abienol and labdenediol in 90% of the varieties. Several labdanoid diterpenes have been reported to exhibit antifungal activity (Sallaud et al. 2012). For this reason, involvement of cis-abienol or related compounds in resistance to the black shank tobacco disease is possible (Vontimitta and Lewis 2012). One major QTL for the resistance to black shank, which is caused by the soilborne pathogen *Phytophthora parasitica* var. *nicotiana*, cosegregates with cis-abienol (Vontimitta and Lewis 2012).

Other resistance loci to soilborne diseases such as black shank and bacterial wilt have also been discovered recently (Drake-Stowe et al. 2017). The two largest QTLs affecting resistance to black shank overlapped almost perfectly with the two QTLs with greatest effects against bacterial wilt. Further studies are necessary to determine whether the same genes affect resistance to both diseases or whether selective breeding produced favorable coupling-phase linkages between resistances. Nonetheless, the number of loci and quantitative nature of the resistance make it difficult to breed for resistance to these diseases. Therefore, it is of foremost importance for tobacco breeding programs to identify the genomic regions involved in disease resistance and their genetic markers. The availability of the genome sequence may allow for identification of the genes involved in resistance.

The development of SNP arrays may contribute to an increase in genetic mapping publications. These arrays are typically cheaper to screen than microsatellite markers. The marker density is also much higher, and SNP markers are linked to contigs in the genome, facilitating the identification of candidate genes, as demonstrated by Edwards et al. (2017) for the chlorophyll deficiency phenotype of Burley tobacco. This trait is conferred by a double homozygous recessive genotype at the *Yellow Burley 1* (*YB1*) and *Yellow Burley 2* (*YB2*) loci (Henika 1932; Clausen and Cameron 1944; Stines and Mann 1960). Although known for long time, the genes controlling this phenotype were only identified by genotypic analysis of three pairs of nearly isogenic lines, differing from their respective parents at genomic regions carrying *YB1* and *YB2* loci (Lewis et al. 2012). These loci were located on chromosomes Nt5 (*N. sylvestris*

origin) and Nt24 (*N. tomentosiformis* origin), as reported previously by Wu et al. (2014). The identification of the genes involved in the Burley phenotype at these loci was conducted using the SNP genetic map, the genome sequence, and a candidate gene approach. Homologs of the peptidase-encoding gene *AtEGY1* from *Arabidopsis thaliana* were identified as strong candidate genes, as they contained sequence polymorphisms predicted to result in truncated proteins in the Burley TN90 alleles but not in flue-cured K326 alleles (Edwards et al. 2017). The synteny between these loci suggests that *NtEGY1* and *NtEGY2* are homeologous. This is the first reported example of genes explaining the differentiation of the Burley tobacco market class.

In the foreseeable future, genetic studies using QTL mapping or association mapping based on improved genome assemblies and SNP arrays may notably increase our understanding of the tobacco genetic resource. This will accelerate aspects of tobacco research that were slowed by the complexity of the genome, with benefits in agronomy for breeders and farmers and in bio-pharmaceutical and biofuel production.

References

Bindler G, van der Hoeven R, Gunduz I et al (2007) A microsatellite marker based linkage map of tobacco. Theor Appl Genet 114:341–349

Bindler G, Plieske J, Bakaher N et al (2011) A high density genetic map of tobacco (*Nicotiana tabacum* L.) obtained from large scale microsatellite marker development. Theor Appl Genet 123:219–230

Clausen RE, Cameron DR (1944) Inheritance in *Nicotiana tabacum.* XVIII. Monosomic analysis. Genetics 29:447–477

Clayton EE (1958) The genetics and breeding progress in tobacco during the last 50 years. Agron J 50:352–356

Davalieva K, Maleva I, Filiposki K, Spiroski O, Efremov GD (2010) Genetic variability of Macedonian tobacco varieties determined by microsatellite marker analysis. Diversity 2:439–449

Drake-Stowe K, Bakaher N, Goepfert S et al (2017) Multiple disease resistance loci affect soilborne disease resistance in tobacco (*Nicotiana tabacum*). Phytopathology 107(9):1055–1061

Edwards KD, Fernandez-Pozo N, Drake-Stowe K et al (2017) A reference genome for *Nicotiana tabacum* enables map-based cloning of homeologous loci implicated in nitrogen utilization efficiency. BMC Genom 18(1):448

Fricano A, Bakaher N, Corvo MD et al (2012) Molecular diversity, population structure, and linkage disequilibrium in a worldwide collection of tobacco (*Nicotiana tabacum* L.) germplasm. BMC Genet 13:18

Gadani F, Hayes A, Opperman CH et al (2003) Large scale genome sequencing and analysis of *Nicotiana tabacum*: the tobacco genome initiative. 5th Bergerac Tobacco Scientific Meeting, Institut du Tabac-Bergerac, Dordogne, France, pp 117–130

Garner WW, Allard AA, Clayton EE (1937) Superior germplasm in tobacco. In: USDA yearbook No. 1580, pp 785–830

Goodman J (ed) (2005) Tobacco in history and culture: an encyclopedia, vol 2. Thomson Gale, Michigan. ISBN 0-684-31453-3

Henika FS (1932) The inheritance of the white burley character in tobacco. J Agric Res 44:477–493

Julio E, Denoyes-Rothan B, Verrier JL, Dorlhac Borne F (2006) Detection of QTLs linked to leaf and smoke properties in *Nicotiana tabacum* based on a study of 114 recombinant inbred lines. Mol Breed 18:69–91

Killebrew JB (1884) Report on the culture and curing of tobacco in the United States. US Government Printing Office, Washington

Lewis RS, Parker RG, Danehower DA et al (2012) Impact of alleles at the *Yellow Burley* (*Yb*) loci and nitrogen fertilization rate on nitrogen utilization efficiency and tobacco-specific nitrosamine (TSNA) formation in air-cured tobacco. J Agric Food Chem 60:6454–6461

Liu KJ, Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21:2128–2129

Moon HS, Nifong JM, Nicholson JS et al (2009a) Microsatellite-based analysis of tobacco (*Nicotiana tabacum L.*) genetic resources. Crop Sci 49:2149–2159

Moon HS, Nicholson JS, Heineman A et al (2009b) Changes in genetic diversity of U.S. flue-cured tobacco germplasm over seven decades of cultivar development. Crop Sci 49:498–508

Murphy JP, Cox TS, Rufty RC, Rodgers DM (1987) A representation of the pedigree relationships among flue-cured tobacco cultivars. Tob Sci 31:70–75

Sallaud C, Giacalone C, Töpfer R et al (2012) Characterization of two genes for the biosynthesis of the labdane diterpene Z-abienol in tobacco (*Nicotiana tabacum*) glandular trichomes. Plant J 72(1):1–17. Erratum in: Plant J (2013) 74(4):713

Sierro N, Battey JND, Ouadi S et al (2014) The tobacco genome sequence and its comparison with those of tomato and potato. Nat Commun 5:3833

Sim SC, Durstewitz G, Plieske J et al (2012) Development of a large SNP genotyping array and generation of high-density genetic maps in tomato. PLoS ONE 7 (7):e40563

Stines BJ, Mann TJ (1960) Diploidization in *Nicotiana tabacum*: a study of the yellow burley character. J Hered 51:222–237

Tong ZJ, Yang ZM, Chen XJ et al (2012) Large-scale development of microsatellite markers in *Nicotiana tabacum* and construction of a genetic map of flue-cured tobacco. Plant Breed 131:674–680

Tong Z, Xiao B, Jiao F et al (2016) Large-scale development of SSR markers in tobacco and construction of a linkage map in flue-cured tobacco. Breed Sci 66(3):381–390

Vontimitta V, Danehower DA, Steede T, Moon HS, Lewis RS (2010) Analysis of a *Nicotiana tabacum* L. genomic region controlling two leaf surface chemistry traits. J Agric Food Chem 58:294–300

Vontimitta V, Lewis RS (2012) Mapping of quantitative trait loci affecting resistance to *Phytophthora nicotianae* in tobacco (*Nicotiana tabacum* L.) line Beinhart-1000. Mol Breed 29:89–98

Vos PG, Uitdewilligen JG, Voorrips RE, Visser RG, van Eck HJ (2015) Development and analysis of a 20K SNP array for potato (*Solanum tuberosum*): an insight into the breeding history. Theor Appl Genet 128 (12):2387–2401

Wang X, Yang S, Chen Y et al (2018) Comparative genome-wide characterization leading to simple sequence repeat marker development for *Nicotiana*. BMC Genom 19:500

Wolf A, Wolf FT (1948) The origin of tobaccos of the oriental type. Bull Torrey Bot Club 75:51–55

Wu F, Eannetta NT, Xu Y et al (2010) COSII genetic maps of two diploid Nicotiana species provide a detailed picture of synteny with tomato and insights into chromosome evolution in tetraploid *N. tabacum*. Theor Appl Genet 120:809–827

Wu Q, Wu X, Zhang X et al (2014) Mapping of two white stem genes in tetraploid common tobacco (*Nicotiana tabacum* L.) Mol Breed 34:1065–1074

Xiao B, Tan Y, Long N et al (2015) SNP-based genetic linkage map of tobacco (*Nicotiana tabacum* L.) using next-generation RAD sequencing. J Biol Res 22:11

Check for updates

4 Characterizing the Genome of *Nicotiana tabacum*

James N. D. Battey, Nicolas Sierro and Nikolai V. Ivanov

Abstract

The tobacco plant is an important crop and model organism, and there is a widespread interest in improving its agronomical properties. Unraveling its genome is necessary for understanding and predicting its biological properties and to ultimately contribute to breeding or engineering efforts for creating new varieties. Here, we discuss the key motivations behind the sequencing of its genome and the current state of genome sequencing efforts, as well as how it has been put to use. We finally speculate on what genomic trends relating to tobacco may be of interest in the near future.

4.1 Interest in the Tobacco Genome Sequence

Crop improvement has been and will remain a central challenge in agriculture, and breeding programs are a necessary way for crop improvement. Not only is simple performance enhancement economically desirable, but it is also necessary to introduce disease and stress resistance in otherwise susceptible varieties, which are necessary to maintain yields in increasingly challenging and ever-changing environments. Plant cultivars with new, desirable properties are in high demand, and growing plants for nontraditional uses such as expressing transgenic proteins has become a recent focus in the industry. Plant genomes have established themselves as a central resource in crop improvement (Bevan et al. 2017). Not only do they allow potential genes to be identified on a large scale by means of homology search, they also allow for the structural grouping of genomic features such as single-nucleotide polymorphisms with functional regions, thus allowing functionally variant sequences to be found. Coupled with a large catalogue of variant loci from a diverse background population, this can be used to accelerate plant breeding.

Tobacco (*Nicotiana tabacum*), in particular, has garnered interest from a broad range of research fields, because not only is it an economically important crop plant but—owing to its comparatively short generation time, biochemical complexity, and ease of manipulation—it has also become a model organism within the Solanaceae (Gebhardt 2016) and the plant kingdom. Its model organism status also stems from it being the source of one of the most used plant cell lines, BY-2 (Nagata et al. 2004). Its ease of manipulation has also made it an attractive plant for investigating the expression of transgenic

J. N. D. Battey (✉) · N. Sierro · N. V. Ivanov
PMI R&D, Philip Morris Products S.A., Quai Jeanrenaud 5, Neuchâtel 2000, Switzerland
e-mail: james.battey@pmi.com

© Springer Nature Switzerland AG 2020

N. V. Ivanov et al. (eds.), *The Tobacco Plant Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-030-29493-9_4

proteins from the earliest stages (Hiatt et al. 1989). Sequencing the tobacco genome allows the identification of trait markers that are essential for breeding purposes and gene-finding exercises that allow the identification of functions for modification, thus leading to new tobacco cultivars with new desirable characteristics.

The tobacco plant is known to be an allotetraploid, with most of its genetic information stemming from two closely related progenitor species, *Nicotiana sylvestris* and *Nicotiana tomentosiformis*. As such, assembling its genome constituted a significant challenge, as there are difficulties not only in bridging highly repetitive repeat regions but also in computationally disentangling closely related homologous sequences from the two progenitor species.

4.2 Non-assembly Based Methods for Investigating the Genome

Prior to full-fledged sequencing efforts, various mapping techniques have been used to pinpoint genetic markers to aid breeding efforts. Most early techniques identified simple sequence repeats (SSR) that could be used as markers for mapping the tobacco genome. An initial microsatellite-based method using a cross of Hicks Broad Leaf and Red Russian has identified close to 300 loci, which could be mapped to 24 tentative linkage groups in the genome, covering a total of 1,920 cM of the genome; this formed the basis for trait mapping in tobacco (Bindler et al. 2007). A later effort, using the same crossing of varieties, identified over 5,000 functional microsatellite markers, yielding a far more detailed view of the tobacco genome (Bindler et al. 2011). Since then, these efforts have been extended to other cultivars, starting with individual varieties (Tong et al. 2012); their subsequent extension to multiple tobacco cultivars has increased the number of known markers and candidate species (Tong et al. 2016).

More advanced techniques have harnessed modern sequencing technologies without performing full genome assembly. The Whole-Genome Profiling (WGP™) technology aims to create a physical map of the genome: The process involves the creation of several bacterial artificial chromosome (BAC) libraries of a target genome by using various restriction enzymes and sequencing the ends of these BACs—the so-called WGP tags. The tag information can be used to assemble contigs into more complex units containing the order of the BACs within each contig. Application of this technology to the tobacco genome not only yielded a minimum tiling path—that is, allowing those BACs to be identified which together constitute the theoretical minimal set of BACs needed to obtain complete coverage of the genome sequence—but also allowed the origin (*N. tomentosiformis* or *N. sylvestris*) of BACs to be determined for these sequences (Sierro et al. 2013b) without the full sequence being available. Information from such a physical map can also be used to match short scaffolds derived from shotgun short-read sequencing to WGP contigs and use this information to join the scaffolds into longer super scaffolds.

4.3 Genome and Transcriptome Assembly Efforts

Preparations for the tobacco genome sequencing effort had already commenced as early as 2003 (Opperman et al. 2003), and, in 2008, portions of the tobacco genome had been sequenced by the Tobacco Genome Initiative (News 2008). While the draft was far from complete, its sequences have nevertheless constituted a useful source of information for biologists (Rabara et al. 2015) and set the stage for later efforts in the area. Other, earlier sequencing efforts focused on obtaining the transcriptomic sequences of *N. tabacum* (Bombarely et al. 2012), because this allows the assembly of transcript models, which contain a large proportion of the protein-coding information in the tobacco genome.

As a precursor to the sequencing of the tobacco genome, its two putative ancestral species, *N. sylvestris* and *N. tomentosiformis* were sequenced and assembled individually (Sierro

et al. 2013a). Both genomes reached 83% and 72% of their expected size of 2.7 and 2.4 GB, respectively. In both species, a higher number of transcripts were reported to be expressed in the flower than in the root or leaf. In *N. sylvestris*, 53,247 flower-expressed transcripts were reported, while only 46,114 leaf- and 46,313 root-expressed transcripts were observed. In *N. tomentosiformis*, 48,043 flower-, 43,743 leaf-, and 44,169 root-expressed transcripts were reported.

This effort formed the basis for assembling the tobacco genome, because it would now be possible to assign most portions of the genome to one of the ancestral genomes. Subsequently, three representatives of the main market classes of tobacco—the flue-cured Kentucky 326 (K326), air-cured Tennessee 90 (TN90), and sun-cured Basma Xanthi—were sequenced (Sierro et al. 2014). These three varieties were sequenced at coverages of 38x, 49x, and 29x, respectively, by using Illumina short paired-end reads. The expected genome sizes were estimated by using k-mer distribution to be 4.60, 4.41, and 4.57 GB, respectively, and the genome assemblies covered 81.1%, 84.3%, and 81.8% of these sizes. A follow-up study employed advances in optical genetic mapping based on the BioNano Genomics technology to obtain new scaffolding information for anchoring *N. tabacum* scaffolds derived from massively parallel short-read sequencing to chromosome-level pseudomolecules (Edwards et al. 2017). The tobacco genome assembly thus obtained presented an improved N50 length of 2.17 Mb, with 64% of the genome anchored on 24 pseudomolecules.

4.4 Findings of Genome Analysis

The *N. tabacum* genome shows a low degree of divergence from those of its progenitor species. Microsynteny with other Solanaceae species was demonstrated for *N. tabacum* genomic regions (Sierro et al. 2014). More than 90,000 gene models were identified in the *N. tabacum* genome (Sierro et al. 2014). Repetitive sequences represent more than 70% of the genomic DNA of cultivated *N. tabacum* genotypes (Kenton et al. 1993, Narayan 1987, Sierro et al. 2014, Zimmerman and Goldberg 1977), which implies that a fairly small proportion of the genome is responsible for phenotypic and genetic variability. The estimated genome size of *N. tabacum* and the sum of the estimates for *N. sylvestris* and *N. tomentosiformis* suggest that a downsizing of 4–8% of the tobacco genome occurred post-polyploidization (Sierro et al. 2014). This is in agreement with an earlier estimate of a 3.7% downsizing on the basis of flow cytometry and Feulgen microdensitometry findings (Leitch et al. 2008). The downsizing is thought to affect the repetitive elements of the genome and to have been more frequent in the paternal-derived T-genomic regions than in S-genomic regions (Renny-Byfield et al. 2011, Sierro et al. 2014). Whole-genome sequence comparison shows that *N. sylvestris* and *N. tomentosiformis* contribute 53% and 47%, respectively, to the complete *N. tabacum* genome, further suggesting that genomic reduction has been biased to the T genome. Whether coding regions have contributed to this genomic reduction has not yet been investigated.

The *N. tabacum* genome sequences confirm the hypothesized ancestry of the species, previously supported by genetic and physical maps (Bindler et al. 2011, Sierro et al. 2013a). *N. sylvestris* (S genome) and *N. tomentosiformis* (T genome) are thought to be the presumed maternal and paternal species, respectively (Bland et al. 1985, Leitch et al. 2008). Whole-genome shotgun sequencing and phylogenetic analysis of chloroplast DNA (cpDNA) sequence data support the argument that the progenitor of *N. tomentosiformis* was the likely paternal genome donor and that of *N. sylvestris* the likely maternal genome donor to *N. tabacum* (Clarkson et al. 2004, Sierro et al. 2014, Yukawa et al. 2006). Accessions of *N. tabacum* share an identical internal transcribed spacer (ITS) sequence with *N. tomentosiformis* (Chase et al. 2003). Molecular cytogenetic evidence and amplified fragment-length polymorphism (AFLP) diversity indicate that *N. tomentosiformis* comprises two lineages differentiated by 45S rDNA distribution,

noncoding tandem repeat sequences, and the tandem repeats GRD53 and GRD3 and that the T genome of *N. tabacum* is genetically closer to one than the other (Murad et al. 2002). Skalická et al. concluded from indirect genetic evidence that a single ancestral polyploid origin for *N. tabacum* is most likely (Skalicka et al. 2005). The genome of *N. sylvestris* has been suggested to be more similar to the ancestral *Nicotiana* genome than to the genomes of other extant species (Chase et al. 2003), but this hypothesis is not supported by its placement in phylogenetic reconstructions from ITS or cpDNA sequence data (Chase et al. 2003, Clarkson et al. 2004).

More than 90% of SSR markers developed for *N. tabacum* are specific to one or two loci of the ancestral genomes; the availability of markers mapping to only one locus indicates that the two progenitor-derived genomes possess a high degree of genetic diversity (Kovarik et al. 2012). The larger genome sizes of *N. sylvestris* (2.4 Gb) and *N. tomentosiformis* (2.7 Gb) compared with those of tomato (900 Mb) and potato (844 Mb) are suggested to be the result of expansion of repeat sequences, mostly transposable elements (Sierro et al. 2013a). Repetitive elements have been characterized in the genomes of *N. sylvestris* and *N. tomentosiformis*, and they show notable differences (Renny-Byfield et al. 2011, Sierro et al. 2013a). The genome of *N. sylvestris* contains considerably more repeats of high homogeneity, whereas the *N. tomentosiformis* genome shows greater heterogeneity in repetitive elements (Renny-Byfield et al. 2011). The *N. sylvestris* genome contains higher proportions of four specific retrotransposons (Tnt1-OL13, Tnt1-OL16, Tnt2d, and Tto1-1R) than the *N. tomentosiformis* genome (Petit et al. 2007). However, certain repetitive elements are more frequent in the *N. tomentosiformis* genome than in the *N. sylvestris* genome (Renny-Byfield et al. 2011). These characteristics indicate that the two genomes have followed markedly different evolutionary paths since the divergence of the *Nicotiana* approximately 15 million years ago (Clarkson et al. 2004).

4.5 Uses of the Tobacco Genome

Derivative bioinformatics resources. The genomes of newly sequenced species need to be annotated in order to be useful for biological research. Some of the key tools in this function are functional annotation tools, which can be linked to more complex downstream analysis for placing this information in a biological context. Biochemical pathways perform such a role because they link enzymatic function information into larger metabolic pathways. Currently, work on tobacco biochemical networks is being performed (Foerster et al. 2018) which involves combining information from automated annotation pipelines with that from the literature.

Gene finding. While a few model plant genomes have been annotated extensively, the tobacco genome remains a relatively new territory for functional annotation owing to the lack of experimental evidence. Gene finding has become one of the primary uses of the genome in recent years, with most gene-finding methods employing homology-based methods for gene identification, usually coupled with comparative evolutionary analysis with other species. A number of recent studies have reported such genome-wide characterization of gene families in tobacco, such as those of the expansin gene family (Ding et al. 2016), the stress-related Hsp70 gene family (Song et al. 2019), autophagy-related genes (Zhou et al. 2015), and the disease-resistance-related R-gene family (Wei et al. 2016; Long et al. 2016).

Genome modification. A new avenue of investigation has been opened with the advent of the CRISPR-Cas9 system for introducing targeted mutations. This system requires genome-wide knowledge for finding the target genes, obtaining their exact sequences, and filtering potential target sites for off-target matches. Other studies have used an as-yet unpublished version of the genome to perform manipulations of phytoene metabolism by targeting the relevant synthesis (PDS) and transport (PDR6) genes

(Gao et al. 2015). Target sequences with no detectable off-site hits were identified by searching this genome; this process was used by the same group that targeted the CCD8 gene (Gao et al. 2018). More recently, the CRISPR Cas9 system was used to re-engineer tobacco to shut down nicotine production by simultaneously targeting all BBL genes for inactivation (Schachtsiek and Stehle 2019); in order to do so, target genes were identified by finding common sequences matching no other parts of the genome.

4.6 Future Possibilities

The current drafts of the genome are hampered by the typical limitations of available technology. Particularly, current short-read technology cannot bridge repeat-rich regions; the sequences of genomes containing many such regions will thus contain long, unresolved stretches. This not only causes the genomes to be fragmented but also means that many of the assembled fragments cannot be anchored to chromosomes. This limits the biological interpretation of scientific results and makes it harder to link markers to actual genes. Third-generation sequencing technologies promise to overcome the limitations of short-read technologies.

Long-read technologies, currently mostly produced by Pacific Biosciences and Oxford Nanopore, offer the ability to obtain far longer read lengths. To date, reads of over 100,000 base pairs in length have been obtained by using Oxford Nanopore's Minion technology (Tyson et al. 2018). These increasingly long reads promise to bridge the gaps in the current short-read-assembled genomes, allowing for much longer read fragments, which, in turn, can be anchored to chromosomes by using known markers. Another advantage of long-read technology is its ability to sequence long stretches of the genome, which allows mutations to be not only identified but also assigned to individual chromosomes. This allows not only the correct deconvolution of polyploid genomes (Jiao and Schneeberger 2017) but also the haplotyping of tobacco variants, thus improving the speed and efficiency of tobacco-breeding projects. As more and more varieties of tobacco become available, the number of known genetic variants will grow, allowing for more highly detailed genetics maps to be established.

The plant microbiome, or the sum of all microbes associated with a given plant, may have strong implications for the survival and health of the plant. While pathological interactions are well known—the tobacco mosaic virus was the first plant virus to be discovered (Lefeuvre et al. 2019)—the beneficial effects of higher microbes such as bacteria have been shown, even in tobacco. For example, it has been shown that certain species of microbes can induce plant immunity against diseases in tobacco (Kim et al. 2015). In recent years, there has been an interest in applying sequencing-based microbiome characterization methods, i.e., microbiomics, for analyzing microbial communities on or even in plants (Knief 2014). As the field becomes more mature, the techniques pioneered on other plants can be expected to be applied to tobacco; indeed multiple reports to this effect have appeared in recent years (e.g., microbiomic investigation of fungal species and their dependence upon environmental factors) (Yuan et al. 2018).

4.7 Conclusions

Sequencing of the tobacco genome has provided invaluable insights into its functional composition and organization. Over the years, increasingly sophisticated techniques have been used to assemble the tobacco genome sequence. This has allowed for large-scale gene finding and, more recently, aided in the selection of targets for CRISPR Cas9-based editing. Future developments will include characterization of genetic variation both within tobacco and in its close relatives in the hope of discovering biologically valuable genetic elements. The recent interest in characterizing the microbiome will presumably lead to many sequencing-based investigations of the immediate microorganismal environment of the tobacco plant.

References

Bevan MW, Uauy C, Wulff BB, Zhou J, Krasileva K, Clark MD (2017) Genomic innovation for crop improvement. Nature 543:346–354

Bindler G, Plieske J, Bakaher N, Gunduz I, Ivanov N, van der Hoeven R, Ganal M, Donini P (2011) A high density genetic map of tobacco (*Nicotiana tabacum* L.) obtained from large scale microsatellite marker development. Theor Appl Genet 123:219–230

Bindler G, van der Hoeven R, Gunduz I, Plieske J, Ganal M, Rossi L, Gadani F, Donini P (2007) A microsatellite marker based linkage map of tobacco. Theor Appl Genet 114:341–349

Bland M, Matzinger D, Levings C (1985) Comparison of the mitochondrial genome of *Nicotiana tabacum* with its progenitor species. Theor Appl Genet 69:535–541

Bombarely A, Edwards KD, Sanchez-Tamburrino J, Mueller LA (2012) Deciphering the complex leaf transcriptome of the allotetraploid species *Nicotiana tabacum*: a phylogenomic perspective. BMC Genom 13:406

Chase MW, Knapp S, Cox AV, Clarkson JJ, Butsko Y, Joseph J, Savolainen V, Parokonny AS (2003) Molecular systematics, GISH and the origin of hybrid taxa in *Nicotiana* (Solanaceae). Ann Bot 92:107–127

Clarkson JJ, Knapp S, Garcia VF, Olmstead RG, Leitch AR, Chase MW (2004) Phylogenetic relationships in *Nicotiana* (Solanaceae) inferred from multiple plastid DNA regions. Mol Phylogenet Evol 33:75–90

Ding A, Marowa P, Kong Y (2016) Genome-wide identification of the expansin gene family in tobacco (*Nicotiana tabacum*). Mol Genet Genomics 291: 1891–1907

Edwards KD, Fernandez-Pozo N, Drake-Stowe K, Humphry M, Evans AD, Bombarely A, Allen F, Hurst R, White B, Kernodle SP, Bromley JR, Sanchez-Tamburrino JP, Lewis RS, Mueller LA (2017) A reference genome for *Nicotiana tabacum* enables map-based cloning of homeologous loci implicated in nitrogen utilization efficiency. BMC Genom 18:448

Foerster H, Bombarely A, Battey JND, Sierro N, Ivanov NV, Mueller LA (2018) SolCyc: a database hub at the Sol Genomics Network (SGN) for the manual curation of metabolic networks in Solanum and Nicotiana specific databases. Database (Oxford)

Gao J, Wang G, Ma S, Xie X, Wu X, Zhang X, Wu Y, Zhao P, Xia Q (2015) CRISPR/Cas9-mediated targeted mutagenesis in *Nicotiana tabacum*. Plant Mol Biol 87:99–110

Gao J, Zhang T, Xu B, Jia L, Xiao B, Liu H, Liu L, Yan H, Xia Q (2018) CRISPR/Cas9-mediated mutagenesis of carotenoid cleavage dioxygenase 8 (CCD8) in tobacco affects shoot and root architecture. Int J Mol Sci 19

Gebhardt C (2016) The historical role of species from the Solanaceae plant family in genetic research. Theor Appl Genet 129:2281–2294

Hiatt A, Cafferkey R, Bowdish K (1989) Production of antibodies in transgenic plants. Nature 342:76–78

Jiao WB, Schneeberger K (2017) The impact of third generation genomic technologies on plant genome assembly. Curr Opin Plant Biol 36:64–70

Kenton A, Parokonny AS, Gleba YY, Bennett MD (1993) Characterization of the *Nicotiana tabacum* L. genome by molecular cytogenetics. Mol Gen Genet 240:159–169

Kim JS, Lee J, Lee CH, Woo SY, Kang H, Seo SG, Kim SH (2015) Activation of pathogenesis-related genes by the Rhizobacterium, Bacillus sp. JS, which induces systemic resistance in tobacco plants. Plant Pathol J 31:195–201

Knief C (2014) Analysis of plant microbe interactions in the era of next generation sequencing technologies. Front Plant Sci 5:216–216

Kovarik A, Renny-Byfield S, Grandbastien M-A, Leitch A (2012) Evolutionary implications of genome and karyotype restructuring in *Nicotiana tabacum* L. In: Soltis PS, Soltis DE (eds) Polyploidy and genome evolution. Springer, Berlin, Heidelberg

Lefeuvre P, Martin DP, Elena SF, Shepherd DN, Roumagnac P, Varsani A (2019) Evolution and ecology of plant viruses. Nat Rev Microbiol 1

Leitch IJ, Hanson L, Lim KY, Kovarik A, Chase MW, Clarkson JJ, Leitch AR (2008) The ups and downs of genome size evolution in polyploid species of *Nicotiana* (Solanaceae). Ann Bot 101:805–814

Long N, Ren X, Xiang Z, Wan W, Dong Y (2016) Sequencing and characterization of leaf transcriptomes of six diploid *Nicotiana* species. J Biol Res Thessalon 23:6

Murad L, Lim KY, Christopodulou V, Matyasek R, Lichtenstein CP, Kovarik A, Leitch AR (2002) The origin of tobacco's T genome is traced to a particular lineage within *Nicotiana tomentosiformis* (Solanaceae). Am J Bot 89:921–928

Nagata T, Hasewa S, Inzé D (eds) (2004) Tobacco BY-2 cells. Springer

Narayan R (1987) Nuclear DNA changes, genome differentiation and evolution in *Nicotiana* (Solanaceae). Plant Syst Evol 157:161–180

News, NCSU (2008) NC State University maps tobacco genome [Online]. https://news.ncsu.edu/2008/06/nc-state-university-maps-tobacco-genome/. Accessed 19 Feb, 2019

Opperman CH, Lommel SA, Sosinski BR, Lakey N, Gadani F (2003) The tobacco genome initiative. In: CORESTA meeting, agronomy/phytopathology. Bucharest

Petit M, Lim KY, Julio E, Poncet C, De Borne FD, Kovarik A, Leitch AR, Grandbastien MA, Mhiri C (2007 July 1) Differential impact of retrotransposon populations onthe genome of allotetraploid tobacco (Nicotiana tabacum). Mol Genet Genomics 278(1): 1–5

Rabara RC, Tripathi P, Reese RN, Rushton DL, Alexander D, Timko MP, Shen QJ, Rushton PJ (2015) Tobacco drought stress responses reveal new targets

for Solanaceae crop improvement. BMC Genom 16:484

Renny-Byfield S, Chester M, Kovarik A, le Comber SC, Grandbastien MA, Deloger M, Nichols RA, Macas J, Novak P, Chase MW, Leitch AR (2011) Next generation sequencing reveals genome downsizing in allotetraploid *Nicotiana tabacum*, predominantly through the elimination of paternally derived repetitive DNAs. Mol Biol Evol 28:2843–2854

Schachtsiek J, Stehle F (2019) Nicotine-free, non-transgenic tobacco (*Nicotiana tabacum* L.) Edited by CRISPR-Cas9. Plant Biotechnol J

Sierro N, Battey JN, Ouadi S, Bakaher N, Bovet L, Willig A, Goepfert S, Peitsch MC, Ivanov NV (2014) The tobacco genome sequence and its comparison with those of tomato and potato. Nat Commun 5:3833

Sierro N, Battey JN, Ouadi S, Bovet L, Goepfert S, Bakaher N, Peitsch MC, Ivanov NV (2013a) Reference genomes and transcriptomes of *Nicotiana sylvestris* and *Nicotiana tomentosiformis*. Genome Biol 14:R60

Sierro N, van Oeveren J, van Eijk MJ, Martin F, Stormo KE, Peitsch MC, Ivanov NV (2013b) Whole genome profiling physical map and ancestral annotation of tobacco Hicks Broadleaf. Plant J 75:880–889

Skalicka K, Lim K, Matyasek R, Matzke M, Leitch A, Kovarik A (2005) Preferential elimination of repeated DNA sequences from the paternal, *Nicotiana tomentosiformis* genome donor of a synthetic, allotetraploid tobacco. New Phytol 166:291–303

Song Z, Pan F, Lou X, Wang D, Yang C, Zhang B, Zhang H (2019) Genome-wide identification and characterization of Hsp70 gene family in *Nicotiana tabacum*. Mol Biol Rep

Tong Z, Xiao B, Jiao F, Fang D, Zeng J, Wu X, Chen X, Yang J, Li Y (2016) Large-scale development of SSR markers in tobacco and construction of a linkage map in flue-cured tobacco. Breed Sci 66:381–390

Tong Z, Yang Z, Chen X, Jiao F, Li X, Wu X, Gao Y, Xiao B, Wu W (2012) Large-scale development of microsatellite markers in *Nicotiana tabacum* and construction of a genetic map of flue-cured tobacco. Plant Breed 131:674–680

Tyson JR, O'Neil NJ, Jain M, Olsen HE, Hieter P, Snutch TP (2018) MinION-based long-read sequencing and assembly extends the *Caenorhabditis elegans* reference genome. Genome Res 28:266–274

Wei C, Chen J, Kuang H (2016) Dramatic number variation of R genes in Solanaceae species accounted for by a few R gene subfamilies. PLoS ONE 11: e0148708

Yuan XL, Cao M, Liu XM, Du YM, Shen GM, Zhang ZF, Li JH, Zhang P (2018) Composition and genetic diversity of the *Nicotiana tabacum* microbiome in different topographic areas and growth periods. Int J Mol Sci 19

Yukawa M, Tsudzuki T, Sugiura M (2006) The chloroplast genome of *Nicotiana sylvestris* and *Nicotiana tomentosiformis*: complete sequencing confirms that the *Nicotiana sylvestris* progenitor is the maternal genome donor of *Nicotiana tabacum*. Mol Genet Genomics 275:367–373

Zhou XM, Zhao P, Wang W, Zou J, Cheng TH, Peng XB, Sun MX (2015) A comprehensive, genome-wide analysis of autophagy-related genes identified in tobacco suggests a central role of autophagy in plant response to various environmental cues. DNA Res 22:245–257

Zimmerman JL, Goldberg RB (1977) DNA sequence organization in the genome of *Nicotiana tabacum*. Chromosoma 59:227–252

Tobacco Resources in the Sol Genomics Network and *Nicotiana* Metabolic Databases

5

Hartmut Foerster and Lukas A. Mueller

Abstract

Genomic databases provide essential information to scientists worldwide, including genome sequences, gene annotations, genetic markers, and phenotypic information. In addition to collecting and disseminating information, many databases also actively curate their data using a number of approaches, to ensure the data represent current knowledge and correspond to accepted quality standards. In this chapter, we review genome databases for the *Nicotiana* clade, and survey databases designed to further understand the metabolism of members of this clade. Although *Nicotiana tabacum* and its relatives, such as *Nicotiana benthamiana*, have been used widely in plant research over the last few decades, relatively few online *Nicotiana* resources exist, especially when compared with Solanaceae model systems such as tomato (*Solanum lycopersicum*). Tobacco plants are a major component of databases such as the Sol Genomics Network (https://solgenomics.net/) and the SolCyc metabolic databases, on which we will largely focus in this chapter.

H. Foerster (✉) · L. A. Mueller
Boyce Thompson Institute, Cornell University, Ithaca, NY, USA
e-mail: hf263@cornell.edu

5.1 Seed Resources

Many *Nicotiana* plants are popular ornamentals and seeds are widely available for multiple species. We focus here on gene banks that are more oriented toward scientists. A *Nicotiana* gene bank at North Carolina State University holds more than 2,000 accessions of *Nicotiana* species from around the world. The catalog is accessible via the United States Department of Agriculture Germplasm Resources Information Network system (https://npgsweb.ars-grin.gov/gringlobal/site.aspx?id=25).

5.2 Available Genome Sequence Datasets for *Nicotiana* Species

Nicotiana genomes available today include *N. tabacum*, *N. tomentosiformis*, *N. sylvestris*, *N. benthamiana*, and *N. attenuata*. The genome information for most of these species is available from both GenBank (Benson et al. 2010) (Table 5.1) and the Sol Genomics Network (SGN; https://solgenomics.net) (Fernandez-Pozo et al. 2015; Mueller et al. 2005). The Sequence Read Archive of the National Center for Biotechnology Information (NCBI; https://www.ncbi.nlm.nih.gov/sra) contains the primary read data for many of the genomes, while SGN contains the assembled genomes and annotations.

The first *Nicotiana* genome that scientists attempted to sequence was that of *N. tabacum*.

© Springer Nature Switzerland AG 2020

N. V. Ivanov et al. (eds.), *The Tobacco Plant Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-030-29493-9_5

Table 5.1 *Nicotiana* genomes in Genbank

Species	ID	Total length (Mb)	Protein count	Submitter
Nicotiana tomentosiformis	12239	1688.47	48,963	Philip Morris International R&D
Nicotiana otophora	32281	2689.35	NP	Philip Morris International R&D
N. tabacum TN90	425	3643.47	84,255	Philip Morris International R&D
N. tabacum K326	425	3732.64	NP	Philip Morris International R&D
N. tabacum Basma Xanthi	425	3735.82	NP	Philip Morris International R&D
N. tabacum K326	425	4646.65	NP	British American Tobacco
Nicotiana obtusifolia	53488	1222.77	NP	Max Planck Institute for Chemical Ecology
Nicotiana sylvestris	13135	2221.99	48,160	Philip Morris International R&D
Nicotiana benthamiana	10940	61.9511	NP	NAAS
Nicotiana attenuata UT	13243	2365.68	44,491	Max Planck Institute for Chemical Ecology
N. attenuata Arizona	13243	1827.78	NP	Max Planck Institute for Chemical Ecology
Nicotiana glauca	67062	3222.83	NP	Saint Petersburg University

NP—information not provided

However, the complexity of the genome and limitations of the sequencing technology available at the time made it clear that special methods would be required. In the early 2000s, a project was launched by Philip Morris International to sequence the gene-rich regions using a methylation filtration approach (Whitelaw et al. 2003) in conjunction with Sanger sequencing. In the methylation filtration approach, highly repetitive peri-centromeric repeats that are predominantly methylated are excluded from sequencing using methylation-sensitive restriction enzymes. These enzymes cut only non-methylated restriction sites, which are then cloned and sequenced, and primarily represent the gene-containing euchromatin. While the methylation filtration approach was successful, the throughput of the underlying Sanger sequencing method proved to be insufficiently high to achieve a contiguous assembly given the genome size, and the work remained unpublished. The sequence data from this effort are available as a BLAST dataset on SGN and can also be downloaded from the FTP site. An extensive database of tobacco transcription factors, TOBFAC, was created from this effort (Rushton et al. 2008), but neither the tobaccogenome.org website for the genome project nor the TOBFAC website appears to be maintained.

More recently, two efforts were undertaken to sequence the *N. tabacum* genome (Edwards et al. 2017; Sierro et al. 2014) using next-generation short-read approaches. Both projects sequenced the tetraploid *N. tabacum* and the diploid parental genomes *N. sylvestris* and *N. tomentosiformis* to varying degrees of completion. For *N. tabacum*, both projects sequenced the K326 variety, a high-yielding, flue-cured tobacco. Sequencing of *N. otophora*, another species that has been proposed as a parent of *N. tabacum*, was also obtained by Sierro et al. (Sierro et al. 2014), but alignment to the tobacco reference sequence showed this species to be an unlikely candidate for a parental genome. Sierro et al. (Sierro et al. 2014) also sequenced TN90, a burley tobacco, and BX, an oriental tobacco.

All sequence and annotation data from these projects are available on SGN for download and BLAST and annotation searches.

A number of other notable *Nicotiana* genomes have been sequenced. *N. benthamiana*'s natural habitat is Australia; it is an important laboratory plant used for a number of assays such as virus-induced gene silencing (Baulcombe 1999), particularly in the field of plant disease resistance. It also has a complex tetraploid genome and, with *N. sylvestris* as the maternal progenitor, it shares one of the parents of *N. tabacum*. The *N. benthamiana* genome was sequenced using short-read next-generation technology (Bombarely et al. 2012). The genome sequence is still being improved and the plant continues to be the focus of genomic, phylogenetic, and molecular genetic research, summarized by Bally et al. (Bally et al. 2018) and in Chap. 14 of this book.

N. attenuata grows in the southwestern United States and is a model plant in ecology. Its plant–insect interactions have been studied extensively (Kessler and Baldwin 2001), And the genome has recently been sequenced (Xu et al. 2017).

5.3 Metabolic Database Resources in SGN

In addition to the genomic and phenotypic data provided on the Solanaceae family (Bombarely et al. 2011), SGN also hosts SolCyc (https://solgenomics.net/tools/solcyc/index.pl), a Solanaceae-specific metabolic database collection (Foerster et al. 2018). The SolCyc databases are species-specific and termed pathway/genome databases (PGDBs). They have been computationally created from annotated genomes or SGN unigene sequences. The automatic build of the PGDBs is performed by the Pathway Tools software suite (Karp et al. 2010) using MetaCyc (http://www.metacyc.org/) as the reference database (Caspi et al. 2016). MetaCyc is a manually curated, multi-species database spanning all domains of life, and aims to be a universal repository that stores representative samples of experimentally verified pathways. These pathways are furnished with enzymes and genes and provide structural data about compounds and in-depth information of pathway-related aspects such as regulation, key enzymes, and rate-limiting steps (Fig. 5.1). MetaCyc derivative databases are equipped with a wide spectrum of search, visualization, and analysis tools, which makes them valuable resources for biochemical and metabolic data storage and display, metabolic engineering, and bioinformatics applications (Paley and Karp 2006; Latendresse and Karp 2011; Toya et al. 2011; Dreher 2014).

SolCyc currently provides specific databases for tobacco (*N. tabacum*), tomato (*S. lycopersicum*), potato (*Solanum tuberosum*), pepper (*Capsicum annuum*), and petunia (*Petunia hybrida*), as well as a database for coffee (*Coffea Arabica*; *Rubiaceae*). All publicly accessible SolCyc databases belong to the lowest tier (tier 3) of BioCyc databases (http://biocyc.org/); this tier includes databases defined as computationally generated and not curated (Caspi et al. 2016). Although more than 13,000 MetaCyc-derived databases have been created to date, only a small fraction of them are maintained and updated on a regular basis or are subject to manual curation. The tier 1 category includes highly curated databases with at least a one-year literature-centered curation. The resulting high accuracy of the metabolic network and associated enzymatic, genomic, and metabolic information makes such databases the most reliable sources for studying complex and often interrelated metabolic processes (Mueller et al. 2003; Zhang et al. 2005; Walsh et al. 2016).

SGN has recently enhanced its metabolic database repertoire in SolCyc by adding five *Nicotiana*-specific PGDBs for *N. tabacum*, *N. tomentosiformis*, *N. sylvestris*, *N. benthamiana*, and *N. attenuata*. Moreover, two taxon-specific databases, *SolanaCyc* and *NicotianaCyc*, have been generated, which contain only curated and experimentally verified pathways of the Solanaceae family and *Nicotiana* genus, respectively. These databases are extensively curated and represent the currently most accurate

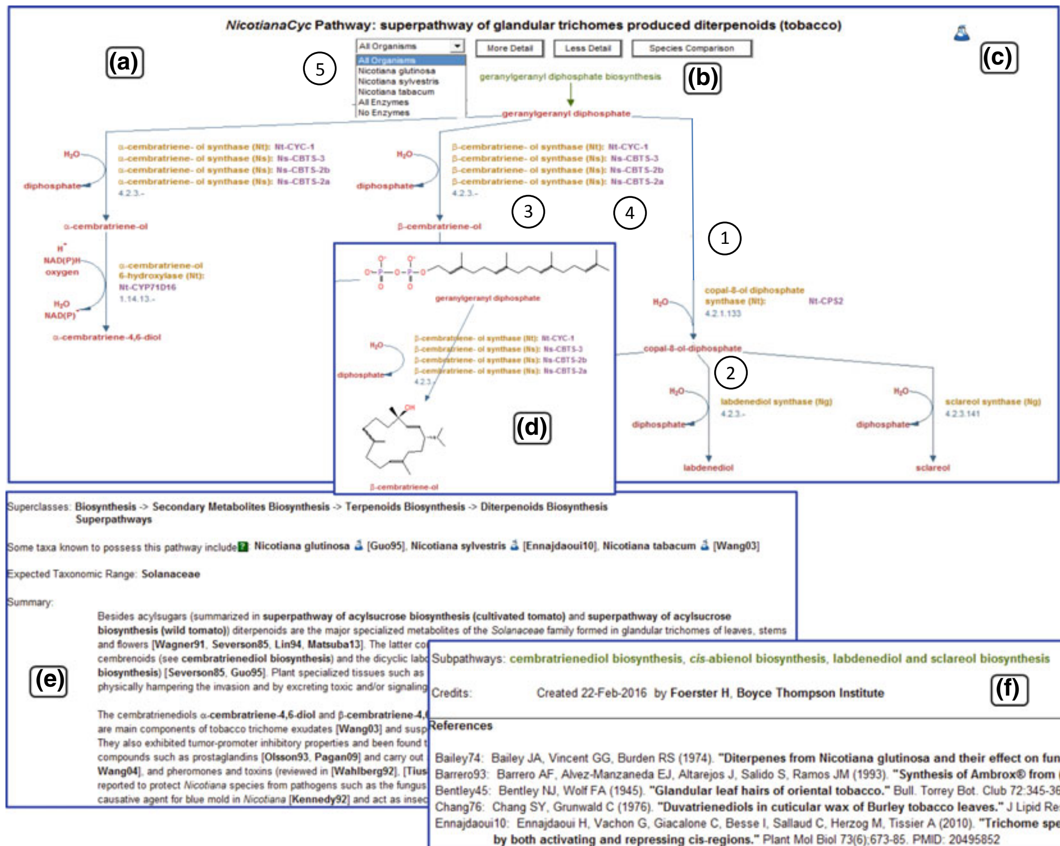

Fig. 5.1 Representation of the superpathway of glandular trichomes produced diterpenoids in *Nicotiana* species. **a** The pathway diagram displays a sequence of reactions (1), in which involved compounds (2) are converted, and catalyzing enzymes (3) and encoding genes (4) are accordingly associated. A dropdown menu (5) allows modifying the displayed information of the pathway. Green hyperlinks connect to related pathways that either feed into **b** or branch off the pathway. **c** The curation status of the pathway is visualized by evidence code icons. **d** Compound structures become available after increasing the detail of the pathway. **e** A comprehensive summary of the pathway is provided including a list of corresponding literature **f** with links to external databases (see also text)

metabolic networks for the species and will be published in SGN, further increasing the wealth of information about *Nicotiana*.

5.4 Pathway Validation and Curation of Metabolic Networks

Curated databases have long been considered invaluable for the analysis of all facets of metabolism (Fernie et al. 2011; Hur et al. 2013; Karp et al. 2013) because they provide the structure and tools required for transforming the large volumes of data into meaningful biological insights (Rhee and Crosby 2005) and pave the way to new discoveries (Baxevanis and Bateman 2015). *SolanaCyc* and *NicotianaCyc* are the primary curation databases to which relevant pathways from MetaCyc are imported and where the curation of new pathways for the respective taxa takes place. Imported pathways from MetaCyc that fit the taxonomic range and list Solanaceae species among the taxa known to possess the pathway are enriched with enzymes and genes published for solanaceous species.

The curation flow of the revised and newly created pathways is finalized by distributing the curated pathways to the respective *Nicotiana*-specific databases. *SolanaCyc* also stores all curated pathways of pathogenic organisms known to attack species of the Solanaceae family. The long-term target at SGN is to transfer pathogen-related data into a separate database, *PathogenCyc*, with the goal of determining metabolic intersections between host and pathogen at the moment of infection and identifying metabolic pathways that maybe exploited as gateways by the attacker.

The taxon-specific curation databases *SolanaCyc* and *NicotianaCyc* have four aims: (i) to elevate the existing SolCyc databases, with current focus on *Nicotiana*-specific databases, to a higher tier; (ii) to constantly improve the metabolic networks through validation and curation of the predicted pathways; (iii) to provide a foundation of metabolic data to future PGDBs for any new Solanaceae species for which an annotated genome has been published; and (iv) to develop *SolanaCyc* and *NicotianaCyc* into the most comprehensive reference databases for the biochemistry and molecular biology of the Solanaceae family and *Nicotiana* genus, respectively. Keeping the databases updated also requires supervised incorporation of relevant new metabolic data provided by new MetaCyc releases and the integration of changes to the species genomes when improved annotations of genomic data become available.

5.4.1 Validation of *Nicotiana*-Specific PGDBs

The five *Nicotiana*-specific PGDBs were created using genomic annotation information of the respective *Nicotiana* species (Sierro et al. 2014; Xu et al. 2017; Sierro et al. 2013) as input for the PathoLogic component of Pathway Tools using MetaCyc as a reference database. Although the various PGDBs have been created with different Pathway Tools and MetaCyc versions, they have all been updated with the most current versions. The PathoLogic prediction of the complement of a metabolic network in species with sufficient genomic information is based on two main factors, the entirety of curated pathways in MetaCyc and the sequence similarity of corresponding annotated proteins. Because of the metabolic diversity within the various kingdoms (Ball and Cherry 2001) and the fact that protein sequence similarity does not necessarily imply similar functions (Pearson 2015), a varying number of false positive predictions occurs and these false positives should be addressed by validating the predicted network.

For the species-specific database of *N. tabacum*, an exhaustive manual validation of all 569 pathways predicted by PathoLogic was carried out (Table 5.2). Utilizing resources and databases such as the EC nomenclature (Nomenclature Committee of the International Union of Biochemistry and Molecular Biology; https://iubmb.org/biochemical-nomenclature/), GenBank (NCBI Resource Coordinators 2018), the Universal Protein Resource (UniProt) (The UniProt Consortium 2017) for confirming genes, enzymes, and reactions, and literature search engines of NCBI's PubMed (https://www.ncbi.nlm.nih.gov/pubmed) and Google Scholar (https://scholar.google.ca/) to extract information about the occurrence of pathways and compounds in *N. tabacum* resulted in the identification of 156 pathways that are likely not valid predictions for this species. The majority of these pathways (112, 72%) were highly specific for bacteria, fungi, or metazoan, with no evidence supporting their occurrence in a higher plant such as tobacco. Other pathways that were rejected by the curator were specialized pathways not reported for the *Nicotiana* genus (Paley et al. 2017) not reported for the Nicotiana genus, plus C4-related photosynthesis pathways [3 unfounded for the C3 plant tobacco]. The validation process was not limited to *N. tabacum* but included the entire Solanaceae family, which allowed the propagation of the validation results to the other four *Nicotiana* species (Table 5.2).

However, not all the pathways labeled invalid for *N. tabacum* were predicted for the metabolic networks of other *Nicotiana*-specific Cyc databases (Table 5.3). This indicates the uniqueness

Table 5.2 Number of pathways (pwys) in Nicotiana-specific databases predicted in the first builds (version 1.0) and approved pathways after validation and curation of the metabolic networks (version 2.0). Note that pathways can be associated with more than one species

N. tabacum				*N. tomentosiformis*				*N. sylvestris*				*N. attenuata*				*N. benthamiana*			
Version 1.0		Version 2.0		Version 1.0		Version 2.0		Version 1.0		Version 2.0		Version 1.0		Version 2.0		Version 1.0		Version 2.0	
All pwys	Curated pwys	All pwys	Curated pwys	All pwys	Curated pwys	All pwys	Curated pwys	All pwys	Curated pwys	All pwys	Curated pwys	All pwys	Curated pwys	All pwys	Curated pwys	All pwys	Curated pwys	All pwys	Curated pwys
569	0	482	99	519	0	424	0	449	0	376	5	297	0	271	12	541	0	458	6

Table 5.3 Number of invalid pathways found in the predicted networks of *Nicotiana*-specific PGDB's in SolCyc

Species	# of invalid pathways after validation
N. tabacum	156
N. sylvestris	94
N. tomentosiformis	116
N. benthamiana	105
N. attenuata	41

of each of those networks, but also reveals differences in the comprehensiveness and annotation status of the corresponding genomes. An important input factor for the PathoLogic metabolic network prediction is the quality of the genome assembly and annotation, which is paramount in determining the number of predicted genes, enzymes, and reactions. The genetic distance of the species is also expected to play a role. Both factors may explain the high accordance of predicted invalid pathways between the allotetraploid *N. tabacum* and its ancestral, diploid parents *N. sylvestris* and *N. tomentosiformis* (Table 5.3) despite the overall lower number of predicted pathways in those species in relation to *N. tabacum* (Table 5.2). The better agreement for invalid pathways curated for *N. tabacum* compared with those for *N. benthamiana* and the much lower correspondence for those pathways in *N. attenuata* (Table 5.3) is certainly a reflection on the number of the originally predicted pathways in these species. For example, *N. benthamiana* holds an advantage of 244 more pathways over *N. attenuata* (Table 5.2), indicating a less complete metabolic network map owing to a less well-annotated genome for the latter.

5.4.2 Curation of *Nicotiana*-Specific PGDBs

Biocuration is a meticulous, time-consuming process that requires expert knowledge of biochemistry and molecular biology of the curated entity, as well as a high familiarity with the functions and features of the target metabolic database. Pathway curation requires the involvement of a broad spectrum of resources to accurately depict and supplement the published research from peer-reviewed sources. For characterizing curated biological processes, molecular functions, and cellular components, associated elements of pathways are assigned the terminology developed by the Gene Ontology Consortium (Ashburner et al. 2000; The Gene Ontology Consortium 2017). In addition to the already mentioned databases for validating and curating genes, enzymes, and reactions, external databases such as PubChem (Kim et al. 2016), ChEBI (Hastings et al. 2013), and ChemSpider (Pence and Williams 2010) are used to confirm compound structure and ontology.

Once a pathway has been revised or newly created in the applicable curator database, furnished with enzymes and genes, and equipped with curator-written summaries and comments, it is then distributed to the relevant species-specific database(s). Curated enzymes with sequence information deposited in NCBI's GenBank or UniProt are used in BLAST searches (https://blast.ncbi.nlm.nih.gov/Blast.cgi) against the local protein databases and matched and merged with respective enzymes and encoding genes of the *Nicotiana*-specific PGDBs.

The results of a curation process are shown in Fig. 5.1, which exemplifies the visual pathway presentation and data commentary based on an instance of the superpathway of diterpenoid production in glandular trichomes in tobacco. Superpathways are used in MetaCyc and derivative databases to provide an extended and more holistic view of certain sections of metabolism by assembling a varying number of connected base pathways (Caspi et al. 2013). The glandular trichome-produced diterpenoid superpathway combines three base pathways, cembratrienediol biosynthesis, *cis*-abienol biosynthesis, and

labdenediol and sclareol biosynthesis, representing specialized compounds produced in *Nicotiana* species primarily in glandular trichomes. The pathway diagram (Fig. 5.1a) shows the reaction sequence (Fig. 5.1-1) in which the compounds (Fig. 5.1-2) are converted. Clicking on the "More Detail" and "Less Detail" buttons will increase or decrease the granularity of the pathway; for example, providing the structure of the compounds (Fig. 5.1d). Corresponding enzymes (Fig. 5.1-3) and genes (Fig. 5.1-4) are associated with the reaction and can be shown or hidden in a customized view (Fig. 5.1-5). All objects of a pathway, such as enzymes, genes, or compounds, have their own detail pages, which provide more in-depth information and links to external databases. In addition, available interconnecting pathways are presented, which hyperlink the feed into (Fig. 5.1b) or out of the displayed pathway. Small evidence icons in the upper right-hand corner (Fig. 5.1c) allow a fast assessment of the curation status of the pathway, indicating experimental evidence for the existence of the pathway (flask symbol) or only computational prediction (computer icon). Comments on all curated enzymes and genes summarize the findings of the relevant literature. A similar abstract is written for the pathway (Fig. 5.1e), concisely summarizing important points, such as key enzymes, rate-limiting steps, the stereochemistry of compounds, and regulation. All pathway information is referenced with the appropriate literature (Fig. 5.1f), links to external databases, and full bibliographic information.

Next to tomato, tobacco is the Solanaceae species with the most published studies covering its biochemical and molecular biology. This is reflected in version 2.0 of the *Nicotiana*-specific databases, which list 99 curated pathways for tobacco (Table 5.2) corresponding to 40.2% of all 246 curated pathways in *SolanaCyc* at this time. This is also observed in the development of *NicotianaCyc* as a taxon-specific database, hosting all curated pathways for the *Nicotiana* genus (Table 5.4). The current version (2.0, March 2018) increases the number of pathways from 18 in version 1.0 (June 2016) to 100. Enzymatic reactions have been enhanced more than five-fold and curated enzymes and compounds have multiplied. The latest version of *NicotianaCyc* increases the number of species with curated data from three to nine. The expansion of *NicotianaCyc* with metabolic information from more *Nicotiana* species enriches the database and renders it more comprehensive and relevant.

5.5 Useful Tools in *Nicotiana* PGDBs

MetaCyc-derived species-specific databases are equipped with a wide range of tools for searching, visualizing, analyzing, and publishing data. MetaCyc is constantly updated and enhanced with new program features, which are conferred to derivative databases with each new release version of the Pathway Tools software. MetaCyc provides capabilities for querying metabolic networks (Karp and Caspi 2011), conducting local database BLAST searches, finding chokepoint and dead-end metabolites, comparing pathways and reactions within the confines of all organism-specific databases in personal database collections, performing metabolic flux analysis (Toya et al. 2011) and omics data analysis (Paley et al. 2017), and storing pathway object data in tabular form (for full instructions see https://yeast.biocyc.org/PToolsWebsiteHowto.shtml#SearchHelp). To stay within the scope of this chapter, only a few user-centered tools are described here.

Table 5.4 Statistics of the content of *NicotianaCyc* from initial installation (version 1.0) to current status (version 2.0)

NicotianaCyc	Pathways	Reactions	Enzymes	Transporter	Compounds	Species
Version						
1.0	18	75	32	0	260	3
2.0	100	407	139	29	646	9

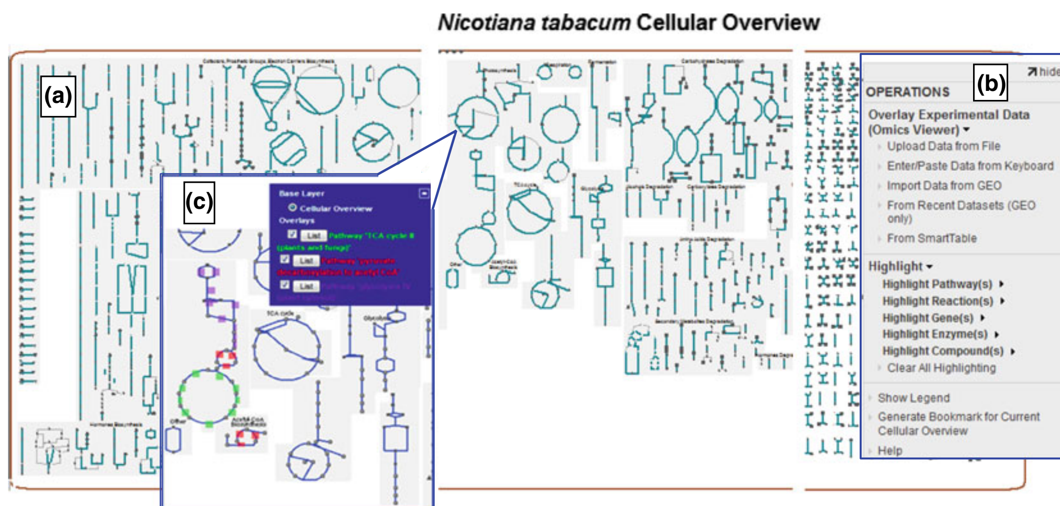

Fig. 5.2 The cellular overview diagram of *N. tabacum*. **a** Sections of the cellular overview **b** menu of operations for the customized use of the cellular overview **c** Enlarged sector of the cellular overview showing highlighted pathways of interest (see also text)

A popular feature of species-specific databases is the cellular overview, a bird's-eye projection of the entire metabolism predicted for a species (Latendresse and Karp 2011; Toya et al. 2011). The cellular overview for *N. tabacum* (Fig. 5.2a) positions pathways of central importance, such as the tricarboxylic acid cycle and glycolysis, in the middle of the schema, with anabolic and intermediary pathways placed to the left and catabolic pathways to the right. Individual reactions that have not yet been associated with pathways are located on the very right of the diagram. At the upper right-hand corner, a menu (Fig. 5.2b) is available for executing miscellaneous operations on the cellular overview, such as uploading and overlaying the map with omics data or highlighting (Fig. 5.2c) pathways, reactions, genes, enzymes, or compounds of interest.

A relatively new feature of PGDBs is the pathway collage tool (Paley et al. 2016). Unlike the algorithm dictating the rigid arrangement of reactions and associated objects in a pathway of stand-alone curator databases, web-based PGDBs allow for a customized look and data display of pathways through the pathway collage tool (Fig. 5.3). The pathway collage generated for the Calvin–Benson–Bassham cycle in *N. tabacum* illustrates the flexibility to re-position reactions, compounds, and associated objects to the requirements of the user. The look and information displayed on the pathway can be customized and edited using the menu in the upper left-hand corner, which also provides a zoom function. Objects in the pathway can be highlighted, a useful feature for indicating metabolic changes in metabolomics studies. Pathway collages can generate customized multi-pathway schemata of interest, on which omics data from metabolic and genomic experiments can be imposed and visualized. The assembly of pathways for pathway collages is accomplished using the SmartTable function in Pathway Tools (Travers et al. 2013). SmartTables can be compiled in both web-based databases and stand-alone desktop PGDBs and exported to produce an image of the selected pathways and their associated objects. Both single pathways (Fig. 5.3) and more complex pathway combinations can be saved as portable network graphics (PNG) files and used as charts in publications or presentations. Saved pathway collages with omics data can also be uploaded into the cellular overview, accentuating the pathways of interest in relation to the metabolic network.

Producing metabolites of high value at low cost in the most time-efficient way is the driving

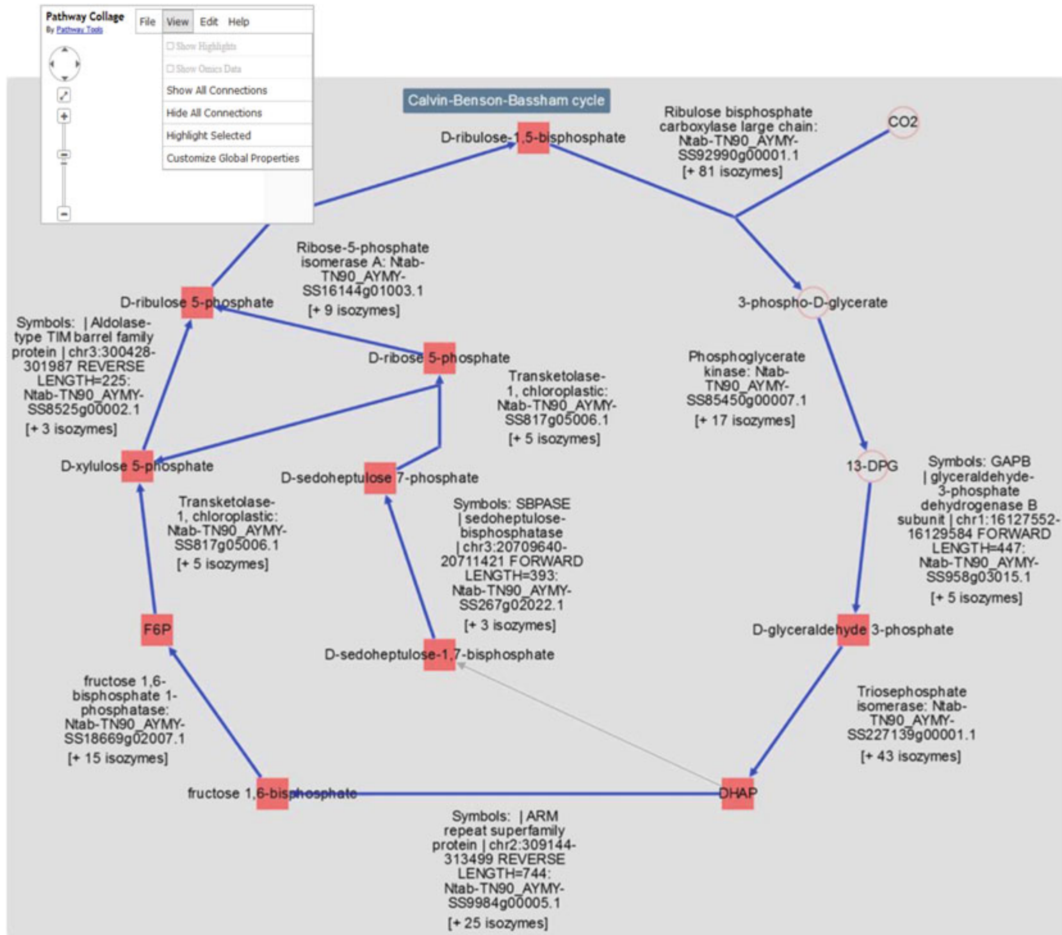

Fig. 5.3 Pathway collages of the Calvin-Benson-Bassham cycle as depicted in the species-specific database of *N. tabacum*. The pathway displays reactions, compounds, enzymes, and genes and can be customized to the user's preferences (see also text)

force behind metabolic engineering (Zadran and Levine 2013). *Nicotiana* species have been used successfully as hosts for engineered metabolic pathways. Increased production of taxadiene, a precursor in the biosynthesis of the anti-cancer drug paclitaxel, has been achieved through ectopic expression of foreign genes and metabolic pathway shunting in *N. benthamiana* (Hasan et al. 2014). The same species was also key in elucidating modification reactions of the indole glucosinolate biosynthesis in *Arabidopsis* by harboring this cruciferous pathway and enabling the characterization of the catalyzing cytochrome P450 monooxygenases (Pfalz et al. 2011). *N. tabacum* and *N. benthamiana* served as transformants for expressing a multigenic construct of *Antirrhinum* transcription factors and *Medicago* genes to achieve coordination of anthocyanin and proanthocyanin biosynthetic pathways (Fresquet-Corrales et al. 2017). Astaxanthin is a pigment of substantial commercial value as a food additive for fish and crustaceans in captivity. The modification of the biosynthetic route towards carotenoids in flowers of *N. tabacum* (Mann et al. 2000) and *N. glauca* (Gerjets et al. 2007) by introducing beta-carotene ketolases from algae or cyanobacteria resulted in the production of this foreign ketocarotenoid

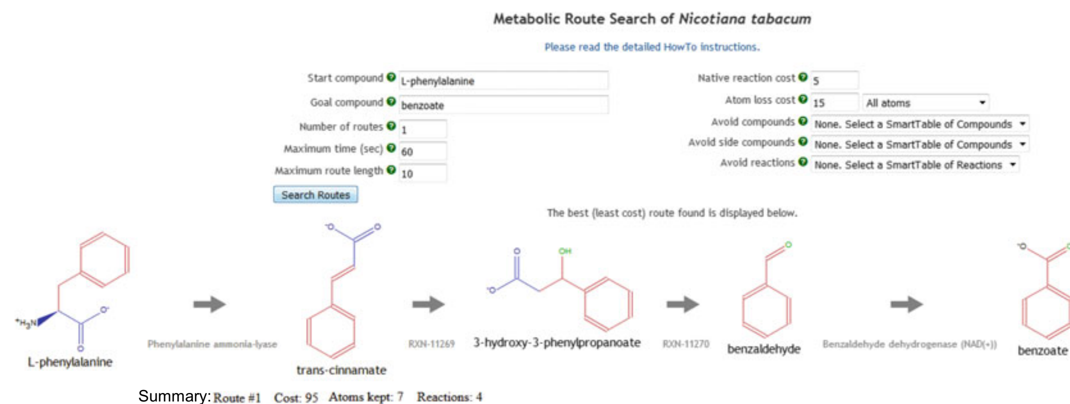

Fig. 5.4 Metabolic route search tool used for extricating the shortest and most cost-effective way to go from compound A, for instance, L-phenylalanine, to compound B, for instance, benzoate, within the confines of the predicted and curated metabolic network in *N. tabacum*

pigment in these *Nicotiana* species. The key for the successful orchestration of metabolic engineering is the access to databases that archive, maintain, display, and update metabolic data from the published literature and allow purposeful mining of information on metabolic diversity (Owen et al. 2017).

The metabolic route search tool (Fig. 5.4) assists in ascertaining the shortest and most cost-effective biosynthetic route between two compounds in an organism's given metabolic network. For example, searching the PGDB of *N. tabacum* with L-phenylalanine as the start compound and benzoate as the goal compound produces a four-step pathway. There are two more possible routes, consisting of 5- and 6-step pathways, which can be disregarded depending on the predetermined search parameters. The search summary provides details on the costs of the selected route with a view to overall losses of atoms and reactions. The selected start compounds, goal compounds, and organisms can be fed into the metabolic route search using SmartTables that enable simultaneous searches of multiple pathway routes. The search can be enhanced by tapping the reaction and compound pools in MetaCyc or any other organism databases listed in the multi-organism selector. Consequently, the metabolic route search not only permits the identification of short and low-metabolic-cost routes in a species but also provides the option to incorporate reactions not found in the local database. In this context, new opportunities for constructing novel pathways for the targeted biosynthesis of natural products arise, further increasing the applications of species-specific databases in metabolic engineering.

5.6 Outlook for *Nicotiana*-Related Databases in SGN

The metabolic information available in SGN for solanaceous plants has increased significantly over the past two years. SolCyc, the collection of PGDBs for the Solanaceae family, has been expanded with five *Nicotiana*-specific and two taxon-specific databases. This development is in accord with the notable impact that important crops of the Solanaceae family have on the daily life of humans. It also reflects the ever-growing demand for databases that store, manage, and curate the wealth of published information and provide the tools for translating the vast amount of omics data into organismal functionalities. It is generally agreed that the quality of data in a metabolic database, when combined with bioinformatics tools and applications, defines the usefulness of the database for mining and analysis (Mueller et al. 2003; Zhang et al. 2005; Karp et al. 2013). The systematic validation and

continuous curation of metabolic data in *Nicotiana*-specific databases in SGN improve metabolic network records and document basic and applied *Nicotiana* research.

Many solanaceous species are representatives of major crops, but their importance is not limited to human food and health. *Nicotiana* species have long been associated with classical and molecular genetic research, and have served as model organisms for somatic cell genetics (*N. tabacum*) (Sussex 2008) and ecological studies (Baldwin 2001) (*N. attenuata*; see also Chap. 13). The fairly high number of curated pathways for *N. tabacum*, the second best-curated organism in any SolCyc database, underscores the central importance of this species within the *Nicotiana* genus. The development of *NicotianaCyc* as a curation hub for the genus is driven by the goal of storing curated data, which have been confirmed by peer-reviewed publications, for all species of the Solanaceae family. The next enhancement of *Nicotiana*-related databases is the continued compilation of metabolic data in *NicotianaCyc* with curated pathways of primary and specialized metabolism. The careful harnessing and curation of data resources will contribute to better understanding of metabolism in the *Nicotiana* genus. Species-specific *Nicotiana* PGDBs will benefit from this curation by elucidating the metabolic networks using experimentally verified data, thus increasing the accuracy, reliability, and usability of this resource in metabolism and bioinformatics analyses.

References

Ashburner M, Ball CA, Blake JA et al (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25:25–29

Baldwin IT (2001) An ecologically motivated analysis of plant-herbivore interactions in native tobacco. Plant Physiol 127:1449–1458

Ball CA, Cherry JM (2001) Genome comparisons highlight similarity and diversity within the eukaryotic kingdoms. Curr Opin Chem Biol 5:86–89

Bally J, Jung H, Mortimer C et al (2018) The rise and rise of *Nicotiana benthamiana*: a plant for all reasons. Annu Rev Phytopathol 56:405–426

Baulcombe DC (1999) Fast forward genetics based on virus-induced gene silencing. Curr Opin Plant Biol 2:109–113

Baxevanis AD, Bateman A (2015) The importance of biological databases in biological discovery. Curr Protoc Bioinforma 50(1):1–8

Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW (2010) GenBank. Nucleic Acids Res 38: D46–D51

Bombarely A, Menda N, Tecle IY et al (2011) The Sol Genomics Network (solgenomics.net): growing tomatoes using Perl. Nucleic Acids Res 39:D1149–D1155

Bombarely A, Rosli HG, Vrebalov J, Moffett P, Mueller LA, Martin GB (2012) A draft genome sequence of *Nicotiana benthamiana* to enhance molecular plant-microbe biology research. Mol Plant-Microbe Interact MPMI 25:1523–1530

Caspi R, Dreher K, Karp PD (2013) The challenge of constructing, classifying, and representing metabolic pathways. FEMS Microbiol Lett 345:85–93

Caspi R, Billington R, Ferrer L et al (2016) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Res 44:D471–D480

Dreher K (2014) Putting The Plant Metabolic Network pathway databases to work: going offline to gain new capabilities. Methods Mol Biol Clifton NJ 1083:151–171

Edwards KD, Fernandez-Pozo N, Drake-Stowe K et al (2017) A reference genome for *Nicotiana tabacum* enables map-based cloning of homeologous loci implicated in nitrogen utilization efficiency. BMC Genom 18:448

Fernandez-Pozo N, Menda N, Edwards JD et al (2015) The Sol Genomics Network (SGN)–from genotype to phenotype to breeding. Nucleic Acids Res 43:D1036–D1041

Fernie AR, Aharoni A, Willmitzer L et al (2011) Recommendations for reporting metabolic data. Plant Cell 23:2477–2482

Foerster H, Bombarely A, Battey JND, Sierro N, Ivanov NV, Mueller LA (2018) SolCyc: a database hub at the Sol Genomics Network (SGN) for the manual curation of metabolic networks in *Solanum* and *Nicotiana* specific databases. Database J Biol Databases Curation

Fresquet-Corrales S, Roque E, Sarrión-Perdigones A et al (2017) Metabolic engineering to simultaneously activate anthocyanin and proanthocyanidin biosynthetic pathways in *Nicotiana spp*. PLoS ONE 12:e0184839

Gerjets T, Sandmann M, Zhu C, Sandmann G (2007) Metabolic engineering of ketocarotenoid biosynthesis in leaves and flowers of tobacco species. Biotechnol J 2:1263–1269

Hasan MM, Kim H-S, Jeon J-H et al (2014) Metabolic engineering of *Nicotiana benthamiana* for the increased production of taxadiene. Plant Cell Rep 33:895–904

Hastings J, de Matos P, Dekker A et al (2013) The ChEBI reference database and ontology for biologically

relevant chemistry: enhancements for 2013. Nucleic Acids Res 41:D456–D463

Hur M, Campbell AA, Almeida-de-Macedo M et al (2013) A global approach to analysis and interpretation of metabolic data for plant natural product discovery. Nat Prod Rep 30:565–583

Karp PD, Caspi R (2011) A survey of metabolic databases emphasizing the MetaCyc family. Arch Toxicol 85:1015–1033

Karp PD, Paley SM, Krummenacker M et al (2010) Pathway Tools version 13.0: integrated software for pathway/genome informatics and systems biology. Brief Bioinform 11:40–79

Karp PD, Paley S, Altman T (2013) Data mining in the MetaCyc family of pathway databases. Methods Mol Biol Clifton NJ 939:183–200

Kessler A, Baldwin IT (2001) Defensive function of herbivore-induced plant volatile emissions in nature. Science 291:2141–2144

Kim S, Thiessen PA, Bolton EE et al (2016) PubChem substance and compound databases. Nucleic Acids Res 44:D1202–D1213

Latendresse M, Karp PD (2011) Web-based metabolic network visualization with a zooming user interface. BMC Bioinform 12:176

Mann V, Harker M, Pecker I, Hirschberg J (2000) Metabolic engineering of astaxanthin production in tobacco flowers. Nat Biotechnol 18:888–892

Mueller LA, Zhang P, Rhee SY (2003) AraCyc: a biochemical pathway database for Arabidopsis. Plant Physiol 132:453–460

Mueller LA, Solow TH, Taylor N et al (2005) The SOL genomics network: a comparative resource for Solanaceae biology and beyond. Plant Physiol 138:1310–1317

NCBI Resource Coordinators (2018) Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 46:D8–D13

Owen C, Patron NJ, Huang A, Osbourn A (2017) Harnessing plant metabolic diversity. Curr Opin Chem Biol 40:24–30

Paley SM, Karp PD (2006) The Pathway Tools cellular overview diagram and Omics Viewer. Nucleic Acids Res 34:3771–3778

Paley S, O'Maille PE, Weaver D, Karp PD (2016) Pathway collages: personalized multi-pathway diagrams. BMC Bioinformatics 17:529

Paley S, Parker K, Spaulding A, Tomb JF, O'Maille P, Karp PD (2017) The Omics Dashboard for interactive exploration of gene-expression data. Nucleic Acids Res 45(21):12113–12124

Pearson WR (2015) Protein function prediction: problems and pitfalls. Curr Protoc Bioinforma 51(1):4–12

Pence HE, Williams A (2010) ChemSpider: an online chemical information resource. J Chem Educ 87:1123–1124

Pfalz M, Mikkelsen MD, Bednarek P, Olsen CE, Halkier BA, Kroymann J (2011) Metabolic engineering in *Nicotiana benthamiana* reveals key enzyme functions in Arabidopsis indole glucosinolate modification. Plant Cell 23:716–729

Rhee SY, Crosby B (2005) Biological databases for plant research. Plant Physiol 138:1–3

Rushton PJ, Bokowiec MT, Laudeman TW, Brannock JF, Chen X, Timko MP (2008) TOBFAC: the database of tobacco transcription factors. BMC Bioinform 9:53

Sierro N, Battey JND, Ouadi S et al (2013) Reference genomes and transcriptomes of *Nicotiana sylvestris* and *Nicotiana tomentosiformis*. Genome Biol 14:R60

Sierro N, Battey JND, Ouadi S et al (2014) The tobacco genome sequence and its comparison with those of tomato and potato. Nat Commun 5:3833

Sussex IM (2008) The scientific roots of modern plant biotechnology. Plant Cell 20:1189–1198

The Gene Ontology Consortium (2017) Expansion of the Gene Ontology knowledgebase and resources. Nucleic Acids Res 45:D331–D338

The UniProt Consortium (2017) UniProt: the universal protein knowledgebase. Nucleic Acids Res 45:D158–D169

Toya Y, Kono N, Arakawa K, Tomita M (2011) Metabolic flux analysis and visualization. J Proteome Res 10:3313–3323

Travers M, Paley SM, Shrager J, Holland TA, Karp PD (2013) Groups: knowledge spreadsheets for symbolic biocomputing. Database J Biol Databases Curation 2013:bat061

Walsh JR, Schaeffer ML, Zhang P, Rhee SY, Dickerson JA, Sen TZ (2016) The quality of metabolic pathway resources depends on initial enzymatic function assignments: a case for maize. BMC Syst Biol 10:129

Whitelaw CA, Barbazuk WB, Pertea G et al (2003) Enrichment of gene-coding sequences in maize by genome filtration. Science 302:2118–2120

Xu S, Brockmöller T, Navarro-Quezada A et al (2017) Wild tobacco genomes reveal the evolution of nicotine biosynthesis. Proc Natl Acad Sci USA 114:6133–6138

Zadran S, Levine RD (2013) Perspectives in metabolic engineering: understanding cellular regulation towards the control of metabolic routes. Appl Biochem Biotechnol 169:55–65

Zhang P, Foerster H, Tissier CP et al (2005) MetaCyc and AraCyc. Metabolic pathway databases for plant research. Plant Physiol 138:27–37

Check for updates

6 Leaf Curing Practices Alter Gene Expression and the Chemical Constituents of Tobacco Leaves

Lucien Bovet, Aurore Hilfiker, James N. D. Battey, Nicolas Sierro and Simon Goepfert

Abstract

Tobacco leaves senesce during the early curing phase, and their constituents are altered via specific enzymatic activities. Levels of the reducing sugars glucose and fructose and those of total free amino acids increase via active protease activity and active synthesis processes. We used transcriptomics to identify the genes up- and downregulated in tobacco leaves from the Burley, Virginia, and Oriental varieties after 48 h of curing. Approximately 8% of the transcripts were upregulated, including 591 genes common to all three varieties, and especially genes associated with oxidation-reduction processes, transcription factors, proteolysis, and hydrolase activity. All of these have been linked to the cellular senescence process. Approximately 12% of the transcripts were downregulated, including 1,486 genes common to all three tobacco types; the genes with the highest fold changes were associated with general metabolic processes. Gene ontology overrepresentation analysis suggested that the three main axes governing the curing of tobacco are dehydration stress, amino acid alteration, and sucrose metabolism. We identified sucrose synthase (*SUS*) as a likely driver of the accumulation of reducing sugars during the water-deprivation response, and aspartate aminotransferase (*AAT*) as a catalyst of amino acid metabolism.

L. Bovet (✉) · A. Hilfiker · J. N. D. Battey · N. Sierro · S. Goepfert
PMI R&D, Philip Morris Products S.A., Quai Jeanrenaud 5, Neuchâtel 2000, Switzerland
e-mail: Lucien.Bovet@pmi.com

6.1 Introduction

Tobacco (*Nicotiana tabacum* L.) single leaves or leaves on stalk are traditionally harvested from the field and then processed with two major successive curing steps: full leaf maturation ("yellowing phase") and lamina-stem drying. Three major curing strategies air-, flue-, and sun-curing are typically associated with the main tobacco types, Burley, Virginia, and Oriental, respectively. During the early curing phase (between 0 and 72 h) and even sometimes just before harvest, green leaves senesce by turning yellow. Air-cured, or Burley, tobacco is stored in an aerated barn for approximately 2–3 months; flue-cured, or Virginia, tobacco leaves are cured in a heated barn for approximately 1 week, and Oriental tobacco leaves are cured on a string in sunlight for approximately 1 week. The curing process significantly alters leaf constituents via specific enzymatic activities. The information presented in this chapter includes data from the Oriental tobacco Katerini33 variety, grown and sun-cured in Katerini, Greece, and the Swiss

© Springer Nature Switzerland AG 2020

N. V. Ivanov et al. (eds.), *The Tobacco Plant Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-030-29493-9_6

Virginia ITB683 and Burley Stella varieties, grown in Payerne (Switzerland) and cured in an oven or in an aerated barn, respectively. Green leaves (upper stalk) and mature leaves (ripe at mid-stalk) were collected at harvest, after 2 days, and at the end of the curing period.

Leaves (three biological bulk replicates from subplots) were sampled in the field or during curing at fixed time points. Leaves were immediately rolled in plastic bags, frozen in liquid nitrogen for 60 s, and kept, first, in dry ice for transportation and at −80 °C until further use. Leaf lamina samples were ground and freeze-dried. Lamina colors are known to change during curing in all tobacco species, partly because of chlorophyll degradation (Leffingwell 1999) (Fig. 6.1a). The impact of the first curing step, involving leaf senescence, can be monitored by gene expression analyses. Senescence-associated gene 12 (*SAG12*) has been previously reported as a specific senescence marker (Noh and Amasino 1999). After 48 h of curing by any curing process, *SAG12* is strongly upregulated. In the same samples, photosynthesis-related genes were downregulated, as exemplified by the expression of the RuBisCO small subunit gene (*RbcS*) (Fig. 6.1b). These alterations in gene expression are consistent with those previously observed in rice (Suzuki and Makino 2013).

6.2 Increase of Reducing Sugars and Free Amino Acids in Cured Leaves

Virginia tobacco cultivars have a high starch content in green leaves, whereas Burley and Oriental tobacco have low and medium starch content, respectively. The glucose polymers are converted into reducing sugars during curing (Leffigwell 1999 and references herein). The reducing sugars glucose and fructose are

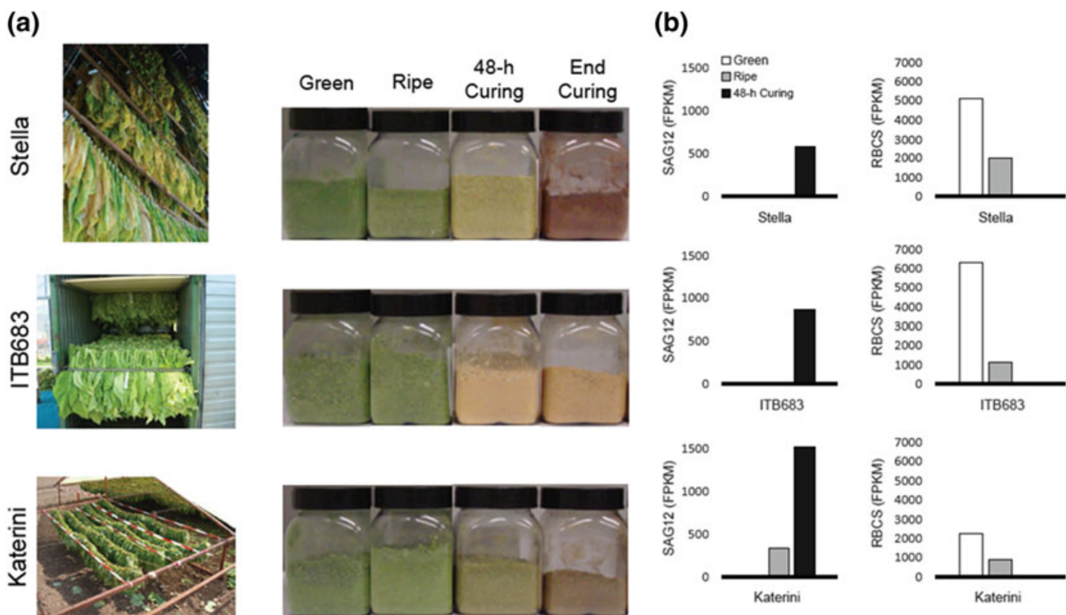

Fig. 6.1 Curing practices of three major tobacco types, Burley (Stella), Virginia (ITB683), and Oriental (Katerini). **a** Tobacco material collected at mid-stalk from field-grown leaves (green) at harvest time (ripe), after 2 days of curing (48 h curing), and at the end of curing was freeze-dried or subjected to RNA sequencing (RNAseq) analyses (except for the material collected at the end of curing). **b** RNAseq analyses (Guo et al., 2013) show that the markers senescence-associated gene 12 (*SAG12*) and RuBisCO small subunit (*RbcS*) were upregulated and downregulated, respectively, after 2 days of curing in all three tobacco varieties (data are expressed in fragments per kilobase million [FPKM])

produced rapidly during the first days of curing, and their content partly decreases during the next phase of the curing process (lamina-stem drying, Fig. 6.2a). Burley tobacco leaves, however, have a low starch reserve and accumulate high levels of nitrate (Legg et al. 1977; Leffingwell 1999); this characteristic of Burley tobacco is the result of a specific mutation within the *yb* loci that confers a chlorotic phenotype (Edwards et al. 2017). Figure 6.2a shows that this low-starch phenotype affected the accumulation of reducing sugars during the air-curing process of the Burley Stella variety. Depending on the environmental conditions and the type of Burley cultivars grown, some reducing sugars may accumulate after 48 h of curing (data not shown), but in all cases, a very low amount remains at the end of curing (Leffigwell 1999). The three main tobacco types cultivated exhibited an increase in total free amino acids (Fig. 6.2b). This accumulation may result from both active protease activity in senescing tissue, as described in Diaz-Mendoza et al. (2016), and active synthesis processes (Masclaux-Daubresse et al. 2005; Uzelac et al. 2016). Redistribution of nitrate from source leaves to sink organs occurs in the form of amino acids, such as glutamine or asparagine (Gaufichon et al. 2016). This may explain the accumulation of free amino acids in yellowing and fully cured leaves observed in the main tobacco types (Fig. 6.2b).

6.3 Genes Up- and Downregulated After 48 h of Curing

Figure 6.2 illustrates the metabolic differences between the three main tobacco types. We used transcriptomics to identify the genes up- and downregulated in early-cured tobacco leaves collected at mid-stalk and to assess the drivers of metabolic variations and common denominators. Among the 81,404 genes in the tobacco genome, approximately 7.6% (6,189 genes, $\log_2$ fold change ≥ 2, adjusted p-value $\leq$ 0.05) of the transcripts in the three tobacco varieties were upregulated in at least one cultivar after 48 h of curing. This indicated that the number of genes activated during the early curing step is not negligible and that these genes contribute to modifications in the tobacco leaf chemistry (Fig. 6.3a). Of note, the number of genes activated in the Stella variety during the first 48 h of curing (2,022) was lower than the number in the Katerini and ITB683 varieties, and corresponded to approximately 61% of the genes activated in the Katerini (3,314) and ITB683 (3,324) varieties. This reduced number of upregulated

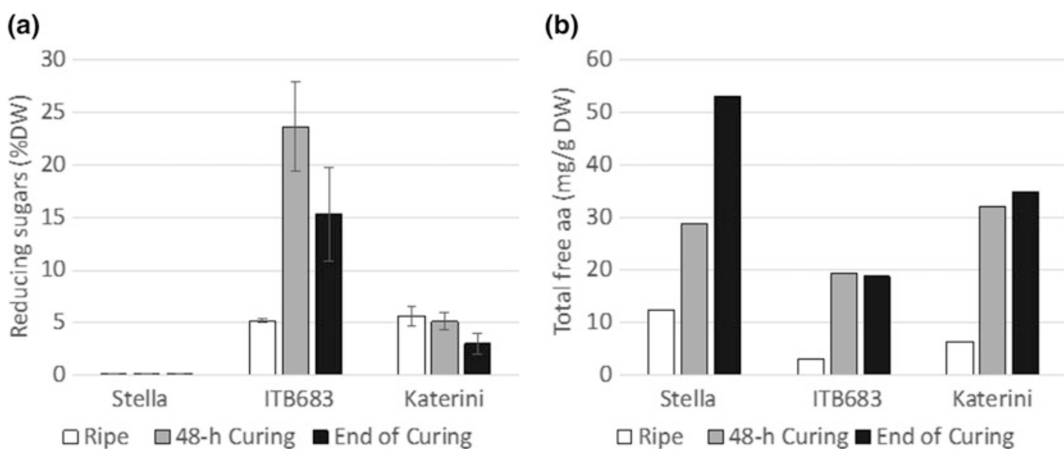

Fig. 6.2 **a** Content of reducing sugars per tobacco variety and **b** total free amino acids after harvest (ripe), 48 h of curing, and at the end of the curing period. Abbreviations: DW, dry weight; h, hours

transcripts in Stella is due, at least in part, to a slower curing process in Burley tobacco. The temperature in an air-curing barn at the beginning of Burley curing does not exceed 25–30 °C and remains constant depending on the ambient variation, whereas for Virginia and Oriental tobaccos, the temperature is typically 30–35 °C in the heated barn or in sunlight and increases during the day, promoting rapid leaf senescence and cellular activities.

Nonetheless and based on gene ontology terms from Sierro et al. 2014, we found that 591 activated genes were common to all three varieties after 48 h of curing, independent of the curing process and the origin of the tobacco cultivars (Fig. 6.3a). Of these 591 upregulated genes, the genes with the highest fold changes (303) belonged to protein families (>10 putative gene products per family) involving oxidation-reduction processes, zinc ion binding, transcription factors, transmembrane transport, protein binding, protein phosphorylation, proteolysis, DNA binding, iron ion binding, and hydrolase activity. Finding genes involved in oxidation-reduction reactions, transcription factor regulation, proteolysis, and hydrolase activity was expected, because these genes have been shown to be activated during the senescence process (Rogers and Munné-Bosch 2016; Kim et al. 2016; Hickman et al. 2013; Diaz-Mendoza et al. 2016; Li et al. 2016). This suggests that the activity of a small number of genes expressed in the three tobacco types contributed to basal activities in the cured leaves, driving notably common senescence processes (Fig. 6.3a).

Among the 81,404 genes expressed in tobacco, approximately 10.8% (8,798) of the transcripts were downregulated ($\log_2$ fold change ≤ -2, adjusted p-value $\leq$ 0.05; Fig. 6.3b) in at least one variety after 48 h of curing, which is slightly lower than the number of genes that were upregulated. The number of genes which are downregulated in all three tobacco types is approximately three times (1,486) that of those which are upregulated (591) in all three tobacco types, presumably as all functions related to photosynthesis are deactivated during senescence (corresponding to early tobacco leaf curing), as reported by Guo and Gan (2005). Among the 1,486 genes downregulated in all three varieties, the genes with the highest fold changes (>10 genes per family) were involved in general metabolic processes such as oxidation-reduction reactions, carbohydrate metabolism, lipid metabolism, and photosynthesis. Photosynthesis genes are downregulated during senescence (Ludewig and Sonnewald 2000), including genes coding for the two RuBisCO subunits, supporting our findings (Fig. 6.1b) for a nuclear-encoded *RbcS* gene. Additionally, high lipid turnover also occurs during senescence (Troncoso-Ponce et al. 2013), possibly explaining why these gene families were downregulated during tobacco leaf curing.

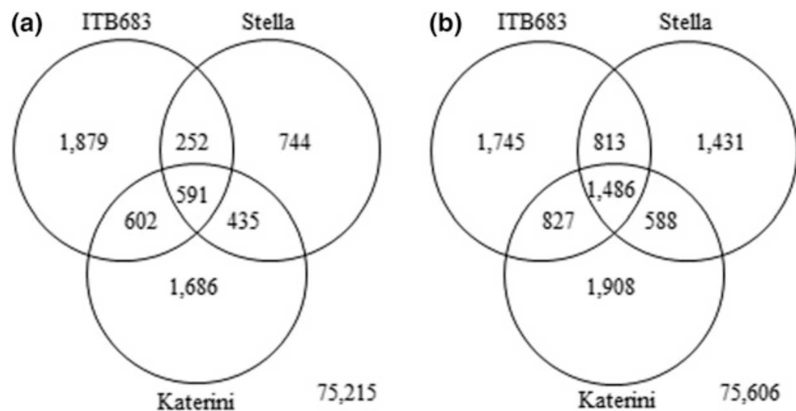

Fig. 6.3 Venn diagrams showing the number of upregulated (**a**) and downregulated (**b**) genes following 48 h of curing in three tobacco leaf varieties. The number of genes in each panel sums to 81,404

6.4 Functional Overrepresentation in Upregulated Genes After 48 h of Curing in Burley, Virginia, and Oriental Tobacco Leaf

Figure 6.2 illustrated that active changes took place during the first days of curing. To identify key functions contributing to metabolic changes during the early curing stage, we analyzed functional overrepresentation of genes upregulated in cured leaves, in comparison to expression in ripe leaves at harvest ($\log_2$ fold change ≥ 2, adjusted p-value $\leq$ 0.05), in all three tobacco varieties (Fig. 6.4). Gene ontology terms previously assigned to tobacco gene models (Sierro et al. 2014) were analyzed for enrichment using the BiNGO tool (version 3.0.3; Maere et al. 2005). The 13 most overrepresented functions were response to water (four genes), the asparagine biosynthetic process (three genes), embryo development (two genes), and the sucrose metabolic process (four genes), although the two transcripts associated with embryo development may be linked to the "response to water" cluster. These two genes code for putative late embryogenesis-abundant proteins, known to accumulate in high levels during the last stages of seed development or in vegetative and reproductive tissues under water-deficit conditions (Olvera-Carrillo et al. 2010). This gene ontology term identification suggests that the three main axes governing the curing of tobacco are water stress, amino acid alteration, and sucrose metabolism.

After leaf detachment at harvest, the decline in photosynthesis observed at the beginning of curing (Leffigwell 1999) is aggravated further by dehydration (Christ et al. 2014; Patro et al. 2014). Moreover, a senescing leaf loses water more rapidly than a non-senescing leaf, and abscisic acid (ABA), the water stress-induced plant hormone, promotes leaf senescence by inducing the production of ethylene (Yang et al. 2002; Zhang et al. 2012). Stomata closing, mediated by strong ABA accumulation, has been shown in senescing oat (*Avena sativa* cv. Victory) leaves when chlorophyll loss is rapid (Gepstein and Thimann 1980). These reports support the concept that a genetic response to water stress occurs at the early curing stage in tobacco leaf (Fig. 6.4) to maintain cell metabolism and prevent loss of cell integrity. As dehydration stress triggers mature leaf senescence, the remobilization of nutrients as a result of carbon/nitrogen imbalance also depends on resistance to drought (Chen et al. 2015). In tobacco, this nutrient remobilization at the beginning of curing results in the accumulation of reducing sugars and free amino acids in the lamina (Fig. 6.2). The accumulation of sugars in senescing leaves and the induction of cytosolic glutamine synthetase and glutamate dehydrogenase, enzymes involved in nitrogen metabolism, have been well documented in the literature (Pourtau et al. 2006; Wingler et al. 2009). The expression of these two enzymes is regulated by sucrose and amino acids (Masclaux-Daubresse et al. 2005).

In barley (*Hordeum vulgare* L.), the expression of glutamine synthetase and asparagine synthetase, which are involved in ammonium assimilation in plants, is also regulated during leaf senescence (Avila-Ospina et al. 2015). Our gene ontology term analysis shows that genes involved in water stress and the production of reducing sugars and free amino acids were active after 48 h of curing, independent of the curing method or tobacco variety. These data are in line with the metabolic changes affecting the carbon-to-nitrogen ratio observed in cured leaves (Fig. 6.2).

6.5 Key Genes Upregulated During Tobacco Curing

Because the leaf concentrations of reducing sugars (with the exception of Stella leaves) and total free amino acids rapidly increase during the first days of curing, we focused on two genes, sucrose synthase (*SUS*) and aspartate aminotransferase (*AAT*), related to the sucrose metabolic process and response to chemicals, respectively (Fig. 6.4). *SUS* encodes a key enzyme involved in sucrose metabolism

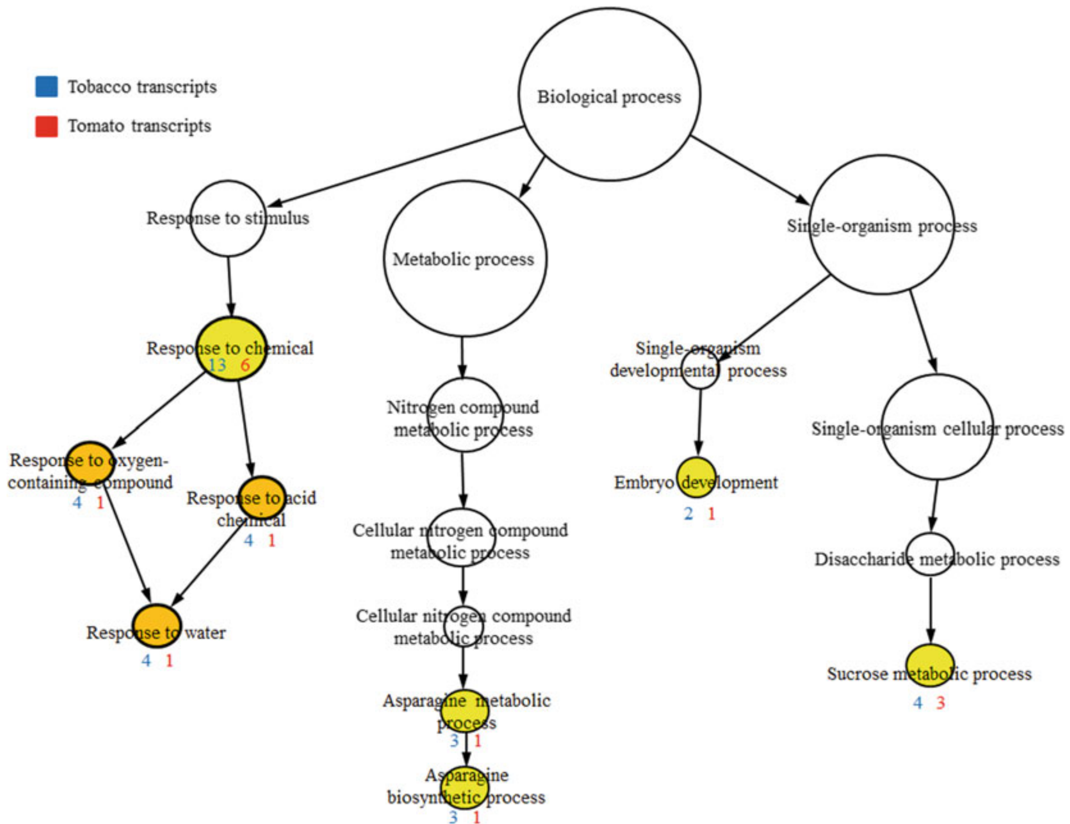

Fig. 6.4 Gene ontology term overrepresentation analysis for transcripts upregulated following curing in all three tobacco varieties. The circle radius is proportional to the number of genes annotated with the given gene ontology term; the color corresponds to the significance of the overrepresentation. Number of tobacco transcripts and corresponding ortholog genes in tomato (*Solanum lycopersicum*) of the most represented gene ontology terms. Tomato ortholog transcript data were obtained from https://solgenomics.net/organism/solanum_lycopersicum/genome

(Stitt et al. 1988; Wingler et al. 2009). Once sucrose is produced during leaf senescence, it can potentially be exported to sink organs and seeds or hydrolyzed into its constituent reducing sugars, glucose and fructose (Sonnewald et al. 1991; Pourtau et al. 2006; Shi et al. 2016). Thus, SUS is likely a key enzyme driving the accumulation of reducing sugars in cured detached leaves.

Figure 6.5 shows a phylogenetic tree based on translated amino acid sequences constructed using the 12 *NtSUS* genes from the ancestral parents of *N. tabacum*, *N. sylvestris* (S), and *N. tomentosiformis* (T), termed NtSUS1-S, NtSUS1-T, NtSUS2-S, NtSUS2-T, NtSUSS3-S, NtSUS3-T, NtSUS4-S, NtSUS4-T, NtSUS5-S, NtSUS5-T, NtSUS6-S, and NtSUS6-T.

Among the four *SUS* transcripts assigned gene ontology terms related to sucrose metabolism (Fig. 6.4), the corresponding gene products NtSUS3-S and NtSUS3-T grouped with Solyc09g098590, NtSUS2-S grouped with Solyc12g009300, and NtSUS4-S grouped with Solyc07g042550 (Fig. 6.5). Surprisingly, *NtSUS2-T* and *NtSUS4-T* were not induced during leaf senescence, suggesting different regulation or function of *NtSUS2* and *NtSUS4* in *N. sylvestris* and *N. tomentosiformis* during that time (Table 6.1). In addition to be induced in senescing leaf, *NtSUS4-S* was approximately six times more expressed in immature flowers and petals compared to *NtSUS4-T*, suggesting a particular role in remobilizing carbon sources in

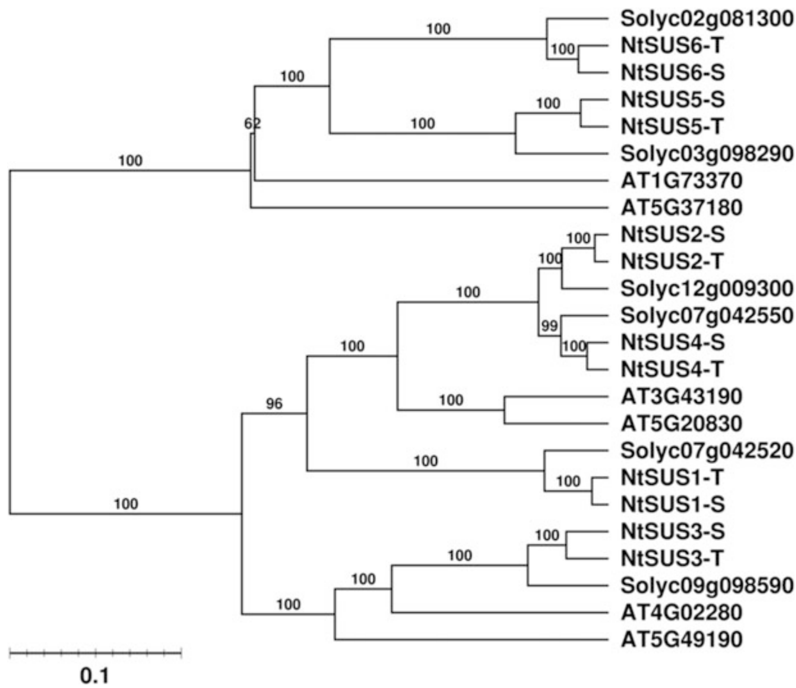

Fig. 6.5 Phylogenetic tree of *SUS* gene products of tobacco, tomato, and *Arabidopsis thaliana*. The translated amino acid sequences of 12 *Nicotiana tabacum SUS* (*NtSUS*) gene products derived from six copies of the ancestral parents, *N. sylvestris* (S) and *N. tomentosiformis* (T), of *N. tabacum* were used to construct the tree. In addition, the amino acid sequences of six tomato (*Solanum lycopersicum*) (Solyc) *SUS* genes and six *A. thaliana SUS* (*AtSUS*) genes were included in the phylogenetic tree. The software used is Phylip and numbers are bootstrap values, the distance between two sequences referred to the 0.1 scale

Table 6.1 Expression of *NtSUS* genes in green, ripe, and 48 h cured leaves of the tobacco varieties Stella, ITB683, and Katerini

	Stella			ITB683			Katerini		
	Green	Ripe	48 h curing	Green	Ripe	48 h curing	Green	Ripe	48 h curing
NtSUS1-S	3.4	0.2	0.0	10.4	0.0	0.4	3.4	0.1	0.5
NtSUS1-T	4.1	0.2	0.0	11.4	0.1	0.2	3.4	0.2	0.6
NtSUS2-S	0.2	0.3	28.5	0.3	0.1	1.7	0.6	2.3	17.8
NtSUS2-T	4.0	1.7	5.0	2.6	0.3	0.5	2.7	2.8	3.7
NtSUS3-S	23.4	65.6	130.4	14.2	57.8	82.3	18.7	42.0	163.9
NtSUS3-T	8.5	12.2	76.5	2.4	3.2	41.2	6.1	10.7	126.7
NtSUS4-S	11.6	4.9	22.8	15.5	3.6	22.9	19.0	18.9	105.4
NtSUS4-T	3.2	2.7	8.1	2.7	3.1	1.9	38.0	56.2	37.9
NtSUS5-S	0.6	0.2	0.1	5.1	0.1	0.5	1.4	0.3	0.8
NtSUS5-T	0.5	0.5	0.3	3.6	0.4	0.6	1.1	0.8	1.6
NtSUS6-S	7.8	9.5	8.2	7.6	7.7	5.4	7.8	8.2	8.3
NtSUS6-T	3.1	4.2	4.7	1.6	4.3	3.2	3.7	4.9	5.0

The transcripts related to sucrose metabolism are in gray. Data are expressed in fragments per kilobase million (FPKM). Abbreviation: h, hours

non-photosynthetic tissue, excluding roots (data not shown).

Three *NtSUS* copies (*NtSUS2-S*, *NtSUS3-S*, and *NtSUS4-S*) originating from *N. sylvestris* were expressed in leaf senescent tissues, unlike only one copy (*NtSUS3-T*) that originated from *N. tomentosiformis*, suggesting differential regulation of sucrose metabolism in these two ancestors. The production of reducing sugars must be ensured via *SUS* activity (upregulated *SUS* transcripts) likely be part of sucrose hydrolysis during senescence (Jin et al. 2009). Interestingly, *NtSUS3-S* and *NtSUS3-T* gene products are close to AtSUS3 from *Arabidopsis thaliana* (At4g02280, Fig. 6.5). AtSUS3 has been shown to be important for sucrose metabolism during the late maturation phase of seeds but also during drastic dehydration conditions (Baud et al. 2004). Additionally, in mature leaves, *AtSUS3* is specifically expressed in stomatal guard cells (Bieniawska et al. 2007). This suggests a role for tobacco *NtSUS3* orthologs in the water-deprivation response known to occur during tobacco leaf senescence and curing.

AAT catalyzes the reversible transfer of the α-amino group between aspartate and glutamate and thus is a key enzyme in amino acid metabolism by channeling nitrogen from glutamine and aspartate. In plants, AAT is present as several isoenzymes located in subcellular compartments such as the cytosol, mitochondria, peroxisomes, and plastids (De la Torre et al. 2014; Gaufichon et al. 2016).

The eight *NtAAT* genes from the ancestral parents of *N. tabacum*, *N. sylvestris* (S) and *N. tomentosiformis* (T) were used to construct the tree with the corresponding protein sequences termed NtAAT1-S, NtAAT1-T, NtAAT2-S, NtAAT2-T, NtAAT3-S, NtAAT3-T, NtAAT4-S, and NtAAT4-T (Fig. 6.6). Similar to the phylogenetic tree for *NtSUS* gene products, the *NtAAT* gene products are equally distributed from both tobacco ancestors and have close orthologs in tomato, suggesting a certain degree of conservation among the Solanaceae for these two gene families.

Among the *AAT* transcripts assigned gene ontology terms related to responses to chemicals (Fig. 6.4), *NtAAT2-S* and *NtAAT2-T* grouped with the tomato gene *Solyc08g068330* (see Fig. 6.5). *NtAAT2-S* and *NtAAT2-T* were the most expressed *NtAAT* genes after 48 h of curing

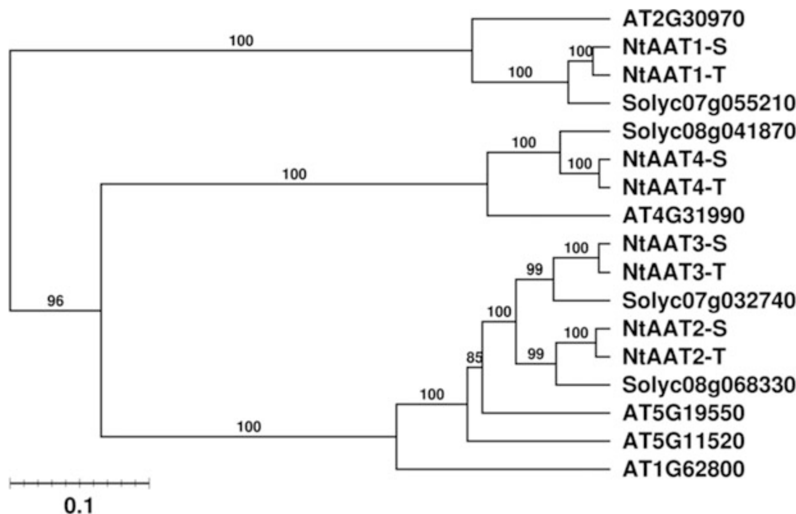

Fig. 6.6 Phylogenetic tree of aspartate aminotransferase (*AAT*) gene products of tobacco, tomato, and *Arabidopsis thaliana*. The translated amino acid sequences of eight *Nicotiana tabacum AAT* (*NtAAT*) gene products derived from four copies of the ancestral parents, *N. sylvestris* (S) and *N. tomentosiformis* (T), of *N. tabacum* were used to construct the tree. The amino acid sequences of four tomato (*Solanum lycopersicum*) (Solyc) *AAT* genes and five *A. thaliana AAT* (*AtAAT*) genes were included in the phylogenetic tree

Table 6.2 Gene expression of *NtAAT* genes in green, ripe, and 48 h cured leaves of the tobacco varieties Stella, ITB683, and Katerini

	Stella			ITB683			Katerini		
	Green	Ripe	48 h curing	Green	Ripe	48 h curing	Green	Ripe	48 h curing
NtAAT1-S	7.8	6.7	9.2	11.4	8.1	15.8	12.9	12.9	13.0
NtAAT1-T	10.5	13.4	28.5	11.0	5.5	28.1	12.4	10.0	38.9
NtAAT2-S	9.8	21.9	234.1	1.9	6.0	65.1	18.5	49.3	665.6
NtAAT2-T	2.7	9.4	105.8	1.1	2.3	12.0	5.1	11.7	80.4
NtAAT3-S	14.5	10.7	17.7	21.2	13.2	11.1	24.5	18.1	22.6
NtAAT3-T	14.7	13.0	21.0	18.2	16.6	9.7	29.4	29.0	22.7
NtAAT4-S	44.5	39.8	29.0	45.6	41.7	28.9	6.4	7.0	3.2
NtAAT4-T	60.7	47.0	21.6	57.3	55.9	19.2	44.9	47.3	25.8

The transcripts related to responses to chemicals are in gray. Data are expressed in fragments per kilobase million (FPKM). Abbreviation: h, hours; *NtAAT*, *Nicotiana tabacum aspartate aminotransferase*

(Table 6.2), with *NtAAT1-S* and *NtAAT1-T* also being upregulated, although to a lesser extent. *NtAAT2-S* and *NtAAT2-T* are specifically expressed in petals but not in roots, similar to *NtSUS4-S* (data not shown). AAT catalyzes the interconversion of aspartate and α-ketoglutarate to oxaloacetate and glutamate and thus is important to amino acid metabolism, bridging the carbon and nitrogen pathways. Aspartate synthesis is also essential for the synthesis of other amino acids, such as asparagine, threonine, isoleucine, cysteine, and methionine (De la Torre et al. 2014). For instance, aspartate is converted to asparagine by asparagine synthetase, and transcripts encoding this enzyme were annotated with gene ontology terms related to the asparagine biosynthetic process (Fig. 6.4). Asparagine and glutamine are also key compounds in nitrogen remobilization in senescent leaves (Avila-Ospina et al. 2015; Li et al. 2016). Moreover, asparagine is the major amino acid produced in cured Burley leaves via specific senescence-activated asparagine synthetases (see patent application: Plants with reduced asparagine content, WO 2017042162 A1). Both *NtAAT2-S* and *NtAAT2-T* transcripts were particularly and strongly expressed after 48 h of curing in all three tobacco varieties, suggesting that *AAT* senescence induction likely triggers the amino acid alteration observed in cured leaves (Fig. 6.2).

In summary, extensive chemical changes occur during tobacco leaf curing, affecting key metabolic compounds such as sugars and amino acids. These metabolic changes are triggered by the activation of a large genetic program that up- and downregulates approximately 20% of the transcriptome during the first curing days. Gene ontology term overrepresentation analysis for transcripts upregulated following curing identified *SUS*, *AAT*, and *asparagine synthetase* as metabolism-related genes actively involved in the chemical changes that occur in the three main tobacco varieties cultivated.

Acknowledgements The authors gratefully acknowledge the support of Delphine Langlet (plant growing and RNA isolation), Hervé Broye (field management), Hélène Laparra (greenhouse management), Florian Martin (co-initiator of the transcriptomic plan), Markus Klein (project support and discussion), Dion Florack (scientific discussion), and Luca Rossi (project support).

References

Avila-Ospina L, Marmagne A, Talbotec T, Krupinska K, Masclaux-Daubresse C (2015) The identification of new cytosolic glutamine synthetase and asparagine synthetase genes in barley (*Hordeum vulgare* L.), and their expression during leaf senescence. J Exp Bot 66 (7):2013–2026. https://doi.org/10.1093/jxb/erv003

Baud S, Vaultier MN, Rochat C (2004) Structure and expression profile of the sucrose synthase multigene family in *Arabidopsis*. J Exp Bot 55(396):397–409. https://doi.org/10.1093/jxb/erh047

Bieniawska Z, Barratt DHP, Garlick AP et al (2007) Analysis of the sucrose synthase gene family in *Arabidopsis*. Plant J 49:810–828. https://doi.org/10.1111/j.1365-313X.2006.03011.x

Chen D, Wang S, Xiong B, Cao B, Deng X (2015) Carbon/Nitrogen imbalance associated with drought-induced leaf senescence in *Sorghum bicolor*. PLoS ONE 10(8):e0137026. https://doi.org/10.1371/journal.pone.0137026

Christ B, Egert A, Süssenbacher I et al (2014) Water deficit induces chlorophyll degradation via the 'PAO/phyllobilin' pathway in leaves of homoio- (*Craterostigma pumilum*) and poikilochlorophyllous (*Xerophyta viscosa*) resurrection plants. Plant Cell Environ 37(11):2521–2531. https://doi.org/10.1111/pce.12308

De la Torre F, El-Azaz J, Ávila C, Cánovas FM (2014) Deciphering the role of aspartate and prephenate aminotransferase activities in plastid nitrogen metabolism. Plant Physiol 164(1):92–104. https://doi.org/10.1104/pp.113.232462

Diaz-Mendoza M, Velasco-Arroyo B, Santamaria ME et al (2016) Plant senescence and proteolysis: two processes with one destiny. Genet Mol Biol 39(3):329–338. https://doi.org/10.1590/1678-4685-GMB-2016-0015

Edwards KD, Ferandez-Poro N, Drake-Stowe K et al (2017) A reference genome for *Nicotiana tabacum* enables map-based cloning of homeologous loci implicated in nitrogen utilization efficiency. BMC Genom 18:448. https://doi.org/10.1186/s12864-017-3791-6

Gaufichon L, Rothstein SJ, Suzuki A (2016) Asparagine metabolic pathways in *Arabidopsis*. Plant Cell Physiol 57(4):675–689. https://doi.org/10.1093/pcp/pcv184

Gepstein S, Thimann KV (1980) Changes in the abscisic acid content of oat leaves during senescence. Proc Natl Acad Sci USA 77(4):2050–2053

Guo Y, Gan S (2005) Leaf senescence: signals, execution, and regulation. Curr Top Dev Biol 71:83–112. https://doi.org/10.1016/S0070-2153(05)71003-6

Guo Y, Li CI, Ye F, Shyr Y (2013) Evaluation of read count based RNAseq analysis methods. BMC Genom 14(8):S2. https://doi.org/10.1186/1471-2164-14-S8-S2

Hickman R, Hill C, Penfold CA et al (2013) A local regulatory network around three NAC transcription factors in stress responses and senescence in *Arabidopsis* leaves. Plant J 75(1):26–39. https://doi.org/10.1111/tpj.12194

Jin Y, Ni D-A, Ruan Y-L (2009) Posttranslational elevation of cell wall invertase activity by silencing its inhibitor in tomato delays leaf senescence and increases seed weight and fruit hexose level. Plant Cell 21(7):2072–2089. https://doi.org/10.1105/tpc.108.063719

Kim HJ, Nam HG, Lim PO (2016) Regulatory network of NAC transcription factors in leaf senescence. Curr Opin Plant Biol 33:48–56. https://doi.org/10.1016/j.pbi.2016.06.002

Leffingwell JC (1999) Leaf chemistry. In: Davis DL, Nieslson MT (eds) Tobacco: production, chemistry, and technology. Blackwell Science

Legg PD, Chaplin JF, Williamson RE (1977) Genetic diversity in Burley and flue-cured tobacco. Crop Sci 17:943–947. https://doi.org/10.2135/cropsci1977.0011183X001700060032x

Li L, Zhao J, Zhao Y et al (2016) Comprehensive investigation of tobacco leaves during natural early senescence via multi-platform metabolomics analyses. Sci Rep 6:37976. https://doi.org/10.1038/srep37976

Ludewig F, Sonnewald U (2000) High CO_2-mediated down-regulation of photosynthetic gene transcripts is caused by accelerated leaf senescence rather than sugar accumulation. FEBS Lett 479(1–2):19–24. https://doi.org/10.1016/S0014-5793(00)01873-1

Maere S, Heymans K, Kuiper M (2005) BiNGO: a cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics 21(16):3448–3449. https://doi.org/10.1093/bioinformatics/bti551

Masclaux-Daubresse C, Carrayol E, Valadier MH (2005) The two nitrogen mobilisation- and senescence-associated *GS1* and *GDH* genes are controlled by C and N metabolites. Planta 221(4):580–588. https://doi.org/10.1007/s00425-004-1468-2

Noh YS, Amasino RM (1999) Identification of a promoter region responsible for the senescence-specific expression of SAG12. Mol Biol 41(2):181–194. https://doi.org/10.1023/A:1006342412688

Olvera-Carrillo Y, Campos F, Reyes JL, Garciarrubio A, Covarrubias AA (2010) Functional analysis of the group 4 late embryogenesis abundant proteins reveals their relevance in the adaptive response during water deficit in *Arabidopsis*. Plant Physiol 154(1):373–390. https://doi.org/10.1104/pp.110.158964

Patro L, Mohapatra PK, Biswal UC, Biswal B (2014) Dehydration induced loss of photosynthesis in *Arabidopsis* leaves during senescence is accompanied by the reversible enhancement in the activity of cell wall β-glucosidase. J Photochem Photobiol B 137:49–54. https://doi.org/10.1016/j.jphotobiol.2014.03.018

Pourtau N, Jennings R, Pelzer E, Pallas J, Wingler A (2006) Effect of sugar-induced senescence on gene expression and implications for the regulation of senescence in *Arabidopsis*. Planta 224(3):556–568. https://doi.org/10.1007/s00425-006-0243

Rogers H, Munné-Bosch S (2016) Production and scavenging of reactive oxygen species and redox signaling during leaf and flower senescence: similar but different. Plant Physiol 171(3):1560–1568. https://doi.org/10.1104/pp.16.00163

Shi H, Wang B, Yang P, Li Y, Miao F (2016) Differences in sugar accumulation and mobilization between sequential and non-sequential senescence wheat cultivars under natural and drought conditions. PLoS ONE 11(11):e0166155. https://doi.org/10.1371/journal.pone.0166155

Sierro N, Battey JND, Ouadi S et al (2014) The tobacco genome sequence and its comparison with those of

tomato and potato. Nature Com 5:3833. https://doi.org/10.1038/ncomms4833

Sonnewald U, Brauer M, von Schaewen A, Stitt M, Willmitzer L (1991) Transgenic tobacco plants expressing yeast-derived invertase in either the cytosol, vacuole or apoplast: a powerful tool for studying sucrose metabolism and sink/source interactions. Plant J 1(1):95–106. https://doi.org/10.1111/j.1365-313X.1991.00095.x

Stitt M, Wilke I, Feil R, Heldt HW (1988) Coarse control of sucrose-phosphate synthase in leaves: alterations of the kinetic properties in response to the rate of photosynthesis and the accumulation of sucrose. Planta 174(2):217–230. https://doi.org/10.1007/BF00394774

Suzuki Y, Makino A (2013) Translational downregulation of RBCL is operative in the coordinated expression of Rubisco genes in senescent leaves in rice. J Exp Bot 64(4):1145–1152. https://doi.org/10.1093/jxb/ers398

Troncoso-Ponce MA, Cao X, Yang Z, Ohlrogge JB (2013) Lipid turnover during senescence. Plant Sci 205–206:13–19. https://doi.org/10.1016/j.plantsci.2013.01.004

Uzelac B, Janošević D, Simonović A et al (2016) Characterization of natural leaf senescence in tobacco (*Nicotiana tabacum*) plants grown in vitro. Protoplasma 2:259–275. https://doi.org/10.1007/s00709-015-0802-9

Wingler A, Masclaux-Daubresse C, Fischer AM (2009) Sugars, senescence, and ageing in plants and heterotrophic organisms. J Exp Bot 60(4):1063–1066. https://doi.org/10.1093/jxb/erp067

Yang J, Zhang J, Wang Z, Zhu Q, Liu L (2002) Abscisic acid and cytokinins in the root exudates and leaves and their relationship to senescence and remobilization of carbon reserves in rice subjected to water stress during grain filling. Planta 215(4):645–652. https://doi.org/10.1007/s00425-002-0789-2

Zhang K, Xia X, Zhang Y, Gan S-S (2012) An ABA-regulated and Golgi-localized protein phosphatase controls water loss during leaf senescence in *Arabidopsis*. Plant J 69(4):667–678. https://doi.org/10.1111/j.1365-313X.2011.04821

Check for updates

7 Repetitive DNA Dynamics and Polyploidization in the Genus *Nicotiana* (Solanaceae)

Steven Dodsworth, Ales Kovarik, Marie-Angèle Grandbastien, Ilia J. Leitch, and Andrew R. Leitch

Abstract

Large variations in genome size are observed in angiosperms as a result of whole-genome duplications and the balance between amplification and deletion of repetitive DNA, together explaining the observed variation in plant genome size. In the genus *Nicotiana*, there are 42 cytogenetically diploid species that have been classified into eight sections. There are also six allopolyploid *Nicotiana* sections that have evolved from species in different diploid sections. The phylogenetic relationships among these *Nicotiana* species, along with recurrent polyploidization events, permits the divergence of repetitive content in both diploid and allopolyploid genomes to be compared through evolutionary time. In this chapter, we review genome size variation in *Nicotiana* that reveals both genome upsizing and genome downsizing in different polyploid species. We discuss the divergence of specific repetitive elements, including tandemly repeated satellite DNAs, retroelements, and intergenic spacers as well as the sub-repeats contained in 35S rDNA. The lag-phase hypothesis, which describes post-polyploid radiations, is posed as a potentially important mechanism of evolution in *Nicotiana* section *Suaveolentes*, the largest polyploid section that consists of over half the current species diversity.

S. Dodsworth
School of Life Sciences, University of Bedfordshire, Luton LU1 3JU, UK

A. Kovarik
Institute of Biophysics, Academy of Sciences of the Czech Republic, Brno 61265, Czech Republic

M.-A. Grandbastien
Institut Jean-Pierre Bourgin, INRA, AgroParisTech, CNRS, University Paris-Saclay, 78000 Versailles, France

I. J. Leitch
Department of Comparative Plant and Fungal Biology, Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3DS, UK

A. R. Leitch (✉)
School of Biological and Chemical Sciences, Queen Mary University of London, Mile End Road, London E1 4NS, UK
e-mail: a.r.leitch@qmul.ac.uk

7.1 Introduction to Plant Repetitive DNA

There is, approximately, a 2,400-fold range in genome size (GS; measured as the amount of DNA in giga base pairs [Gbp] in the haploid unreplicated 1C nucleus) across angiosperms, ranging from ~0.061 Gbp/1C in *Genlisea aurea* to ~148 Gbp/1C in *Paris japonica* (Pellicer et al. 2010), a larger range than that observed for any comparable group of eukaryotes (Leitch and Leitch 2012; Pellicer et al. 2018). This variation in GS arises from two sources: polyploidy

© Springer Nature Switzerland AG 2020

N. V. Ivanov et al. (eds.), *The Tobacco Plant Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-030-29493-9_7

(whole-genome duplication) and the balance between processes that amplify repetitive DNA (i.e., (retro)transposition and recombination-based amplification) and those that delete repeats and other types of non-essential DNA (i.e., recombination-based deletion) (Kejnovsky et al. 2012).

Despite the large range in GS found across angiosperms, most species have a small GS (modal GS 0.587 Gbp/1C, mean GS 5.020 Gbp/1C, n = 10,768 species, Pellicer et al. 2018). This is interesting given the context of multiple ancestral polyploid events in most, if not all, angiosperm lineages. For example, within seed plants, the lineage leading to Solanaceae is predicted to have been through multiple rounds of whole-genome duplication, such that all species are at least paleo-36-ploid (Wendel 2015). Such high ploidal levels are not generally reflected in large genomes, most likely as a consequence of genome downsizing that typically follows polyploidy and involves the deletion of both genes and repeats during the process of post-polyploidization diploidization (Leitch and Bennett 2004; Dodsworth et al. 2016). Another general feature reported for repeats in plants with a small to medium GS (<10 Gbp/1C) is that many are highly dynamic, with amplification and deletion processes occurring so rapidly such that the half-life of any particular repeat is reported to be only 3–4 million years in Poaceae, Brassicaceae, Fabaceae, and Vitaceae (Ma et al. 2004; El Baidouri and Panaud 2013).

The proportion of a plant genome that is repetitive is highly variable between taxa. This partly reflects differences in the balance between amplification and deletion processes, which in turn influence the occurrence and longevity of (micro)satellite repeats, (retro)transposable elements, and their truncated derivatives. For example, conifers typically have larger GS than most lineages of angiosperms, probably reflecting contrasting genome dynamics between gymnosperms and angiosperms (Leitch and Leitch 2012). There are also differences in repetitive content of the genome depending on GS, such that species with the smallest GS have the smallest genome proportion occupied by repeats. For example, the estimated genome proportion in *Utricularia gibba* with a GS of 0.088 Gbp/1C is just 3% (Ibarra-Laclette et al. 2013) versus a reported 91.6% genome proportion in *Aegilops tauschii*, with a GS of 4.3 Gbp/1C (Li et al. 2004).

7.2 Introduction to *Nicotiana*

There are 42 species of *Nicotiana* that are cytogenetically diploid, falling into eight sections: *Alatae* with eight species (haploid chromosome count, n = 9–10), *Noctiflorae* with six species (n = 12), *Petunioides* with eight species (n = 12), *Paniculatae* with seven species (n = 12), *Sylvestres* with a single species (n = 12), *Tomentosae* with five species (n = 12), *Trigonophyllae* with two species (n = 12), and *Undulatae* with five species (n = 12). Homoploid hybrids have been reported to form between some of the sections, particularly in species belonging to *Petunioides* and *Noctiflorae* (Knapp et al. 2004; Kelly et al. 2010).

In addition, there are six allopolyploid *Nicotiana* sections, thought to involve species from different diploid sections, and providing an opportunity to study and compare repeat divergence in diploid lineages and the genomic melting pot generated by allopolyploidy itself. All the allopolyploids retain a signature of polyploidy in chromosome counts, and evidence of interspecific hybridization is seen in phylogenetic patterns from multiple nuclear markers, including the internally transcribed spacer sequences of ribosomal DNA (rDNA) (Clarkson et al. 2010, 2017), paralogs of the nuclear genes *ADH*, *GS*, *FLO/LFY*, *WAXY*, and *MADS1/FUL* (Kelly et al. 2010), and comparisons between these and plastid sequences from the maternal genome donor (Clarkson et al. 2004, 2010). Figure 7.1 summarizes the phylogenetic relationships in genus *Nicotiana*.

(1) The oldest allopolyploid section, *Nicotiana* section *Suaveolentes*, contains at least 40 species (Knapp et al. 2004). This section has radiated substantially in Australia, and there are likely to be many more taxa than are currently

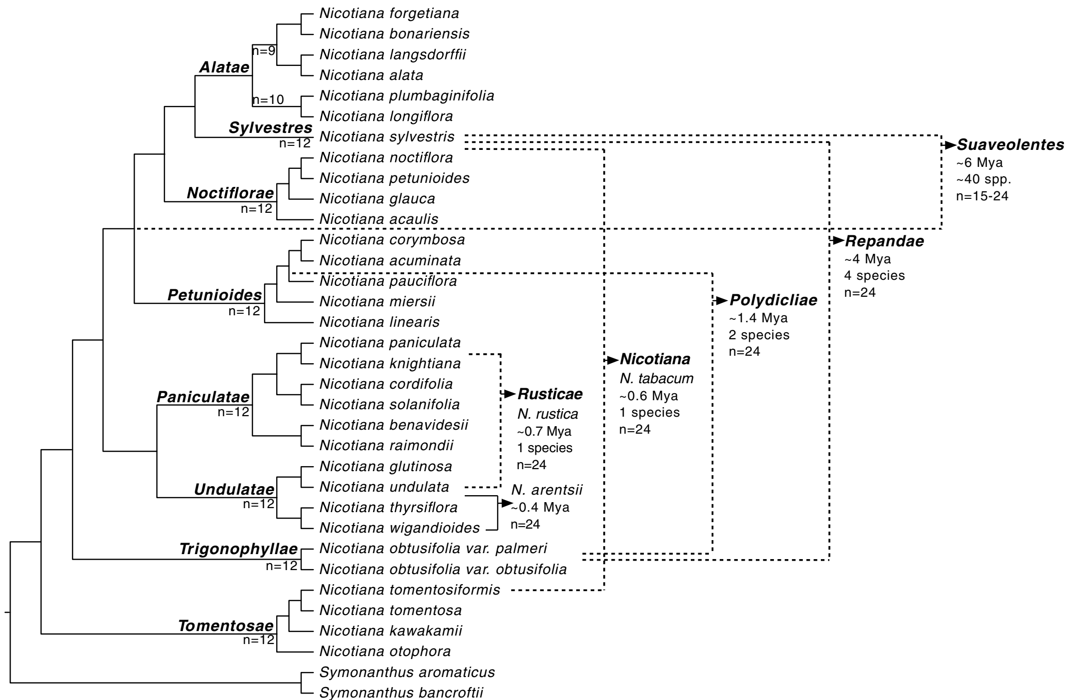

Fig. 7.1 Summary of phylogenetic relationships in the genus *Nicotiana*, including the dates of formation (Clarkson et al. 2017) and number of species in each of the allopolyploid sections (Knapp et al. 2004). The diploid phylogenetic tree was reconstructed from combined plastid and nrITS data and is summarized from Clarkson et al. (2004) and Kelly et al. (2010). Haploid chromosome numbers (*n*) are given for each section. Section names in bold have been given. There are three polyploid sections containing multiple species (*Nicotiana* sections *Suaveolentes*, *Repandae* and *Polydicliae* and two monotypic polyploid sections, *Nicotiana* section *Nicotiana* (*N. tabacum*) and *Rusticae* (*N. rustica*)). The polyploid species *N. arentsii* is in *Nicotiana* section *Undulatae*, which also includes diploid species

recognized (four new species were described recently by Chase et al. 2018). The section includes *N. benthamiana*, a species widely used in studies of plant–pathogen interactions, and whose genome has been fully sequenced (Bombarely et al. 2012). Time-calibrated phylogenetic reconstructions suggest that the section probably formed about 6 million years ago (Mya). The diploid progenitors of section *Suaveolentes* are difficult to determine, likely because of their formation from one or more diploid hybrid taxa. Phylogenetic evidence suggests the involvement of several diploid sections (*Noctiflorae*, *Alatae*, and *Petunioides*) as the maternal genome donor, with *N. sylvestris* (section *Sylvestres*) as the most closely related extant relative of the paternal genome donor (Clarkson et al. 2010, 2017; Schiavinato et al. 2020). Chromosome numbers vary from the expected n = 24 (the sum of the parental chromosome counts, for an allotetraploid) down to n = 15, depending on the species. The species with lower chromosome numbers are presumed to have arisen from post-polyploidization dysploidy involving chromosome rearrangements, as part of a broader diploidization process.

(2) *Nicotiana* section *Repandae* has four allotetraploid species, each with n = 24 chromosomes, with two species that are endemic to the Revillagigedo Islands. The section is thought to have formed about 4 Mya (Clarkson et al. 2017) from a maternal parent most closely related to extant *N. sylvestris* and a paternal species related to *Nicotiana* section *Trigonophyllae* (e.g., *N. obtusifolia* var. *obtusifolia*/*N. obtusifolia* var. *palmeri*).

(3) *Nicotiana* section *Polydicliae* has two species, each with n = 24 chromosomes, and one of which, *N. quadrivalvis*, was cultivated by native North Americans, presumably for recreational use. The section formed about 1.4 Mya (Clarkson et al. 2017) from progenitors of *N. obtusifolia* (section *Trigonophyllae*; maternal genome donor) and a member of section *Petunioides* as the paternal genome donor.

(4) *Nicotiana rustica* (section *Rusticae*, n = 24) formed about 0.7 Mya (Clarkson et al. 2017) from progenitors of *N. undulata* (section *Undulatae*; paternal) and section *Paniculatae* (either *N. paniculata* or *N. knightiana*) as the maternal progenitor.

(5) *Nicotiana tabacum* (section *Nicotiana*, n = 24), the best-known *Nicotiana* allopolyploid, formed about 0.6 Mya (Clarkson et al. 2017) from progenitors of *N. sylvestris* (section *Sylvestris*; maternal genome donor) and *N. tomentosiformis* (section *Tomentosae*; paternal).

(6) *Nicotiana arentsii* (section *Undulatae*, n = 24) formed about 0.4 Mya (Clarkson et al. 2017) from two other species in section *Undulatae*, most closely related to extant *N. undulata* (section *Undulatae*; maternal parent) and *N. wigandioides* (section *Undulatae*; paternal parent).

7.3 Genome Size Variation in *Nicotiana*

An analysis of angiosperm GSs, grouped into presumed ploidal levels based on chromosome counts (from diploid to octoploid), indicated that higher ploidal levels were not associated with larger GS, as might be predicted (Leitch and Bennett 2004). These findings suggested that following polyploidy, there is a tendency toward the loss of genomic DNA, termed genome downsizing, perhaps caused by selection against large GS in angiosperms (Leitch and Leitch 2012; Guignard et al. 2016). Genome downsizing is predicted as a likely occurrence given enough time post-polyploidization. Serial genome rearrangements associated with genome downsizing lead to the so-called "wondrous cycles of polyploidy" reported for many angiosperm lineages (Wendel 2015).

To predict the direction of GS change following ancient polyploidy, it is necessary to reconstruct the size of ancestral diploid genomes at the point of polyploid formation. Such ancestral GSs can be estimated by summing the GSs of the most closely related extant diploids (Leitch et al. 2008). Alternatively, they can be derived by reconstructing the ancestral diploid GSs of diploid lineages at the point of allopolyploidy. This is achieved using ancestral reconstruction approaches, which requires a confident phylogenetic assignment of parental lineages. With estimates of the ages of polyploid events, rates of genome size change can be estimated. The predicted GS of ancestral polyploids can then be compared with the actual GS of the extant polyploid taxa. When such an analysis is conducted on *Nicotiana*, some polyploid species are predicted to have experienced genome upsizing, while most experienced genome downsizing.

Given the GS of the progenitor diploid taxa, the allotetraploids *N. tabacum* and *N. arentsii* are thought to have undergone genome downsizing of ~3.7% and ~3.9%, respectively, over ~0.4–0.6 million years. Similarly, *N. rustica* is likely to have downsized by ~1.9–5.4% over a similar time frame. In contrast, in section *Polydicliae*, divergence over a longer time frame of ~1.4 million years is thought to have given rise to approximately 2.5% (*N. clevelandii*) and 7.5% (*N. quadrivalvis*) increases in GS. Both genome upsizing and downsizing have been reported in species in section *Repandae*, over ~4 million years, with genome upsizing in *N. repanda* (~26.6%), *N. nesophila* (~19.1%), and *N. stocktonii* (~19.1%), and downsizing in *N. nudicaulis* (~14.3%) (Leitch et al. 2008). In the oldest section, *Suaveolentes*, which originated ~6 Mya, almost all species have undergone substantial genome downsizing associated with their divergence (Fig. 7.2). This has generated genomes that are similar in size to the mean GS of extant *Nicotiana* diploids (i.e., ~3.2 Gbp/1C, 34 species) and not much larger than *N. sylvestris* (2.6 Gbp/1C, paternal parent) and the average GS of species in sections *Noctiflorae*,

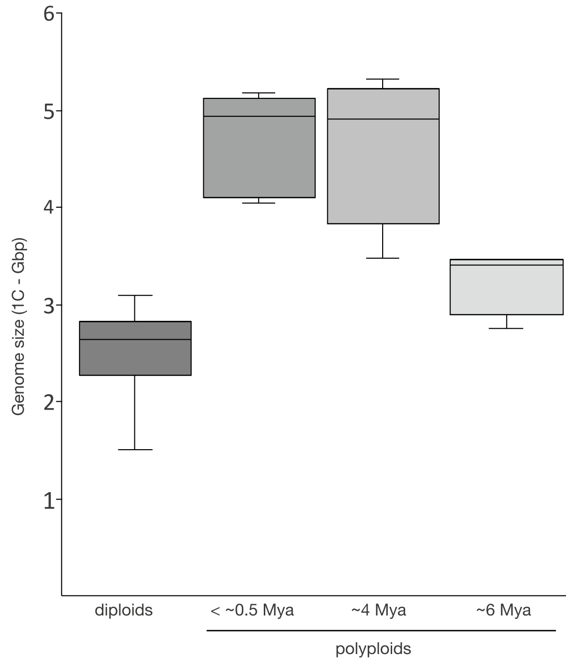

Fig. 7.2 Box plots showing the distribution of genome sizes (1C-values) across diploid and polyploid *Nicotiana* taxa. Young polyploids show near-additive and/or genome upsizing, whereas older polyploid groups tend to show genome downsizing. Analysis based on data from Leitch et al. (2008) for diploid (nine taxa), young polyploid taxa (*N. tabacum*, *N. rustica*, *N. arentsii*, *N. clevelandii*, and *N. quadrivalvis*), and the 4-Mya section *Repandae* (all four taxa), and unpublished data for the 6-Mya section *Suaveolentes* (five taxa from the core Australian group: *N. simulans*, *N. velutina*, *N. maritima*, *N. truncata,* and *N. goodspeedii*)

Alatae, and *Petunioides* (about 3.4 Gbp/1C, 14 species) of which an extinct relative or a homoploid hybrid is likely to have been the maternal parent. Such genome downsizing likely arises as part of the diploidization process that includes chromosome rearrangements and dysploidy (reduction from n = 24 to n = 15 chromosomes in some taxa).

7.4 Repeat Divergence in Diploid *Nicotiana*

In concert with GS variation, diploid *Nicotiana* genomes contain variable amounts of repetitive elements. The types, abundance, and patterns of element accumulation within the genomes of *Nicotiana* are, however, fairly consistent across the different diploid taxa thus far investigated. *Nicotiana* genomes are dominated by retrotransposons, particularly Ty3/Gypsy and Ty1/Copia elements. The most abundant retroelements in all *Nicotiana* genomes studied are Ty3/Gypsy elements, most of which are chromoviruses, particularly Tekay elements (Fig. 7.3). Most of the retrotransposon families found in *Nicotiana* are also present in genus *Symonanthus*, which is thought to be the closest relative to *Nicotiana*, based on phylogenetic evidence available to date (Clarkson et al. 2004, 2010; Kelly et al. 2010; Särkinen et al. 2013). Nevertheless, some elements are unique to *Symonanthus* (e.g., Ty1/Copia elements of the Ikaros family) and do not appear to be present in *Nicotiana* genomes (Fig. 7.3).

Xu et al. (2017) examined repeats in four diploid *Nicotiana* species, as well as in potato (*Solanum tuberosum*) and tomato (*Solanum lycopersicum*), and suggested there was an expansion of Ty3/Gypsy retroelements in *Nicotiana* that correlated with its increased GS. In Fig. 7.3, we present further evidence that GS is

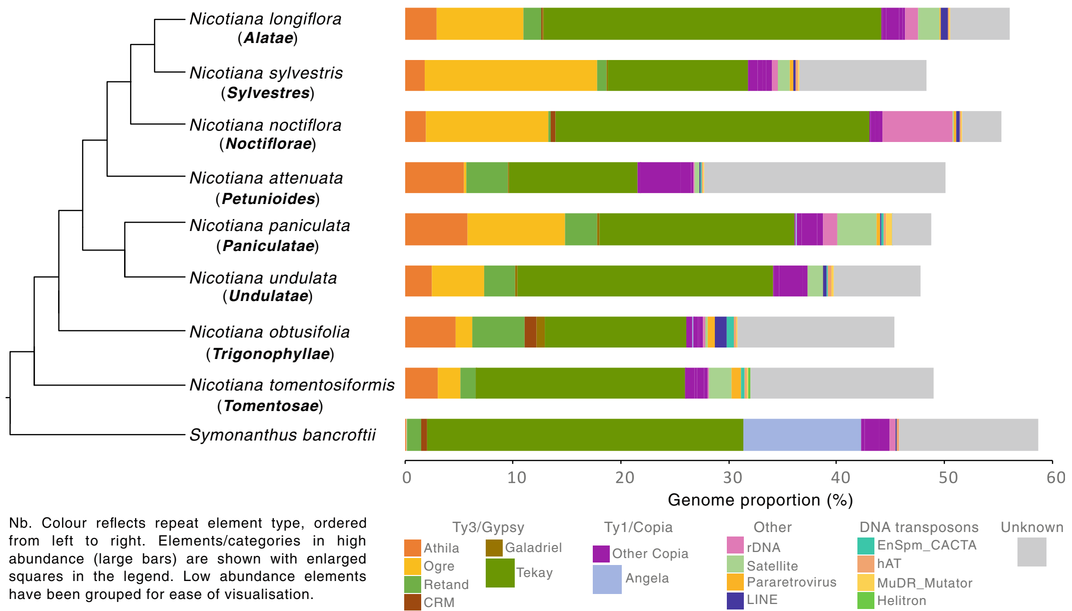

Fig. 7.3 Comparison of the repetitive DNA content in diploid *Nicotiana* species (with *Symonanthus* as the outgroup) as analyzed using high-throughput sequence data (Illumina) and RepeatExplorer2 de novo clustering (Novak et al. 2010; Novák et al. 2013). The abundance of repetitive elements is shown as a proportion of the genome and annotations are shown for major repeat types using REXdb (Neumann et al. 2019) as implemented in RepeatExplorer2. Abundant categories are shown with enlarged squares in the legend. Phylogenetic relationships are summarized and section names are shown in bold

correlated with the abundance of retroelements, particularly Ty3/Gypsy retrotransposons, across a broader set of eight *Nicotiana* diploid species (including representatives from all diploid sections). Xu et al. (2017) also suggested there was an expansion of a Solanaceae-specific subgroup of MITE elements (called DTT-NIC1) in *N. attenuata* that were enriched within a 1-Kbp region upstream of the genes involved in nicotine biosynthesis. They proposed that these elements may have been involved in the up-regulation of genes involved in nicotine biosynthesis. Other class II DNA transposons that are present in all *Nicotiana* genomes analysed to date include hAT and MuDR_Mutator elements (Fig. 7.3), while EnSpm/CACTA elements were found in the genomes of about half of the species analysed.

Using molecular and cytogenetic approaches, much research has been conducted on tandemly repeated satellite DNAs of *Nicotiana*, including well-studied families such as GRS (Gazdová et al. 1995), NTRS (Matyasek et al. 1997), geminivirus-related DNA (GRD) (Bejarano et al. 1996), and HRS60 (e.g., Fajkus et al. 1995b; Gazdová et al. 1995) (Table 7.1). These repetitive elements were isolated predominantly from telomeric and subtelomeric domains (Fajkus et al. 1995a) and centromeric and pericentromeric domains (Shibata et al. 2013), and their overall structure and typically high levels of cytosine methylation studied (Kovarik et al. 2000). In addition, there has been considerable focus on sequence divergence, activity, cytosine methylation, and distribution of 5S (Fulneček et al. 2002; Matyasek et al. 2002) and 35S rDNA loci (Kovarik et al. 2004).

The structure of the intergenic spacer and the sub-repeats contained in 35S rDNA have been further studied because they are thought to be involved in driving homogenization of rDNA unit arrays and the regulation of rDNA expression (Borisjuk et al. 1997; Volkov et al. 1999; Lim et al. 2004b; Kovarik et al. 2008). One 135-bp sub-repeat termed A1/A2, which was isolated

Table 7.1 Repeats that characterize the chromosome domains of *Nicotiana tabacum* (common tobacco)

Repeat family	Feature	Genome proportion (%)	Chromosome localization	Subgenome specificity		Reference
				S	T	
GRD3	Endogenous geminivirus	<0.01	Interstitial	−	+	(Bejarano et al. 1996; Murad et al. 2002)
GRD5	Endogenous geminivirus	<0.01	Interstitial, subtelomeric	+	+	(Bejarano et al. 1996; Murad et al. 2002)
NsEPRV	Endogenous pararetrovirus	n.d.	Dispersed	+	−	(Matzke et al. 2004)
NtoEPRV	Endogenous pararetrovirus	n.d.	Dispersed	−	+	(Gregor et al. 2004)
Tnt1	Retroelement Ty1/Copia	<0.01	Dispersed	+	+	(Melayah et al. 2004)
NTS9	Satellite	0.075	Pericentromeric, interstitial	+	−	(Moscone et al. 1996)
HRS60	Satellite	2	Subtelomeric	+	−	(Koukalova et al. 1990, 2010; Kenton et al. 1993)
GRS	Satellite	0.15	Pericentromeric	−	+	(Lim et al. 2000b)
NTRS	Satellite	0.75	Interstitial	−	+	(Matyasek et al. 1997)
A1/A2	Satellite, related to the intergenic spacer of rDNA	0.022	Interstitial, subtelomeric	−	+	(Lim et al. 2004b)
NicCL3	Satellite	0.1	Subtelomeric	−	+	(Renny-Byfield et al. 2012)
TAS49	Dispersed	0.15	Interstitial, subtelomeric,	±	+	(Horakova and Fajkus 2000)
R8.1	Retroelement	0.3	Dispersed	−	+	(Kuhrova et al. 1991)
NtoCR	Retroelement, TY3/Gypsy	n.d.	Centromeric, pericentromeric	−	+	(Nagaki et al. 2011)
NtCR	Retroelement, TY3/Gypsy	n.d.	Centromeric, pericentromeric	+	+	(Nagaki et al. 2011)
RETS	Tandem	n.d.	Interstitial, subtelomeric	+	−	(Shibata et al. 2013)
NTA11	Tandem	n.d.	Interstitial	−	+	(Shibata et al. 2013)
HT3E06	Tandem	n.d.	Centromeric	+	−	(Shibata et al. 2013)
HT1H04	Tandem	n.d.	Centromeric, pericentromeric	+	−	(Shibata et al. 2013)
HT3G02	Tandem	n.d.	Centromeric, pericentromeric	−	+	(Shibata et al. 2013)
Telomeric repeats	TTTAGGG	0.053	Mostly telomeric, a few interstitial signals	+	+	(Kenton et al. 1993)

Subgenomes are *N. sylvestris* (S) and *N. tomentosiformis* (T); ± signs indicate presence and absence, respectively; n.d., none detected

from the 26–18S rDNA intergenic spacer of *N. tomentosiformis*, was shown to be dispersed across several chromosomes (Lim et al. 2004b). A comparison of the distribution of A1/A2 repeats in species of section *Tomentosae* showed that they were present in low copy numbers in all species except *N. tomentosiformis* and likely expanded in the lineage leading to this species.

The HRS60 family of repeats occurs at a subtelomeric location in all diploid *Nicotiana* sections examined (i.e., *Alatae*, *Sylvestres*, *Undulatae*, and *Paniculatae*), but not in section *Tomentosae* where the HRS60 repeats are interstitial; the significance of this is currently unclear (Lim et al. 2000b, 2004b, 2005, 2006). All the HRS60 sequences are clearly related, and variants have diverged and undergone sequence homogenization such that each *Nicotiana* section has its own characteristic member of this family (Koukalova et al. 2010). The predominantly subtelomeric location of the HRS60 family of sequences and rDNA loci in *Nicotiana* means that the sequences interface directly with telomeric and degenerate telomeric motifs. The colocalisation of these apparently distinct domains may have consequences for plant genome evolution, and certainly they share some interesting biology; for example, they can generate non-coding transcripts, promote large-scale chromosomal rearrangements (e.g., Robertsonian fusions), and can be associated with nucleoli (Dvořáčková et al. 2015).

The GRD tandem repeat family is of particular interest and is thought to have derived from a free-living geminivirus (with which there is much sequence similarity) that integrated into the genome. This integration is proposed to have occurred in an ancestor of diploid species in section *Tomentosae*, where it amplified in a tandem array (GRD5 sequences on homologous chromosome 2). Later, specifically in the lineage leading to *N. tomentosiformis*, but before the formation of *N. tabacum*, a related sequence integrated *de novo* at a new chromosomal locus (GRD3 on chromosome 2) where it too amplified in a tandem array. Sequence analysis suggests that a large number of synonymous changes occurred in GRD3 compared with the number of synonymous changes in GRD5, indicative of purifying selection between integration events (Murad et al. 2002, 2004). The best explanation of these data is that there was a recombination event involving GRD5 and another free-living geminivirus, which generated the new GRD3 sequence that was subsequently re-integrated into the genome in a process that may have involved a helitron transposable element (Murad et al. 2004). Thus, *N. tabacum* inherited both GRD3 and GRD5 from *N. tomentosiformis* at its formation, and subsequently a variant sequence (GRD53) integrated at a third locus in the tobacco genome, potentially also involving a helitron (Murad et al. 2004).

7.5 Divergence of Repeats in Allopolyploid *Nicotiana*

Allopolyploids have formed in *Nicotiana* over widely different timescales (<0.5–6 Mya), frequently involving the same diploid parental lineages, and creating multiple species in section *Suaveolentes*. In addition, multiple synthetic polyploid lines have been developed. This makes the *Nicotiana* genus an ideal model to determine the fate of repetitive sequences and genome restructuring subsequent to polyploidization over a range of timescales. Using sequence-specific amplified polymorphisms to characterize the distribution of Copia-like retrotransposons, Petit et al. (2007) compared *N. tabacum* with its progenitor diploids and found that, in *N. tabacum*, the different Copia-like elements studied showed unique patterns that arose subsequent to the formation of *N. tabacum*. The frequencies of losses and gains of the observed insertion sites reflect retroelement mobility and sequence losses. Such changes are also apparent in the S4 generation of synthetic tobacco polyploids, where significant amplifications of Tnt1 Copia elements have been observed (Petit et al. 2010). As first proposed by McClintock (1984), the genomic shock of polyploidy, which is thought to arise from the merger of two distinct parental genomes in the nucleus, potentially stimulated rapid and dynamic

genomic changes early in *N. tabacum* divergence (Petit et al. 2010).

In addition to changes in the mobility of retroelements, the targeted loss of tandemly repeated and dispersed repeats in *N. tabacum*, particularly from the *N. tomentosiformis*-derived subgenome, has been reported from tobacco accessions (Skalicka et al. 2005; Renny-Byfield et al. 2011) and, to a lesser extent, from synthetic tobacco lines (Skalicka et al. 2003). For example, the NicCL3 tandem repeat sequence has a high copy number in *N. tomentosiformis* as in other diploids in section *Tomentosae* (Renny-Byfield et al. 2012). In contrast, in *N. tabacum* the NicCL3 tandem repeat sequence has a much lower copy number than in *N. tomentosiformis*, indicative of its loss from the *N. tomentosiformis*-derived subgenome of *N. tabacum* (Renny-Byfield et al. 2012). Similarly, there are fewer copies of an endogenous pararetrovirus-like sequence (Matzke et al. 2004) and of A1/A2 tandem repeats (Lim et al. 2004b) in *N. tabacum* than in *N. tomentosiformis.* The loss of the repeats from the paternally derived *N. tomentosiformis* progenitor's genome is an example of asymmetric loss of DNA from parental subgenomes, a phenomenon called biased fractionation (cf. Wendel 2015). This may have arisen from incompatibilities between the maternally derived cytoplasm and the biparentally derived nucleus at the time of allopolyploid formation (Leitch et al. 2006), as proposed by the nuclear–cytoplasmic interaction hypothesis (Gill and Friebe 2013).

A comparison of six transposable elements using sequence-specific amplified polymorphisms in synthetic polyploids of *N. tabacum*, *N. rustica*, and *N. arentsii* and multiple accessions of their diploid progenitors revealed that although element losses were evident in newly formed polyploids (Mhiri et al. 1997, 2019), element mobility is not apparent until later generations, suggesting that meiosis stimulates mobility. Nevertheless, it should be noted that the dynamics of the elements were not the same in the three synthetic polyploids and that the most active elements in the diploid progenitors were those that had the greatest impact on the early genome divergence of the polyploids (Mhiri et al. 2019).

Unlike the repeats described in the previous paragraphs, fluorescent in situ hybridization to metaphase spreads of *N. tabacum*, *N. rustica*, and *N. arentsii* revealed complete additivity in the number and distribution of rDNA loci to those found in the parental diploids (Lim et al. 2004a). In some S4 generation synthetic tobacco lines, there was perfect additivity of rDNA loci and rDNA units to that found in the parents; however, in other lines, new loci had evolved and new units amplified. Furthermore, in some of these lines, new units with close similarities to those in *N. tomentosiformis* have completely overwritten parental units. This process represents astonishingly rapid homogenization that has affected thousands of genes in just a few generations (Skalicka et al. 2003). In natural allopolyploids, rDNA units show varying degrees of rDNA homogenization depending on the allopolyploid species studied. *Nicotiana arentsii* (formed ~0.4 Mya) shows complete homogenization toward an rDNA unit that is similar to that observed in the *N. undulata* parental genome, this process having thus replaced all the units derived from the *N. wigandioides* parent. In contrast, in the older polyploid *N. rustica* (~0.7 Mya), we see the least change in the rDNA unit structure, with the different accessions analyzed showing only different ratios of parental rDNA unit types (Dadejová et al. 2007). In *N. tabacum*, there is large-scale, but incomplete, replacement of *N. sylvestris*-type units by units that are most similar to the *N. tomentosiformis* parent (Lim et al. 2000a). More recently, an analysis of rDNA diversity in *N. tabacum* using high-throughput sequence read data revealed that sequences in the 35S rDNA genic component of the rDNA unit are most homogenous, whereas those of the intergenic spacer are more diverse, indicating a more complete homogenization of the functional domains, or less subsequent divergence in these regions following homogenization (Lunerová et al. 2017).

There is no evidence to suggest that nuclear–cytoplasmic interactions affect the direction of rDNA unit homogenization, because the direction of replacement/bias is not always toward the maternal progenitor (Volkov et al. 1999; Lim et al. 2000a). This suggests that these processes involve different mechanisms to those that affect the loss of other high-copy repeats, as reported by Renny-Byfield et al. (2011) for *N. tabacum*. Thus, as described previously in nucleolar dominance studies, there is no indication that the parental direction of the cross has any effect on homogenization processes (Chen and Pikaard 1997). Instead, based on an analysis of rRNA transcripts in *N. tabacum* and *N. rustica*, it appears that the rDNA units that are most actively transcribed are also those that were amplified (Lunerová et al. 2017). Kovarik et al. (2008) proposed that epigenetic changes triggered by polyploid formation establish patterns of nuclear dominance that affect the expression of rDNA units, rendering the active units vulnerable to homogenization and unit conversion. Conversely, silent units, which are mostly condensed at interphase, may be less vulnerable to recombination processes, that may occur in the nucleolus and are most likely to mutate and be lost. In this way, variant types that are active are prone to amplify across the rDNA loci, first across intralocus tandem arrays, then less frequently between arrays (Kovarik et al. 2008; Lunerová et al. 2017).

In older polyploids in sections *Polydicliae* and *Repandae*, which formed approximately 1 and 4 Mya, respectively, the numbers of rDNA loci have dropped to two or three, as found in the progenitor diploids (Clarkson et al. 2005). In section *Polydicliae*, subtelomeric satellite DNAs from both parents can be found on the same chromosomes, a characteristic that is not observed in younger polyploids, which tend to maintain subtelomeric satellites intact (Kenton et al. 1993; Lim et al. 2004a, 2005; Matyasek et al. 2011). The distribution in section *Polydicliae* could only have arisen if recombination-based processes occurred between both parental subgenomes (Koukalova et al. 2010). In the even older polyploid section *Repandae* (~4 Mya), three species (*N. nesophila*, *N. stocktonii*, and *N. repanda*) have a clade-specific satellite (NNES10) that has replaced other existing satellites found in the progenitor diploids, revealing the emergence, amplification, and homogenization of sequences over time frames of a few million years (Koukalova et al. 2010; Dodsworth et al. 2017).

Genomic in situ hybridization used to identify the parental origin of the chromosomes of polyploid species in section *Repandae* largely failed in early experiments (Lim et al. 2007). This was presumed to be due to large-scale replacement of the high copy repeats across the genome. Such genome turnover, in under 5 million years, is consistent with the half-life reported for repeats in rice, of ~6 million years (Ma et al. 2004). Modifications to genomic in situ hybridization protocols have enabled the discrimination of parental subgenomes in all *Repandae* species (Dodsworth et al. 2017), albeit with differences between taxa. These experiments probably localize the lower copy repeat fractions more effectively and, in doing so, reveal intergenomic translocations (Dodsworth et al. 2017). As noted in Sect. 7.3, three of the species in section *Repandae* (*N. nesophila, N. stocktonii*, and *N. repanda*) show genome upsizing, largely arising from the expansion of some high-copy repeats, whereas a fourth species (*N. nudicaulis*) shows genome downsizing and is sister to the rest of the species in this section (Renny-Byfield et al. 2013; Dodsworth et al. 2017). The divergence of section *Repandae* is thus associated with a changing profile of repeats (Dodsworth et al. 2017) that reflects the evolutionary history of this and other sections of *Nicotiana* (Koukalova et al. 2010).

7.6 Lag-Phase Hypothesis and *Nicotiana*

Much has been written about the advantages of polyploidy, including, for example, fixed heterozygosity, the release from selection of redundant gene duplicates that can then take up new functions (subfunctionalization/neofunctionalization), and the mix-and-match of

Table 7.2 A hypothesis for genome evolution in polyploids, with reference, where possible, to *Nicotiana* species

Early polyploidy (first few 10–100s of generations)	
Chromosome pairing and meiotic instability	• Low fertility in many synthetic polyploids, as seen in Burk's synthetic *N. tabacum*
	• Strong selection pressure to establish regular meiosis expected, as seen in Burk's synthetic *N. tabacum*
	• Fixed heterozygosity of multiple alleles on different homologs leading to heterosis, which will be maintained with disomic inheritance of homologs
"Genomic shock" causing altered epigenetic patterns and triggering mobility of (retro)transposons	• Early-generation synthetic *Nicotiana* polyploids can show novel patterns of DNA methylation and deletions and insertions of rDNA loci and (retro)transposable elements
	• Genetic and epigenetic heterogeneity in the offspring creates a great deal of genetic variation upon which selection can act
Diploidization (occurring over a few million years)	
Genome downsizing	• A general trend following polyploidy and predicted to have occurred in some but not all *Nicotiana* polyploids. It is thought to be an evolutionary response to reduce the nutrient costs associated with additional DNA arising from polyploidy, and/or help overcome disadvantages associated with larger cells and/or cell cycle times
	• Preferential loss of DNA sequences, particularly high-copy repeats, from one of the progenitor genomes (e.g., loss of repetitive DNA, especially from the *N. tomentosiformis*-derived genome in *N. tabacum*)
	• Loss of low- and middle-copy repeat sequences reflecting diploidization and downsizing
Genome upsizing	• Amplification of high-copy repeats that can lead to genome turnover
rDNA locus number reduction	• rDNA locus number drops to a diploid-like number and is fixed for species-specific alleles within a few million years
Neofunctionalization/subfunctionalization and the evolution of new biochemistry/phenology	• Seen in the evolution of nicotine, a potent insecticide, in the genus *Nicotiana*
Lag phase and species radiation (5–10 million years)	
Rapid diversification	• Observed to have occurred in section *Suaveolentes* over the last 2–3 Mya, potentially triggered through novel traits that have advantages in multiple ecological environments
Dysploidy and chromosomal rearrangements	• Chromosomal reduction processes that lead to lower chromosome numbers, as occurred in section *Suaveolentes* from 24 chromosomes (tetraploid) down to 14 chromosomes

induced redundancy in biochemical pathways allows for the generation of new biochemistry (Soltis and Soltis 2000; Wendel 2015). In the polyploid events ancestral to Solanaceae and *Nicotiana*, there were duplications of the polyamine and nicotinamide adenine dinucleotide pathways that led to gene redundancy, reducing selection pressure on these pathways, and the evolution of a new alkaloid, nicotine, which characterizes the genus (Xu et al. 2017).

Despite the advantages of polyploidy, there may also be both severe and subtle disadvantages. For example, in the early stages of polyploid formation, there is a potential for reduced

fitness, perhaps associated with nuclear–cytoplasmic interactions and the aberrant segregation of chromosomes at meiosis, the latter seen in the synthetic tetraploid *N. tabacum* (Burk, 1973; Leitch et al. 2006) and in other well-studied tetraploids such as *Arabidopsis arenosa* (Yant et al. 2013). It has also been proposed that the extra burden of DNA associated with polyploidy may lead to resource limitation, particularly nitrogen and phosphorus, because nucleic acids are particularly demanding for these elements and many soils contain only small amounts of these nutrients (Šmarda et al. 2013; Guignard et al. 2016). Furthermore, increased GS is associated with larger cell sizes, longer cell cycle times, and negative impacts on photosynthesis (Greilhuber and Leitch 2013), all of which may act as a selection pressure against larger GSs, potentially leading to genome downsizing.

Given the high occurrence of polyploidy in the ancestry of most angiosperm lineages at a variety of nested phylogenetic levels (Wood et al. 2009), it has been proposed that the advantages likely outweighed disadvantages, especially from a long-term evolutionary viewpoint. Nonetheless, in the short term, polyploids are likely to show higher extinction rates (Mayrose et al. 2011), and, possibly, higher diversification rates than diploids (Soltis et al. 2014; Mayrose et al. 2015). Further analyses of diversification rates among angiosperms suggest nested shifts in diversification that do not correlate perfectly with predicted paleopolyploidy events, instead increased diversification rates appear to follow a lag phase (Tank et al. 2015; Landis et al. 2018). The lag-phase hypothesis (Schranz et al. 2012) posits that over several millions or tens of millions of years, polyploid species can undergo rapid radiations and speciation depending on subsequent evolutionary events. These events may include the diploidization process itself (e.g., genome downsizing and gene neo- and sub-functionalization) that occurs subsequent to polyploidy (Dodsworth et al. 2016). Potentially, such diploidization processes also facilitate adaptation to new environments and the radiation of diploidized taxa into new niches. This phenomenon is certainly occurring in section *Suaveolentes*, where, despite the origin of the section being ~6 Mya (Clarkson et al. 2017), most species have evolved more recently (2–3 Mya; Chase et al. 2018), concomitant with moves into new habitats across the arid zones of Australia. A summary of the events that are likely to have occurred in that lag phase from a genomic viewpoint is given in Table 7.2, including an outline of the genomic processes affecting repeat dynamics over shorter timescales in polyploid *Nicotiana*.

References

Bejarano ER, Khashoggi A, Witty M, Lichtenstein C (1996) Integration of multiple repeats of geminiviral DNA into the nuclear genome of tobacco during evolution. Proc Natl Acad Sci 93(2):759–764

Bombarely A, Rosli HG, Vrebalov J, Moffett P, Mueller LA, Martin GB (2012) A draft genome sequence of *Nicotiana benthamiana* to enhance molecular plant-microbe biology research. Mol Plant Microbe Interact 25(12):1523–1530

Borisjuk NV, Davidjuk YM, Kostishin SS, Miroshnichenco GP, Velasco R, Hemleben V, Volkov RA (1997) Structural analysis of rDNA in the genus *Nicotiana*. Plant Mol Biol 35(5):655–660

Burk LG (1973) Partial self fertility in a theoretical amphiploid progenitor of *N. tabacum*. J Hered 64 (6):348–350

Chase MW, Christenhusz MJ, Conran JG, Dodsworth S, Medeiros de Assis FN, Felix LP, Fay MF (2018) Unexpected diversity of Australian tobacco species (*Nicotiana* section Suaveolentes, Solanaceae) Curtis's Bot Mag 35(3):212–227

Chen ZJ, Pikaard CS (1997) Transcriptional analysis of nucleolar dominance in polyploid plants: biased expression/silencing of progenitor rRNA genes is developmentally regulated in *Brassica*. Proc Natl Acad Sci 94(7):3442–3447

Clarkson JJ, Dodsworth S, Chase MW (2017) Time-calibrated phylogenetic trees establish a lag between polyploidisation and diversification in *Nicotiana* (Solanaceae). Plant Syst Evol 303:1001–1012

Clarkson JJ, Kelly LJ, Leitch AR, Knapp S, Chase MW (2010) Nuclear glutamine synthetase evolution in *Nicotiana*: phylogenetics and the origins of allotetraploid and homoploid (diploid) hybrids. Mol Phylogenet Evol 55(1):99–112

Clarkson JJ, Knapp S, Garcia VF, Olmstead RG, Leitch AR, Chase MW (2004) Phylogenetic relationships in *Nicotiana* (Solanaceae) inferred from multiple plastid DNA regions. Mol Phylogenet Evol 33(1): 75–90

Clarkson JJ, Lim KY, Kovarik A, Chase MW, Knapp S, Leitch AR (2005) Long-term genome diploidization in

allopolyploid *Nicotiana* section *Repandae* (Solanaceae). New Phytol 168(1):241–252

Dadejová M, Lim KY, Soucková-Skalická K, Matyášek R, Grandbastien M-A, Leitch A, Kovařík A (2007) Transcription activity of rRNA genes correlates with a tendency towards intergenomic homogenization in *Nicotiana* allotetraploids. New Phytol 174(3):658–668

Dodsworth S, Chase MW, Leitch AR (2016) Is post-polyploidization diploidization the key to the evolutionary success of angiosperms? Bot J Linn Soc 180 (1):1–5

Dodsworth S, Jang T-S, Struebig M, Chase MW, Weiss-Schneeweiss H, Leitch AR (2017) Genome-wide repeat dynamics reflect phylogenetic distance in closely related allotetraploid *Nicotiana* (Solanaceae). Plant Syst Evol 303(8):1013–1020

Dvořáčková M, Fojtová M, Fajkus J (2015) Chromatin dynamics of plant telomeres and ribosomal genes. Plant J 83(1):18–37

El Baidouri M, Panaud O (2013) Comparative genomic paleontology across plant kingdom reveals the dynamics of TE-driven genome evolution. Genome Biol Evol 5(5):954–965

Fajkus J, Kovařík A, mKrálovics R, Bezděk M (1995a) Organization of telomeric and subtelomeric chromatin in the higher plant *Nicotiana tabacum*. Mol Gen Genet MGG 247(5):633–638

Fajkus J, Královics R, Kovařík A, Fajkusová L (1995b) The telomeric sequence is directly attached to the HRS60 subtelomeric tandem repeat in tobacco chromosomes. FEBS Lett 364(1):33–35

Fulneček J, Lim KY, Leitch AR, Kovařík A, Matyášek R (2002) Evolution and structure of 5S rDNA loci in allotetraploid *Nicotiana tabacum* and its putative parental species. Heredity 88:19–25

Gazdová B, Široký J, Fajkus J, Brzobohatý B, Kenton A, Parokonny A, Heslop-Harrison JS, Palme K, Bezděk M (1995) Characterization of a new family of tobacco highly repetitive DNA, GRS, specific for the *Nicotiana tomentosiformis* genomic component. Chromosom Res 3(4):245–254

Gill BS, Friebe B (2013) Nucleocytoplasmic interaction hypothesis of genome evolution and speciation in polyploid plants revisited: polyploid species-specific chromosomal polymorphisms in wheat. In: Chen AJ, Birchler JA (eds) Polyploid and hybrid genomics, pp 213–221

Gregor W, Mette MF, Staginnus C, Matzke MA, Matzke AJM (2004) A distinct endogenous pararetrovirus family in *Nicotiana tomentosiformis*, a diploid progenitor of polyploid tobacco. Plant Physiol 134 (3):1191–1199

Greilhuber J, Leitch IJ (2013) Genome size and the phenotype. In: Leitch IJ, Greilhuber J, Doležel J, Wendel JF (eds) Plant genome diversity, vol 2, Physical structure, behaviour and evolution of plant genomes. Springer, Wien, pp 323–344

Guignard MS, Nichols RA, Knell RJ, Macdonald A, Romila C-A, Trimmer M, Leitch IJ, Leitch AR (2016) Genome size and ploidy influence angiosperm species' biomass under nitrogen and phosphorus limitation. New Phytol 210:1195–1206

Horakova M, Fajkus J (2000) TAS49-a dispersed repetitive sequence isolated from subtelomeric regions of *Nicotiana tomentosiformis* chromosomes. Genome 43 (2):273–284

Ibarra-Laclette E, Lyons E, Hernandez-Guzman G, Perez-Torres CA, Carretero-Paulet L, Chang T-H, Lan T, Welch AJ, Juarez MJA, Simpson J et al (2013) Architecture and evolution of a minute plant genome. Nature 498(7452):94–98

Kejnovsky E, Hawkins JS, Feschotte C (2012) Plant transposable elements: biology and evolution. In: Wendel JF, Greilhuber J, Doležel J, Leitch IJ (eds) Plant genome diversity, vol 1, Plant genomes, their residents, and their evolutionary dynamics. Springer-Verlag, Wien, pp 17–34

Kelly LJ, Leitch AR, Clarkson JJ, Hunter RB, Knapp S, Chase MW (2010) Intragenic recombination events and evidence for hybrid speciation in *Nicotiana* (Solanaceae). Mol Biol Evol 27(4):781–799

Kenton A, Parokonny AS, Gleba YY, Bennett MD (1993) Characterization of the *Nicotiana tabacum* L. genome by molecular cytogenetics. Mol Gen Genet 240 (2):159–169

Knapp S, Chase MW, Clarkson JJ (2004) Nomenclatural changes and a new sectional classification in *Nicotiana* (Solanaceae). Taxon 53(1):73–82

Koukalova B, Moraes AP, Renny-Byfield S, Matyasek R, Leitch AR, Kovarik A (2010) Fall and rise of satellite repeats in allopolyploids of *Nicotiana* over c. 5 million years. New Phytol 186(1):148–160

Koukalova B, Reich J, Bezdek M (1990) A BamHI family of tobacco highly repeated DNA—a study about its species-specificity. Biol Plant 32(6):445–449

Kovarik A, Dadejova M, Lim YK, Chase MW, Clarkson JJ, Knapp S, Leitch AR (2008) Evolution of rDNA in *Nicotiana* allopolyploids: a potential link between rDNA homogenization and epigenetics. Ann Bot 101(6):815–823

Kovarik A, Koukalova B, Lim KY, Matyasek R, Lichtenstein CP, Leitch AR, Bezdek M (2000) Comparative analysis of DNA methylation in tobacco heterochromatic sequences. Chromosom Res 8(6):527–541

Kovarik A, Matyasek R, Lim KY, Skalicka K, Koukalova B, Knapp S, Chase M, Leitch AR (2004) Concerted evolution of 18-5.8-26S rDNA repeats in *Nicotiana* allotetraploids. Biol J Linn Soc 82(4):615–625

Kuhrova V, Bezdek M, Vyskot B, Koukalova B, Fajkus J (1991) Isolation and characterization of two middle repetitive DNA sequences of nuclear tobacco genome. Theor Appl Genet 81:740–744

Landis JB, Soltis DE, Li Z, Marx HE, Barker MS, Tank DC, Soltis PS (2018) Impact of whole-genome duplication events on diversification rates in angiosperms. Am J Bot 105(3):348–363

Leitch AR, Leitch IJ (2012) Ecological and genetic factors linked to contrasting genome dynamics in seed plants. New Phytol 194(3):629–646

Leitch AR, Lim KY, Skalicka K, Kovarik A (2006) Nuclear cytoplasmic interaction hypothesis and the role of translocations in *Nicotiana* allopolyploids. NATO Security through Science Series 319–326

Leitch IJ, Bennett MD (2004) Genome downsizing in polyploid plants. Biol J Lin Soc 82:651–663

Leitch IJ, Hanson L, Lim KY, Kovarik A, Chase MW, Clarkson JJ, Leitch AR (2008) The ups and downs of genome size evolution in polyploid species of *Nicotiana* (Solanaceae). Ann Bot 101(6):805–814

Li W, Zhang P, Fellers JP, Friebe B, Gill BS (2004) Sequence composition, organization, and evolution of the core Triticeae genome. Plant J 40(4):500–511

Lim KY, Kovarik A, Matyasek R, Bezdek M, Lichtenstein CP, Leitch AR (2000a) Gene conversion of ribosomal DNA in *Nicotiana tabacum* is associated with undermethylated, decondensed and probably active gene units. Chromosoma 109(3):161–172

Lim KY, Kovarik A, Matyasek R, Chase MW, Clarkson JJ, Grandbastien MA, Leitch AR (2007) Sequence of events leading to near-complete genome turnover in allopolyploid *Nicotiana* within five million years. New Phytol 175(4):756–763

Lim KY, Kovarik A, Matyasek R, Chase MW, Knapp S, McCarthy E, Clarkson JJ, Leitch AR (2006) Comparative genomics and repetitive sequence divergence in the species of diploid *Nicotiana* section *Alatae*. Plant J 48(6):907–919

Lim KY, Matyasek R, Kovarik A, Fulnecek J, Leitch AR (2005) Molecular cytogenetics and tandem repeat sequence evolution in the allopolyploid *Nicotiana rustica* compared with diploid progenitors *N. paniculata* and *N. undulata*. Cytogenet Genome Res 109(1–3):298–309

Lim KY, Matyasek R, Kovarik A, Leitch AR (2004a) Genome evolution in allotetraploid *Nicotiana*. Biol J Lin Soc 82(4):599–606

Lim KY, Matyasek R, Lichtenstein CP, Leitch AR (2000b) Molecular cytogenetic analyses and phylogenetic studies in the *Nicotiana* section *Tomentosae*. Chromosoma 109(4):245–258

Lim KY, Skalicka K, Koukalova B, Volkov RA, Matyasek R, Hemleben V, Leitch AR, Kovarik A (2004b) Dynamic changes in the distribution of a satellite homologous to intergenic 26-18S rDNA spacer in the evolution of *Nicotiana*. Genetics 166 (4):1935–1946

Lunerová J, Renny-Byfield S, Matyášek R, Leitch AR, Kovařík A (2017) Concerted evolution rapidly eliminates sequence variation in rDNA coding regions but not in intergenic spacers in *Nicotiana tabacum* allotetraploid. Plant Syst Evol 303(8):1043–1060

Ma JX, Devos KM, Bennetzen JL (2004) Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice. Genome Res 14(5):860–869

Matyasek R, Fulnecek J, Leitch AR, Kovarik A (2011) Analysis of two abundant, highly related satellites in the allotetraploid *Nicotiana arentsii* using double-strand conformation polymorphism analysis and sequencing. New Phytol 192(3):747–759

Matyasek R, Fulnecek J, Y Lim K, Leitch A, Kovarik A (2002) Evolution of 5S rDNA unit arrays in the plant genus *Nicotiana* (Solanaceae)

Matyasek R, Gazdová B, Fajkus J, Bezdek M (1997) NTRS, a new family of highly repetitive DNAs specific for the T1 chromosome of tobacco. Chromosoma 106:369–379

Matzke M, Gregor W, Mette MF, Aufsatz W, Kanno T, Jakowitsch J, Matzke AJM (2004) Endogenous pararetroviruses of allotetraploid *Nicotiana tabacum* and its diploid progenitors, *N. sylvestris* and *N. tomentosiformis*. Biol J Linn Soc 82(4):627–638

Mayrose I, Zhan SH, Rothfels CJ, Arrigo N, Barker MS, Rieseberg LH, Otto SP (2015) Methods for studying polyploid diversification and the dead end hypothesis: a reply to Soltis et al. (2014). New Phytol 206(1):27–35

Mayrose I, Zhan SH, Rothfels CJ, Magnuson-Ford K, Barker MS, Rieseberg LH, Otto SP (2011) Recently formed polyploid plants diversify at lower rates. Science 333(6047):1257

McClintock B (1984) The significance of responses of the genome to challenge. Science 226(4676):792–801

Melayah D, Lim KY, Bonnivard B, Chalhoub B, Dorlac de Borne F, Mhiri C, Leitch AR, Grandbastien MA (2004) Distribution of the Tnt1 retrotransposon family in the amphidiploid tobacco (*Nicotiana tabacum*) and its wild *Nicotiana* relatives. Biol J Linn Soc 82:639–649

Mhiri C, Morel J-B, Vernhettes S, Casacuberta JM, Lucas H, Grandbastien M-A (1997) The promoter of the tobacco Tnt1 retrotransposon is induced by wounding and by abiotic stress. Plant Mol Biol 33 (2):257–266

Mhiri C, Parisod C, Daniel J, Petit M, Lim KY, Dorlhac de Borne F, Kovarik A, Leitch AR, Grandbastien M-A (2019) Parental transposable element loads influence their dynamics in young *Nicotiana* hybrids and allotetraploids. New Phytol 221(3):1619–1633

Moscone EA, Matzke MA, Matzke AJ (1996) The use of combined FISH/GISH in conjunction with DAPI counterstaining to identify chromosomes containing transgene inserts in amphidiploid tobacco. Chromosoma 105(5):231–236

Murad L, Bielawski JP, Matyasek R, Kovarik A, Nichols RA, Leitch AR, Lichtenstein CP (2004) The origin and evolution of geminivirus-related DNA sequences in *Nicotiana*. Heredity 92(4):352

Murad L, Lim KY, Christopodulou V, Matyasek R, Lichtenstein CP, Kovarik A, Leitch AR (2002) The origin of tobacco's T genome is traced to a particular lineage within *Nicotiana tomentosiformis* (Solanaceae). Am J Bot 89(6):921–928

Nagaki K, Shibata F, Suzuki G, Kanatani A, Ozaki S, Hironaka A, Kashihara K, Murata M (2011) Coexistence of NtCENH3 and two retrotransposons in tobacco centromeres. Chromosom Res 19(5):591–605

Neumann P, Novák P, Hoštáková N, Macas J (2019) Systematic survey of plant LTR-retrotransposons elucidates phylogenetic relationships of their polyprotein domains and provides a reference for element classification. Mobile DNA 10(1):1

Novak P, Neumann P, Macas J (2010) Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. BMC Bioinform 11(1):378

Novák P, Neumann P, Pech J, Steinhaisl J, Macas J (2013) RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. Bioinformatics 29(6):792–793

Pellicer J, Fay MF, Leitch IJ (2010) The largest eukaryotic genome of them all? Bot J Linn Soc 164(1):10–15

Pellicer J, Hidalgo O, Dodsworth S, Leitch IJ (2018) Genome size diversity and its impact on the evolution of land plants. Genes 9(2):88

Petit M, Guidat C, Daniel J, Denis E, Montoriol E, Bui QT, Lim KY, Kovarik A, Leitch AR, Grandbastien MA et al (2010) Mobilization of retrotransposons in synthetic allotetraploid tobacco. New Phytol 186(1):135–147

Petit M, Lim K, Julio E, Poncet C, de Borne Dorlhac, Fo Kovarik A, Leitch AR, Grandbastien M-Al, Mhiri C (2007) Differential impact of retrotransposon populations on the genome of allotetraploid tobacco (*Nicotiana tabacum*). Mol Genet Genomics 278(1):1–15

Renny-Byfield S, Chester M, Kovařík A, Le Comber SC, Grandbastien M-A, Deloger M, Nichols RA, Macas J, Novák P, Chase MW et al (2011) Next generation sequencing reveals genome downsizing in allotetraploid *Nicotiana tabacum*, predominantly through the elimination of paternally derived repetitive DNAs. Mol Biol Evol 28(10):2843–2854

Renny-Byfield S, Kovarik A, Chester M, Nichols RA, Macas J, Novak P, Leitch AR (2012) Independent, rapid and targeted loss of highly repetitive DNA in natural and synthetic allopolyploids of *Nicotiana tabacum*. PLoS ONE 7(5):e36963

Renny-Byfield S, Kovarik A, Kelly LJ, Macas J, Novak P, Chase MW, Nichols RA, Pancholi MR, Grandbastien MA, Leitch AR (2013) Diploidization and genome size change in allopolyploids is associated with differential dynamics of low- and high-copy sequences. Plant J 74(5):829–839

Särkinen T, Bohs L, Olmstead RG, Knapp S (2013) A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): a dated 1000-tip tree. BMC Evol Biol 13(1):214

Schiavinato M, Marcet-Houben M, Dohm JC, Gabaldón T, Himmelbauer H (2020) Parental origin of the allotetraploid tobacco. Plant J

Schranz EM, Mohammadin S, Edger PP (2012) Ancient whole genome duplications, novelty and diversification: the WGD radiation lag-time model. Curr Opin Plant Biol 15(2):147–153

Shibata F, Nagaki K, Yokota E, Murata M (2013) Tobacco karyotyping by accurate centromere identification and novel repetitive DNA localization. Chromosom Res 21(4):375–381

Skalicka K, Lim KY, Matyasek R, Koukalova B, Leitch AR, Kovarik A (2003) Rapid evolution of parental rDNA in a synthetic tobacco allotetraploid line. Am J Bot 90(7):988–996

Skalicka K, Lim KY, Matyasek R, Matzke M, Leitch AR, Kovarik A (2005) Preferential elimination of repeated DNA sequences from the paternal, *Nicotiana tomentosiformis* genome donor of a synthetic, allotetraploid tobacco. New Phytol 166(1):291–303

Šmarda P, Hejcman M, Březinová A, Horová L, Steigerová H, Zedek F, Bureš P, Hejcmanová P, Schellberg J (2013) Effect of phosphorus availability on the selection of species with different ploidy levels and genome sizes in a long-term grassland fertilization experiment. New Phytol 200:911–921

Soltis DE, Segovia-Salcedo MC, Jordon-Thaden I, Majure L, Miles NM, Mavrodiev EV, Mei W, Cortez MB, Soltis PS, Gitzendanner MA (2014) Are polyploids really evolutionary dead-ends (again)? A critical reappraisal of Mayrose et al. (2011). New Phytol 202(4):1105–1117

Soltis PS, Soltis DE (2000) The role of genetic and genomic attributes in the success of polyploids. Proc Natl Acad Sci (USA) 97(13):7051–7057

Tank DC, Eastman JM, Pennell MW, Soltis PS, Soltis DE, Hinchliff CE, Brown JW, Sessa EB, Harmon LJ (2015) Nested radiations and the pulse of angiosperm diversification: increased diversification rates often follow whole genome duplications. New Phytol 207:454–467

Volkov RA, Borisjuk NV, Panchuk II, Schweizer D, Hemleben V (1999) Elimination and rearrangement of parental rDNA in the allotetraploid *Nicotiana tabacum*. Mol Biol Evol 16(3):311–320

Wendel JF (2015) The wondrous cycles of polyploidy in plants. Am J Bot 102(11):1753–1756

Wood TE, Takebayashi N, Barker MS, Mayrose I, Greenspoon PB, Rieseberg LH (2009) The frequency of polyploid speciation in vascular plants. Proc Natl Acad Sci (USA) 106(33):13875–13879

Xu S, Brockmöller T, Navarro-Quezada A, Kuhl H, Gase K, Ling Z, Zhou W, Kreitzer C, Stanke M, Tang H et al (2017) Wild tobacco genomes reveal the evolution of nicotine biosynthesis. Proc Natl Acad Sci 114(23):6133–6138

Yant L, Hollister Jesse D, Wright Kevin M, Arnold Brian J, Higgins James D, Franklin FChris H, Bomblies K (2013) Meiotic adaptation to genome duplication in *Arabidopsis arenosa*. Curr Biol 23 (21):2151–2156

8 The Use of *Nicotiana* Species in Tobacco Improvement

Apoloniusz Berbeć and Teresa Doroszewska

Abstract

There are more than 80 naturally occurring relatives of cultivated tobacco (*Nicotiana tabacum* L.) in the genus *Nicotiana*. In this review, we examine how and to what extent these natural germplasm resources have been utilized in hybridization and introgression experiments over the past century. To date, more than 400 interspecific *Nicotiana* hybrids have been reported. We focus on individual *Nicotiana* species involved in interspecific hybrids with cultivated tobacco produced by sexual and asexual methods, including the recently discovered grafting method. Problems related to the hybridization of *N. tabacum* with other species, namely, cross-incompatibility, maternal phenotypes in hybrid offspring, interspecific incongruity, lethality of juvenile hybrids, and sterility of viable hybrids, are reviewed. Among the 58 interspecific hybrids involving *N. tabacum* reported thus far, 25 were also reported as somatic hybrids and two were obtained only as somatic hybrids. Thirty-six sterile sexual F_1 hybrids were converted to fertile or partly fertile allopolyploids. Sixteen *Nicotiana* species have been deployed as a source of usable traits that were introgressed into *N. tabacum*, offering resistance to or tolerance of pathogens or pests. The mechanisms of introgression, such as alien addition and substitution, as well as the barriers and limitations of introgression, including erratic inheritance and adverse linkages, are discussed. Thirty-one *Nicotiana* species used as sources of cytoplasmic male sterility in *N. tabacum* have produced multiple alloplasmics; most showed negative effects of alien cytoplasm but a few have been deployed successfully in hybrid cultivars of *N. tabacum*.

8.1 Introduction

Tobacco continues to be the most important non-food crop in the world that provides a widely used nicotine-based stimulant. Cultivated tobacco (*Nicotiana tabacum*) has many close relatives, most of which are known from natural habitats, although some have been cultivated either as economic crops (primarily *N. rustica*) or as ornamentals (e.g., *N. alata* and *N. sylvestris*). To date, depending on classification criteria, more than 80 species have been identified and classified within the genus *Nicotiana*.

The interest in *Nicotiana* hybrids began in the eighteenth century. The major driving motive behind the production of interspecific *Nicotiana* crosses was related to *Nicotiana* phylesis and

A. Berbeć (✉) · T. Doroszewska
Institute of Soil Science and Plant Cultivation, State Research Institute, Pulawy, Poland
e-mail: berbec@iung.pulawy.pl

© Springer Nature Switzerland AG 2020

N. V. Ivanov et al. (eds.), *The Tobacco Plant Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-030-29493-9_8

systematics. Initially, the study of affinities between various *Nicotiana* species drew, to a large extent, on the behavior of meiotic chromosomes and on other aspects of cytogenetics in interspecific hybrids, and until the early 1970s, the vast majority of *Nicotiana* hybrids were probably created with that sole purpose in mind. The interest of biologists in interspecific hybrids waned considerably as new biochemical and molecular methods began to replace cytogenetic tools.

However, wild *Nicotiana* and its hybrids with cultivated tobacco have continued to attract the interest of agronomists and breeders. That interest is primarily focused on the resistance of a large number of these species to common *N. tabacum* diseases. Another important asset of wild *Nicotiana* species is their cytoplasmic genomes, which have provided the source of cytoplasmic male sterility (CMS) for developing male-sterile isolines of inbred lines and cultivars. CMS is a prerequisite for technically feasible and economically viable seed production of hybrid cultivars.

Many excellent reviews and monographic essays have been written on *Nicotiana* and hybridization in the genus. Most of them focus on issues and problems encountered and on solutions to those problems rather than on the species themselves. Far from neglecting those issues and problems, this review of *Nicotiana* literature shifts the perspective to emphasize individual species as objects of research aimed at improving *N. tabacum*.

8.2 The Genus *Nicotiana*

The genus *Nicotiana* is one of the six largest genera in the family Solanaceae (Knapp et al. 2004). Most *Nicotiana* species are native to the New World (South, Central, and North America), although several occur naturally in Australia and in South Pacific islands, and one was found in southern Africa (Namibia) (Table 8.1). The genus includes two cultivated species: *N. tabacum* and *N. rustica*. The former is the major source of tobacco, an important nicotine-based stimulant available in various presentations such as smoking, chewing, and snuffing materials.

Nicotiana systematics began with the eighteenth-century description of four species by Carl Linnaeus. The foundations for the present-day classification of *Nicotiana* were laid by Kostoff (1943) and Goodspeed (1954). Kostoff divided the genus into 8 sections and 47 species known to him. Goodspeed's taxonomical division of the genus, while not differing much in its essentials from Kostoff's, was much larger and included 60 species divided into 3 subgenera and 14 sections, and provided the basic framework for subsequent additions and revisions. Knapp et al. (2004) proposed some significant modifications to Goodspeed's classification. The current number of reported *Nicotiana* species is considerably larger, and known *Nicotiana* species now number 83 (Table 8.1). This is a fairly liberal classification because it includes taxa that are known from collections and natural sites as well as taxa for which only herbarium specimens exist. The status of the latter is also liable to change; for example, *N. ameghinoi* was recently rediscovered in the wild (Knapp 2013). The new classification has been accepted as standard by most researchers of *Nicotiana* (Lewis and Nicholson 2007; Doroszewska et al. 2009; Lewis 2011). Other lists of *Nicotiana* species and varieties are available. For example, there are currently 97 *Nicotiana* species and varieties listed in NCBI's Taxonomy Browser (https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi) and 307 records of *Nicotiana* species and varieties in The Plant List database (http://www.theplantlist.org/tpl1.1/search?q=Nicotiana), Searching the International Plant Name Index (http://www.ipni.org/ipni/plantnamesearchpage.do) under genus with 'Nicotiana' as the keyword found 549 records.

8.3 General Notes on Experimentally Produced Interspecific *Nicotiana* Hybrids

According to Kostoff (1943), the first experimental crossing of two *Nicotiana* species was made by J. G. Kölreuter in 1760. Kölreuter's classical work demonstrated the equivalency of

Table 8.1 Species of the genus *Nicotiana*, including their somatic number, region of origin, systematic status, and hybridization record with *N. tabacum*

Section	Species	2n[a]	Native to[b]	Description provided by	Hybrids with *N. tabacum*[c]
Alatae	*N. alata*	18	SAm	Goodspeed (1954)	Y
	N. azambujae	?	SAm	X[d]	N
	N. bonariensis	18	SAm	Goodspeed (1954)	Y
	N. forgetiana	18	SAm	Goodspeed (1954)	Y
	N. langsdorffii	18	SAm	Goodspeed (1954)	Y
	N. longiflora	20	SAm	Goodspeed (1954)	Y
	N. mutabilis	18 (?)	SAm	X[e]	N
	N. plumbaginifolia	20	SAm	Goodspeed (1954)	Y
	N. sp. 'Rastroensis'	?	SAm	X[f]	N
	N. sanderae	18	SAm	X[g]	Y
Nicotiana	*N. tabacum*	48	SAm	Goodspeed (1954)	x
Noctiflorae	*N. acaulis*	24	SAm	Goodspeed (1954)	N
	N. ameghinoi	?	SAm	Goodspeed (1954)	N
	N. glauca	24	SAm	Goodspeed (1954)	Y
	N. noctiflora	24	SAm	Goodspeed (1954)	Y
	N. paa	?	SAm	Martinez-Croveto (1978)	N
	N. petunioides	24	SAm	Goodspeed (1954)	N
Paniculatae	*N. benavidesii*	24	SAm	Goodspeed (1954)	Y
	N. cordifolia	24	SAm	Goodspeed (1954)	Y
	N. cutleri	?	SAm	D'Arcy (1976)	N
	N. knightiana	24	SAm	Goodspeed (1954)	Y
	N. paniculata	24	SAm	Goodspeed (1954)	Y
	N. raimondii	24	SAm	Goodspeed (1954)	Y
	N. solanifolia	24	SAm	Goodspeed (1954)	Y
Acuminatae	*N. acuminata*	24	SAm	Goodspeed (1954)	Y
	N. attenuata	24	CAm, NAm	Goodspeed (1954)	N
	N. corymbosa	24	SAm	Goodspeed (1954)	N
	N. longibracteata	?	SAm	Goodspeed (1954)	N
	N. linearis	24	SAm	Goodspeed (1954)	N
	N. miersii	24	SAm	Goodspeed (1954)	N
	N. pauciflora	24	SAm	Goodspeed (1954)	Y
	N. spegazzinii	?	SAm	Goodspeed (1954)	N
Polydicliae	*N. clevelandii*	48	Nam	Goodspeed (1954)	Y
	N. quadrivalvis	48	Nam	Goodspeed (1954)	Y

(continued)

Table 8.1 (continued)

Section	Species	2n[a]	Native to[b]	Description provided by	Hybrids with *N. tabacum*[c]
Repandae	*N. nesophila*	48	CAm (I)	Goodspeed (1954)	Y
	N. nudicaulis	48	Cam	Goodspeed (1954)	Y
	N. repanda	48	NAm	Goodspeed (1954)	Y
	N. stocktonii	48	CAm (I)	Goodspeed (1954)	Y
Rusticae	*N. rustica*	48	SAm	Goodspeed (1954)	Y
Suaveolentes	*N. africana*	46	Afr	Merxmueller and Buttler (1975)	Y
	N. amplexicaulis	36	Au	Burbidge (1960)	Y
	N. benthamiana	38	Au	Goodspeed (1954)	Y
	N. burbidgeae	42	Au	Symon (1984)	N
	N. cavicola	40[h]	Su	Burbidge (1960)	Y
	N. debneyi (*N.forsteri*)[i]	48	Au	Goodspeed (1954) and Marks (2010)	Y
	N. eastii	64	SAm	Kostoff (1943)	Y
	N. excelsior	38	Au	Goodspeed (1954)	Y
	N. exigua	32	Au	Goodspeed (1954)	Y
	N. fatuhivensis	?	Au (M)	Marks (2010)	N
	N. faucicola	?	Au	X[j]	N
	N. fragrans	48	Au (NC)	Goodspeed (1954)	Y
	N. goodspeedii	40[k]	Au	Goodspeed (1954)	Y
	N. gossei	36	Au	Goodspeed (1954)	Y
	N. hesperis	42	Au	Burbidge (1960)	N
	N. heterantha	48[l]	Au	Symon and Kennealy (1994)	N
	N. ingulba	32	Au	Goodspeed (1954)	Y
	N. maritima	32[m]	Au	Goodspeed (1954)	Y
	N. megalosiphon		Au	Goodspeed (1954)	Y
	N. monoschizocarpa	48[n]	Au	Horton (1981)	Y
	N. occidentalis	42	Au	Burbidge (1960)	Y
	N. rosulata	20	Au	Goodspeed (1954)	Y
	N. rotundifolia	44[o]	Au	Goodspeed (1954)	Y
	N. simulans	40	Au	Burbidge (1960)	N
	N. stenocarpa	?	Au	Goodspeed (1954)	Y
	N. suaveolens	30	Au	Goodspeed (1954)	Y
	N. truncata	36	Au	Symon (1998)	N
	N. umbratica	46	Au	Burbidge (1960)	Y
	N. velutina	32	Au	Goodspeed (1954)	Y
	N. wuttkei	28, 32[p]	Au	Clarkson and Symon (1991)	Y
	N. sp. ‘Corunna’ (*N. symonii*)	32[q]	Au	X[r]	N

(continued)

Table 8.1 (continued)

Section	Species	2n[a]	Native to[b]	Description provided by	Hybrids with *N. tabacum*[c]
Sylvestres	*N. sylvestris*	24	SAm	Goodspeed (1954)	Y
Tomentosae	*N. kawakamii*	24	SAm	Ohashi (1976)	Y
	N. otophora	24	SAm	Goodspeed (1954)	Y
	N. setchellii	24	SAm	Goodspeed (1954)	Y
	N. tomentosa	24	SAm	Goodspeed (1954)	Y
	N. tomentosiformis	24	SAm	Goodspeed (1954)	Y
Trigonophyllae	*N. obtusifolia*	24	Cam, Nam	Goodspeed (1954)	Y
	N. palmeri	24	NAm	Goodspeed (1954)	Y
Undulatae	*N. arentsii*	48	SAm	Goodspeed (1954)	Y
	N. glutinosa	24	SAm	Goodspeed (1954)	Y
	N. thyrsiflora	24	SAm	Goodspeed (1954)	N
	N. undulata	24	SAm	Goodspeed (1954)	Y
	N. wigandioides	24	SAm	Goodspeed (1954)	N

[a]Somatic number of chromosomes (two haploid genomes) reported in the description referred to in column IV unless indicated otherwise by the reference after the chromosome number
[b]SAm (South and Central America including offshore islands), NAm (North America), CAm (Central America), CAm (I) (Revillagigedo islands southwest of Baja California) Au (Australia), Au (NC) (New Caledonia archipelago, Au (M) (Marquesas archipelago)
[c](Y) the species was hybridized with *N. tabacum;* (N) there is no hybrid with *N. tabacum* on record (Berbeć, unpublished list of interspecific *Nicotiana* hybrids)
[d]No valid description; reported by Smith and Downs (1964), included in the list of *Nicotiana* species by Knapp et al. (2004)
[e]No valid description; reported by Stehman et al. (2002), included in the list of *Nicotiana* species by Knapp et al. (2004), available as an ornamental
[f]Described by Murfett et al. (2005) as closely related to *N. bonariensis* and listed in several other studies on hybrids within the section *Alatae*
[g]No recognized status; in most publications treated as an artificial horticultural hybrid of *N. alata* with *N. forgetiana* (*N. × sanderae*), in some varieties of *N. sanderae* an involvement of *N. langsdorffii* and other 18-chromosome members of the section Alatae is also possible
[h]Originally reported by Burbidge (1960) as 2n = 46, corrected by Williams (1975) to 2n = 40, correction confirmed by Tatemichi (1990)
[i]According to Marks (2010) the name *N. forsteri* takes rightful precedence over *N. debneyi*
[j]Raised from seeds supplied by C. Marks and studied by Dodsworth (2015)
[k]As reported Goodspeed (1945) and confirmed by Tatemichi (1990), Marks et al. (2011) report the number as 2n = 32
[l]Reported by Marks et al. (2011)
[m]Reported by Goodspeed (1954) and confirmed by Tatemichi (1990), Marks et al. (2011) report the number as 2n = 30
[n]Reported by Marks et al. (2011)
[o]Reported by Goodspeed (1954) and confirmed by Tatemichi (1990), Marks et al. (2011) report the number as 2n = 32
[p]Reported by Clarkson and Symon (1991) and Laskowska and Berbeć (2003), respectively
[q]Chromosome number reported by Marks et al. (2011)
[r]The name *N. symonii* proposed by Dodsworth (2015)

maternal and paternal forms by obtaining plants resembling *N. paniculata* after a series of backcrosses of the hybrid *N. paniculata* × *N. rustica* to *N. paniculata*. Other interspecific combinations were then obtained. East (1928) listed 65 interspecific hybrids, not including reciprocals. Kostoff (1943) described 181 interspecific hybrids, and Goodspeed (1945) and Goodspeed and Bradley (1942) described 243. Smith (1968) estimated that more than 300 interspecific

hybrids had been reported in genus *Nicotiana*, and this number has been cited in many publications, including recent ones. References to 445 interspecific hybrid combinations within the genus *Nicotiana* produced until 2019 were identified in the literature, including several that may have never survived beyond the seedling stage (Berbeć, unpublished list of interspecific *Nicotiana* hybrids). The majority of *Nicotiana* species (58) have been hybridized with both *N. tabacum* and at least one other sister *Nicotiana* species (Table 8.1). Some (*N. mutabilis*, *N. petunioides*, *N. attenuata*, *N. corymbosa*, *N. linearis*, *N. burbidgeae*, *N. thyrsiflora*, and *N. wigandioides*) were hybridized with at least one other *Nicotiana* but not with *N. tabacum*. Fourteen species (*N. azambujae*, *N. acaulis*, *N. ameghinoi*, *N. paa*, *N. cutleri*, *N. longibracteata*, *N. spegazzini*, *N. faucicola, N. fatuhivensis, N. heterantha*, *N. monoschizocarpa*, *N. stenocarpa*, *N. truncata*, and *N.* sp. 'Corunna', also termed *N. symonii*) had no hybridization records (Table 8.1, column 6).

8.4 Pre- and Post-fertilization Barriers to Production of Viable *Nicotiana* Hybrid Seeds

8.4.1 Causes of Interspecific Cross-incompatibility in *Nicotiana*

The number of interspecific hybrids produced between various *Nicotiana* species, although impressive, is still only a small fraction of more than 3,000 theoretically possible combinations. Not all hybrids are obtainable with equal ease. The speciation process depends on the development of different blocks preventing plants from diverging taxa from crossing with each other and producing viable offspring. In *Nicotiana*, those blocks have been evolving relatively slowly, unlike gene and chromosome alterations, which have progressed more rapidly (Goodspeed 1945). The causes of cross-incongruity between *Nicotiana* species were discussed in detail by Kostoff (1943) who pointed out three major causes of the failure to obtain viable hybrid seeds: failure of the pollen tube to reach the ovary, failure of the sperm to fuse with the egg, and abortion of the hybrid embryo.

8.4.2 Unilateral Cross-incompatibility

The first cause, failure of the pollen tube to reach the ovary, may be attributed to factors such as differences in length of styles between the maternal and paternal species (Swaminathan and Murthy 1957; Lee et al. 2008). The pollen of one species may be rejected by the pistil of another species through physiological inhibition of pollen tube growth (Kuboyama et al. 1994). There are many records of a hybrid produced successfully one way but not as the reciprocal cross. This came to be known as unilateral incompatibility. Early workers tried to manage the problem by cutting the style of the maternal parent to a length of 2–3 cm, placing a drop of sucrose agar on the cut surface, and dusting the pollen of the male parent on the cut surface (Swaminathan and Murthy 1957). Ternovsky et al. (1976) applied in vitro fertilization of cultured ovules as a more efficient method to overcome the incompatibility of the crosses *N. tabacum* × *N. rosulata* and *N. tabacum* × *N. debneyi*. Several other hybrids were successfully produced using that approach.

8.4.3 Inviability of Hybrid Embryos

Interspecific matings in *Nicotiana* can result in the actual fusion of the gametes, but the resulting embryo aborts after fertilization or develops poorly, resulting in unviable, non-germinating seeds. Such embryos can be rescued, and hybrid plants can be regenerated by pollinating field- or greenhouse-grown plants (in situ) and culturing the excised fertilized ovules in vitro. Alternatively, in vitro cultured ovules can be pollinated. A number of hybrids have been obtained using those two methods (Reed and Collins 1978; Larkina 1980; Subhashini et al. 1986; DeVerna et al. 1987; Nikova et al. 1998b, 2006).

8.4.4 Maternal and Paternal Plants in the Progeny of Interspecific Crosses

Both gynogenetic and androgenetic haploids were reported in the offspring of interspecific *Nicotiana* crosses (Clausen and Mann 1924; Kostoff 1943; Kehr 1951). Kumashiro and Oinuma (1985) used irradiated pollen from *N. alata* to produce gynogenetic haploids and doubled haploids from *N. tabacum* and compared them with analogous doubled haploids from another culture.

One of these crosses, *N. tabacum* × *N. africana*, produced maternal haploids regularly, and this approach became a method to generate haploid plants in *N. tabacum* for experimental and breeding purposes (Burk et al. 1979; Nielsen and Collins 1989). Haploids found in the progenies of *N. tabacum* × *N. africana* crosses, however, were not all necessarily purely gynogenetic. At least some may have arisen as products of selective elimination of *N. africana* chromosomes from the original hybrid embryos (Chimoyo and Pupert 1988; Hancock et al. 2015).

The divisions of the sperm of *N. tabacum* in the nucleus of *N. eastii* followed by chromosome doubling in the developing embryo may have accounted for the appearance of a cytoplasmic male-sterile (CMS) *N. tabacum* plant as the result of crossing maternal *N. eastii* with *N. tabacum* as the pollen parent (Berbeć and Berbeć 1992).

Maternal diploids in the offspring of interspecific matings in *Nicotiana* have long been a controversial topic. Goodspeed (1915) dismissed diploid maternals (apomictic or otherwise) as products of experimental errors in disagreement with East (1930) but later changed his view (Goodspeed 1954). Alleged apomictics continued to be reported either as induced by pollination with X-ray-irradiated pollen (Pandey and Phung 1982; Kumashiro and Oinuma 1985) or as spontaneous byproducts or even as sole products of interspecific matings (Berbeć and Doroszewska 1981; Sarychev 1986; Murthy and Subbarao 2004; Laskowska et al. 2015). According to Naumenko (2012), diploid maternals described as 'pseudogamic' and resulting from fertilizing intraspecific F_1 hybrids of *N. tabacum* with pollen from *N. alata* could be generated regularly and were not much different from the regular selfed progeny of those plants, except that the former displayed improved vegetative vigor. Liao et al. (2017) recovered self-fertile maternal phenotypes along with apparently regular interspecific hybrids by fertilizing senescent flowers of CMS *N. tabacum* with the pollen of *N. alata*. The two accounts, by Naumenko (2012) and Liao et al. (2017), bear intriguing similarities although it is not clear whether Naumenko used normally self-fertile or male-sterile *N. tabacum* hybrids in her study.

8.5 Lethality of Interspecific *Nicotiana* Hybrids

Necrosis, decay, and death of juvenile hybrid *Nicotiana* plants are well-known phenomena. Until the mid-1960s and even much later, investigators and breeders of *Nicotiana* managed the lethality of hybrid seedlings by the simple expedient of pollinating as many plants and flowers and sowing as many seeds as possible, hoping to find rare phenotypes that would survive to maturity. This policy bores fruit (e.g., Clayton et al. 1967), but it was inefficient and time- and labor-intensive, and success was a matter of luck.

Butenko and Luneva (1966) were probably the first to overcome hybrid lethality in *Nicotiana* by resorting to in vitro culturing. With time, different variants of in vitro cultures on solidified media became a standard practice for rescuing dying hybrid seedlings. Although these methods proved successful in the majority of reported cases, the mechanisms underlying hybrid lethality remained obscure.

It was only in the late 1980s that an in-depth investigation into the causes of seedling inviability in interspecific hybrids of *Nicotiana* was undertaken, primarily by Japanese researchers. Tezuka (2012) summed up the results in an exhaustive review of the work that had been carried out by him and by his associates and colleagues.

Five types of hybrid lethality were distinguished by the external symptoms of inviable seedlings:

- Type I: browning of shoot apex and root tip,
- Type II: browning of hypocotyl and roots,
- Type III: yellowing of true leaves,
- Type IV: formation of multiple shoots, and
- Type V: fading of shoot color.

Type II lethality appears to be prevalent in crosses involving *N. tabacum*. It is shown by 19 hybrids of *N. tabacum* with species of section Suaveolentes, regardless of cross direction (Tezuka et al. 2010; Tezuka 2012). Using monosomic analysis, Marubashi and Onosato (2002) established the factor(s) responsible for the death of seedlings in the hybrid *N. tabacum* × *N. suaveolens* to be located on the Q chromosome on the T subgenome of *N. tabacum*. Further studies identified the causative factor of type II lethality in *N. debneyi* that interacts with the factors of the Q chromosome in *N. tabacum* as a single dominant factor designated *Hla1-1* (Tezuka 2012). Type V lethality was found to be related to factors located on both subgenomes (S and T) of *N. tabacum*. Lethality types II and V are also peculiar in that the symptoms are temperature-dependent; that is, they develop at temperatures of 28 °C and below, are completely suppressed at 34–36 °C, and recur when the temperature drops again below 30 °C (Manabe et al. 1989). The temperature-dependent lethality of hybrid seedlings observed in type II is characterized by condensation of chromatin, fragmentation of nuclei, and fragmentation of DNA in the dying cells; this phenomenon is termed apoptosis or programmed cell death. It is one of the mechanisms of reproductive isolation and also occurs in other plant genera (Tezuka and Marubashi 2004). According to Bomblies (2009), apoptotic hybrid lethality is caused by the hyperactivation of plant defense responses and thus may share a common mechanism with the resistance to some viral pathogens (e.g., tobacco mosaic virus [TMV] or tomato spotted wilt virus [TSWV]) that is based on a hypersensitive response.

8.6 Sterility and Fertility of Viable *Nicotiana* Interspecific Hybrids

8.6.1 Causes of Sterility in Amphihaploid Hybrids of *Nicotiana*

The genus *Nicotiana* is unique among plant genera in that its species show high degree of cross-compatibility (more than 400 hybrid combinations have been produced so far), and the vast majority of those hybrids are practically self- and cross-sterile, that is, are not capable of producing seeds if pollinated by their own pollen or by the pollen of another species. In other genera, approximately 75% of successful interspecific hybrids show at least some degree of self-fertility (Stebbins 1950). According to Goodspeed (1954), chromosomal rearrangements and other structural changes accumulated faster in the genus *Nicotiana* than did barriers to hybridization. This led to deranged micro- and macrosporogenesis in hybrids, resulting in chromosomal imbalance and thus inviable or dysfunctional gametes and, consequently, aborted or unviable ovules and pollen. All known interspecific hybrids involving *N. tabacum* (approximately 60) are sterile, even most of the hybrids in which *N. tabacum* was crossed with its direct progenitor species or their relatives; exceptions and reservations are indicated below. Sterility of interspecific hybrids makes them of little use, especially if further generations are contemplated (e.g., for breeding purposes). In some cases, hybrid sterility may be of potential advantage. Ling et al. (2012) and Rice et al. (2013) considered the applicability of the hybrid *N. tabacum* × *N. glauca* for the production of transgene-induced pharmaceuticals because of its high biomass and other properties. They pointed to the sterility of the hybrid as a safeguard against unintended escape of the transgene to the environment, although they noted that the sterility was not complete (Rice et al. 2013).

8.6.2 Conversion of Sterile Amphihaploids to Fertile Amphidiploids or Partly Fertile Sesquidiploids Using Natural Processes in Hybrid Plants

Early *Nicotiana* researchers exploited large-scale pollination of sterile hybrids to overcome their sterility. During the aberrant events of gametogenesis, one of both reductional divisions may fail, resulting in the formation of unreduced gametes that are usually viable (Kostoff 1943; Goodspeed 1954). If by chance a female unreduced gamete fuses with its unreduced male counterpart, a sterile amphihaploid may give rise to a fertile amphidiploid. There are several cases on record in which otherwise-sterile hybrids of various *Nicotiana* species with *N. tabacum* yielded viable offspring through assiduous effort of selfing the amphihaploid plants (Ternovsky 1934). If an amphihaploid was backcrossed to the *N. tabacum* parent, partly fertile sesquidiploids or near-sesquidiploid plants were the typical outcome (Clausen and Goodspeed 1925).

8.6.3 Conversion to Fertility Using External Agents

Conversion to fertility that relies on the vagaries of nature was both time-consuming and highly unreliable. Therefore, researchers sought to identify substances that can act as antimitotic drugs and thus induce the doubling of chromosomes in the cells of an interspecific hybrid, interfering with the formation of the spindle during cell division. Warmke and Blakeslee (1939) experimented with the hybrid *N. tabacum* × *N. glutinosa* and proposed colchicine as the antimitotic drug of choice to be used in restoring fertility to sterile *Nicotiana* hybrids. Colchicine is used in various formulations; for example, as water solutions of different strengths, water solutions with agar, or mixtures with lanolin. It was applied both in situ, on field- or greenhouse-grown plants, and in vitro in various types of aseptic cultures. Chromosome doubling is induced at various growth stages and in different plant parts, including in cultured embryos, germinating seeds, seedlings, and growing plants (Wark 1970; Ternowsky and Larkina 1978; Subhashini et al. 1986; Berbeć 1988). The dividing meristems to which the drug is applied include apices and ancillary buds. Another antimitotic drug, acenaphthene, was used by some investigators (Kostoff 1943). More recently, oryzalin, an herbicidal substance, has also been used to induce chromosome doubling (Lim et al. 2006).

8.6.4 Synthetic Species

Synthetized amphidiploids in *Nicotiana* are known to differ from one another for the extent of variation they exhibit in successive selfed generations. Some lineages of the new amphidiploid 4n *N. wuttkei* × *N. tabacum* are stable enough to deserve the name of a 'synthetic species' (Laskowska et al. 2015). Such stable, self-perpetuating 'synthetic species' were even given specific names. *N.* × *didebta* for 4n *N. debneyi* × *N. tabacum* (Clayton et al. 1967), *N.* × *digluta* for 4n *N. glutinosa* × *N. tabacum* (Clausen and Goodspeed 1925), and *N.* × *ditagla* for 4n *N. tabacum* × *N. glauca* (Ternovsky 1934) are among 'synthetic' species whose origin involves *N. tabacum*.

8.7 *Nicotiana* Parasexual Interspecific Hybrids

8.7.1 Parasexual Hybrids by Fusion of Isolated Somatic Protoplasts

The idea of circumventing pre-fertilization barriers to crossability at the gametic level prompted attempts to fuse the somatic cells that lacked such obstructions. The first interspecific somatic hybrid in *Nicotiana* appeared to confirm the expectations. *N. langsdorffii* × *N. glauca* was reported by Carlson et al. (1972) as a regular fertile amphidiploid (4× = 42), but it soon

became apparent that the method had serious limitations. The number of interspecific somatic combinations, mostly those involving *N. tabacum*, that yielded genetically stable, fertile amphidiploids was disappointingly small and limited to hybrids of *N. tabacum* with *N. glauca, N. nesophila, N. debneyi*, *N. megalosiphon*, and *N. otophora* (Ilcheva and San 1997; Berbeć and Doroszewska, monograph under preparation). The other hybrids of *N. tabacum* synthetized by protoplast fusion showed high variability in external morphology and mostly aneuploid chromosome numbers because of chromosome elimination, either random or preferential, resulting in asymmetric hybrids, that is, those with a predominance of genetic material from one parental species (e.g., Donaldson et al. 1995).

Asymmetric hybrids have also been synthetized consciously using chemical agents (iodoacetate) or irradiation (gamma or X-rays) to inactivate all or part of the nuclear genome of one parent prior to fusion (e.g., Bates 1990). Highly asymmetric hybrids to which the irradiated parent almost exclusively contributed the cytoplasmic DNA and the other parent donated both cytoplasmic and nuclear DNA (i.e., the so-called cybrids, obtained by the donor–recipient method) proved to be a useful tool in producing interspecific mitochondrial recombinants and provided a fast method of transferring whole plasmons or selected cytoplasmic traits from one species to another (Kumashiro and Kubo 1986; Kumashiro et al. 1989). Cytoplasmic recombinants are practically impossible to obtain by sexual hybridization because in *Nicotiana*, as in most other genera, the cytoplasmic DNA is inherited unilaterally through the maternal lineage, save for some rare exceptions (Medgyesy et al. 1985; Horlow et al. 1990).

8.7.2 Gametosomatic Hybrids

Some mixed approaches have been recorded where protoplasts from immature tetrad microsporocytes or from mature pollen grains were fused with somatic protoplasts. Gametosomatic hybrids of *N. tabacum* (2n) × *N. rustica* (1n) obtained by Pirrie and Power (1986) and by Giddings and Rees (1992) cytogenetically mimicked sesquidiploids (pentaploids) from sexual matings. Other gametosomatic regenerants (*N. tabacum* [1n] × *N. plumbaginifolia* [2n]) created by Desprez et al. (1992) represented an array of sterile aneuploid forms that also showed mixoploidy. According to those investigators, androgenetic somatic regenerants obtained by asymmetric gametosomatic fusion can be used to transfer mitochondrial genomes separately from chloroplast genomes.

A concise and informative review of parasexual hybridization by protoplast fusion in *Nicotiana* was published by Ilcheva and San (1997).

8.7.3 Parasexual Hybrids by Grafting

A singular case of a somatic hybrid was reported by Fuentes et al. (2014). To demonstrate that the horizontal transfer of whole genomes can be a factor in the evolution of allopolyploid plants, the authors grafted the stems of *N. glauca* and *N. tabacum* growing alongside each other, imitating grafting events that occur in nature. After the fusion of stock and scion had taken place, fragments of tissue of the fusion zone were excised and cultured in vitro. Callus culture and plant regeneration were performed according to the genetic complementation protocol based on double selectable markers, just as it is routinely conducted in somatic hybridization by protoplast fusion. A fully fertile, regular 72-chromosome amphidiploid 4n *N. glauca* × *N. tabacum* was obtained and given the name '*Nicotiana tabauca*', a new species that arose by natural fusion of somatic cells.

Fuentes et al. (2014) presented their discovery as a tool for crop improvement mostly because, as they argued, grafting is less technically demanding than protoplast fusion. Although this is true, the sexual method is even less demanding. Indeed, the allopolyploids *N. tabacum* × *N. glauca* and their reciprocals have been obtained repeatedly by conventional crossing for nearly a

century and a name for the 'artificial species' was proposed as '*Nicotiana ditagla*', preceding '*N.* Tabauca' by several decades (see Sect. 8.6.4). The feasibility of obtaining an interspecific hybrid by grafting is reminiscent of the long-forgotten controversy about 'vegetative hybrids', a significant part of the so-called 'revolutionary biology' created in opposition to Mendelian genetics in Soviet Russia in the 1940s. The historical context of the achievement of Ignacia Fuentes and her colleagues was recalled by Zhou and Liu (2015).

8.8 Ending Notes on Sexual and Asexual Interspecific Hybrids Involving *N. tabacum*

A list of all interspecific hybrids involving cultivated tobacco, regardless of the method by which they were obtained, is presented in Table 8.2. Information was found on a total of 58 hybrids involving *N. tabacum*; reciprocals were not included in the count. Thirty-six of those hybrids produced amphidiploids or near-amphidiploids, mostly with at least partially restored self-fertility, one female-sterile amphidiploid (*N. tabacum* × *N. setchellii*), and one nearly female-sterile amphidiploid (*N. obtusifolia* × *N. tabacum*). Female sterility was also observed in the trigenomic allohexaploid 6n (*N. tabacum* × *N. setchellii* × *N. otophora*) (Berbeć et al. 1982). Two allopolyploid combinations were obtained directly by protoplast fusion (*N. rotundifolia* × *N. tabacum* and *N. arentsii* × *N. tabacum*), of which the alloaneuploids *N. rotundifolia* × *N. tabacum* were partly self-fertile (Ilcheva et al. 2001). We found information on 25 *Nicotiana* species that were hybridized with *N. tabacum* by protoplast fusion, fewer than half the number of reported sexual hybrids. Among these 25 somatic hybrids, only a fraction represents those in which whole genomes of both parents became united (Berbeć and Doroszewska, monograph under preparation). To our knowledge, only two somatic hybrids with *N. tabacum* obtained by somatic fusion, *N. rotundifolia* × *N. tabacum* and, possibly, *N. arentsii* *N. tabacum*, have not been obtained by the conventional sexual method. It appears that, although the parasexual approach circumvents the pre-fertilization barriers that separate species, it does little to alleviate the incongruities that exist between the fused genomes (Zenkteler and Melchers 1978).

8.9 Interspecific Gene Transfer and Introgression in *Nicotiana*

8.9.1 Goals and Types of Interspecific Introgression

The germplasm available in wild *Nicotiana* species can potentially be utilized for improving cultivated *N. tabacum* in a twofold manner:

(a) Polygenic introgression from wild species can broaden the germplasm basis of *N. tabacum*, which has tended to become increasingly narrow during the process of domestication and breeding manipulations.
(b) Specific usable traits can be transmitted to and incorporated in the *N. tabacum* germplasm.

8.9.1.1 Polygenic Introgression

Two papers that tackled the problem were concerned with *N. sylvestris* and *N. otophora* as potential sources of polygenic additive variation for improving flue-cured tobacco (Wernsman et al. 1976; Oupadissakon and Wernsman 1977). Some possibilities for yield improvement were found but at the expense of other elements of agronomic performance, such as timely maturity and cured-leaf quality. Recently, Hancock and Lewis (2017) re-examined the closest relatives of *N. tabacum* (*N. sylvestris*, *N. tomentosiformis*, and *N. otophora*) for their potential to improve *N. tabacum*. Synthetic tobaccos 4 × (*N. sylvestris* *N. otophora*) and 4 × (*N. sylvestris* × *N. tomentosiformis*) were found to be preferable vehicles for genetic exchange between tobacco and its wild progenitors to the direct hybrids with each of those species. Heterotic effects were found for yield and growth rate in hybrids of

Table 8.2 Interspecific hybrids of *Nicotiana* species with cultivated tobacco (*Nicotiana tabacum*)

Section	Species hybridized with *N. tabacum*	Sexual F_1 hybrid (amphihaploid) reported by	Amphipolyploid (sexual or somatic) reported by
Alatae	*N. alata*	Pre-Mendelian (East 1928)	Dorossiev et al. (1978)
	N. bonariensis	Busconi et al. (2010)	Busconi et al. (2010) (?)[a]
	N. forgetiana	Ahuja (1962) and Takenaka (1963)	
	N. langsdorffii	East (1928) and Hu (1956)	
	N. longiflora	Malloch and Malloch (1924)	Clayton (1947)
	N. plumbaginifolia	Pal and Nath (1936)	Moav and Cameron (1960)
	N. sanderae	Christoff (1928)	Skucińska et al. (1977) and Malecka (1977)
Noctiflorae	*N. glauca*	Pre-Mendelian (East 1928)	Ternovsky (1934)
	N. noctiflora	Dorossiev et al. (1978)	Dorossiev et al. (1978)
Paniculatae	*N. benavidesii*	Goodspeed (1945)	
	N. cordifolia	Durbin and Uchytil (1977) and Burk and Durbin (1978)	Burk and Durbin (1978)
	N. knightiana	Goodspeed (1945)	Berbeć and Doroszewska (1992)
	N. paniculata	Pre-Mendelian (East 1928)	Nikova et al. (1991) (partly fertile aneuploids)
	N. raimondii	Kostoff (1943)	Burk et al. (1982) and Berbeć (1987)
	N. solanifolia	Goodspeed (1945)	
Acuminatae	*N. acuminata*	Kostoff (1943)	
	N. pauciflora	Gentscheff (1931)	
Polydicliae	*N. clevelandii*	Kaul (1988)	
	N. quadrivalvis	Pre-Mendelian (East 1928)	Burk (1960)
Repandae	*N. nesophila*	Reed and Collins (1978)	Reed and Collins (1978)
	N. nudicaulis	Gentscheff (1931) and Kostoff (1943)	Burk and Neas (1966)
	N. repanda	Kincaid (1949) and Pittarelli and Stavely (1975)	Pittarelli and Stavely (1975)
	N. stocktonii	Wong (1975) and Reed and Collins (1978)	Reed and Collins (1978)
Rusticae	*N. rustica*	Pre-Mendelian (East 1928)	Eghis (1927)
Suaveolentes	*N. africana*	Gerstel et al. (1979)	Doroszewska and Berbeć (1990) and Nikova and Zagorska (1990)
	N. amplexicaulis	Wark (1970)	Wark (1970)
	N. benthamiana	Subhashini et al. (1986) and DeVerna et al. (1987)	Subhashini et al. (1986), DeVerna et al. (1987), and Dorossiev et al. (1990)
	N. cavicola	Nikova et al. (2006)	Nikova et al. 2006 (?)
	N. debneyi	Kostoff (1943) and Clayton (1958)	Bailov et al. (1964) and Clayton (1968)[b]
	N. excelsior	Wark (1970)	Wark (1970)
	N. exigua	Wark (1970) and Kobus (1971)	Wark (1970) and Kobus (1971) (sesquidiploids ExiTT)

(continued)

Table 8.2 (continued)

Section	Species hybridized with *N. tabacum*	Sexual F_1 hybrid (amphihaploid) reported by	Amphipolyploid (sexual or somatic) reported by
	N. fragrans	Durbin and Uchytil (1977)	
	N. goodspeedii	Wark (1970) and Palakarcheva et al. (1978)	Wark (1970) and Palakarcheva et al. (1978)
	N. gossei	Takenaka (1962) and Wark (1970)	Wark (1970) and Burk and Dean (1975)
	N. hesperis	Kubo (1985)[c]	Kubo (1985)[c]
	N. ingulba	Nikova et al. (1998b)	Nikova et al. (1998b)
	N. maritima	Wark (1970)	Wark (1970)
	N. megalosiphon	Clayton (1950)	Palakarcheva and Bailov (1976)
	N. occidentalis	Ternovsky et al. (1972) and Wong (1975)	Ternovsky et al. (1972)
	N. rosulata	Ternovsky et al. (1976), and Larkina (1980)	Ternovsky and Larkina (1978)
	N. rotundifolia		Ilcheva et al. (2001)
	N. simulans	Kubo (1985)[c]	Kubo (1985)[c]
	N. suaveolens	Pre-Mendelian (East 1928)	Izard and Hitier (1955)
	N. umbratica	Murthy et al. (2014)	Murthy et al. (2014)[d]
	N. velutina	Wark (1970)	Wark (1970)
	N. wuttkei	Laskowska and Berbeć (2012)	Laskowska et al. (2015)
	N. eastii	Chaplin and Mann (1961) (direct sesquidiploid EaTT)[e]	
Sylvestres	*N. sylvestris*	Malinowski (1916), Goodspeed and Clausen (1917), and Sachs-Skalińska (1917)	Rybin (1929) and Eghis (1930)
Tomentosae	*N. kawakamii*	three-specific hybrid only (*sylvestris* × *kawakamii*) × *tabacum*; Ohashi (1985)	
	N. otophora	Goodspeed (1945)	Goodspeed and Bradley (1942
	N. setchellii	Greenleaf (1941) and Goodspeed (1945)	Larkina (1983)
	N. tomentosa	Goodspeed and Clausen (1928)	Goodspeed and Bradley (1942)
	N. tomentosiformis	Brieger (1928) and Lehmann (1936)	Fardy and Hitier (1945)
Trigonophyllae	*N. obtusifolia*	Takenaka (1956) and Tanaka (1961)	Chung et al. (1996)
	N. palmeri	Goodspeed (1945)	
Undulatae	*N. arentsii*		DeVerna (1984) (somatic)
	N. glutinosa	Pre-Mendelian (East 1928)	Clausen and Goodspeed (1925)
	N. undulata	Kehr and Smith (1952)	Kehr and Smith (1952)

[a]The account of converting several F_1 hybrids, including *N. tabacum* × *N. bonariensis*, to fertile amphidiploids is ambiguous
[b]Clayton (1968) reported the existence of amphidiploid 4n (*N. debneyi* × *N. tabacum*) in 1938
[c]Only circumstantial evidence exists for the hybrids *N. hesperis* × *N. tabacum* and *N. simulans* × *N. tabacum* having been synthesized based on the report on new cytoplasmic male-sterile lines, cms *N. hesperis,* and cms *N. simulans* (Kubo 1985)
[d]There is a hint to fertile hybrids having been obtained, but it is not explicit enough for certainty
[e]From crossing autotetraploid *N. tabacum* with *N. langsdorffii*

synthetic tobaccos and tobacco. Genetic recombination was found to be reduced but germplasm exchange between chromosomes was relatively unrestricted, owing to the elimination of preferential pairing that hinders genetic exchange in direct crosses with alien diploid relatives. The authors concluded that synthetic tobaccos offer a convenient system for introgressing genetic diversity into *N. tabacum*.

8.9.1.2 Oligogenic Introgression

The majority of documented transfers of traits controlled by several genes appear to be confined to resistance to blue mold caused by *Peronospora hyoscyami*. In a series of successive backcrosses of the amphidiploid 4n *N. debneyi* × *N. tabacum* to *N. tabacum*, Clayton (1958) transferred resistance to blue mold from *N. debneyi*. In the backcross progenies, he recovered resistant plants that consistently contained the full chromosome complement of *N. tabacum* and varying numbers of univalents from *N. debneyi*. Further assiduous selection must have resulted in several alien translocations to the genome of *N. tabacum* because an agronomically acceptable stable-resistant breeding line was ultimately developed (Smith 1968). That apparently unique accomplishment, considering the distance between *N. debneyi* and *N. tabacum*, corresponds well with the much-later findings of Kitamura et al. (2005), who reported that *N. debneyi* and *N. tabacum* had more intergenomic affinities than their respective taxonomic positions would suggest. Oligogenic patterns of inheritance were also reported for blue mold resistance transferred from *N. goodspeedii* (Wuttke 1969) and from *N. velutina* (Wark 1963).

Attempts to transfer resistance to potato virus Y (PVY) from *N. africana* to *N. tabacum* were carried out independently in two laboratories. This resulted in breeding lines that, although tolerant of the virus, did not express the full resistance of the wild species, likely because more than one gene controlled the resistance in *N. africana* (Lewis 2007; Korbecka-Glinka et al. 2017).

8.9.1.3 Monogenic Introgression

With the exception of the above examples, controlled introgression in *Nicotiana* has been limited to simply inherited monogenic dominant traits, primarily to disease resistance factors. *Nicotiana* species have been screened for resistance to various diseases and pests. These studies indicated that the genus contains a vast pool of resistant germplasm that for various reasons remains largely unexploited. The major problem is that the reservoir is difficult to access for reasons discussed in multiple reviews, including a relatively recent review by Lewis (2011). In the following sections, we briefly discuss the proportion of this potentially inexhaustible pool that has been utilized.

Unlike the genetic exchange between varieties of the same species, which is practically unrestricted and predictable because it is governed by simple Mendelian principles of segregation and independent assortment, the transmission and distribution of genetic material between different species is far more complex with no fixed rules that extend over the whole range of potential introgression events.

8.9.2 Transfer of Hypersensitive Response to TMV from *N. glutinosa* to *N. tabacum*

Here, we describe the transmission and incorporation of the gene conferring resistance to TMV from *N. glutinosa* to *N. tabacum* because it best illustrates the mechanics and problems inherent to interspecific breeding in *Nicotiana* (Holmes 1938). It was also the first successful transfer of a usable trait from an alien species to tobacco, and one of the few that have continued to have a weighty and lasting impact, both academically and practically.

(a) the amphihaploid hybrid (TG) and its conversion to fertile amphidiploid (TTGG),
(b) backcrossing the amphidiploid (TTGG) to the recipient parent *N. tabacum* (TT), and
(c) subsequent backcross generations to the recipient *N. tabacum* parent.

'Holmes Samsun', also known as Samsun H, was later demonstrated to contain a regular double complement of 24 chromosome pairs in its somatic cells that is characteristic of *N. tabacum*, but one pair came from *N. glutinosa* (Mallach 1943; Gerstel 1943). It was precisely that pair of chromosomes carried resistance to TMV. In other words, a pair of *N. glutinosa* chromosomes was substituted for a pair of *N. tabacum* chromosomes in the genome of *N. tabacum*. The majority of interspecific transfers that have since been accomplished by investigators and breeders essentially copied this scheme, but there were instances where this protocol could not be followed strictly because of inherent difficulties and barriers caused by particular hybrid combinations. These obstacles were overcome or bypassed by introducing modifications to the basic scheme and they are discussed in the subsequent two sections.

8.9.3 Bridge-Cross method

Chromosomal and genic incongruities between the donor and the recipient, both pre- and post-fertilization, were the primary hindrances that had to be overcome to accomplish the interspecific transfer of a desired trait, and different tools were used to this end. Failing this, an intermediary between the two incongruent species could be deployed. The main quality expected of such an intermediary species was that it ought to cross readily with both the recipient and the donor. This approach was first attempted experimentally by Burk (1967). *N. repanda*, a valuable source of germplasm resistant to many tobacco diseases, is difficult to hybridize directly with *N. tabacum*. Burk's goal was to transfer resistance to TMV from *N. repanda* to *N. tabacum*. This procedure, which came to be called the bridge-cross method, was deployed by several other investigators, who used different combinations of donors, recipients, and bridging species to achieve their goals. Notably, *N. otophora* was used as a bridging species in a successful attempt to transfer hypersensitive response to TSWV from *N. alata* to *N. tabacum* (Gajos 1981). The amphihaploid *N. tabacum* × *N. alata* (TA) was converted to a fertile amphidiploid (TTAA). Sesquidiploids obtained by backcrossing TTAA × TT (TTA sesquidiploids) were self- and cross-sterile. Seeds were obtained by crossing TTA with another hybrid amphidiploid (4n *N. tabacum* × *N. otophora*) obtained earlier (Gajos 1979). In the offspring, one plant showed hypersensitive reaction to TSWV inoculation. The plant was self-sterile because of pollen sterility but produced offspring when backcrossed as the female to *N. tabacum*. Self-fertility was restored, and several lines that showed hypersensitivity to TSWV, characteristic of the donor parent (*N. alata*), were developed. The hypersensitivity, however, was linked to morphological malformations (see Sect. 8.10.2). In another study, two rather than one bridging species (*N. sylvestris* and *N. longiflora*) were used in transferring resistance to root-knot disease from *N. repanda* to *N. tabacum* (Schweppenhauser 1974).

8.9.4 The Use of the Autotetraploid Form of the Recipient Species

Occasionally, the bridge-cross method also circumvented the crossability barrier between two species. Chaplin and Mann (1961) found that using the autotetraploid of *N. tabacum* helped overcome the incompatibility between the cultivated species and *N. rustica* and *N. alata*. The method was also found effective to produce viable hybrids of *N. tabacum* with *N. alata* by Berbeć (1987) and Laskowska and Berbeć (2005).

Another benefit of using an autotetraploid *N. tabacum* parent to produce the starting hybrid is that it bypasses the amphidiploid stage by yielding directly sesquidiploid hybrids. The trade-off is the lessened likelihood of chromosomal interchange in sesquidiploids compared with that in amphidiploids (Patel and Gerstel 1961).

8.9.5 Mechanics of the Incorporation and Integration of Genes from Alien *Nicotiana* Species into the Tobacco Genome

8.9.5.1 Alien Chromosome Substitution or Addition

The mechanism of what had happened in 'Holmes Samsun', a TMV-resistant variety derived from the cross *N. tabacum* × *N. glutinosa*, was explained by Gerstel (1946). Several possible chains of events during gametogenesis all eventually ended in the fusion of two gametes, male and female, each containing 23 chromosomes of the recipient and 1 chromosome of the donor. Such gametes may arise by chance as a result of random and irregular distribution of unpaired chromosomes of the donor species in the sesquidiploid and breakdown generations. If the two fused gametes happen to contain the same homolog chromosome of the donor, barring some unexpected genic imbalances, a stable self-perpetuating line is the result. This was termed an alien substitution line, and the mechanism was described as alien chromosome substitution.

The simplest explanation of the origin of alien substitution plants is the fusion of two gametes, each containing a haploid complement of 24 chromosomes from *N. tabacum* and an additional monosome from *N. glutinosa*. A 50-chromosome alien addition line was a transition stage in integrating the factor of resistance to PVY from *N. africana* in the genome of *N. tabacum* (Lewis 2005). Black root rot-resistant addition lines of *N. tabacum* were obtained by Bai et al. (1996) among backcross derivatives of the somatic hybrid *N. tabacum* × *N. debneyi*. Addition lines containing more than one pair of alien chromosomes were also developed (Burk 1960).

8.9.5.2 Alien Segmental Substitution

Alien substitution or addition lines that involve whole foreign chromosomes are a poor choice in tobacco breeding. Traits introgressed in this manner are difficult to transfer to other varieties because of irregular inheritance, and the varieties thus modified are generally below agronomic standards because they carry a heavy load of alien chromatin derived from the wild donor species, making them practically useless as cultivars. Supernumerary chromosomes are also not likely to recombine with the chromosomes of the recipient genome (Lewis 2005). However, the acceptable TMV-resistant variety Burley 21 was developed by Valleau (1952) through repeated backcross and selection cycles that involved recipient commercial varieties to be improved with the agronomically flawed 'Holmes Samsun' as a donor (Gerstel and Burk 1960). At least some residual homology between the chromosomes involved in the gene transfer was recommended to facilitate the substitution of a whole chromosome, for example, by trivalent formation (Gerstel 1946), but translocation-via-recombination can only occur through chromatid exchange between paired homeologous chromosomes (Clausen and Cameron 1957). Opportunities for the translocation of a chromosome fragment between alien chromosomes are highest in the amphidiploid generation and diminish with subsequent transfer generations (Patel and Gerstel 1961). A translocated segment can be large enough to make the rearranged chromosome visually recognizable as a heteromorphic bivalent when it pairs in meiosis with its normal unchanged homolog (Ramavarma et al. 1991). The translocated fragment itself can be visualized using the genomic in situ hybridization method (Laskowska et al. 2015).

Experimental results of the transfer of a marker locus from *N. plumbaginifolia* to *N. tabacum* obtained by Moav (1958) showed that the incorporation of the *N. plumbaginifolia* locus in *N. tabacum* was clearly non-random. Eight out of 14 transfers involved the same chromosome and 6 were distributed equally among 3 others. Lewis (2002) found that the introgression of resistance to PVY from *N. africana* to *N. tabacum* was likewise non-random, and in three out of seven events the resistance locus was integrated with the same *N. tabacum* chromosome.

Resistance to *Phytophthora nicotianae* (previously *Phytophthora parasitica* var. *nicotianae*) was transferred to *N. tabacum* from *N. longiflora*

(Valleau et al. 1960) and from *N. plumbaginifolia* (Chaplin 1962). The single dominant genes for resistance to the black shank disease, *Php* and *Phl* from *N. plumbaginifolia* and *N. longiflora*, respectively, were found to be located on the same chromosome of *N. tabacum* but were not allelic. The two loci, however, were very close to one another, such that double recessive recombinants were produced in F_2 *Php* × *Phl* at a frequency just over 0.05%, and testcrosses had a frequency of about 1.5% (Johnson et al. 2002b).

Interestingly, a similar situation was found with two single dominant genes for resistance to *Thielaviopsis basicola* (black root rot), one originally from *N. debneyi* and transferred to local cultivars via AC Gayed, a Canadian flue-cured cultivar, and the other introgressed directly from *N. glauca* (Trojak-Goluch and Berbeć 2009). The F_2 population of approximately 600 plants from selfing the F_1 hybrid between the two resistant genotypes contained no susceptible double recessive recombinants (Berbeć, unpublished observation). Unlike *N. longiflora* and *N. plumbaginifolia*, which are the closest relatives within the same section, *N. debneyi* and *N. glauca* have been classified into two distantly related sections (Chase et al. 2003). However, the analysis of 5S rDNA spacer sequences (Kitamura et al. 2005) revealed that 24 of the 48 chromosomes of *N. debneyi* showed a high degree of homology with the genomic DNA of *N. glauca*. A sequence-characterized amplified region marker linked to the black root rot-resistant region from *N. debneyi* (Julio et al. 2006) was amplified in the accessions of *N. debneyi* and *N. glauca*, as well as in two black root rot-resistant lines of *N. tabacum,* one with resistance from *N. debneyi* and the other from *N. glauca* (Dr G. Korbecka, private communication). Random breakage and reunion of chromosomes, either spontaneous (Moav 1958) or induced (Niwa 1969), also may play a part in translocating chromosome fragments, but because of the random nature of such translocations, they typically result in biologically ill-compensated and agronomically inferior products (Lewis 2011).

8.9.6 Overcoming or Circumventing Barriers to Interspecific Introgression

8.9.6.1 Chromosome Homology/Homeology

Recombination involving the exchange of chromosome segments (alien substitution) in first hybrid generations was reported for crosses with the progenitor species of *N. tabacum* or its close relatives, *N. otophora*, *N. tomentosiformis*, and *N. sylvestris*, that retained a considerable degree of chromosome homeology with their cultivated descendant (Chaplin and Mann 1961). Laskowska et al. (2015) showed that translocations of chromosome fragments can also occur in the amphidiploid stage of considerably more remotely related hybrids, in which only vestigial homeology was conserved. In some other closely tracked introgression events, the transfer of a whole chromosome either as an addition or as a substitution was the first stage of gene transfer (Gerstel 1945; Burk 1960; Lewis 2005).

8.9.6.2 Preferential Pairing

Preferential pairing in interspecific hybrids is the propensity of a chromosome to associate with its counterpart from the same species to the exclusion of an alien chromosome, even if the latter shows some degree of homology with the former (Gerstel 1961). Preferential pairing further restricts already scant opportunities for chromosomal interchanges between alien species, and the condition becomes increasingly worse in sesquidiploid and breakdown generations. Unlike other genera where preferential pairing is controlled by specific genes, in *Nicotiana* it is controlled chromosomally (Gerstel 1961).

8.9.6.3 Sterility and Instability of Amphidiploids

Chromosome elimination is frequently observed in amphihaploid hybrids and other allopolyploids. For example, it has been reported in the amphidiploids of *N. tabacum* × *N. glauca* (Doroszewska and Berbeć 2000; Szilagyi 1975),

N. tabacum × *N. glutinosa* (Patel and Gerstel 1961), *N. tabacum* × *N. sylvestris*, *N. tabacum* *N. otophora*, *N. tabacum* × *N. tomentosiformis* (Yang 1960), *N. raimondii* × *N. tabacum* (Berbeć 1988), and *N. knightiana* × *N. tabacum* (Berbeć, unpublished observation). Chromosome elimination also has been found in amphidiploids produced via somatic protoplast fusion. The self-pollinated progeny of a true somatic amphidiploid *N. tabacum* × *N. debneyi* (2n = 96) ranged from an amphidiploid to an aneuploid with 60 somatic chromosomes (Sproule et al. 1991). The chromosome loss in amphidiploids may involve a chromosome carrying the gene required in the interspecific transfer. Therefore, if amphidiploids are maintained as seed stocks, they should be verified for the presence of the desired trait prior to their deployment in introgression projects (Burk and Chaplin 1979).

8.9.6.4 Sterility of the Sesquidiploid Generation

At least partial fertility of sesquidiploids or near-sesquidiploids is essential for conventional introgression. The sesquidiploid condition is the bottleneck of introgression because, in theory at least, all interspecific gene transfers have to pass through that stage, regardless of how the transfer actually began: by backcrossing of the amphihaploid hybrid to the recipient parent (chancing restitution gametes being formed by the former), by converting the sterile amphihaploid to fertile amphidiploid and backcrossing it to the recurrent parent, or by crossing the autotetraploid recipient with the diploid donor. Fortunately, fully sterile sesquidiploids happen relatively rarely. The sesquidiploid involving *N. tabacum* and *N. alata* has been consistently reported as sterile by several investigators, regardless of how it was developed (Chaplin and Mann 1961; Gajos 1981; Berbeć 1987).

8.9.6.5 Elimination of Somatic Chromosomes

Elimination of somatic chromosomes can also be a source of unpredictable inheritance patterns. It commonly occurs in somatic hybrids, as discussed in Sect. 8.7.1, and is the source of somaclonal variability in tissue culture. Variable chromosome numbers were recorded in regenerants from embryo rescue cultures of *N. glutinosa* × *N. megalosiphon* (Subhashini et al. 1986), *N. tabacum* × *N. knightiana* (Slusarkiewicz-Jarzina and Zenkteler 1983), and *N. repanda* × *N. tabacum* (Iwai et al. 1985). Explant cultures used to induce chromosome doubling also resulted in chromosome loss, aneuploidy, and mixoploidy of the regenerants of *N. velutina* × *N. tabacum*, *N. maritima* × *N. tabacum*, and *N. benthamiana* × *N. tabacum* (Nikova et al. 1991). A particularly highly variable population of regenerants was produced in the stem pith culture of *N. paniculata* × *N. tabacum*, ranging from 42 to 64 chromosomes (Nikova et al. 1991).

8.9.6.6 Erratic Inheritance in Early Transfer Generations

Lack of homology between the species involved in interspecific transfer not only blocks the necessary chromosome interchanges and recombination but also results in a substantial number of unassociated chromosomes (univalents) and their random elimination and/or uneven distribution to the gametes. Unpaired chromosomes were included in only 20–25% of the gametes (Gerstel 1943) on the female side, and on the male side the transmission rate was much lower. Lower monosome transmission rates through the pollen than through the ovule were also reported by Lewis (2005) in his account of the transfer of the resistance to PVY from *N. africana* to *N. tabacum*.

8.9.6.7 Erratic Inheritance in Advanced Introgression Stages

The hypersensitive response to TSWV transferred from *N. alata* and incorporated into the tobacco variety 'Polalta' (Gajos 1981), the sole known source of resistance to that virus, has become notorious in introgressive breeding. In addition to a peculiar kind of genetic drag, which is the main issue reported (see Sect. 8.10.2), this resistance has other flaws related to inheritance. In some cases, perfect ratios for the monogenic dominant pattern were observed (Moon and Nicholson 2007), whereas in others, the

proportion of resistant-to-susceptible plants was found to be close to that expected for a digenic type of inheritance (Kennedy and Nielsen 1993). Resistance is now considered to be controlled by the single dominant 'Polalta'-derived *RSTV-al* gene (Moon and Nicholson 2007; Laskowska et al. 2013). The resistant-to-susceptible ratios, however, tend to be skewed toward the susceptible genotypes if male gametes are involved in the transmission of the resistance factor (Laskowska and Berbeć 2010). The *RSTV-al*-mediated hypersensitive response to TSWV becomes highly unstable once introgressed in a new genetic background (Laskowska et al. 2013).

8.9.6.8 Genetic Tumors in Interspecific Hybrids of *Nicotiana*

Formation of tumorous growth in some hybrid combinations, first described by Kostoff (1930), is another peculiar manifestation of hybrid instability. By studying the tumorization process in the hybrid *N. glauca* × *N. langsdorffii*, Ichikawa et al. (1990) explained tumor growth as controlled by the *rol* genes that conferred increased sensitivity to endogenous auxins in tumorous hybrids.

Sequences of cellular transferred cT-DNA (*rol* genes) were acquired horizontally from *Agrobacterium rhizogenes* by some ancestral *Nicotiana* species in their phylogenetic past and were subsequently dispersed sexually across the genus. They have been detected in several present-day species (Matveeva and Lutova 2014). Tumorization may be induced by complex gene interactions that affect hormone levels in plants and sensitivity to hormones such as indole-3-acetic acid and cytokinins (Matveeva and Lutova 2014). Tumors in hybrids involving *N. tabacum* were reported for several trispecific combinations of *N. debneyi* × *N. tabacum* as female parents, with members of section Alatae (*N. alata*, *N. longiflora*, *N. langsdorffii*, *N. plumbaginifolia*, and *N. sanderae*) as males (Kehr 1951; Ahuja 1962). Genetic tumors typically appear in the late stages of plant life, generally after cessation of active growth and flowering, and they have not yet been reported as interfering with interspecific transfer and introgression processes per se. This phenomenon, however, may be implicated in the negative effects associated with the *RSTV-al* gene conferring resistance to TSWV transferred from *N. alata* (see Sect. 8.10.2).

8.10 Effects of Introgressing Alien Genes into *N. tabacum*

8.10.1 Side Effects of Interspecific Introgression in *Nicotiana*

Four major classes of side effects may be produced by an alien gene transferred to the genetic background of a cultivated variety:

(a) The chromosome segment introgressed in the genome of *N. tabacum* carries alien chromatin along with the desired gene (linkage drag);
(b) The expression of the alien gene may be affected by the genetic background in which it has been introgressed;
(c) The introgressed gene may affect more than one trait (pleiotropy); and
(d) The performance of the introgressed trait varies depending on the site of insertion (position effect).

Pleiotropic effects have been demonstrated to be of little, if any, importance in TMV resistance conferred by the *N* gene from *N. glutinosa*. By comparing the performance of two isogenic lines, one containing the *N* gene introgressed by conventional interspecific transfer and the other transformed with the same *N* gene in its 'pure' cloned form, the agronomically negative effects associated with resistance to TMV were found to be practically absent in the transgenic line (Lewis et al. 2007a).

Position effects also appeared to have little effect in another case study of resistance to TMV introgressed from *N. glutinosa*. An array of breeding lines of tobacco carrying the *N* gene on different chromosomes did not have significantly different agronomic performances (Lewis and Rose 2010).

The failure to recover the full expression of the resistance trait in its native genetic background once it had been transferred to the genome of *N. tabacum* was reported for resistance to PVY from *N. africana* (Lewis 2007; Doroszewska 2010; Korbecka-Glinka et al. 2017). The level of *N. africana*-derived tolerance of PVY also varied among the tolerant genotypes (Lewis 2007).

A well-known case of reduced expression in *N. tabacum* compared with expression in the native species is the resistance to blue mold from *N. debneyi,* as many as four levels of expression were defined by Clayton (1968) depending on each individual introgression, indicating the involvement of several genes from both the donor and the recipient genomes. The incomplete recovery of the desired trait in the majority of interspecific transfers is typically best explained by the oligogenic inheritance of those traits.

8.10.2 Linkage Drag

Phenotype-modifying effects of the introgressed alien gene in different genetic environments vary by individual case. In two independent studies of *N. debneyi*-derived resistance to black root rot (*Thielaviopsis basicola*), one by Legg et al. (1981) in burley and the other by Haji et al. (2003) in flue-cured tobacco, the negative impact on yield was similar. Interactions with genetic background were found for resistance to wildfire race 0 from *N. longiflora* (Legg et al. 1982; Nielsen et al. 1985) and resistance to TMV (Nielsen et al. 1985). Generally, alien introgressions performed better in air-cured dark tobacco and burley genotypes than in flue-cured tobacco (Chaplin and Burk 1971).

Rao and Stokes (1963) reported that TMV-resistant varieties were more liable to develop calcium deficiency symptoms than TMV-susceptible varieties. They also established that the H chromosome of *N. glutinosa* that carried the hypersensitive response to TMV was the most likely bearer of the genes that influence the expression of such symptoms. An increased sensitivity to calcium deficiency seemingly related to TMV resistance was also observed by Berbeć (unpublished data).

Linkage drag associated with resistance to TSWV from *N. alata* is a peculiar instance of adverse effects related to interspecific introgression. The resistance gene, *RSTV-al*, was transferred from *N. alata* with the possible involvement of a third species, *N. otophora* (Gajos 1979, 1981, 1984). As mentioned in Sect. 8.9.6.7, the resistance factor itself behaved in a rather unpredictable manner, but persistent selection for homozygous-resistant genotypes apparently largely stabilized its mode of inheritance and regularity of expression (Moon and Nicholson 2007; Laskowska and Berbeć 2010; Trojak-Goluch et al. 2011; Trojak-Goluch et al. 2016b). The most common side effects associated with the *RSTV-al* gene are deformed leaf blade and leaf venation and tumorous outgrowths, although the latter usually appear in aged plants (Moon and Nicholson 2007, and our own unpublished observations). Another peculiar quality of the malformations linked to TSWV resistance is that they are most pronounced in heterozygous conditions, diminish or disappear altogether once the homozygosity of *RSTV-al* is restored, and recur in the heterozygous offspring of TSWV-resistant × non-resistant lines (Laskowska and Berbeć 2010; Trojak-Goluch et al. 2011; Trojak-Goluch et al. 2016a). In all those features, the linkage drag associated with the *RSTV-al* gene bears a striking similarity to genetic tumors developed by some interspecific hybrid combinations in *Nicotiana*, including the paradigmatic hybrid *N. glauca* × *N. langsdorffii*, as discussed briefly in Sect. 8.9.6.8. According to the 'plus × minus' hypothesis (Näf 1958), the hybrid *N. tabacum* × *N. alata* can develop tumors; such tumors have been reported, but only by one author who described them as occurring 'rarely' (Kostoff 1943). Other investigators who reported on that hybrid did not observe these malformations (Nagao 1979; Gajos 1981; Berbeć 1987; Berbeć and Laskowska 1997). These reports suggest that the interaction between factors from *N. alata* and *N. tabacum* alone was not likely to produce the massive teratological changes associated with the *RSTV-al* gene.

However, it may be added in this context that abnormal morphology was observed in transgenic cotton that expressed a Thielaviopsis basicola-inhibiting protein, NaD1, encoded by the nad1 gene derived from *N. alata* (Pereg 2013). According to Gajos (1984), both *N. alata* and TSWV-resistant derivatives from crossing *N. tabacum* with *N. alata* also expressed resistance to black root rot.

Therefore, one may hypothesize that a series of rearrangements and recombinations within the genomes of *N. otophora*, *N. tabacum*, and *N. alata* was involved in the transfer of the *RSTV-al* gene and led to the formation of a block of tightly linked genes that included, along with the *RSTV-al* gene, some genes or/and gene deletions that affected hormonal balance, especially in genotypes heterozygous/hemizygous for the genes within that chromatin segment.

Instances of introgression from *Nicotiana* species that resulted in detrimental effects on the agronomic performance of cultivated tobacco are listed in Table 8.3.

8.10.3 The Role of 'Synthetic Tobaccos' in Facilitating Interspecific Gene Transfer and Alleviating Linkage Drag

In some gene transfers, 'synthetic' rather than 'natural' *N. tabacum* was used with success as an intermediary to eliminate undesirable linkages. 'Beltsville 771', a blue mold-resistant line of *N. tabacum* with polygenic resistance factors from *N. debneyi*, continued to show irregular inheritance and inferior agronomic performance despite assiduous selection efforts (Clayton 1968). A considerable improvement in resistance stability and in agronomic performance was obtained when 'Beltsville 771' was crossed with the so-called Kostoff's hybrid, at that time believed to be a pure amphidiploid 4n *N. sylvestris* × *N. tomentosiformis*, but later demonstrated to carry a considerable amount of introgression from *N. tabacum* (Sheen 1972; Lim et al. 2006). 'Synthetic tobacco' was also used by Clayton (1968) and by Schweppenhauser (1975a) to solve the problem of adverse linkages derived from *N. tabacum* sources and associated with the resistance to root-knot nematodes. Present-day commercial *N. tabacum* cultivars and breeding lines have undergone considerable downsizing of their functional genomes and narrowing of their genetic base through the evolutionary process (Renny Byfield et al. 2011; Bombarely et al. 2012) and as a consequence of purposeful selection (Hancock and Lewis 2017). In contrast, 'synthetic tobaccos' (4n *N. sylvestris* × *N. tomentosiformis*, 4n *N. sylvestris* × *N. otophora*), especially those synthetized de novo (Skalicka et al. 2005), retain the full genetic potential of their ancestors and preserve all homeologous genes in duplicate. This makes them more receptive to gene exchanges and better buffered against adverse effects of chromosomal rearrangements and loss of biologically essential chromosome regions involved in interspecific translocations. Because of these qualities, 'synthetic tobaccos' offer a convenient tool to facilitate gene flow between *N. tabacum* and other *Nicotiana* species (Chaplin and Mann 1978; Hancock and Lewis 2017).

8.10.4 In Vitro Culture-Aided and Marker-Assisted Recovery of Introgressed Genes with Less Alien Chromatin

Breeders aim for the introgressed chromatin fragment to be as small as possible while retaining the essential gene. The simplest approach is to resort to repeated assiduous backcrossing, hoping for a rare recombination to occur within the translocated region. Opportunities for recombination, however, are considered limited within and around the introgressed region. Johnson et al. (2002a) found that recombination was highly suppressed in the region around the *Ph* gene, which is allelic to the *Php* gene for resistance to black shank; the *Php* gene was introgressed from *N. plumbaginifolia* (Johnson et al. 2002b).

Table 8.3 Instances of linkage drag effects in introgression from alien *Nicotiana* species to *Nicotiana tabacum*

Source of introgressed factor	Introgressed factor	Side effects caused by linkage drag	Reported by
N. alata	Resistance to tomato spotted wilt virus	Morphological malformations and tumors	Gajos (1984) and Kennedy, Nielsen (1993), Moon and Nicholson (2007), and Laskowska and Berbeć (2010)
N. longiflora	Resistance to *Phytophthora nicotianae* race 0 (black shank) (*Phl* locus)	Reduced yield, reduced cured-leaf quality	Valleau et al. (1960) and Apple (1967)
	Resistance to *Globodera javanica* (root knot)	Reduced cured-leaf quality	Crowder et al. (2003)
	Resistance to *Pseudomonas syringae* pv. *tabaci* race 0 (wildfire)	No side effects	Nielsen et al. (1985)
		Reduced leaf number, reduced total alkaloids, reduced yields; other agronomic parameters vary with genetic background	Legg et al. (1982)
N. plumbaginifolia	Resistance to *Phytophthora nicotianae* race 0 (black shank) (*Php* locus)	Reduced yield, reduced cured-leaf quality	Chaplin (1962) and Lewis (2011)
N. rustica var. *brasilea*	Resistance to *Phytophthora nicotianae* race 0 and race 1 (black shank); resistance to *Pseudomonas syringae* pv. *tabaci* race 0 and race 1 (wildfire) and *P. syringae* pv. *angulata* (angular leaf spot)	No side effects	Drake et al. (2015)
N. africana	Resistance to potato virus Y	Reduced yield	Lewis (2007)
N. debneyi	Resistance to *Thielaviopsis basicola* (black root rot)	Delayed flowering, delayed leaf maturity, reduced cured-leaf quality, reduced total nitrogen, reduced total alkaloids	Legg et al. (1981), Haji et al. (2003), and Bai et al. (1996)
	Resistance to *Peronospora hyoscyami* f.pv. *tabacina* (blue mold)	Increased alkaloid content (burley tobacco), reduced yield, reduced cured-leaf quality (flue-cured tobacco)	Verrier et al. (2016)
N. tomentosa	Quantitative trait locus affecting leaf number and days to flower	Decreases in percent total alkaloids and increases in percent reducing sugars	Eickholt and Lewis (2014)
N. glutinosa	Resistance to tobacco mosaic virus	Slow growth, reduced leaf size, reduced yield, changed green leaf appearance, reduced leaf quality	Chaplin et al. (1961), Chaplin et al. (1966), Chaplin and Mann (1978), Legg et al. (1979), Lewis et al. (2007a) and Rao and Stokes (1963)
		Increased sensitivity to calcium deficiency	Rao and Stokes (1963), Berbeć (unpublished)

The *Ph*/*Php* genes confer resistance to black shank (*Phytophthora nicotianae*).

Another advocated remedy is to re-transfer the gene of interest from its original source species in the expectation that more beneficial and smaller translocations would occur during the process (Chaplin and Mann 1978; Bai et al. 1996).

A selection scheme for small-size translocations in the progenies of a 50-chromosome addition line resistant to PVY, derived from the hybrid *N. tabacum* × *N. africana*, and assisted by tissue culture and DNA sequence markers was devised by Lewis (2005). The stimulation effect of in vitro culture on the rate of translocations between *N. africana* chromosomes and the *N. tabacum* genome was manifested by six out of seven alien substitution and/or translocation events being identified in backcross progenies from the tissue culture regenerants.

Several DNA markers linked to traits introgressed into *N. tabacum* from different *Nicotiana* species have been developed (Lewis 2011). The list may be supplemented with sequence-characterized amplified region markers for black root rot and blue mold resistances from *N. debneyi* (Julio et al. 2006) and single-nucleotide polymorphism markers for black root rot resistance, also from *N. debneyi* (Qin et al. 2016). Amplified fragment length polymorphism and simple sequence repeat markers linked to the gene conferring resistance to blue mold in a race of *N. langsdorffii*, reportedly a new potential alternative to section Suaveolentes species as a source of resistance to that disease (Zhang and Zaitlin 2008), were also developed (Zhang et al. 2012).

8.10.5 'Genetic Drag' Associated with Beneficial Effects

Some results presented by Lewis (2007) on the effects of the gene transferred from *N. africana* that conferred tolerance of PVY indicated that the alien chromatin introgressed with the resistance gene may actually contribute to improved cured-leaf quality of resistant genotypes. There are several flue-cured varieties originally bred for resistance to black shank that carry the resistance gene *Php* derived from *N. plumbaginifolia*. These varieties were subsequently found to exhibit a high degree of tolerance toward *Globodera tabacum*, the cyst nematode (Johnson et al. 2009). The two resistances were closely linked and apparently carried on the same chromosome segment with suppressed recombination capabilities (Johnson et al. 2009). The resistance from *N. rustica* var. *brasilea* was bred into a line termed WZ. The *Wz* gene was found to segregate independently of the resistance conferred by *N. longiflora* (Woodend and Mudzengerere 1992). Drake and Lewis (2013), however, demonstrated than the introgressed chromatin region surrounding *Wz* also contained a resistance factor that was closely linked to black shank, and effective against two main races (0 and 1) of that pathogen. They concluded that the added benefit of black shank resistance had been accomplished inadvertently by Woodend and Mudzengerere (1992) while transferring the resistance to wildfire. In another example, the TSWV-resistant derivatives from crossing *N. tabacum* with *N. alata* were also found to resist black root rot (Gajos 1984). The observed added benefit was, in the researcher's account, attributable to unconscious concomitant selection for two closely linked resistance factors derived from an accession of *N. alata* that also showed full resistance to black root rot. Generally, *N. alata* is not listed among the sources of resistance to black root rot (Burk and Heggestad 1966, Doroszewska and Przybyś 2007), although different lineages of that allogamous and polymorphic may vary in their response to inoculation with *Thielaviopsis basicola* (Doroszewska and Przybyś 2007). Interestingly, the defensin-encoding gene, nad1, cloned from *N. alata,* was experimented with to confer protection against *Th. basicola* to some other crops (Pereg 2013).

8.10.6 Designer or Shuttle Chromosome

Suppressed recombination and inheritance *en bloc* that plagues interspecific introgression was intended to be leveraged by constructing an

artificial chromosome onto which gene constructs could be gradually added to create a whole package of agronomically useful genes. Such a package would segregate as a single unit and could be shuttled from one breeding line to another in intact form, thus facilitating the development of improved cultivars (Campbell et al. 1994; Lewis and Wernsman 2001). Campbell and colleagues incorporated a pair of chromosomes from *N. africana* into a 50-chromosome breeding line of *N. tabacum*. A linkage block was created that included resistance to PVY and tobacco etch virus native to the *N. africana* chromosome, the cloned *N* gene originally from *N. glutinosa* that imparts resistance to TMV, and a *dhfr* transgene that confers resistance to the antibiotic methotrexate. The occasional loss of the *N* gene from the linkage package, possibly through recombination with the *N. tabacum* genome, was an unexpected setback. Some recombination involving the *N. africana* chromosome and *N. tabacum* chromosomes in the 50-chromosome addition line occurred in other experiments that were not directly related to the construction of the gene shuttle (Lewis 2005).

8.10.7 Linkage Drag in Heterozygous Condition

Save for the unusual case of the linkage drag associated with resistance to TSWV, many of the adverse effects of introgression were eliminated or alleviated when the introgressed chromatin was in the heterozygous rather than homozygous condition. Further, vital genes lost in the translocation region of the rearranged chromosome could be replaced by those present in its unchanged counterpart. Because multiple alien genes inserted along with a transferred gene of interest can exhibit additive action, however, their effects would be diminished in the heterozygous condition (Lewis 2011). This was confirmed in F_1 hybrids between parents resistant and susceptible to black root rot (*Thielaviopsis basicola*) from *N. debneyi* (Haji et al. 2006), resistant to black shank (*Phytophthora nicotianae*) from *N. longiflora* (Wernsman and Rufty 1987), resistant to TMV from *N. glutinosa* (Chaplin et al. 1966; Lewis and Rose 2010; Lewis 2011), and resistant to PVY from *N. africana* (Lewis 2007). Commercial utilization of genes resistant to diseases and possibly other genes from interspecific sources is largely possible because in F_1 hybrids and in hybrids involving more than two breeding lines (Berbeć 2017), the trade-offs of interspecific introgression are considerably mitigated.

8.10.8 Summary of Gene Transfer from *Nicotiana* Species to Tobacco

Table 8.4 lists *Nicotiana* species from which introgression to tobacco was reported. All accounts of interspecific introgression found in the literature were included. They range from very detailed and well-documented reports to accounts with only fragmentary information available in abstracts or even to brief citations. Full descriptions were not always available, because many of the reports published in journals were difficult to access and/or were in languages and scripts that were difficult to translate. Therefore, we almost certainly have missed some important details and perhaps also whole projects, possibly involving other species and other introgressed traits.

8.11 *Nicotiana* Species as Sources of CMS in *N. tabacum*

8.11.1 Genetic Background and Phenotypic Manifestation of CMS

CMS is a specific case of interaction between the nuclear and cytoplasmic DNA (Gerstel et al. 1978; Gerstel 1980; Hanson and Conde 1985). More specifically, the interaction was found to be dependent on the complementarity of specific chromosomal and cytoplasmic genes and on the influence of specific nuclear genes on the

Table 8.4 *Nicotiana* species as sources of resistance or tolerance traits introgressed in *Nicotiana tabacum*

Section	Introgression source species	Introgressed trait	Reported by	Notes
Alatae	*N. alata*	Resistance to tomato spotted wilt virus	Gajos (1981, 1984)	Transfer of resistance performed through the mediation of the hybrid *N. tabacum* × *N. otophora* (Gajos 1981)
			Atanassov et al. (1991)	Via asymmetric protoplast fusion
			Patrascu et al. (1999)	Via asymmetric protoplast fusion
		Resistance to *Thielaviopsis basicola* (black root rot)	Gajos (1984)	Allegedly, introgressed from *N. alata* together with resistance to TSWV because of the close linkage between the two resistance factors
	N. longiflora	Resistance to *Phytophthora nicotianae* race 0 (black shank)	Valleau et al. (1960) and Collins and Legg (1969)	
		Resistance to *Pseudomonas syringae* pv. *tabaci* race 0 (wildfire)	Clayton (1947)	
		Resistance to *Meloidogyne javanica* (root-knot nematode)	Schweppenhauser (1975b), Mackenzie et al. (1986), Ternouth et al. (1986), Venkateswarlu et al. (1998)	
	N. plumbaginifolia	Resistance to *Phytophthora nicotianae* race 0	Cameron (1958), Chaplin (1962), and Apple (1962)	
		Tolerance of *Globodera tabacum* (cyst nematode)	Johnson et al. (2009)	Found to be inadvertently introgressed from *N. plumbaginifolia* because it was closely linked with resistance to *Phytophthora nicotianae*
	N. sanderae	Resistance to tobacco mosaic virus	Palakarcheva (1984) and Dorossiev et al. (1990)	Reported without details
		Resistance to *Erysiphe cichoracearum* (powdery mildew)	Dorossiev et al. (1990)	Reported without details
Noctiflorae	*N. glauca*	Resistance to *Thielaviopsis basicola* (black root rot)	Berbeć (1966) and Trojak-Goluch and Berbeć (2009)	Report by Berbeć (1966) lacks details
		Resistance to potato virus Y	Berbeć (1966)	Reported without details
		Tolerance of tobacco mosaic virus	Berbeć (1966)	Reported without details
	N. noctiflora	Resistance to tobacco mosaic virus	Palakarcheva (1984)	Reported without details
Paniculatae	*N. benavidesii*	Tolerance of potato virus Y	Berbeć and Głażewska (1988)	Symptomless carrier type
	N. knightiana	Resistance to *Peronospora hyoscyami* f.pv. *tabacina*	Corbaz (1962)	Reported without details
Repandae	*N. repanda*	Resistance to *Meloidogyne javanica* (root-knot nematode)	Schweppenhauser (1974), Gwynn et al. (1986), Mackenzie et al. (1986), Ternouth et al. (1986), and Mudzengerere (1994)	
		Resistance to *Globodera tabacum* (cyst nematode)	Gwynn et al. (1986)	
		Resistance to *Pseudomonas syringae* pv. *tabaci* (wildfire, race not specified)	Gwynn et al. (1986)	
		Resistance to tobacco mosaic virus	Gwynn et al. (1986)	Not completed, unstable heterozygous substitution
			Bates (1990)	By asymmetric protoplast fusion followed by sexual backcross to *N. tabacum*; resistant *N. tabacum*-like phenotypes recovered
		Resistance to *Cercospora nicotianae* (frog eye)	Wan et al. (1971)	Carried to advanced generations but not completed

(continued)

Table 8.4 (continued)

Section	Introgression source species	Introgressed trait	Reported by	Notes
Rusticae	*N. rustica* var. *pumila*	Race 0 and race 1 (wildfire) and *Pseudomonas. syringae* pv. *angulata* (angular leaf spot)	Stavely and Skoog (1976, 1978)	
	N. rustica var. *brasilea*	Resistance to *Phytophthora nicotianae* race 0 and race 1 (black shank)	Bukuta (2002)	Introgressed fragment of the *N. rustica* chromosome conferring resistance to wildfire and angular leaf spot linked to resistance factor to black shank (Drake and Lewis 2013). Resistance to wildfire and angular leaf spot also mentioned as linked to resistance to black shank by Bukuta (2002)
		Resistance to *Pseudomonas syringae* pv. *tabaci* race 0 and race 1 (wildfire) and *P. syringae* pv. *angulata* (angular leaf spot)	Woodend and Mudzengerere (1992) and Bukuta (2002)	
	N. rustica var. *chlorotica*	Resistance to *Peronospora hyoscyami* pv. *tabacina*	Pandeya et al. (1986)	Resistance identified in progenies from a somatic *N. rustica* × *N. tabacum* hybrid. A singular case because *N. rustica* is not known to carry resistance to blue mold. Explained by possible complementation or interactions of nuclear and/or cytoplasmic genomes of parental species. Validity of results doubted by Rufty (1989)
	N. rustica	General agronomic performance: yield, alkaloid content, tolerance of diseases	White et al. (1979) and Pandeya and White (1984)	Cultivar 'Delgold', a commercially viable flue-cured variety
Suaveolentes	*N. africana*	Resistance to potato virus Y	Doroszewska and Berbeć (1999) and Doroszewska (2010)	Recessive substitution, tolerance (symptomless host response, covers a wide spectrum of potato virus Y strains) Inherited in monogenic recessive fashion (Korbecka-Glinka et al. 2017)
		Resistance to potato virus Y and tobacco etch virus	Lewis (2005) and Lewis et al. (2007a, b)	Partially dominant segmental substitution
	N. amplexicaulis	Resistance to *Meloidogyne javanica* (root-knot nematode)	Venkateswarlu et al. (1998)	Introgression reported as in progress
	N. benthamiana	Resistance to *Spodoptera litura* (tobacco caterpillar)	Ramavarma et al. (1980)	Visual evidence for segmental substitution as a heteromorphic bivalent in meiosis
		Resistance to aphids	Murthy et al. (2014)	
	N. debneyi	Resistance to *Peronospora hyoscyami* f.pv. *tabacina* (blue mold)	Clayton (1958) and Smith (1968)	Oligogenic resistance transferred by means of several translocations involving different *N. debneyi* chromosomes
			Palakarcheva (1981)	Dominant, oligogenic (three-factorial) resistance
			Lea (1963), Wark (1963, 1970), Clayton et al. (1967), and Marani et al. (1971)	Monogenic resistance conditioned by a single major gene
		Resistance to *Thielaviopsis basicola* (black root rot)	Clayton (1958, 1969); Brandle et al. (1992); Bai et al. (1996); Kenward et al. (1999)	Resistance transferred from a sexual *N. tabacum* × *N. debneyi* hybrid Resistance transferred from a somatic *N. debneyi* × *N. tabacum* hybrid Tnd-1 retrotransposon identified as either participating in or closely linked to black root rot resistance
		Resistance to *Erysiphe cichoracearum* (powdery mildew)	Palakarcheva and Krusteva (1978) and Palakarcheva (1981)	Dominant resistance factor(s)
			Smeeton and Ternouth (1992)	
	N. excelsior	Resistance to *Peronospora hyoscyami* f.pv. *tabacina*	Gillham et al. (1977)	
		Resistance to aphids	Murthy et al. (2014)	
	N. exigua	Resistance to *Peronospora hyoscyami* f.pv. *tabacina*	Wark (1975), Gillham et al. (1977), and Manolov (1980)	

(continued)

Table 8.4 (continued)

Section	Introgression source species	Introgressed trait	Reported by	Notes
	N. goodspeedii	Resistance to *Peronospora hyoscyami* f.pv. *tabacina*	Wark (1963) and Wuttke (1969)	Dominant monogenic or oligogenic patterns of resistance depending on lineage (Wuttke 1969) The validity of that introgression doubted by Milla et al. (2005)
			Palakarcheva (1981)	Dominant resistance factor(s)
		Resistance to *Erysiphe cichoracearum* (powdery mildew)	Palakarcheva (1981)	Dominant resistance factor (s)
		Resistance to aphids	Palakarcheva (1984)	
	N. gossei	Resistance to *Spodoptera litura* (tobacco caterpillar)	Rao et al. (1980)	
		Resistance to aphids	Krusteva et al. (2003)	Resistance bred into oriental lines; no details of the actual transfer reported
	N. umbratica	Tolerance of leaf curl virus	Murthy et al. (2014)	
		Resistance to aphids	Murthy et al. (2014)	
	N. velutina	Resistance to *Peronospora hyoscyami* f.pv. *tabacina*	Wark (1963)	Oligogenic pattern of inheritance
			Corbaz (1962) and Gillham et al. (1977)	
Tomentosae	*N. kawakamii*	Resistance to potato virus Y	Ohashi (1985)	Reported without details
	N. setchellii	Resistance to *Erysiphe cichoracearum* (powdery mildew)	Shabanov et al. (1974)	No details available
	N. tomentosa	Resistance to *Meloidogyne incognita* (root-knot nematode)	Slana et al. (1977) and Legg and Smeeton (1999)	
		Quantitative trait locus affecting leaf number and days to flower	Eickholt and Lewis (2014)	
		Wh-p 'pale white flower' and Pp 'purple plant' factors	Clausen and Cameron (1944)	
	N. tomentosiformis	Resistance to *Erysiphe cichoracearum* (powdery mildew)	Ternovsky (1941), Ohashi (1985), and Smeeton and Ternouth (1992)	
		Resistance to *Meloidogyne javanica* (root-knot nematode)	Mackenzie et el al (1986) and Ternouth et al. (1986)	
		Resistance to tobacco vein mottling virus	Legg and Smeeton (1999)	
	N. glutinosa	Resistance to tobacco mosaic virus	Holmes (1938), Gerstel (1943, 1946), Ternovsky (1941), Oka (1961)	

mitochondrial genome (Hakansson et al. 1988; Hakansson and Glimelius 1991). In many allogamous species, mutations are the main source of CMS. In autogamous tobacco, potential male-sterile mutations are rapidly eliminated from the population and very few have been reported (Berbeć 1974; Kobus 1978). Overwhelmingly, male-sterile forms in tobacco are produced by moving the entire nuclear genome from one species into the cytoplasmic milieu of another species, thereby disrupting the complementarity between the nucleus and the cytoplasm (Gerstel 1980).

In the majority of alloplasmics in *Nicotiana*, staminal male sterility accompanied by deformed or missing stamens is the most frequent form (Gerstel 1980). A few alloplasmics of *N. tabacum*, notably those involving some species of section Paniculatae (cms *N. knightiana*, cms *N. raimondii*, cms *N. paniculata*) but also *N. rustica* (Hart 1965), represent the postmeiotic type of CMS in which microspores are produced but are either aborted or develop into dysfunctional pollen grains.

Some alloplasmic lineages with postmeiotic CMS (cms *N. raimondii* and cms *N. knightiana*)

produce apparently normal pollen grains with vestigial germination ability that makes them partially self-fertile (Berbeć 1994a, 1994b, 2001).

In cms *N. glauca*, expression of the CMS trait may vary from staminal (absence of stamens) to postmeiotic: 1–3 stamens with normal anthers on shortened filament, or with stamens modified into secondary pistils or into secondary stigmas (Nikova and Vladova 2002).

8.11.2 Transfer of Alien Cytoplasm and Production of CMS Alloplasmics in Tobacco

The first transfer of alien male-sterile cytoplasm into the nuclear background of *N. tabacum* was performed by Clayton (1950). His method became the paradigm for the majority of subsequent CMS transfers and essentially mimicked the classic introgression scheme for chromosome-borne traits in that it began with the amphidiploid 4n *N. debneyi* × *N. tabacum* and progressed through the sesquidiploid and breakdown generations that segregated for male-sterile, partly fertile, and male-fertile plants, reflecting the segregation of compatible and incompatible factors from *N. tabacum* in hybrid nuclei. The CMS lines are perpetuated, and their genomic integrity is preserved by backcrossing them to their male fertile counterparts (isolines).

Apart from this standard method, there are two other approaches that were used to transfer alien cytoplasm into the genome of *N. tabacum*. The first involves direct transfer of cytoplasm through somatic fusion of protoplasts, of which one donates the cytoplasmic factors by having its nuclear genome inactivated by irradiation and the other supplies the nuclear genome (Kubo 1985; Kubo et al. 1988; Kumashiro et al. 1989). The second is a mixed approach that consists of producing a somatic fertile hybrid and transferring CMS factors by sexual backcrossing to the pollen parent (Chen et al. 2012).

The use of protoplast fusion offered some unique opportunities for CMS transfer. A potential benefit, unattainable in conventional sexual introgression, is recombination occurring between mitochondrial genomes from different species. By fusing different alloplasmics of tobacco, novel CMS phenotypes can be created and tested for their practical value (Kofer et al. 1990). Complementary recombination events can restore the disrupted compatibility between the nucleus and the cytoplasm of fused CMS forms with different sources of alien cytoplasm. Kofer et al. (1991) observed plants with restored stamen morphology and pollen production among regenerants from somatic hybridization between *N. tabacum* cms *N. undulata* and *N. tabacum* cms *N. quadrivalvis*.

Table 8.5 lists alloplasmic forms of tobacco that have been reported in the literature and the researchers who were the first to report them; there are typically no more than two reports per alloplasmic CMS combination. Of the 29 species that have been recorded to be involved in alloplasmics with *N. tabacum*, 18 belong to section Suaveolentes and the remaining 17 are dispersed across eight other sections. Twelve alloplasmics of *N. tabacum* with different *Nicotiana* species were recorded by Gerstel (1980). The number of alloplasmic combinations has since grown considerably, as has the proportion of those involving Suaveolentes species as CMS donors. Some CMS lineages were synthesized via somatic hybridization, but those account for a relatively small proportion of all alloplasmics involving *N. tabacum*, much as somatic hybrids account for only a small proportion of all interspecific combinations in the genus.

8.11.3 Side Effects Associated with the Introgression of Alien Cytoplasm in Tobacco ('Cytoplasmic Drag') and *Nicotiana* Species as Sources of CMS for Tobacco Improvement

Berbeć and Laskowska (2005) studied 14 alloplasmic CMS isolines and their male fertile cultivated flue-cured counterpart cultivar 'Wiślica'

Table 8.5 *Nicotiana* species reported as sources of cytoplasmic male sterility (cms) bred into *Nicotiana tabacum*

Section	Cytoplasmic male sterility introgressed from	cms type[a]	Reported by	Remarks
Alatae	*N. longiflora*	PM	Nikova et al. (2001) and Nikova and Vladova (2002)	
	N. plumbaginifolia	ST	Chaplin (1964)	
Nicotiana	*N. tabacum* (cms *N. tabacum* mutant)	ST	Berbeć (1967, 1974)	Spontaneous cms mutation discovered in the field of a normal male fertile *N. tabacum* variety. Morphologically close to cms *N. glauca* but differing from it in some aspects of growth and development (Berbeć and Berbeć 1976, Berbeć and Laskowska 2005)
	N. tabacum (*alata*)	?	Atanassov et al. (1998)	Obtained by asymmetric protoplast fusion (*N. tabacum* × *N. alata*), mix of mitochondrial DNA from both parents
		ST	Nikova et al. (1999)	cms found in the backcross progenies of the cross of maternal *N. tabacum* × *N. alata*
	N. tabacum (*glutinosa*)	ST	Burk (1960) and Chaplin (1964)	cms found in the backcross progenies of the cross of maternal *N. tabacum* × *N. glutinosa*; pistil protrudes above the shortened corolla (Chaplin 1964)
	N. tabacum (*plumbaginifolia*)	ST	Burk (1960) and Chaplin (1964)	cms found in the backcross progenies of the cross of maternal *N. tabacum* × *N. plumbaginifolia*; pistil protrudes above the shortened corolla (Chaplin 1964)
	N. tabacum (*glauca*)	ST	Stoyanova (1972)	cms found in the backcross progenies of the cross of maternal *N. tabacum* × *N. glauca*
Noctiflorae	*N. glauca*	ST/PM	Berbeć (1966) and Nikova and Vladova (2002)	Developed from sexual hybrid *N. glauca* × *N. tabacum*
			Chen et al. (2012)	Developed from somatic hybrid *N. glauca* × *N. tabacum*
Paniculatae	*N. knightiana*	PM	Berbeć and Doroszewska (1992), Nikova and Shabanov (1992), and Berbeć (1994a, b)	Alloplasmic forms developed by Berbeć (1994a, b) were partly fertile; those obtained by Nikova and Shabanov (1992) showed full male sterility
	N. paniculata	PM	Nikova et al. (1991) and Nikova and Vladova (2002)	Morphologically normal anthers void of pollen
	N. raimondii	PM	Berbeć and Doroszewska (1992) and Berbeć (1994a, b)	Lineages with aborted pollen and apparently normal pollen obtained (Berbeć 2001)
Polydicliae	*N. clevelandii*	Premeiotic staminal (?)	Kaul (1988)	Premeiotic breakdown of pollen mother cell development, feminization of male organs (Kaul 1988)
	N. quadrivalvis	ST	Chaplin (1959) and Burk (1960)	
Repandae	*N. repanda*	ST/occasionally PM	Burk (1967), Stavely et al. (1973), and Reed and Burns (1986)	
Rusticae	*N. rustica*	ST/PM	Hart (1965) and Nikova et al. (1997)	Partly male fertile *N. rustica* alloplasmics reported by Hart (1965); those reported by Nikova et al. (1997) were fully male-sterile

(continued)

Table 8.5 (continued)

Section	Cytoplasmic male sterility introgressed from	cms type[a]	Reported by	Remarks
Suaveolentes	*N. africana*	PM	Kumashiro et al. (1988), Nikova and Zagorska (1990), and Nikova and Vladova (2002)	Morphologically normal anthers void of pollen
	N. amplexicaulis	ST	Nikova and Shabanov (1988) and Berbeć et al. (1990)	Split corolla
	N. benthamiana	ST	Ramavarma et al. (1978), Nikova (1984), Nikova et al. (1991), and Nikova and Vladova (2002)	Nikova et al. (1991) transferred cms factors directly from *N. benthamiana*; Ramavarma et al. (1978) used *N. glutinosa* as a bridging species
	N. debneyi	ST	Clayton (1950) and Chaplin (1964)	
	N. excelsior	ST	Nikova (1986), Nikova and Vladova (2002), and Nikova et al. (2004)	
	N. eastii	ST	Berbeć and Berbeć (1992)	cms supposedly of androgenetic origin
	N. exigua	ST	Berbeć (1966, 1972)	
	N. goodspeedii	ST	Palakarcheva et al. (1980) and Tsikov et al. (1977)	
	N. gossei	?	Hanson and Conde (1985) and Tsikov and Tsikova (1986)	
	N. hesperis		Kubo (1985)	
	N. maritima	ST/PM	Nikova et al. (1991) and Nikova and Vladova (2002)	Both stamenless flowers and normally developed stamens with anthers void of pollen reported
	N. megalosiphon	ST	Clayton (1950) and Chaplin (1964)	
	N. occidentalis	ST	Ternovsky et al. (1973)	
	N. rotundifolia	?	Ilcheva et al. (2001)	Male-sterile alloplasmics obtained among asymmetric regenerants of protoplast fusion
	N. simulans	?	Kubo (1985)	
	N. suaveolens	ST	Izard and Hitier (1955) and Schweppenhauser and Mann (1968)	Fertility restorer factor introgressed into *N. tabacum* cms *N. suaveolens* (Schweppenhauser and Mann 1968)
	N. velutina	ST	Nikova et al. (1991) and Nikova and Vladova (2002)	Both stamenless flowers and normally developed stamens with anthers void of pollen reported
	N. wuttkei	ST	Laskowska and Berbeć (2007)	Stamenless flowers, shortened and split corolla, normal pistil
Undulatae	*N. undulata*	ST	Chaplin (1964)	Pistil protrudes above the shortened corolla (Chaplin 1964)
	N. glutinosa		Naumenko (2012)	cms *glutinosa* listed as distinct from cms *tabacum* (*glutinosa*), no further details except female fertility

[a]ST, staminal type (stamens absent or transformed into petaloid or stigmatoid structures); PM, postmeiotic (stamens normally developed; anthers morphologically normal with aborted or non-germinating pollen to shrunken with vestigial aborted pollen or void of pollen)

for several indices of agronomic performance, and found significant differences that could be attributable to the effect of alien cytoplasm. Among the studied alloplasmics, cms *N. bigelovii*, cms *N. tabacum*, and cms *N. suaveolens* were in no way inferior to their autoplasmic isoline (cultivar Wiślica) in terms of the selected growth parameters: plant height, days to flower, leaf number, leaf area combined with cured-leaf yield, quality, and economic return. cms *N. undulata* and cms *N. glauca* were also acceptable but slightly inferior to the autoplasmic variety in terms of yield indices and economic returns. Negative effects related to alien cytoplasm were particularly evident in some CMS isolines as reduced leaf size in cms *N. raimondii*, cms *N. megalosiphon*, and cms *N. occidentalis*, considerably shorter plants in cms *N. plumbaginfolia*, cms *N. tabacum* (*glutinosa*), and cms *N. eastii*, and delayed date of anthesis in cms *N. knightiana*, cms *N. raimondii*, cms *N. eastii*, cms *N. tabacum* (*glutinosa*), and cms *N. plumbaginifolia*. Save for the alloplasmics mentioned above as agronomically acceptable, the other cms analogues exhibited reductions in cured-leaf yield. These findings differ somewhat from those reported by Lawson et al. (2002) and Amankwa et al. (2014) for cms *N. glauca* and cms *N. suaveolens*; they found no important differences between the two alloplasmics in terms of agronomic suitability. The findings are also in disagreement with those for cms *N. goodspeedii* (Iancheva and Palakarcheva 1990): cms *N. goodspeedii* was found to exert no detrimental effects on seed production and agronomic performance of Virginia, burley, and oriental tobacco hybrids. Table 8.6 lists the CMS systems currently available and what is known about their agronomic potential based on findings by different researchers.

8.11.4 Currently Deployed CMS Systems in Commercial Hybrid Cultivars of Tobacco

Although information on the CMS source used in present-day hybrid cultivars is often lacking, the *N. suaveolens* system appears to be universally deployed because of its merits in yield and crop quality (Lawson et al. 2002; Amankwa et al. 2014). However, alternatives to cms *N. suaveolens* have been intensively sought because of difficulties in seed production, owing to depressed ovule fertility in some genotypes. In the study by Berbeć and Laskowska (2005) involving cv. 'Wiślica' as both the male and female parents, it did not pose much of a problem, although it can be a serious issue with some other genotypes, to the extent of making hybrid seed production impracticable (Berbeć, unpublished observations).

Another issue connected with cms *N. suaveolens* is its increased susceptibility to some diseases. Under natural field pressure from undetermined strains of PVY, cultivar 'Wiślica', a carrier of the *va* locus for PVY resistance, showed less than 1% of individuals with PVY symptoms (Berbeć and Laskowska 2005). Under the same conditions, 8% of cms *N. suaveolens* alloplasmics of 'Wiślica' developed PVY symptoms but none of cms *N. bigelovii*. In the majority of cases, the alloplasmic condition made cultivar 'Wiślica' prone to white leaf spots, mostly attributable to *Cercospora* sp. Only the plants of autoplasmic 'Wiślica' and four alloplasmics (cms *N. debneyi*, cms *N. tabacum* (*glutinosa*), cms *N. knightiana*, and cms *N. plumbaginifolia*) were free of those spots. In some of the alloplasmics, including cms *N. bigelovii* and cms *N. suaveolens*, the percentage of plants with white spots was fairly high (14% in cms *N. suaveolens* and 17% in cms *N. bigelovii*).

N. glauca has been studied for its potential as an alternative CMS source to replace the *N. suaveolens* system (Lawson et al. 2002; Amankwa et al. 2014; Zheng et al. 2018). The apparently closely related cms *N. tabacum* mutant system was formerly deployed in probably the world's first commercial flue-cured tobacco hybrids (Berbeć 1966). It is still used in Poland in the breeding of some flue-cured and burley varieties and is probably also deployed in Hungarian flue-cured hybrids. The cms *N. bigelovii* system is currently the major CMS source in

Table 8.6 Cytoplasmic male sterility (cms) systems in tobacco (*Nicotiana tabacum* L.) and their potential for tobacco improvement

Alloplasmic	Potential for tobacco improvement
cms *N. suaveolens*	See Table 8.7
cms *N. undulata*	Indicated as suitable for breeding and seed production (Chaplin and Ford 1965; Tsikov and Nikova 1980); no negative impact on plant growth and yield indices (Maktari 1991; Berbeć and Laskowska 2005); increased susceptibility to frog eye (Berbeć and Laskowska 2005); reduced seed yield (Kubo 1981)
cms *N. glauca* (see also cms *N. tabacum* mutant in Table 8.7)	No negative impact on plant growth and chemical composition (Maktari 1991; Chen et al. 2012), yield of hybrid seeds (Maktari 1991), or yield and quality of cured leaves (Lawson et al. 2002; Amankwa et al. 2014); slight yield depression compared with that of the isogenic autoplasmic counterpart (Berbeć and Laskowska 2005)
cms *N. benthamiana*	No significant impact on agronomically important traits versus its autoplasmic isogenic counterpart (Nikova et al. 2004)
cms *N. amplexicaulis*	Generally indicated as having potential for tobacco breeding and seed production (Nikova et al. 2004); no adverse effect on days to flower or leaf size (Czubacka et al. 2016)
cms *N. longiflora*, cms *N. paniculata*, cms *N. velutina*, cms *N. maritima*	Generally indicated as having a potential for tobacco breeding and seed production (Nikova et al. 2004)
cms *N. tabacum* (*plumbaginifolia*)	Small yield reduction versus autoplasmic isogenic counterpart reported by Chaplin and Ford (1965); indicated as suitable for tobacco breeding by Naumenko (2012)
cms *N. tabacum* (*glutinosa*)	Probably unsuitable for agronomic purposes: small yield reduction versus autoplasmic isogenic counterpart (Chaplin and Ford 1965); reduced female fertility and substantially reduced internode length, plant height, leaf size, cured-leaf yield, and economic return versus autoplasmic isogenic counterpart (Berbeć and Laskowska 2005)
cms *N. megalosiphon*	Depressed seed yield (Chaplin 1964; Maktari 1991), delayed growth rate, and lowered leaf number (Chaplin and Ford 1965); decreased leaf size and leaf number (Biskup et al. 1972); conflicting reports on yield and crop quality: no negative effect (Chaplin and Ford 1965); both yield and crop quality negatively affected (Berbeć and Laskowska 2005)
cms *N. rustica*	Unusable because of the occurrence of lineages with vestigial male fertility (Hart 1965; Kubo 1981, 1985); no significant impact on agronomically important traits versus the autoplasmic isogenic counterpart (Nikova et al. 2004)
cms *N. knightiana*, cms *N. raimondii*	Unusable because of the occurrence of lineages with vestigial male fertility (Berbeć and Laskowska 2005); significant growth retardation and leaf yield reduction versus the isogenic autoplasmic counterpart (Berbeć and Laskowska 2005; Czubacka et al. 2016)
cms *N. bigelovii*/*N. quadrivalvis* (see also cms *N. bigelovii* in Table 8.7)	Unusable because of significant negative impact on cured-leaf yield and quality (Chaplin and Ford 1965); indicated as usable by Tsikov and Nikova (1980) and Maktari (1991). The best performing entry is in the study of isogenomic alloplasmics by Berbeć and Laskowska (2005) but the particular accession is tentatively treated in this review as a possible variant of cms *N. suaveolens* (cms *N. bigelovii*) on account of its flower morphology

(continued)

Table 8.6 (continued)

Alloplasmic	Potential for tobacco improvement
cms *N. plumbaginifolia*	Unusable because of seed yield reduction (Chaplin 1964), reduced number of leaves and leaf size, delayed flowering (Czubacka et al. 2016), and negative impact on leaf yield (Chaplin and Ford 1965; Berbeć and Laskowska 2005); increased susceptibility to blue mold (Berbeć 2001); indicated as suitable for seed production by Maktari (1991) and Naumenko (2012)
cms *N. africana*, cms *N. excelsior*	Unusable because of heavy depression of growth, development, and leaf yield indices versus the autoplasmic isogenic counterpart (Nikova et al. 2004)
cms *N. debneyi*	In most studied cases, unusable because of depressed vegetative vigor, reduced seed yield (Maktari 1991), delayed flowering, and delayed leaf maturity (Dudek 1971; Kubo 1985; Maktari 1991); negative impact on cured-leaf yield, quality, and economic return to growers (Berbeć and Laskowska 2005)
cms *N. eastii*	Depressed cured-leaf yield and economic return versus the autoplasmic isogenic counterpart (Bezova and Skula 1980); lowered plant height and leaf size (Naumenko 2012); increased susceptibility to blue mold (Berbeć 2001)
cms *N. exigua*	Unusable because of depressed seed yield and delayed growth rate and flowering (Maktari 1991), reduced cured-leaf yield and quality versus the autoplasmic isogenic counterpart (Berbeć and Laskowska 2005), and reduced plant height and heavily depressed seed yield (Naumenko 2012)
cms *N. goodspeedii*	Unusable because of depressed seed yield (Maktari 1991), delayed growth rate, delayed flowering, and reduced cured-leaf yield and quality versus the autoplasmic isogenic counterpart (Berbeć and Laskowska 2005); reported as suitable for commercial hybrids by Iancheva and Palakarcheva (1990)
cms *N. gossei*	Unusable because of depressed female fertility (Tsikov and Tsikova 1986)
cms *N. occidentalis*	Unusable because of substantial reduction in most indices of agronomic performance such as seed yield, growth rate, and cured-leaf yield (Maktari 1991; Berbeć and Laskowska 2005; Naumenko 2012)
cms *N. wuttkei*	Unusable because of a substantial negative impact on plant height, leaf number, leaf size, and days to flower (Laskowska and Berbeć 2007)
cms *N. clevelandii*, cms *N. hesperis*, cms *N. repanda*, cms *N. simulans*, cms *N. rotundifolia, N. tabacum (alata), N. tabacum (glauca), N. glutinosa*	No data on agronomic usability, except unimpaired seed yield in cms *N. glutinosa* (Naumenko 2012)

Polish flue-cured hybrids (Berbeć 2007, 2017), and cms *N. bigelovii* is reportedly also used in Cuba (Lopez et al. 2008). The advantages and disadvantages of three major CMS systems used in commercialized cultivars are compiled in Table 8.7.

Table 8.7 Primary sources of cytoplasmic male sterility (cms) systems deployed in the development and seed production of tobacco hybrid cultivars

Alloplasmic	Advantages and defects
cms *N. suaveolens*	Universally recognized as agronomically suitable and the most extensively used system in hybrid breeding and seed production (Amankwa et al. 2014; Haji et al. 2000, 2007; Kubo 1981; Lawson et al. 2002); generally no negative impact on agronomic performance (Chaplin and Ford 1965; Kubo 1981,1985, 2005); increased susceptibility to potato virus Y and to frog eye versus alloplasmic isogenic counterparts (Berbeć and Laskowska 2005); heavily depressed seed yield (Naumenko 2012) and reduced female fertility in combination with some nuclear genotypes caused by malformations of entire carpels (Berbeć and Doroszewska, unpublished observations)
cms *N. bigelovii*	Most likely, a distinct lineage of cms *N. suaveolens*. Chief morphological differences: corolla slightly elongated in cms *N. bigelovii* and slightly shortened in cms *N. suaveolens*; alterations of the pistil (shortened or fasciated style with multiple stigma lobes, malformed ovary) more frequent in cms *N. suaveolens* than in cms *N.* bigelovii (Berbeć 2001, and Berbeć's unpublished observations). No depressive effect on growth, yield quality, or economic return. Unlike cms *N. suaveolens*, cms *N. 'bigelovii'* showed no impaired resistance to potato virus Y infection but was more prone to develop frog-eye leaf spots versus its autoplasmic counterpart and cms *N. suaveolens* (Berbeć and Laskowska 2005). Unlike in some cms *N. suaveolens* genotypes, impaired female fertility is apparently not an issue in cms *N. 'bigelovii'* (Berbeć, unpublished observations). At present, the most frequently used type of cms in Polish-bred hybrid cultivars, chiefly Virginia-type tobacco
cms *N tabacum* mutant	Possibly a distinct lineage of cms *N. glauca*. No adverse effect on cured-leaf yield and quality versus its autoplasmic male fertile counterpart. Outyielded cms *N. glauca* (Berbeć and Laskowska 2005). Both cms *N. glauca* and cms *N. tabacum* mutant are reliable seed producers. Indicated as suitable for hybrid development and seed production (Berbeć 1975; Tsikov and Nikova 1980; Maktari 1991; Naumenko 2012). Reduced number of leaves, hastened flowering, and reduced leaf size reported by Czubacka et al. (2016). Used for breeding and commercial purposes in Poland since the late 1960s (Berbeć 1967, 1974) and probably deployed in other countries as well (Berbeć 2017).

8.12 Concluding Notes on Interspecific Hybridization in Tobacco Improvement

The so-called heyday of interspecific cytogenetics and gene introgression in *Nicotiana* occurred between 1930 and 1979. It is in those years that most *Nicotiana* hybrids were produced and studied, the most significant books on *Nicotiana* were written (Kostoff 1943; Goodspeed 1954), and the fundamental rules of interspecific gene transfer and introgression were elucidated. It was also in those years that some of the most momentous gene transfers in the history of interspecific introgression in *Nicotiana* were accomplished, such as *N* gene from *N. glutinosa* conferring resistance to TMV, *Phl* and *Php* genes from *N. longiflora* and *N. plumbaginifolia*, respectively, for resistance to black shank, resistance to wildfire, also from *N. longiflora*, and resistance to blue mold and black root rot from *N. debneyi*. Shifts to more virulent races made some of these resistances obsolete. The resistances to TMV from *N. glutinosa* and to black root rot from *N. debneyi* have remained unchallenged for decades. What these transfers have in common is that most of them were accomplished by simple (though not crude) means, before the advent of sophisticated biotechnological and molecular methods.

This is not to say that introgressive breeding in tobacco stopped altogether after that early fruitful period. Significant advances were made toward the end of the twentieth century and are still being made. Resistance factors from *N. rustica* that combined effectiveness against the whole range of important races of wildfire, angular leaf spot, and black shank without the penalty of undesirable linkages were bred into

tobacco cultivars (Woodend and Mudzengerere 1992; Drake et al. 2015). Resources of PVY-resistant germplasm were enriched and complemented by the tolerance response from *N. africana*, covering the entire range of PVY isolates, including those against which the *va* alleles had become ineffective (Lewis 2007, Doroszewska 2010, Korbecka-Glinka et al. 2017). Research is in progress to leverage the resistance to TSWV from *N. alata* by minimizing the adverse linkage (Laskowska and Berbeć 2010; Trojak-Goluch et al. 2011, 2016a, b).

We may not even be fully aware of how and to what extent *Nicotiana* species have contributed to the genotypes of present-day tobacco cultivated worldwide. Wild *Nicotiana*, especially the most popular species, were frequently part of in-house germplasm collections of tobacco breeding units. Hints dispersed in both the literature and personal accounts suggest that many breeders worked with wild species. Only a few were expert geneticists or academicians contributing to scientific journals. Unfamiliar with the intricacies of interspecific introgression, they might have treated the process as they would normal intervarietal crosses. Quite a few breeding lines and local cultivars may have been developed as a result, and the alien germplasm dispersed in the pedigrees of new cultivars. The impact of what may be called fortuitous introgression that has passed unnoticed and unlabeled is difficult to evaluate. The molecular tools to do this are there and have been used with success to trace the events from the remote past of wild *Nicotiana* (see Sect. 8.5). They have also helped to clarify some introgression events in the pedigrees of tobacco cultivars (Johnson et al. 2002b; Milla et al. 2005). Much work is yet to be done, however, to fully assess the impact of *Nicotiana* species in tobacco improvement. The poorly advanced genetic mapping of *Nicotiana* species (Lewis 2011) makes such an assessment difficult at present.

Both state-of-the-art biotechnology and genetic engineering methods have apparently made a moderate contribution to the genetic makeup of present-day tobacco cultivars, including introgression from wild *Nicotiana* species. It is noteworthy only two genes (N gene from N. glutinosa and nad1 gene from N. alata) are the only resistance-related genes from any *Nicotiana* known to be cloned (Whitham et al. 1996; Lewis 2011; Pereg 2013). The transfer of tobacco breeding from the public sector to private enterprises with their confidentiality policies may account in part for that dearth of information on advances in *Nicotiana* breeding techniques. There are various other reasons accounting for the decline—perhaps irreversible—in interest in *Nicotiana* species and interspecific introgression. These reasons have been covered in detail by Lewis (2011) in his review and need not be repeated here.

References

Ahuja MR (1962) A cytogenetic study of heritable tumors in *Nicotiana* species hybrids. Genetics 47:865–880

Amankwa GA, Mishra S, Shearer AD, Van Hooren DL (2014) Evaluation of two flue-cured tobacco F1 hybrids with different sources of male sterile cytoplasm. CORESTA Congress Quebec, Canada APPOST02

Apple JL (1962) Transfer of resistance to black shank (Phytophthora parasitica var. Nicotianae) from Nicotiana. Phytopathology 52:1

Apple JL (1967) Occurrence of race 1 *Phytophthora parasitic*a var. *nicotianae* in North Carolina and its implications. Tob Sci 11:79–83

Atanassov A, Dimantov D, Atanassov I et al (1991) Transfer of resistance to tomato spotted wilt virus/TSWV/from wild *Nicotiana* species to *N. tabacum* via somatic hybridisation. Physiol Plant 82: A23

Atanassov II, Atanassova SA, Dragoeva AI, Atanassov AI (1998) A new CMS source in *Nicotiana* developed via somatic cybridization between *N. tabacum* and *N. alata*. Theor Appl Genet 97:982–985

Bai DP, Reeleder R, Brandle JE (1996) Production and characterization of tobacco addition lines carrying *Nicotiana debneyi* chromosomes with a gene for resistance to black root rot. Crop Sci 36:852–857

Bailov D, Palakarcheva M, Daskalov S (1964) Novi amfidiplodi *Nicotiana tabacum* × *N. debneyi* Dom. Rastienievod Nauki 7:3–16

Bates GW (1990) Asymmetric hybridization between *Nicotiana tabacum* and *N. repanda* by donor recipient protoplast fusion: transfer of TMV resistance. Theor Appl Genet 80:481–487

Berbeć A (1987) Cytogenetical study on *Nicotiana tabacum* L. cv. Nadwiślański Mały (2× and 4×) × *Nicotiana alata* Link et Otto hybrids. Genet Pol 28:251–261

Berbeć A (1994a) Effect of *Nicotiana knightiana* cytoplasm on the stability of a true-breeding tobacco variety increased by selfing. CORESTA Congress Inf Bull Harare: 61

Berbeć A (1994b) Microsporogenesis in two alloplasmic isonuclear strains of tobacco *Nicotiana tabacum* L. cv. Zamojska 4 with the cytoplasms of *N. knightiana* Goodspeed and *N. raimondii* Macbride. Genet Pol 35:23–31

Berbeć A (2001) Floral morphology and some other characteristics of iso-genomic alloplasmics of *Nicotiana tabacum* L. Beitrage zur Tabakforschung International 19:309–314

Berbeć A (2017) Three-way crosses vs. single crosses in tobacco: first agronomic assessment. Crop Sci 57:1363–1372. https://doi.org/10.2135/cropsci2016.07.0624

Berbeć A, Berbeć J, Doroszewska T (1990) Graded effects of four alien cytoplasms in *Nicotiana tabacum* L. Bul Spec CORESTA, 1990, Symposium Kallithea: 143 A34

Berbeć A, Doroszewska T (1981) Investigations of the interspecific hybrid Nicotiana amplexicaulis Burbidge × *Nicotiana tabacum*. Genet Pol 22:197–207

Berbeć A, Doroszewska T (1992) Alloplazmatyczne formy tytoniu uprawnego (*Nicotiana tabacum* L.) z podstawioną cytoplazmą gatunków *Nicotiana amplexicaulis*, *N. knightiana* i *N. raimondii*. Pam Puł 100:141–150

Berbeć A, Głażewska Z (1988) Transfer of resistance to Potato Virus Y from Nicotiana benavidesii Goodspeed to *N. tabacum* L. Genet Pol 29:323–333

Berbeć A, Laskowska D (1997) Compatibility and tentoxin sensitivity in the off-spring of near-amphihaploid hybrid *Nicotiana tabacum* × *N. alata*. J Appl Genet 38:143–150

Berbeć A, Laskowska D (2005) Investigations of isogenomic alloplasmics of flue-cured tobacco *Nicotiana tabacum* cv. Wislica. Beitrage zur Tabakforschung Int 21:258–263

Berbeć J (1966) Investigations on the cross *Nicotiana glauca* × *N. tabacum*. Badania nad krzyżówką *Nicotiana glauca* × *N. tabacum* (in Polish) Inf PAN 7–213:340

Berbeć J (1967) Cytoplasmic male sterile variety of cultivated tobacco (*Nicotiana tabacum* L.) Cytoplazmatycznie męskosterylna odmiana tytoniu uprawnego (Nicotiana tabacum L.) (in Polish) Inf. PAN, I 7–195:299–300

Berbeć J (1972) Investigations into male sterility of tobacco and its utilisation in hybrid seed production (in Russian). Intl Symp Mutual Economic Assistance Council (RWPG), Warsaw, 9–13 May 1972, pp 1–5

Berbeć J (1974) A cytoplasmic male sterile mutation form of *Nicotiana tabacum* L. Zeitschrift fur Pflanzenzuchtung 73:204–216

Berbeć J (1975) Investigations on the utilization of the cytoplasmically inherited male sterile strain of tobacco in the F1 hybrid seed production. Z Pflanzenziicht 74:28–38

Berbeć J, Berbeć A (1976) Growth and development of *Nicotiana tabacum* L. form reconstituted on the cytoplasm of *Nicotiana glauca* Grah. Genet Pol 17:309–318

Berbeć J, Berbeć A (1992) Męska jałowość u tytoniu (*Nicotiana tabacum* L.) uzyskana drogą jednoetapowego podstawienia cytoplazmy gatunku *Nicotiana eastii* Kostoff. Pam Puł 100:136–139

Berbeć J, Berbeć A, Glazewska Z, Opoka B, Kobus I, Czop T (1982) Investigations on obtaining new varieties and breeding lines of *Nicotiana tabacum* resistant to virus leaf rib necrosis (Potato Virus Y) and virus degeneration of plants by Lycopersicum virus 3. USDA Grant no. FG-Po-341 (JB-6) Final Report, Institute of Soils Science and Plant Cultivation, Pulawy, Poland (unpublished)

Bezova K, Skula K (1980) Fertile and sterile forms of VP9. Bul Tabak Priem 23:19–31

Biskup J, Mazur M, Roman T (1972) Ocena wzrostu i rozwoju oraz niektorych cech uzytkowych odmian i mieszancow plodnych w porownaniu z analogicznymi formami meskojalowymi. Biul Centralnego Laboratorium Przemysłu Tytoniowego (3-4):29–40

Bombarely A, Edwards KD, Sanchez-Tamburrino J, Mueller LA (2012) Deciphering the complex leaf transcriptome of the allotetraploid species *Nicotiana tabacum*: a phylogenomic perspective. Genomics 13:406

Bomblies K (2009) Too much of a good thing? Hybrid necrosis as a by-product of plant immune system diversification. Botany 87:1013–1022

Brandle JE, Court WA, Gleddie S (1992) The application of somatic hybridization to a tobacco breeding program. Tob Chem Res Conf 46:30, abstr. 17

Brieger F (1928) Uber die Vermehrung der Chromosomenzahl bei dem Bastard *Nicotiana tabacum* L. × Rusbyi Britt. Zeitschrift Inductive Abstam u Vererbungsl 47:1–53

Bukuta L (2002) Determination of allelism between different sources of Black Shank race 0 (Phytophthora Nicotianae) resistance and Linkage studies between Black Shank race 0 and Angular Leaf Spot race 1 (Pseudomonas syringae pv. tabaci Tox-) resistance in tobacco (*Nicotiana tabacum*). CORESTA Study Grant Report, CORESTA Congress New Orleans

Burbidge NT (1960) The Australian species of Nicotiana. Aust J Bot 8:342–380

Burk LG (1960) Male-sterile flower anomalies in interspecific tobacco hybrids. J Hered:27–31

Burk LG (1967) An interspecific bridge-cross—*Nicotiana repanda* through *N. sylvestris* to *N. tabacum*. J Hered 58:215–218

Burk LG, Durbin RD (1978) Reaction of Nicotiana species to tentoxin. J Hered 69:117–120

Burk LG, Chaplin JF (1979) Hybridization (in) Nicotiana. Procedures for experimental use (ed. RD Durbin) USDA Tech Bul 1586:23–27

Burk LG, Heggestad HE (1966) The genus Nicotiana as a source of resistance to diseases of cultivated tobacco. Econ Bot 20:76–88

Burk LG, Neas MO (1966) 4n (*N. tabacum* × *N. nudicaulis*, a colchicine induced fertile hybrid. Tob Sci 8:65–66

Burk LG, Dean CE (1975) Hybrid fertility and aphid resistance in the cross *Nicotiana tabacum* × *N. gossei*. Euphytica 24:59–63

Burk LG, Gerstel DU, Wernsman EA (1979) Maternal haploids of *Nicotiana tabacum* L. Science 206:586

Burk LG, Gooding GV, Chaplin JF (1982) Reaction of *Nicotiana* species and cultivars or breeding lines of *Nicotiana tabacum* to three strains of Potato Virus Y. Tob Sci 26:85–88

Busconi M, Reggi S, Lorenzoni C, Fogher C (2010) Interspecific crosses in the genus *Nicotiana*. In: Proceedings of the 54th Italian Society of Agricultural Genetics Annu Congress, Post 7.02

Butenko RG, Luneva MZ (1966) Application of the sterile culture technique for the culturing of distant hybrids of *Nicotiana*. Применение метода стерильных культур для выращивания отдаленных гибридов Nicotiana (in Russian) Fizjologija Rastenij 13:733–736

Cameron DR (1958) Alien substitution of a locus affecting immunity to black shank in *Nicotiana tabacum*. Proc X Int Congr Genet Univ Toronto 2:41

Campbell KG, Wernsmann EA, Fitzmaurice WP, Burns JA (1994) Construction of a designer chromosome in tobacco. Theor Appl Genet 87:837–842

Carlson PS, Smith HH, Dearing RD (1972) Parasexual interspecific plant hybridization. Proc Natl Acad Sci 69:2292–2294

Chaplin JF (1959) Interspecific hybridization, gene transfer and chromosomal substitution in Nicotiana. PhD thesis, North Carolina State University, Raleigh

Chaplin JF (1962) Transfer of black shank resistance from *Nicotiana plumbaginifolia* to flue-cured *N. tabacum*. Tob Sci 6:184–189

Chaplin JF (1964) Use of male sterile tobaccos in the production of hybrid seed. Tob Sci 8:105–109

Chaplin JF, Burk LG (1971) Interspecific hybridization and gene transfer in Nicotiana: problems and possible solutions. Fifth Int Tob Sci Congr, Hamburg, Germany Proc. 59:67

Chaplin JF, Ford ZT (1965) Agronomic and chemical characteristics of male-sterile flue-cured tobacco as influenced by cytoplasms of different *Nicotiana* species. Crop Sci 5:436–438

Chaplin JF, Mann TJ (1961) Interspecific hybridization, gene transfer and chromosomal substitution in *Nicotiana*. North Carolina Agric Exp Sta Tech Bull 128

Chaplin JF, Mann TJ (1978) Evaluation of tobacco mosaic resistance factor transferred from burley to flue-cured tobacco. J Hered 69:175–178

Chaplin JF, Mann TJ, Apple JL (1961) Some effects of the *Nicotiana glutinosa* type of mosaic resistance on agronomic characters of flue-cured tobacco. Tob Sci 5:80–83

Chaplin JF, Matzinger DF, Mann TJ (1966) Influence of the homozygous and heterozygous mosaic-resistance factor on quantitative character of flue-cured tobacco. Tob Sci 10:81–84

Chase MW, Knapp S, Cox AV et al (2003) Molecular systematics, GISH and the origin of hybrid taxa in *Nicotiana* (Solanaceae). Ann Bot 92:107–127

Chen X, Luo J, Li W, Xiao B, Li Y (2012) Flow cytometry assisted selection on a CMS source (Nta (gla.)S) of flue-cured tobacco via somatic hybridisation. CORESTA Congress, Sapporo, 2012 AP 06

Chimoyo HM, Pupert EA (1988) The effect of UV irradiation, toluidine blue, and environment on maternal haploid frequencies from the cross between *Nicotiana tabacum* and *N. africana*. Bull Spec

Christoff M (1928) Cytological studies in the genus *Nicotiana*. Genetics 13:233–277

Chung SC, Yoshida KT, Takeda G (1996) Reproductive characteristics of amphidiploids derived from an interspecific hybrid between *Nicotiana trigonophylla* Dun. and *N. tabacum* L. Jpn J Breed 46:29–33

Clarkson and Symon (1991) Nicotiana wuttkei (Solanaceae), a new species from north-eastern Queensland with an unusual chromosome number. Austrobaileya 3(3): 389-392

Clausen RE, Cameron DR (1944) Inheritance in *Nicotiana tabacum*. XVIII. Monosomic analysis. Genetics 29:447–477

Clausen RE, Cameron DR (1957) Inheritance in *Nicotiana tabacum*. XXVIII. The cytogenetics of introgression. Proc Nat Acad Sci USA 43:908–913

Clausen RE, Goodspeed TH (1925) Interspecific hybridization in *Nicotiana*. II. A tetraploid *glutinosa-tabacum* hybrid, an experimental verification of Winge's hypothesis. Genetics 10:278–284

Clausen RE, Mann MC (1924) Inheritance in *Nicotiana tabacum*. V. The occurrence of haploid plants in interspecific progenies. Proc Nat Acad Sci 1O:121–124

Clayton EE (1947) A wildfire resistant tobacco. J Hered 38:35–40

Clayton EE (1950) Male sterile tobacco. J Hered 4:171–175

Clayton EE (1958) The genetics and breeding progress in tobacco during the last 50 years. Agron J 50:352–356

Clayton EE (1968) The transfer of blue mold resistance to tobacco from *Nicotiana debneyi*. Part IV. Breeding progress 1957–1967. Tob Sci 12:112–124

Clayton EE (1969) The study of resistance to the black root disease of tobacco. Tob Sci 13:30–37

Clayton EE, Heggestad HE, Grosso JJ, Burk LG (1967) The transfer of blue mold resistance to tobacco from *Nicotiana debneyi*. Part I. Breeding Progress 1937–1954. Tob Sci 11:91–99

Collins GB, Legg PD (1969) Cytogenetics of a Burley tobacco breeding line derived from a cross with *Nicotiana longiflora* Cav. Can J Genet Cytol 11(2): 382–388

Corbaz R (1962) Research on the control of blue mould of tobacco. Bull Inf. CORESTA:9–19, CORESTA Congress Guangzhou:71 AP-2

Crowder BJ, Wilkinson CA, Johnson CS, Eisenbach JD (2003) Inheritance of resistance to tobacco cyst nematode in flue-cured tobacco. Crop Sci 43:1305–1312

Czubacka A, Depta A, Doroszewska T (2016) Agronomic performance of cytoplasmic male sterile forms of flue-cured tobacco. CORESTA Congress, Berlin, APPOST, p 08

D'Arcy WG (1976) New names and taxa in Solanaceae. Ann Mo Bot Gard 63: 363–369

Desprez B, Chupeau MK., Vermeulen A et al (1992) A Nicotiana gametosomatic hybrid and its progenies. J Exp Bot 43(248):419–425

DeVerna JW (1984) In vitro-facilitated wide hybridization in *Nicotiana*. PhD Dissertation, University of Kentucky, Lexington

DeVerna JW, Myers JR, Collins GB (1987) Bypassing prefertilization barriers to hybridization in *Nicotiana* using in vitro pollination and fertilization. Theor Appl Genet 73:665–671

Dodsworth SA (2015) Genome skimming for phylogenomics. PhD thesis, School of Biological and Chemical Sciences, Queen Mary University of London

Donaldson PA, Bevis E, Pandeya R, Gleddie SC (1995) Rare symmetric and asymmetric *Nicotiana tabacum* (+) *N. megalosiphon* somatic hybrids recovered by selection for nuclear-encoded resistance genes and in the absence of genome inactivation. Theor Appl Genet 91:747–755

Dorossiev L, Palakarcheva M, Jancheva A (1990) Application of *in vitro* methods in the development of disease resistant Oriental tobacco hybrids and lines. Genet Breed 23:306–315

Dorossiev L, Palakarcheva M, Stanoyeva M Petkova M (1978) Overcoming the sterility in F1 of interspecific hybrids of the genus *Nicotiana* using the methods of tissue culture. Bul Spéc CORESTA, Symposium Sofia: 80–81 (P03)

Doroszewska T (2010) Transfer of tolerance to different Potato virus Y (PVY) isolates from *Nicotiana africana* Merxm. to *Nicotiana tabacum* L. Plant Breeding 129:76–81

Doroszewska T, Berbeć A (1990) Investigations of the hybrids of *Nicotiana tabacum* L. × *Nicotiana africana* Merxm. Bul. Spec. CORESTA Symposium Kallithea, Greece:176, P19

Doroszewska T, Berbeć A (1999) Transfer of resistance to Potato Virus Y to tobacco cultivars. CORESTA Meet. Agro-Phyto Groups, Suzhou, China AP4

Doroszewska T, Berbeć A (2000) Cytogenetical investigations of poliploid interspecific hybrids of *Nicotiana africana* with different cultivars of *N. tabacum*. J Genet Breed 54:77–82

Doroszewska T, Krasnodębska-Depta A, Czubacka A (2009) Album of *Nicotiana* species. Puławy Instytut Uprawy Nawożenia i Gleboznawstwa—Państwowy Instytut Badawczy

Doroszewska T, Przybyś M (2007) Characterization of *Nicotiana* species resistance to black root rot (*Thielaviopsis basicola* (Berk. and Broome) Ferr). Charakterystyka odporności gatunków *Nicotiana* na czarną zgniliznę korzeni (*Thielaviopsis basicola* (Berk. and Broome) Ferr. Zeszyty Problemowe Nauk Rolniczych 517:253–266

Drake K, Lewis RS (2013) An Introgressed *Nicotiana rustica* genomic region confers resistance to Phytophthora *Nicotianae* in Cultivated Tobacco. Crop Sci 53:1366–1374

Drake KE, Moore JM, Bertrand P, Fortnum B, Peterson P, Lewis RS (2015) Black shank resistance and agronomic performance of flue-cured tobacco lines and hybrids carrying the introgressed *Nicotiana rustica* region, Wz. Crop Sci 55:79–86

Dudek M (1971) Effect of the *debneyi* i *megalosiphon* sterile cytoplasms on some traits of the hybrids in backcrosses to several tobacco varieties (in Polish). Biul CLPT 1–2:29–44

Durbin RD, Uchytil TF (1977) A survey of plant insensitivity to tentoxin. Phytopathology 67:602–603

East EM (1928) The genetics of the genus *Nicotiana*. Bibliograph Genet 4:243–318

East EM (1930) The origin of the plants of maternal type which occur in connection with interspecific hybridizations. Proc Natl Acad Sci USA 16(6):377–380

Eghis SA (1927) Experiments on interspecific hybridation in the genus *Nicotiana* (in Russian with English summary). Bull Appl Bot Leningrad 17:151–190

Eghis SA (1930) Experiments on interspecific hybridization in the genus *Nicotiana*. II. The fertile hybrids between *N. tabacum* L. and *N. sylvestris* Speg. and Games. Proc USSR Cong Genet Plant & Animal Breed 2

Eickholt DP, Lewis RS (2014) Effect of an introgressed *Nicotiana tomentosa* leaf number QTL on yield and quality characteristics in flue-Cured tobacco. Crop Sci 54:1–9. https://doi.org/10.2135/cropsci2013.07.0464

Fardy A, Hitier H (1945) Espèces tétraploides et hybrides interspécifiques amphidiploides et triples dipoloides ed *Nicotiana*, obtenus par l'action de la colchicine. Mémorial du Serv d'Expl Industr des Tabacs et des Allumettes (SEITA) Ser B: 1–117

Fuentes I, Stegemann S, Golczyk H, Karcher D, Bock R (2014) Horizontal genome transfer as an asexual path to the formation of new species. Nature 511:232–235

Gajos Z (1979) Attempt to use hybrids of *Nicotiana tabacum* L. × *N. otophora* Gris for breeding tobacco resistant to Peronospora tabacina Adam (PT-2) and other diseases. Próby wykorzystania mieszańców *Nicotiana tabacum* L. × *Nicotiana otophora* Gris. w hodowli tytoniu odpornego na Peronospora tabacina Adam (P-2) i inne choroby (in Polish) Biul Inf Centr Lab Przem Tyton (1–2):11–23

Gajos Z (1981) Transfer of resistance to tomato spotted wilt virus from *N. Nicotiana alata* Link et Otto to tobacco. Przeniesienie odporności na wirus brązowej plamistości pomidora (Tomatto Spotted Wilt Virus) z Nicotiana alata Link. et Otto. do tytoniu szlachetnego przez skrzyżowanie obu gatunków (in Polish) Biul Inf Centr Lab Przem Tyton (1–2):3–24

Gajos Z (1984) Genetic resistance to black root rot (Thielaviopsis basicola ferr.) of the interspecific hybrids of *Nicotiana tabacum* L. × *Nicotiana alata* link. et Otto. Genetyczna odporność na czarną zgniliznę korzeni (*Thielaviopsis basicola* Ferr) roślin międzygatunkowego mieszańca *Nicotiana tabacum* L. × *Nicotiana alata* Link. et Otto (in Polish). Biul Inf Centr Lab Przem Tyton 1–4:3–14

Gentscheff G (1931) Investigations on interspecific hybridisation in the genus *Nicotiana*. Проучвания врьху междувидовата хибридизация в рода *Nicotiana*. Zemedlska Misl 2:91–111 (in Bulgarian)

Gerstel DU (1943) Inheritance in *Nicotiana tabacum*. XVII. Cytogenetical analysis of *glutinosa*-type resistance to mosaic disease. Genetics 28:533–556

Gerstel DU (1945) Inheritance in *Nicotiana tabacum*. XIX. Identification of the *tabacum* chromosome replaced by one from *N. glutinosa* in mosaic-resistant Holmes Samsoun tobacco. Genetics 30:448–454

Gerstel DU (1946) Inheritance in *Nicotiana tabacum* XXI. The mechanism of chromosome substitution. Genetics 31:421–427

Gerstel DU (1961) Chromosomal control of preferential pairing in *Nicotiana*. Science 133:579–580

Gerstel DU, Burk LG (1960) Controlled introgression in *Nicotiana*. A cytological study. Tob Sci 4:147–150

Gerstel DU (1980) Cytoplasmic male sterility in *Nicotiana* (A review). North Carolina Agric Res Service Tech Bull 263

Gerstel DU, Burns JA, Burk LG (1978) Cytoplasmic male sterility in *Nicotiana*, restoration of fertility and the nucleolus. Genetics 89:157–169

Gerstel DU, Burns JA, Burk LG (1979) Interspecific hybridizations with an African tobacco, *Nicotiana africana* Merxm. J Hered 70:342–344

Giddings GD, Rees H (1992) A Nicotiana gametosomatic hybrid and its progenies. A *Nicotiana* gametosomatic hybrid and its progenies. J Exp Bot 43:419–425

Gillham FEM, Wark DC, Harrigan EKS (1977) Disease resistant flue-cured tobacco breeding lines for north Queensland. I. Resistance to blue mould, Peronospora tabacina. Aust J Exp Agric Anim Husb 17:652–658

Goodspeed T, Bradley M (1942) Amphidiploidy. Bot Rev 8:272–316

Goodspeed TH (1915) Parthenogenesis, parthenocarpy and phenospermy in *Nicotiana*. Calif Univ Publ Bot 5 (8):249–272

Goodspeed TH (1945) Cytotaxonomy of *Nicotiana*. Bot Rev 11:533–592

Goodspeed TH (1954) The genus *Nicotiana*. Chronica Botanica Co, Waltham, Mass

Goodspeed TH, Clausen RE (1917) Mendelian factor differences versus reaction system contrasts in heredity. Amer Nat 51:31–46

Goodspeed TH, Clausen RE (1928) Interspecific hybridization in *Nicotiana*. VIII. The *sylvestris-tomentosa-tabacum* triangle and its bearing on the origin of tobacco. Univ Calif Pub Bot 11:127–140

Greenleaf WH (1941) Sterile amphidiploids: Their possible relation to the origin of *Nicotiana tabacum*. Am Nat 75:394–399

Gwynn GR, Barker KR, Reilly JJ, Komm DA (1986) Genetic resistance to tobacco mosaic virus, cyst nematodes, root-knot nematodes, and wildfire from Nicotiana repanda incorporated into *Nicotiana tabacum*. Plant Dis 70:958–962

Haji HM, Mishra S, DeVos M (2000) CT681 flue-cured tobacco. Can J Plant Sci 80:167–168

Haji HM, Brammall RA, VanHooren DL (2003) Effect of the black root rot resistance gene on the yield and quality characteristics of flue-cured tobacco in Ontario. Can J Plant Sci 83(4):939–942

Haji HM, Misha S, Devos M (2006) Host plant resistance management strategies for control of black root rot resistance. CORESTA Meeting Agro-Phyto Groups. Paris. PPost 08

Haji HM, Mishra S, DeVos M (2007) CTH2 flue-cured tobacco F1 hybrid. Can J Plant Sci 87:383–384

Hakansson G, van der M, Bonnett HT, Glimelius K (1988) Variant mitochondrial protein and DNA patterns associated with cytoplasmic male-sterile lines of Nicotiana. Theor Appl Genet 76:431–437

Hakansson G, Glimelius K (1991) Extensive nuclear influence on mitochondrial transcription and genome structure in male-fertile and male-sterile alloplasmic Nicotiana materials. Mol Gen Genet 229:380–388

Hancock WG, Kuraparthy V, Kernodie SP, Lewis RS (2015) Identification of maternal haploids of *Nicotiana tabacum* aided by transgenic expression of green fluorescent protein: evidence for chromosome elimination in the *N. tabacum* × *N. africana* interspecific cross. Mol Breed 35:179

Hancock WG, Lewis RS (2017) Heterosis, transmission genetics, and selection for increased growth rate in a *N. tabacum* × synthetic tobacco cross. Mol Breeding 37:53 https://doi.org/10.1007/s11032-017-0654-4

Hanson MR, Conde MF (1985) Functioning and variation of cytoplasmic genomes: lessons from cytoplasmic nuclear interactions affecting male sterility in plants. Int Rev Cytol 94:213–267

Hart GE (1965) Studies on extrachromosomal male sterility in *Nicotiana*. Dissertation Abstracts 26: 3595 (Tobacco Abstracts 10: 2108) Ph.D. Thesis Univ California, Berkley USA

Holmes FO (1938) Inheritance of resistance to tobacco-mosaic disease in tobacco. Phytopathol 28:553–561

Horlow C, Goujaud J, Lepingle A, Missionier C, Bourgin JP (1990) Transmission of paternal chloroplasts in tobacco (*Nicotiana tabacum*). Plant Cell Rep 9:249–252

Horton P (1981) A taxonomic revision of Nicotiana (Solanaceae) in Australia. J Adel Bot Gard 3:1–56

Hu C (1956) Cytogenetic Studies in *Nicotiana*, XII. Reduction divisions in hybrids between *N. langsdorffii* and three other species. Jpn J Breed 6:117–121

Iancheva A, Palakarcheva M (1990) Cytoplasmic male sterile hybrids with Nicotiana goodspeedii cytoplasm. Genet Breed 4:316–321

Ichikawa T, Ozeki Y, Syöno K (1990) Evidence for the expression of the rol genes of *Nicotiana glauca* in genetic tumors of *N. glauca* × *N. langsdorflli*. Mol Gen Genet (MGG) 220(2):177–180

Ilcheva V, San LH (1997) Hybridation somatique chez le gendre *Nicotiana*—revue bibliographique. Ann du Tabac, Sect 2(29):9–25

Ilcheva V, San LH, Zagorska N, Dimitrov B (2001) Production of male sterile interspecific somatic hybrids between transgenic *N. tabacum* (bar) and *N. rotundifolia* (npt ii) and their identification by AFLP analysis. Vitro Cell Dev-Pl 37:496–502

Iwai S, Kishi C, Nakata K, Kubo S (1985) Production of a hybrid of *N. repanda* × *N. tabacum* by ovule culture. Plant Sci 41:175–178

Izard C, Hitier H (1955) Observations sur un hybride complexe susceptible de produire des plantes a sterilite male. Ann Du Tabac Seita, Sect 2(2):1–13

Johnson ES, Wernsman EA, LaMondia JA (2009) Effect of a chromosome fragment marked by the *Php* gene for resistance to *Phytophtora nicotianae* on reproduction of tobacco cyst nematodes. Plant Dis 93:309–315

Johnson ES, Wolff MF, Wernsman EA (2002a) Marker-assisted selection for resistance to black shank disease in tobacco. Plant Dis 86:1303–1309

Johnson ES, Wolff MF, Wernsman EA, Atchley WE, Shrew HD (2002b) Origin of the black shank resistance gene, Ph, in tobacco cultivar Coker 371-Gold. Plant Dis 86:1080–1084

Julio E, Verrier JL, Dorlhac de Borne F (2006) Development of SCAR markers linked to three disease resistances based on AFLP within *Nicotiana tabacum* L. Theor Appl Genet 112:335–346

Kaul MHL (1988) Male sterility in higher plants. Monogr Theor Appl Genet Springer Verlag Berlin 10

Kehr AE (1951) Genetic tumors in *Nicotiana*. Am Nat 99:73–79

Kehr AE, Smith HH (1952) Multiple genome relationships in *Nicotiana*. Cornell Univ Agric Exp Sta Memoir 311

Kennedy BS, Nielsen MT (1993) Characterization of tomato spotted wilt virus (TSWV) resistance in the tobacco cultivar 'Polalta' (Abstr.). Phytopathology 83 (12):1420

Kenward KD, Bai D, Ban MR, Brandle JE (1999) Isolation and characterization of Tnd-1, a retrotransposon marker linked to black root rot resistance in tobacco. Theor Appl Genet 98:387–395

Kincaid RR (1949) Three interspecific hybrids of tobacco. Phytopathol 39:284–287

Kitamura S, Tanaka A, Inoue M (2005) Genomic relationships among *Nicotiana* species with different ploidy levels revealed by 5S rDNA spacer sequences and FISH/GISH. Genes Genet Syst 80:251–260

Knapp S, Chase MW, Clarkson JJ (2004) Nomenclatural changes and a new section classification in *Nicotiana* (Solanaceae) Taxon 52:73–82

Knapp S. 2013. Nat.Hist. Museum NaturePLus. Seeking nightshades in South America. (blog) www.nhm.ac.uk

Kobus I (1978) Wstępne badania wartości hodowlanej męskosterylnych mutantów Nicotiana tabacum uzyskanych w wyniku działania promieni gamma Biul. Inst Hod Rośl 134:249–259

Kobus I (1971) Investigations in the polyploids of *Nicotiana tabacum* L. and their hybrids with wild tobacco species. Genet Pol 12:323–328

Kofer W, Glimelius K, Bonnet HT (1990) Modifications of floral development in tobacco induced by fusion of protoplasts of different male-sterile cultivars. Theor Appl Genet 79:97–102

Kofer W, Glimelius K, Bonnett HT (1991) Restoration of normal stamen development and pollen formation by fusion of different cytoplasmic male-sterile cultivars of *Nicotiana tabacum*. Theor Appl Genet 81:390–396

Korbecka-Glinka G, Czubacka A, Przybys M, Doroszewska T (2017) Resistance vs. tolerance to *Potato virus* Y in tobacco—comparing effectiveness using virus isolates from Central Europe. Breed Sci 67 (5):459–465

Kostoff D (1930) Tumours and other malformations on certain *Nicotiana* hybrids. Zbl Bakt II Abt 81:244–260

Kostoff D (1943) Cytogenetics of the genus *Nicotiana*. State Printing House, Sofia

Krusteva D, Dimitrov B, Nikova V (2003) *Nicotiana gossei* Domin as a source of aphid resistance in *Nicotiana tabacum*. Tiutun (Tobacco) 53:217–220

Kubo T (1979) Effect of alien cytoplasms on agronomic characters in flue-cured tobacco. Bull Iwata Tob Exp Stn 11:83–90

Kubo T (1981) Agronomic characteristics of male sterile flue-cured tobacco hybrids with the cytoplasm from *Nicotiana suaveolens*. Bull Iwata Tob Exp Stn 13:61–71

Kubo T (1985) Studies on hybrid breeding by means of cytoplasmic male sterility in tobacco. Bull Iwata Tob Exp Stn 17:69–138

Kubo T, Kumashiro T, Saito Y (1988) Cytoplasmic male sterile lines of a tobacco variety, Tsukuba 1, developed by asymmetric protoplast fusion. Jpn J Breed 38:153–164

Kuboyama T, Chung CS, Takeda G (1994) The diversity of interspecific pollen-pistil incongruity in *Nicotiana*. Sex Plant Reprod 7:250–258

Kumashiro T, Asahi T, Komari T (1988) A new source of cytoplasmic male sterile tobacco obtained by fusion between *N. tabacum* and X-irradiated *N. africana* protoplasts. Plant Sci 55:247–254 (CORESTA Inf Bul 1989:79)

Kumashiro T, Asahi T, Nakakido F (1989) Transfer of cytoplasmic factors by asymmetric fusion from a cross-incompatible species, *N. repanda* to *N. tabacum* and characterization of cytoplasmic genomes. Plant Sci 1:137–144

Kumashiro T, Kubo T (1986) Cytoplasm transfer of *N. debneyi* to *N. tabacum* by protoplast fusion. Jpn J Breed 36:39–48

Kumashiro T, Oinuma T (1985) Comparison of genetic variability among anther derived and ovule derived

doubled haploid lines of tobacco. Jpn J Breed 35:301–310

Larkina NI (1980) Overcoming incompatibility between *Nicotiana* species by means of *in vitro* pollination. Plant Breed Abstr 50:690 (8067)

Larkina NI (1983) Development of starting material for tobacco improvement by using interspecific hybridisation Создание путем межвидовой гибридизации нового исходного материала для селекции табака DSc thesis. WITIM, Krasnodar

Laskowska D, Doroszewska T, Depta A, Kursa K, Olszak-Przybys H, Czubacka A (2013) A survey of Nicotiana germplasm for resistance to Tomato Spotted Wilt Virus (TSWV) spotted wilt virus (TSWV). Euphytica 193:207–219

Laskowska D, Berbeć A (2005) Cytology and fertility of viable hybrids of Nicotiana tabacum L. cv. TB-566 with N. alata Link et Otto. J Appl Genet 46:11–18

Laskowska D, Berbeć A (2007) The new alloplasmic *Nicotiana tabacum* L. line of with *Nicotiana wuttkei* Clarkson et Symon cytoplasm Nowa męskosterylna linia *Nicotiana tabacum* L. z cytoplazmą *N. wuttkei* Clarkson et Symon (in Polish). Biul IHAR 244:289–296

Laskowska D, Berbeć A (2010) TSWV resistance in DH lines of tobacco (*Nicotiana tabacum* L.) obtained from a hybrid between Polalta and Wiślica. Plant Breed 129:731–733

Laskowska D, Berbeć A, Van Laere K, Kirov I, Czubacka A, Trojak-Goluch A (2015) Cytology and fertility of amphidiploid hybrids between *Nicotiana wuttkei* Clarkson et Symon and *N. tabacum* L. Euphytica 206:597–608. https://doi.org/10.1007/s10681-015-1459-3

Laskowska D, Berbeć A (2012) Production and characterization of amphihaploid hybrids between *Nicotiana wuttkei* Clarkson et Symon and *N. tabacum*. Euphytica 183:75–82

Lawson DM, Schaeffer S, Wernsman E, Nielsen MT (2002) A comparison of *Nicotiana suaveolens* and *Nicotiana glauca* cytoplasmic male sterility systems in flue-cured tobacco. CORESTA Congress, New Orleans, Agro-Phyto Groups, AP11

Lea HW (1963) The transfer of resistance against blue mold (*Peronospora tabacina* Adam) from *Nicotiana debneyi* to cultivated tobacco. CORESTA Inf Bull (3):13–15

Lee CB, Page LE, McClure BA, Holtsford TP (2008) Post-pollination hybridization barriers in *Nicotiana* section *Alata*e. Sex Plant Reprod 21:183–195

Legg PD, Collins GB, Litton CC (1979) Effects of the N mosaic-resistance factor on agronomic and chemical traits in Burley tobacco. Crop Sci 19:455–457

Legg PD, Litton CC, Collins GB (1982) Effects of *Nicotiana longiflora* Cav. resistance to race 0 Phytophthora parasitica var. *Nicotianae* on agronomic and chemical traits in burley tobacco. Crop Sci 22:35–38

Legg PD, Litton CC, Collins GB (1981) Effects of *N. debneyi* black root rot resistance factor on agronomic and chemical traits of burley tobacco. Theor Appl Genet 60:365–368

Legg PD, Smeeton BW (1999) II. Breeding and Genetics. (in) Layten Davis and MT Nielsen ed. Tobacco: production, chemistry and technology. Oxford, UK, Blackwell Science Ltd

Lehmann H (1936) On interspecific cross sterility in the genus Nicotiana. Plant Breed. Abst. 7 (1936), abst. 1023 (in German)

Lewis RS (2002) Tissue culture and molecular marker-assisted introgression of PVY resistance from *N. africana* to *N. tabacum*. CORESTA Congress, New Orleans, Agro-Phyto Groups, AP28

Lewis RS (2005) Transfer of resistance to Potato Virus Y (PVY) from *Nicotiana africana* to *Nicotiana tabacum*: possible influence of tissue culture on the rate of introgression. Theor Appl Genet 110:678–687

Lewis RS (2007) Evaluation of *Nicotiana tabacum* genotypes possessing *Nicotiana africana*-derived genetic tolerance to Potato Virus Y. Crop Sci 47:1975–1984

Lewis RS (2011) *Nicotiana*. In: Kole C (ed) Wild crop relatives: genomic and breeding resources, plantation and ornamental crops. Springer, Berlin Heidelberg. https://doi.org/10.1007/978-3-642-21201-7_10

Lewis RS, Linger LR, Wolff MF, Wernsman EA (2007a) The negative influence of N-mediated TMV resistance on yield in tobacco: linkage drag versus pleiotropy. Theor Appl Genet 115(2):169–178

Lewis RS, Milla SR, Kernodle SP (2007b) Analysis of an introgressed *Nicotiana tomentosa* genomic region affecting leaf number and correlated traits in *Nicotiana tabacum*. Theor Appl Genet 114:841–854

Lewis RS, Nicholson JS (2007) Aspects of the evolution of *Nicotiana tabacum* L. and the status of the United States *Nicotiana* germplasm collection. Genet Resources Crop Evol 54:727–740

Lewis RS, Rose C (2010) Agronomic performance of tobacco mosaic virus-resistant tobacco lines and hybrids possessing the resistance gene N introgressed on different chromosomes. Crop Sci 50:1339–1347

Lewis RS, Wernsman EA (2001) Efforts to initiate construction of a disease resistance package on a designer chromosome in tobacco. Crop Sci 41:1420–1427

Liao J, Dai J, Yang S, Zhou X, Ren L, Chen Z, He H, Chen S (2017) Interspecific cross-hybrids of *Nicotiana tabacum* L. cv. (gla.) S 'K326' with *Nicotiana alata*. Plant Breed 136:427–435. https://doi.org/10.1111/pbr.12474

Lim KY, Souckova-Skalicka K, Sarasan V, Clarkson JJ, Chase MW, Kovarik A, Leitch AR (2006) A genetic appraisal of a new synthetic *Nicotiana tabacum* (Solanaceae) and the Kostoff synthetic tobacco. Am J Bot 93(6):875–883

Ling H-Y, Edwards AM, Gantier MP, DeBoer K, Neale AD, Hamil JD, Walmsley AM (2012) An interspecific *Nicotiana* hybrid as a useful and cost-effective platform for production of animal vaccines.

PLoS ONE 7(4):e35688. https://doi.org/10.1371/journal.pone.0035688
Lopez M, Espino E, Garcia H (2008) Informe de nuevas variedades 'Capero-1': primer híbrido androestéril comercial de tabaco negro cubano (*Nicotiana tabacum* L.). Cultivos Trop 29:51
Mackenzie J, Smeeton BW, Jack AM, Ternouth RAF (1986) Review on breeding for resistance to root knot, Meloidogyne javanica in flue-cured tobacco in Zimbabwe. Bul Spec CORESTA Symposium Taormina: 64, abstr P05
Maktari ATA (1991) Use of cytoplasmic male sterility in the production of hybrid seed in tobacco Использование цитоплазматической мужской стерильности в производстве гибридных семян табака (in Russian). Doctoral thesis. Kubanskij Gosudarstwiennyj Agrarnyj Institut
Malecka J (1977) Cyto-embryological studies in *Nicotiana* hybrids. Acta Biol Cracov Bot 20:89–101
Malinowski E (1916) O wystepowaniu nowych Form w potomstwie mieszańcow *Nicotiana atropurpurea* × *Nicotiana* silvestris. Comptes Rendues de la Societe des Sciences de Varsovie 9:827–864
Mallach GS (1943) Inheritance in Nicotiana tabacum. XVI. Structural differences among the chromosomes of a selected group of varieties. Genetics 28:525–532
Malloch WS, Malloch FW (1924) Species crosses in Nicotiana with special reference to *N. longiflora* × *N. tabacum*, *N. longiflora* × *N. sanderae*, *N. tabacum* × *N. glauca*. Genetics 9:261–291
Manabe, T, Marubashi W, Onozawa Y (1989) Temperature-dependent conditional lethality in interspecific hybrids between *Nicotiana suaveolens* Lehm and *N. tabacum* L. In: proceedings of the 6th International Congress of SABRAO, pp 459–462
Manolov A (1980) Transfer of Peronospora resistance from *Nicotiana exigua* to Oriental Tobacco. Bulgarski Tiutun 25:11–18
Marks CE (2010) Definition of South Pacific taxa of *Nicotiana* section Suaveolentes (Solanaceae). Muelleria 28:74–84
Marks CE, Ladiges PY, Newbigin E (2011) Karyotypic variation in *Nicotiana* section Suaveolentes. Genet Resour Crop Ev 58:797–803
Martinez-Croveto R (1978) Una nueve especie de *Nicotiana* de la flora argentina. Bonplandia 5:7–10
Marubashi W, Onosato K (2002) Q Chromosome controls the lethality of interspecific hybrids between *Nicotiana tabacum* and *Nicotiana suaveolens*. Breed Sci 52:137–142
Matveeva TV, Lutova LA (2014) Horizontal gene transfer from Agrobacterium to plants. Front Plant Sci 5:article 326
Medgyesy P, Fejes E, Maliga P (1985) Interspecific chloroplast recombination in a *Nicotiana* somatic hybrid. Proc Natl Acad Sci USA 82:6960–6964
Merxmueller H, Buttler KP (1975) *Nicotiana* in der Afrikanischen Namib ein pflanzengeographisches und phytogenetisches Ratsel. Mitt. Bot. Munchen 12:91–104
Milla SR, Levin JS, Lewis RS, Rufty RC (2005) RAPD and SCAR markers linked to an introgressed gene conditioning resistance to Peronospora tabacina D.B. Adam. in Tobacco. Crop Sci 45:2346–2354
Moav R (1958) Inheritance in Nicotiana tabacum XXIX: relationship of residual-chromosome homology to interspecific gene transfer. Am Nat 92:267–278
Moav R, Cameron DR (1960) Genetic instability in *Nicotiana* hybrids. I. The expression of instability in *N. tabacum* × *N. plumbaginifolia*. Am J Bot 47:87–93
Moon H, Nicholson JS (2007) AFLP and SCAR markers linked to tomato spotted wilt virus resistance in tobacco. Crop Sci 47:1887–1894
Mudzengerere ET (1994) Breeding for resistance to Meloidogyne javanica in Burley tobacco in Zimbabwe. Bull Spec CORESTA Congress Harare 114: P30
Murfett J, Lee C, Page L, Yates A, Ippolito A, Holtsford, TP (2005) Phylogenetics of Nicotiana Section Alatae and description of *N. rastroensis*. University of Texas, Botany 2005 Annual Meeting, Systematics Section/ASPT, Abstr. 359, Austin, Texas
Murthy TGK, Subbarao IV (2004) Some new interspecific hybrids in the genus *Nicotiana* - characterization and utilization. Tob Res 30:33–41
Murthy TGK, Sredhaar U, Siva Raju K (2014) Research achievements. I. Tobacco cultivar development. Incorporation of aphid resistance from *N. gossei*. *N. repanda*, *N* × . *umbratica-nesophila* and *N* × *benthamiana–repanda*. CTRI Annual Report 2013–2014, Rajamundry, India, pp 17–20
Näf U (1958) Studies on tumour formation in *Nicotiana* hybrids I. The classification of parents into two etiologically significant groups. Growth 22:167–180
Nagao T (1979) Somatic hybridization by fusion of protoplasts. II. Combinations of *Nicotiana tabacum* with *N. glutinosa* and of *N. tabacum* with *N. alata*. Jpn J Crop Sci 48:385–392
Naumenko SA (2012) Particular qualities of the development of flue-cured and burley tobacco varieties in Russia Особенности селекции сортов табака сортотипов Вирджиния и Берлей в России. Particular qualities of the development of flue-cured and Burley tobacco in Russia (in Russian). DSc thesis, Russian Timiryazev State Agrarian University, Moscow
Nielsen MT, Collins GB (1989) Variation among androgenic and gynogenic doubled haploids of tobacco (*Nicotiana tabacum*). Euphytica 43:263–267
Nielsen MT, Legg PD, Litton CC (1985) Effects of two introgressed disease resistance factors on agronomic characteristics and certain chemical components in burley tobacco. Crop Sci 25:698–701
Nikova V (1984) Cytoplasmic male sterility in *Nicotiana* with cytoplasm from *Nicotiana benthamiana* Domin. Compt Rend Acad Sci Bulg 37:1253–1256
Nikova V, Iancheva AM, Vladova R,Pandeva R, Petkova A (2006) Genetic improvement of *Nicotiana tabacum* by applying remote hybridization and biotechnological methods. CORESTA Congress, Paris, Agro-Phyto Groups, APOST 07

Nikova V, Iantcheva AM, Pandeva R, Petkova A (2004) Investigation of alloplasmic tobacco lines by certain economically indices. CORESTA Congress, Kyoto, Agro-Phyto Groups, PPOST9

Nikova V, Pundeva R, Vladova R, Petkova A (2001) A new tobacco cytoplasmic male sterile source from the hybrid combination *Nicotiana longiflora* Gav. and *N. tabacum* L. using in vitro techniques. Isr J Plant Sci 49:9–13

Nikova V, Vladova R (2002) Wild Nicotiana species as a source of cytoplasmic male sterility in *Nicotiana tabacum*. Beiträge Zur Tab Int 20:301–311

Nikova V, Vladova R, Pundeva R, Petkova A, Vladovska A (1998b) Study of *Nicotiana ingulba* Black and its crossability with *N. tabacum* L. Bull Spec. CORESTA Congress Brighton:150, PPOST 12 (b)

Nikova V, Vladova R, Pundeva R, Shabanov D (1997) Cytoplasmic male sterility in *Nicotiana tabacum* L. obtained through interspecific hybridization. Euphytica 94:375–378

Nikova VM (1986) *Nicotiana excelsior* Black as a source of cytoplasmic male sterility in *Nicotiana tabacum* L. Compt Rend Acad Bulg Sci 39:105–107

Nikova VM, Shabanov D (1992) Development of male sterile tobacco forms by crossing *Nicotiana knightiana* × *N. tabacum*. CORESTA Inf Bul Special Congress Jerez de La Frontera:132

Nikova VM, Shabanov DS (1988) *Nicotiana amplexicaulis* Burbidge as a source of cytoplasm male sterility in tobacco (*Nicotiana tabacum* L.). Compt Rend Acad Bulg Sci 41(3):83–86

Nikova VM, Zagorska NA (1990) Overcoming hybrid incompatibility between *Nicotiana africana* Merxm and *Nicotiana tabacum* and development of cytoplasmically male sterile tobacco forms. Plant Cell, Tissue Organ Cult 32:71–75

Nikova VM, Zagorska NA, Pundeva RS (1991) Development of four tobacco male sterile sources using *in vitro* techniques. Plant Cell, Tissue Organ Cult 27:289–295

Nikova V, Pundeva R, Petkova A (1999) *Nicotiana tabacum* L. as a source of cytoplasmic male sterility in interspecific cross with *N. alata* Link & Otto. Euphytica 107:9–12

Niwa M (1969) Radiation induced interspecific transfer of ws(pbg) gene from *Nicotiana plumbaginifolia* to *N. tabacum*: II. Estimation of frequency of the gene transfer induced by irradiation at the gametogenetic stage. Jap Breed 19:84–88

Ohashi Y (1976) *Nicotiana kawakamii*: a new species of the genus Nicotiana. Bul Spec CORESTA, Congress Tokyo: 71 (abstr) P001

Ohashi Y (1985) Breeding studies of wild species related to tobacco, with special reference to disease resistance. Bull Iwata Tob Exp Stn 17:1–62

Oka H (1961) A breeding study on interspecic transfer of disease resistance in tobacco. Hatano Tobacco Expt, St. Bull, p 49

Oupadissakoon S, Wernsman EA (1977) Agronomic performance and nature of gene effects in Progenitor Species-derived Genotypes of Tobacco. Crop Sci 17 (6):843–847

Pal BP, Nath P (1936) A note on the sterile hybrid between *N. tabacum* and *N. plumbaginifolia*. Indian J Agric Sci 6:828–832

Palakarcheva M (1981) Genetic peculiarities of tobacco breeding for disease resistance (in Bulgarian). Genetika Selektsiya 14:277–284

Palakarcheva M (1984) Interspecific hybridization in the genus *Nicotiana*. Bul Spec CORESTA Congress Vienne: 142 PP31

Palakarcheva M, Bailov D (1976) Disease resistant hybrids and varieties of tobacco obtained from interspecific hybridization. Rasteniev. Nauki 13:35–42 (Review of Plant Pathology 56: 3705)

Palakarcheva M, Edreva A, Cholakova N, Noveva S (1978) Morphological, cytological, and biochemical studies of the amphidiploid *Nicotiana goodspeedii* W. × *Nicotiana tabacum* L. (2n = 88). Z Pflanzenzuechtg 80:49–63

Palakarcheva MT, Edreva AM, Cholakova NI (1980) Study of the backcrossing effect of *Nicotiana tabacum* in hybrids obtained with the amphidiploids *N. tabacum* L. × *N. debneyi* D. (2n = 96) and *N. goodspeedii* × *N. tabacum* L. (2n = 88). Bul Spec CORESTA Congress Manila:102

Palakarcheva M, Krusteva D (1978) Study on the inheritance of powdery mildew (Erysiphe cichoracearum D.C.) resistance in intercultivar tobacco hybrids. Genet Sel 11:126–134

Pandey KK, Phung M (1982) "Hertwig effect" in plants: induced parthenogenesis through the use of irradiated pollen. Theor Appl Genet 62:295–300

Pandeya RS, White FH (1984) Delgold: A new flue-cured tobacco. Can J Plant Sci 64:233–236

Pandeya RS, Douglas GC, Keller WA, Setterfield G, Patrick ZA (1986) Somatic hybridization between *Nicotiana rustica* and *N. tabacum*: Development of tobacco breeding strains with disease resistance and elevated nicotine content. Z Pflanzerzüchtg 96:346–352

Patel KA, Gerstel DU (1961) Additional information on the mechanism of chromosome substitution in *Nicotiana*. Tob Sci 5:18–20

Patrascu M, Paunescu D, Ciuperca A (1999) Transfer of the resistance to TSWV from *Nicotiana alata* to some tobacco cultivars, using the in vitro culture technique. CORESTA Meet Agro-Phyto Groups Suzhou, China, POST8

Pereg LL (2013) Black root rot of cotton in Australia: the host, the pathogen and disease management. Crop Pasture Sci 64(12):1112

Pirrie A, Power JB (1986) The production of fertile, triploid somatic hybrid plants (*Nicotiana glutinosa* (n) + *N. tabacum* (2 N) via gametic: somatic protoplast fusion. Theor Appl Genet 72:48–52

Pittarelli GW, Stavely JR (1975) Direct hybridization of *Nicotiana repanda* × *N. tabacum*. J Hered 66:281–284

Qin Q, Li Y, Martinez N, Miller R, Ding N, Wu X, Li D, Yang S (2016) Development of user-friendly markers for disease resistance to black root rot through genotyping by sequencing. TSRC, Tob Sci Res Conf 70, Abstr. 42

Ramavarma KT, Appa Rao K, Narayanan AI (1978) Male sterile flower amomalies in an interspecific *Nicotiana* hybrid. Tobacco Research 4:29–33

Ramavarma KT, Apparao K, Narayanan AI, Sitaramaiah S, Joshi BG (1991) Evolving tobacco varieties resistant to the leaf-eating caterpillar Spodoptera litura F. through interspecific hybridization. Tobacco Research 17:29–32

Ramavarma KT, Apparao K, Sitharamaiah S, Narayanan AI (1980) Interspecific transfer of resistance to tobacco caterpillar (Spodoptera litura F.) from *Nicotiana benthamiana* to *Nicotiana tabacum*. Cytologia 45:103–111

Rao KA, Ramavarma KT, Joshi BG (1980) Interspecific hybridization and breeding for pest resistance in tobacco. Tob Sci 24:46–48

Rao PN, Stokes GW (1963) Role of chromosome H in the development of calcium deficiency symptoms in burley tobacco. Crop Sci 3:261–264

Reed S, Collins GB (1978) Interspecific hybrids in *Nicotiana* through in vitro culture of fertilized ovules. J Hered 69:311–315

Reed SM, Burns JA (1986) Cross-restoration between Nicotiana cytoplasmic male-sterile and restored lines. J Hered 77(3):159–163

Renny Byfield S, Chester M, Kovarik A et al (2011) Next generation sequencing reveals genome downsizing in allotetraploid *Nicotiana tabacum*, predominantly through the elimination of paternally derived repetitive DNAs. Mol Biol Evol 28:2843–2854

Rice JH, Mundell RE, Millwood RJ, Chambers OD, Stewart CN, Davies HM (2013) Assessing the bioconfinement potential of a *Nicotiana* hybrid platform for use in plant molecular farming applications. BMC Biotechnol 13(1):63

Rufty RC (1989) Genetics of host resistance to tobacco blue mold. (in) Blue mold of tobacco. APS Press, USA:141–164

Rybin VA (1929) Uber einen allotetraploiden Bastard von *Nicotiana tabacum* × *Nicotiana sylvestris*. Ber Deut Bot Ges 37:385–394

Sachs-Skalińska M (1917) Przyczynek do cytologii bezpłodnego mieszańca *Nicotiana atropurpurea* × *Nicotiana silvestris* = Beitrag zur Cytologie des sterilen Bastardes *Nicotiana atropurpurea* × *Nicotiana silvestris*. Comptes Rendus de la Société des Sciences de Varsovie 10

Sarychev YF (1986) A new method of inducing diploid apomixis in *Nicotiana tabacum* L. Genetika 22:1142

Shabanov D, Popchristev W, Tomov N (1974) Interspecific hybrid *Nicotiana tabacum* × *N. setchellii* (in Bulgarian). Nauch Trud 4:12–22

Schweppenhauser MA (1975a) A source of *Nicotiana tabacum* resistance to *Meloidogyne javanica*. Tob Sci 19:43–46

Schweppenhauser MA (1975b) Rootknot resistance from *Nicotiana longiflora*. Tob Sci 19:26–29

Schweppenhauser MA (1974) Interspecific bridge transfer in *Nicotiana* of resistance to *Meloidogyne javanica*. S Afr J Sci 70:312–314

Schweppenhauser MA, Mann TJ (1968) Restoration of staminal fertility in *Nicotiana* by introgression. Can J Genet Cytol 10:401–411

Sheen SJ (1972) Isozyme evidence bearing on the origin of *Nicotiana tabacum*. Evolution 26:143–154

Skalicka K, Lim KY, Matyasek R, Matzke M, Leitch AR, Kovarik A (2005) Preferential elimination of repeated DNA sequences from the paternal, *Nicotiana tomentosiformis* genome donor of a synthetic allotetraploid. New Phytol 166:291–303

Skucińska B, Miszke W, Kruczkowska H (1977) Studies on the use of interspecific hybrids in tobacco breeding. Obtaining of fertile hybrids by propagation in vitro. Acta Biol Cracov 20:81–88

Slana LJ, Stavely JR, Grosso JJ, Golden AM (1977) Probable source of *Meloidogyne incognita* resistance in tobacco as indicated by reactions to five Meloidogyne isolates. Phytopathology 67:537–543

Slusarkiewicz-Jarzina A, Zenkteler M (1983) Development of hybrid plants from ovules of *Nicotiana tabacum* pollinated in vitro with pollen grains of *Nicotiana knightiana*. Experimentia 39:1399–1400

Smeeton BW, Ternouth RAF (1992) Sources of resistance to powdery mildew, wildfire, angular leaf spot, and Alternaria. CORESTA Info Bull 1992–1993:127–135

Smith HH (1968) Recent cytogenetic studies in the genus *Nicotiana*. Adv Genet 14:1–43

Smith LB, Downs R (1964) Solanaceas. Flora ilustrada Catarinense. 321. Phytologia 10:438

Sproule A, Donaldson P, Dijak M, Bevis E, Pandeya R, Keller WA, Gleddie S (1991) Fertile somatic hybrids between transgenic *Nicotiana tabacum* and transgenic *N. debneyi* selected by dual antibiotic resistance. Theor Appl Genet 82:450–456

Stavely JR, Skoog HA (1976) Transfer of resistance to a virulent strain of Pseudomonas tabaci from *Nicotiana rustica* to *Nicotiana tabacum* breeding lines. Proc Am Phytopath Soc 3:231

Stavely JR, Skoog HA (1978) Stabilization in *Nicotiana tabacum* of a dominant gene from *N. rustica* for resistance to a virulent strain of Pseudomonas tabaci. Phytopathol. News 12: 181

Stavely JR, Pittarelli GW, Burk LG (1973) Nicotiana repanda as a potential source for disease resistance in *N. tabacum*. J Hered 64(5):265–271

Stehmann JR, Semir J, Ippolito A (2002) *Nicotiana mutabilis* (Solanaceae), a new species from southern Brazil. Kew Bull 57:639–646

Stebbins GL (1950) Variation and evolution in plants. Columbia University Press, New York

Stoyanova M (1972) Investigations of the hybrids between the species *Nicotiana tabacum* and *N. glauca* (in Bulgarian). Otdal Hybridiz Rast C BAN:127–136

Subhashini U, Venkateswarlu T, Anjani K (1986) Embryo rescue in Nicotiana hybrids by in vitro culture. Plant Sci 43:219–222

Swaminathan MS, Murthy BR (1957) One-way incompatibility in some species crosses in the genus *Nicotiana*. Indian Journal of Genetics and Plant Breeding 17:23–26

Symon DE (1998) A new Nicotiana species from near Coober Pedy. South Australia. J. Adelaide Bot. Gard. 18:1–4

Symon DE (1984) A new species of Nicotiana (Solanaceae) from Dalhousie Springs. South Australia J Adelaide Bot Gard. 7:117–121

Symon DE, Kennealy KF (1994) A new species of Nicotiana (Solanaceae) from near Broome, Western Australia. Nuytsia 9:421–425

Szilagyi L (1975) Elimination of chromosomes in an alloploid hybrid of *Nicotiana tabacum* × *N. glauca*. Acta Bot Acad Sci Hung 21:433–441

Takenaka Y (1956) Cytogenetic studies in *Nicotiana*. XIV. Reduction divisions in five interspecific hybrids. Jpn J Genet 31:155–161

Takenaka Y (1962) Cytogenetic studies in *Nicotiana* XVI: Reduction divisions in six interspecific hybrids between *N. tabacum* and other six species. Shokubutsugaku Zasshi 75(888): 237–241

Takenaka Y (1963) Cytogenetic studies in *Nicotiana*. XV. Reduction divisions in three tnterspecific hybrids and one amphidiploid. Jpn J Genet 38:135–140

Tanaka M (1961) The effect of irradiated pollen grains on species crosses of *Nicotiana*. Bulletin of the Hatano Experiment Station 51:1–38

Tatemichi Y (edit) (2005) Illustrated Book of the Genus Nicotiana. Plant Breeding and Genetics Research Laboratory, Japan Tobacco Inc., Tokyo, Japan

Ternouth RAP, MacKenzie J, Shepherd JA (1986) Introduction of *Meloidogyne javanica* resistance into flue-cured tobacco in Zimbabwe. CORESTA Inf Bull Symposium, Taormina, Italy

Ternovsky MF (1934) Die Fragen der Immunitat bei Vertretern der Gattung *Nicotiana*. Der Zuchter 6:140–144

Ternovsky MF (1941) Methods of breeding tobacco varieties resistant to tobacco mosaic and powdery mildew. Методика селекции невосприимчивых сортов табака к табачной мозаике и мучнистой росе Sbornik issl-nauch rabot WITIM, Сб. работ по селек., генет. и семеновод, табака и махорки. ВИТИМ. Краснодар, в. 66, с. 143–146

Ternovsky MF, Butenko RG, Moiseeva ME (1972) The use of tissue culture to overcome the barrier of incompatibility between species and sterility of interspecific hybrids. Sov Genet 8:27–33

Ternovsky MF, Larkina NI (1978) Hybridization of native tobacco cultivars with the wild species *Nicotiana rosulata*. Гибридизация отечественных сортов табака с диким видом N.rosulata Sov Genet 14:1039–1045

Ternovsky MF, Moiseyeva ME, Grebenkin A (1973) Experimentally produced new type of male sterility in interspecies *Nicotiana* hybrids (in Russian). Sov Genet 9:693–702

Ternovsky MF, Shinkareva IK, Larkina NI (1976) Production of interspecific tobacco hybrids by the pollination of ovules *in vitro*. Sov Genet 12:1209–1213

Tezuka T (2012) Hybrid Lethality in the Genus *Nicotiana*. Mworia JK, editor Botany. Rijeka InTech:191–210

Tezuka T, Kuboyama T, Matsuda T, Marubashi W (2010) Seven of eight species in *Nicotiana* section Suaveolentes have common factors leading to hybrid lethality in crosses with Nicotiana tabacum. Ann Bot 106:267–276

Tezuka T, Marubashi W (2004) Apoptotic cell death observed during the expression of hybrid lethality in interspecific hybrids between *Nicotiana tabacum* and *N. suaveolens*. Breed Sci 54:59–66

Trojak Goluch A, Laskowska D, Agacka M, Czarnecka D, Kawka M, Czubacka A (2011) Effectiveness of combining resistance to *Thielaviopsis basicola* and Tomato spotted wilt virus in haploid tobacco genotypes. Breed Sci 61:389–393

Trojak-Goluch A, Berbeć A (2009) Growth, development and chemical characteristics of tobacco lines carrying black root rot resistance derived from *Nicotiana glauca* (Grah.) Plant Breed 130:92–95. https://doi.org/10.1111/j.1439-0523.2009.01755.x

Trojak-Goluch A, Korbecka-Glinka G, Goepfert S, Doroszewska T, Berbec A (2016a) Tobacco breeding for TSWV resistance using RTSW-al factor derived from cultivar Polalta. CORESTA Congress, Berlin, 2016, Agronomy/Phytopathology Groups, APPOST 24

Trojak-Goluch A, Laskowska D, Kursa K (2016b) Morphological and chemical characteristics of doubled haploids of flue-cured tobacco combining resistance to Thielaviopsis basicola and TSWV. Breed Sci 66:293–299

Tsikov D, Nikova V (1980) Male sterility in tobacco (*Nicotiana tabacum*). IV. Seed productivity. Genet Plant Breed 13:198–205

Tsikov D, Tsikova E (1986) Male sterility in tobacco. VI. Cytoplasm of *N. gossei* Мъска стерилност при тютюна (in Bulgarian). Genet Plant Breed 19:121–131 [CORESTA Inf Bul 1986(2):51–52]

Tsikov D, Tsikova E, Nikova V (1977) Male sterility in tobacco. III. On anther feminization Мъска стерилност при тютюна. За феминизациата тичинките (in Bulgarian). Генет и СелGenet Plant Breed 10:129–140

Valleau WD (1952) Breeding tobacco for disease resistance. Econ Bot 6:69–102

Valleau WD, Stokes GW, Johnson E (1960) Nine years' experience with the *Nicotiana longiflora* factor for resistance to Phytophthora parasitica var *Nicotiana*e in the control of black shank. Tob Sci 4:92–94

Venkateswarlu T, Nagarajan K, Subhashini U (1998) Root-knot nematode resistance through interspecific hybridization. Tob Symp 7:B20

Verrier JL, Malpica A, L'Humeau J (2016) Breeding for blue-mould resistance in Burley and flue-cured tobacco: an overview of results. CORESTA Congress, AP, p 07

Wan H, Ueng IN, Lin TT, Chen CH (1971) Interspecific transfer of frogeye resistance from *N. repanda* to *N. tabacum*. Taiwan Tob Wine Monop Bur Tob Res Annu Rep 34–35

Wark DC (1963) *Nicotiana* species as sources of resistance to blue mold (Peronospora tabacina Adam) for cultivated tobacco. Proceedings of Third World Tobacco Scientific Congress Salisbury, Southern Rhodesia, Tobacco Research Board:252–259

Wark DC (1970) Development of flue cured tobacco cultivars resistant to common strain of blue mold. Tob Sci 14:147–150

Wark DC (1975) The development of blue mold resistant cultivars in Australia. CSIRO Div. of Plant Industry Annual Report 1975:31–33

Warmke HE, Blakeslee AF (1939) Induction of tetraploidy in *Nicotiana sanderae* and in the sterile hybrid *N. tabacum* × *N. glutinosa*. Genetics 24:109–110

Wernsman EA, Matzinger DF, Mann TJ (1976) Use of progenitor species germplasm for the Improvement of a cultivated allotetraploid. Crop Sci 16:800–803

Wernsman EA, Rufty RC (1987) Tobacco. In: Fehr WW (ed) Principles of cultivar development, vol 2. Crop species. Macmillian, New York, pp 669–698

White FH, Pandeya RS, Dirks VA (1979) Correlation studies among and between agronomic, chemical, physical and smoke characteristics in flue-cured tobacco (*Nicotiana tabacum* L.). Can J Plant Sci 59:111–120

Whitham S, McCormick, S Baker B (1996) The N gene of tobacco confers resistance to tobacco mosaic virus in transgenic tomato. Proc Natl Acad Sci 93:8776–8781

Williams E (1975) Meiotic chromosome pairing in interspecific hybrids of Nicotiana. Comparative pairing of hybrids in section Suaveolentes. N Z J Bot 13:601–609

Wong YN (1975) Genome analysis of tobacco interspecific hybrids. Taiwan Tob Wine Monop Bur Tob Res Inst Res Rep 2:22–31

Woodend JJ, Mudzengerere E (1992) Inheritance of resistance to wildfire and angular leaf spot derived from *Nicotiana rustica* var. Brasilea Euphytica 64:149–156

Wuttke HH (1969) Different levels of resistance in blue mold resistant tobacco. Austr J Agric Anim Husb 9:545–548

Yang SJ (1960) Numerical chromosome instability in *Nicotiana* hybrids. I. Interplant variation among offspring of amphiploids. Genetics 50:745–756

Zenkteler M, Melchers G (1978) In vitro hybridization by sexual methods and by fusion of somatic protoplasts: Experiments with *Nicotiana tabacum* × *Petunia hybrida*, *N. tabacum* × *Hyoscyamus niger*, *H. niger* × *P. hybrida*, *Melandrium album* × *P. hybrida*. Theor Appl Genet 52:81–90

Zhang S, Gao M, Zaitlin D (2012) Molecular linkage mapping and marker-trait associations with NlRPT, a downy mildew resistance gene in *Nicotiana langsdorffii*. Front Plant Sci 3:185. https://doi.org/10.3389/fpls.2012.00185

Zhang S, Zaitlin D (2008) Genetic Resistance to Peronospora tabacina in *Nicotiana langsdorffii*, a South American Wild Tobacco. Phytopathology 98:519–528

Zheng Y, Liu Z, Sun Y, Liu G, Yang A, Li F (2018) Characterization of genes specific to sua-CMS in *Nicotiana tabacum*. Plant Cell Rep 37(9):1245–1255

Zhou X, Liu Y (2015) Hybridization by grafting: A new perspective? HortScience 50:520–521

Check for updates

9 Nicotine Biosynthesis, Transport, and Regulation in Tobacco: Insights into the Evolution of a Metabolic Pathway

Tsubasa Shoji

Abstract

In *Nicotiana tabacum* (tobacco), nicotine and related pyridine alkaloids are produced in the roots and accumulate mainly in the leaves. Molecular analyses of nicotine biosynthesis, especially of the steps involved in pyrrolidine and pyridine formation, suggest that this specialized pathway evolved through repeated duplication of primary pathways, followed by the recruitment of the metabolic genes into a regulon. In tobacco, jasmonates elicit nicotine formation via a conserved signaling cascade anchored to the downstream nicotine biosynthesis pathway by master transcription factors of the ERF family, particularly ERF189 and its homolog ERF199. ERF transcription factors upregulate metabolic and transport genes directly involved in the pathway by recognizing *cis*-elements in the promoters of target genes. A pair of homologous clusters of related ERF genes, including *ERF189* and *ERF199*, occurs in the tobacco genome. *ERF189* corresponds to the nicotine-controlling *NIC2* locus. A large chromosomal deletion of the cluster that includes *ERF189*, as found in the *nic2* mutant allele, has been exploited to breed low-nicotine tobacco.

T. Shoji (✉)
Graduate School of Biological Sciences, Nara Institute of Science and Technology, Takayama 8916-5, Ikoma, Nara 630-0101, Japan
e-mail: t-shouji@bs.naist.jp

9.1 Introduction

Alkaloids are a large group of nitrogen-containing specialized metabolites, typically with bioactive properties, that are produced by multiple plant species (Shoji 2016). Nicotine and its derivatives, such as nornicotine, anabasine, and anatabine, are pyridine alkaloids found in *Nicotiana* species, including cultivated *Nicotiana tabacum* (tobacco) (Saitoh et al. 1985) (Fig. 9.1). Even though smoking is detrimental to human health, the stimulatory and addictive properties of tobacco alkaloids, which act on nicotinic acetylcholine receptors essential for a range of neuronal activities, account for the widespread consumption of tobacco products.

In tobacco, pyridine alkaloids are synthesized exclusively in underground roots and are largely stored in the leaves as defenses against insects and other predators (Steppuhn et al. 2004). Herbivory can result in increased accumulation of nicotine and other toxic alkaloids (Baldwin 1989). Jasmonates are key signaling molecules that trigger nicotine formation via transcription factors that direct the coordinated expression of genes involved in the nicotine biosynthesis pathway (Shoji et al. 2010). Molecular and genomic analyses have identified genes responsible for nicotine biosynthesis, transport, and regulation (Shoji and Hashimoto 2011a; Dewey and Xie 2013; Shoji and Hashimoto 2013), providing insight into the evolution of this metabolic pathway (Kajikawa et al. 2017a. Shoji 2019).

© Springer Nature Switzerland AG 2020

N. V. Ivanov et al. (eds.), *The Tobacco Plant Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-030-29493-9_9

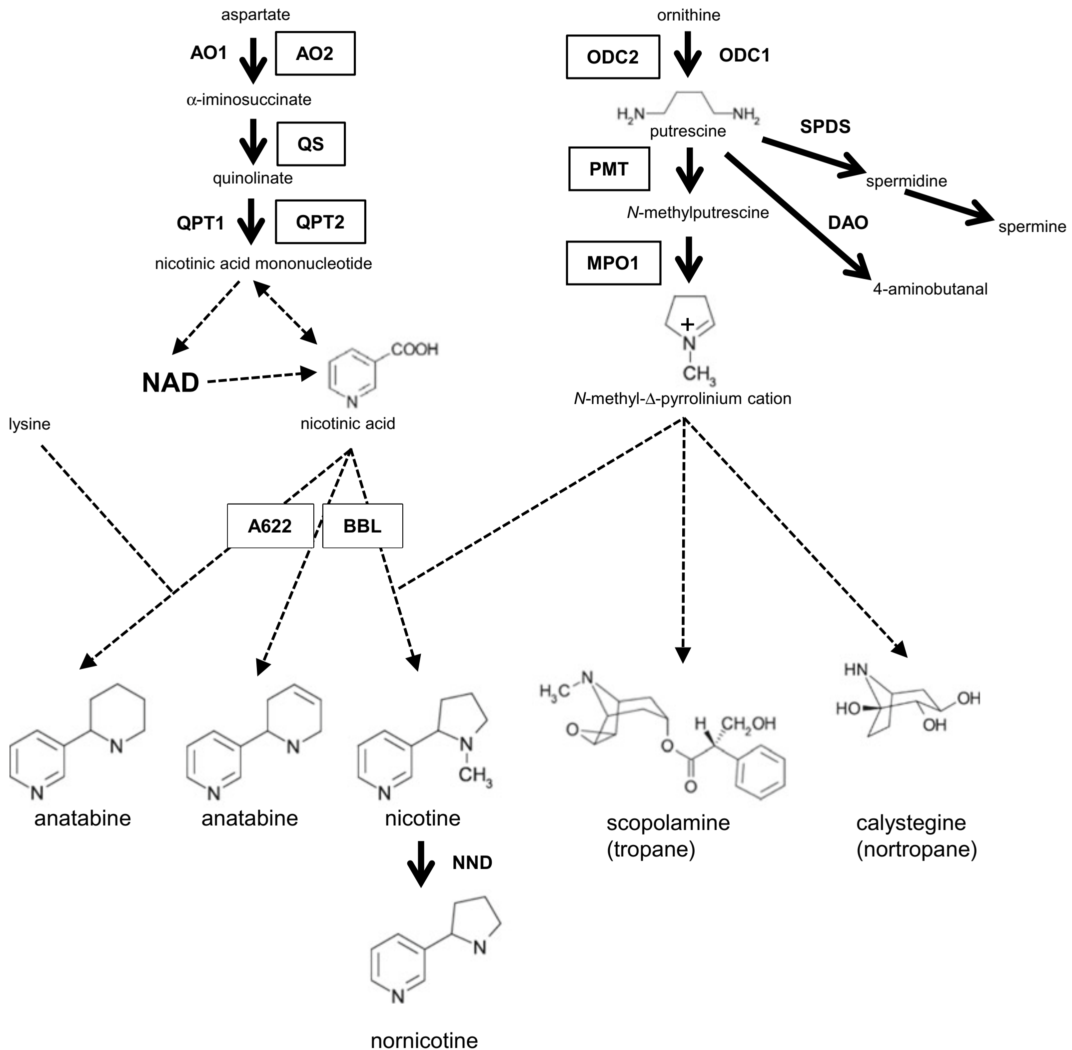

Fig. 9.1 Biosynthesis pathways of nicotine and related alkaloids. Defined steps are shown with arrows and enzyme names; dashed arrows represent undefined or multiple steps. Boxes represent enzymes regulated by ethylene response factor (ERF) transcription factors and thus included in the nicotine biosynthesis regulon. Putrescine *N*-methyltransferase (PMT) and *N*-methylputrescine oxidase (MPO1) have been proposed to evolve from spermidine synthase (SPDS) and diamine oxidase (DAO), respectively (Hibi et al. 1994; Heim et al. 2007; Katoh et al. 2007). AO, aspartate oxidase; BBL, berberine bridge enzyme-like protein; NND, nicotine *N*-demethylase; ODC, ornithine decarboxylase; QPT, quinolinate phosphoribosyltransferase; QS, quinolinate synthase

9.2 Biosynthesis

Nicotine is composed of heterocyclic pyrrolidine and pyridine rings; these rings are formed early in the pathway and are coupled together in later steps (Fig. 9.1). A five-member pyrrolidine ring is formed from ornithine via a symmetric diamine, putrescine, through three consecutive reactions catalyzed by ornithine decarboxylase (ODC) (Imanishi et al. 1998), putrescine *N*-methyltransferase (PMT) (Hibi et al. 1994), and *N*-methylputrescine oxidase (MPO) (Heim et al. 2007; Katoh et al. 2007) (Fig. 9.1). An alternative

route to putrescine from arginine exists, which includes a step catalyzed by arginine decarboxylase (ADC). However, this route is not considered to contribute greatly to nicotine biosynthesis, because transgenic suppression of *ADC,* but not of *ODC* (De Boer et al. 2011a), failed to alter nicotine content significantly (Chintapakorn and Hamill 1990). Because of their structural similarities, PMT and MPO are thought to have arisen through catalytic innovation from primary enzymes involved in related polyamine metabolism, namely, spermidine synthase and diamine oxidase, respectively (Junker et al. 2013; Naconsie et al. 2014) (Fig. 9.1). The ornithine-derived moiety is also used to produce tropane alkaloids (e.g., the clinically important hyoscyamine and scopolamine) and nortropane alkaloids (e.g., calystegines, which inhibit glycosidase) in various Solanaceae species (Shoji and Hashimoto 2015) (Fig. 9.1). The shared branch for pyrrolidine formation may have developed through duplication of the polyamine pathway. The resulting doubling of *ODC* genes may have enabled increased metabolic flow into this branch, and the subsequent innovative evolution of *PMT* and *MPO1* (Kajikawa et al. 2017a). The establishment of the ring-forming extension before the Solanaceae species diversified is underlined by the common existence of *PMT* and *MPO1* genes in Solanaceae genomes (Kajikawa et al. 2017b; Xu et al. 2017).

A pyridine ring of nicotine and related alkaloids is derived from nicotinic acid, a primary metabolite in the pathway that supplies nicotinamide adenine dinucleotide (NAD), an important co-factor for oxidation–reduction reactions. In the NAD pathway, aspartate is converted to nicotinic acid mononucleotide via quinolinate via steps catalyzed by aspartate oxidase (AO), quinolinate synthase (QS), and quinolinate phosphoribosyltransferase (QPT) (Sinclair et al. 2000; Katoh et al. 2006) (Fig. 9.1). To meet increased metabolic demands to support massive downstream production of nicotine, *AO* and *QPT* genes have been duplicated in *Nicotiana* species, but not in other lineages (Kajikawa et al. 2017b; Xu et al. 2017). Nicotinic acid is supplied as an intermediate of a cyclic pathway for de novo and salvage production of NAD, which starts with nicotinic acid mononucleotide (Noctor et al. 2006) (Fig. 9.1). It is unclear whether nicotinic acid itself or its derivatives are directly incorporated into alkaloids (Shoji and Hashimoto 2011a). To avoid excess accumulation of toxic nicotinic acid, plants have developed mechanisms to convert nicotinic acid to less-toxic derivatives (Li et al. 2015a, b; Li et al. 2017). Using nicotinic acid for alkaloid production may have originally emerged as one such detoxification reaction.

Little is known about the late steps required to couple the heterocyclic rings. It has been proposed that two oxidoreductases, A622 (De Boer et al. 2009; Kajikawa et al. 2009) and berberine bridge enzyme-like protein (BBL) (Kajikawa et al. 2011), catalyze the late stages of the pathway (Fig. 9.1), though biochemical details of their reactions are yet to be defined. A622 and BBL are required to produce not only nicotine, but also other pyridine alkaloids (De Boer et al. 2009; Kajikawa et al. 2009, 2011), implying that A622- and BBL-dependent steps are shared between the pathways that produce the different alkaloids.

Nornicotine is formed from nicotine via demethylation mediated by nicotine *N*-demethylase (NND) (Fig. 9.1), an enzyme belonging to the CYP82E subfamily of cytochrome P450 monooxygenases. Three genes, *CYP82E4*, *CYP82E5*, and *CYP82E10*, encoding functional NND enzymes have been cloned from tobacco (Siminszky et al. 2005; Gavilano and Siminszky 2007; Lewis et al. 2010). Nornicotine typically accounts for 3–5% of the total alkaloids in mature tobacco leaves. *CYP82E5* and *CYP82E10* contribute to this conventional accumulation of nornicotine (Gavilano and Siminszky 2007; Lewis et al. 2010). However, a small number of plants within tobacco populations, especially in Burley cultivars, are termed converters, as they convert over 90% of their nicotine to nornicotine during leaf senescence and curing (Griffith et al. 1955). *CYP82E4* is responsible for this unstable conversion phenotype, which depends on occasional reactivation of normally silenced *CYP82E4* in converters (Siminszky et al. 2005).

Nornicotine is reduced to nominal levels by knockout mutations of all three *CYP82E* genes (Lewis et al. 2010). Because *N′*-nitrosonornicotine is a carcinogen that is more harmful than other tobacco-specific nitrosamines and is readily formed from nornicotine during the curing process (Bush et al. 2001), reducing nornicotine levels is a desirable breeding goal.

9.3 Transport

Many plant metabolites, both end products and intermediates, move within and between cells. Various membrane-localized transporters, such as those belonging to the ATP-binding cassette, multidrug and toxic compound extrusion (MATE), and purine permease families, mediate the active transport of low-molecular-weight compounds across biological membranes, including alkaloids, which are often positively charged and thus membrane-impermeable (Shoji 2014; Shitan et al. 2014a).

To prevent cytotoxicity when accumulated at high concentrations, nicotine is sequestered into storage vacuoles in tobacco cells. Tonoplast-localized MATE transporters, *N. tabacum* jasmonate-inducible alkaloid transporter 1 (Nt-JAT1) (Morita et al. 2009), Nt-JAT2 (Shitan et al. 2014b), NtMATE1, and NtMATE2 (Shoji et al. 2009), are proton antiporters that sequester nicotine in the vacuole. These transporters couple proton gradients across the membrane with the energy-consuming uptake of alkaloids into organelles. The Nt-JAT1 and Nt-JAT2 homologs, which are phylogenetically related to xenobiotic-transporting DTX1 from *Arabidopsis*, mediate the vacuolar sequestration of nicotine in tobacco leaves (Morita et al. 2009; Shitan et al. 2014b). *NtMATE1* and *NtMATE2*, which encode homologs of flavonoid transporters, are co-expressed with nicotine biosynthesis genes, and the encoded proteins are thereby involved in the uptake of nicotine into the vacuoles in alkaloid-producing roots (Shoji et al. 2009). It is noteworthy that *NtMATE2* resides near a gene encoding a late-step enzyme A622 (see "Biosynthesis" section) on chromosome (chr) 12 in the tobacco genome; this is the only example of non-homologous clustering among nicotine pathway genes reported to date (Kajikawa et al. 2017b).

Nicotine uptake permease 1 (NUP1) is a plasma membrane-localized purine permease-family transporter of the pyridine ring-bearing molecules nicotine and vitamin B6 (e.g., pyridoxine) (Hildreth et al. 2011; Kato et al. 2014, 2015). Unlike other nicotine pathway genes expressed in inner cell layers (Shoji et al. 2000, 2002, 2009; Shoji and Hashimoto 2011b; Kajikawa et al. 2017b), *NUP1* is mainly expressed in root epidermal cells (Kato et al. 2014). In addition to its role in metabolite transport, *NUP1* is involved in the genetic regulation of nicotine biosynthesis by way of master transcription factors (see "Regulation" section) and in the regulation of root growth (Hildreth et al. 2011; Kato et al. 2014). The mechanisms underlying this regulation, however, are unclear.

Dawson (1942) conducted a classic grafting experiment that clearly demonstrated root-to-shoot transport of tobacco alkaloids between tobacco rootstock and *Solanum lycopersicum* (tomato) scion. It is known that nicotine moves up through the xylem along the transpiration stream, and that xylem loading and unloading depend on nicotine efflux from root cells and influx into leaf cells, respectively, although the transporters responsible for these processes have yet to be defined. Unlike most species in the *Nicotiana* genus, the flowering tobacco *N. alata* is devoid of alkaloids in aboveground shoots, because it lacks long-distance translocation ability (Pakdeechanuan et al. 2012). Elucidating the genetic basis of this natural variation may provide mechanistic insights into alkaloid transport.

9.4 Regulation

While primary pathways are nearly constitutive, pathways for specialized metabolites are subject to dynamic regulation in developmental and environmental contexts. Metabolic flow through a long, multistep pathway relies on the

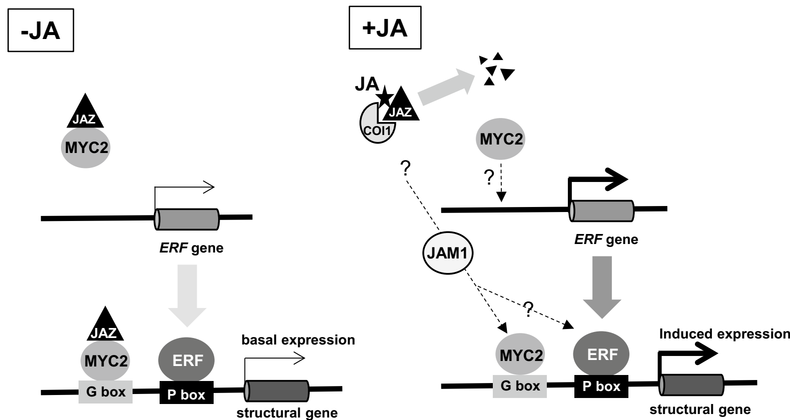

Fig. 9.2 A model for jasmonate (JA)-dependent regulation of nicotine biosynthesis in tobacco. Ethylene response factor (ERF) transcription factors regulate the structural genes of the nicotine pathway by binding at P box elements in their promoters. JA-induced transcription of the *ERF* genes is mediated by a basic helix-loop-helix (bHLH)-family MYC2 transcription factor, a direct target of upstream jasmonate ZIM-domain (JAZ) repressors. MYC2 regulates structural genes along with ERF by directly binding to G box elements. JAZ proteins are degraded when a co-receptor complex comprised of coronatine-insensitive 1 (COI1) and JAZ proteins perceives a JA signal. ERF and MYC2 transcription factors are stimulated by a phosphorylation cascade that includes the JA factor-stimulating MAPKK (JAM1). Other details of the regulation of *ERF* by MYC2 and the phosphorylation cascade are unclear

coordinated expression of metabolic and transport genes, or structural genes. Transcription factors typically control this coordination at the transcriptional level, often forming a multigene network, or regulon, with downstream structural genes (Shoji 2019).

A few transcription factors in the ethylene response factor (ERF) family, particularly ERF189 and its closest homolog ERF199, are involved in the master transcriptional regulation of nicotine pathway genes in tobacco (Dewey and Xie 2013; Shoji and Hashimoto 2013a) (Fig. 9.2). *ERF189* and *ERF199* are expressed strongly, but not exclusively, in the roots and are induced by jasmonates along with structural genes at the transcript level (Shoji et al. 2010; Kajikawa et al. 2017b). ERF transcription factors upregulate nearly the entire series of nicotine metabolic and transport genes, including *ODC2*, *PMT*, *MPO1*, *AO2*, *QS*, *QPT2*, *A622*, *BBL*, *NtMATE1*, and *NtMATE2* (Fig. 9.1), but not *NUP1*, *NtJAT1*, or *NtJAT2*, by directly recognizing specific *cis*-regulatory elements, termed P boxes, in their promoter regions (Fig. 9.2) (Todd et al. 2010; Shoji et al. 2010; De Boer et al. 2011b; Shoji and Hashimoto 2011b; Kajikawa et al. 2017b). The GC-rich P box resembles a canonical GCC box, and a few amino acid residues critical for its recognition have been identified within the DNA-binding domain of the ERF transcription factors (Shoji et al. 2013).

In the allotetraploid *N. tabacum* genome, there are two homologous clusters of related *ERF* genes, including *ERF189* and *ERF 199*. In one cluster, *ERF189* is located on chr 19, and in the other, *ERF199* is located on chr 7 (Fig. 9.3). These clusters may have originated from the diploid ancestral parents of *N. tabacum*, *N. tomentosiformis,* and *N. sylvestris*, respectively (Shoji et al. 2010; Kajikawa et al. 2017b). The *ERF189*-containing cluster corresponds to the *NIC2* locus, one of two genetic loci controlling nicotine content in tobacco (Shoji et al. 2010). A substantial chromosomal deletion (ca. 650 kb) encompassing a large portion of the cluster, including *ERF189* (Fig. 9.3), was found in a *nic2*

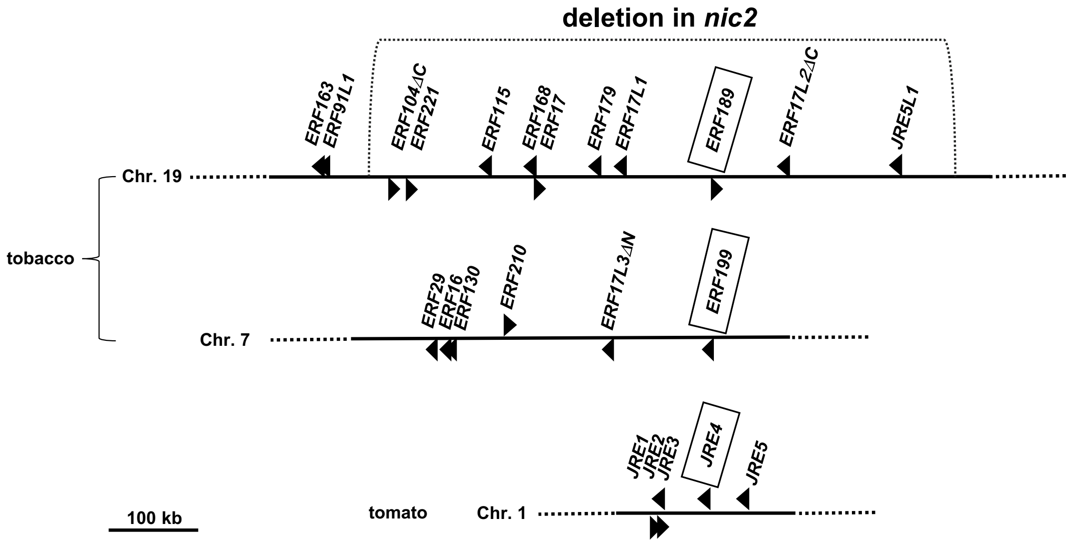

Fig. 9.3 Clustered ethylene response factor (ERF) transcription factor genes in tobacco and tomato. Arrowheads indicate positions and orientations of predicted open reading frames of *ERF* genes. A chromosomal region of tobacco chr. 19 deleted in *nic2* mutant is indicated. Boxes indicate *ERF189* and *ERF199* from tobacco, which are involved in nicotine regulation, and *JRE4* from tomato, which is involved in steroidal glycoalkaloid regulation. Tobacco *ERF* genes, denoted by Δ, possibly encode non-functional transcription factors that lack full-length DNA-binding domains

mutant allele (Kajikawa et al. 2017b) used to breed low-nicotine tobacco cultivars (Legg and Collins 1971).

Given that *ERF* genes exhibit low basal levels of expression and are induced by salt stress (Shoji and Hashimoto 2015; Kajikawa et al. 2017b), and considering the limited effects of overexpression on nicotine biosynthesis (Shoji et al. 2010), the contribution to nicotine regulation of clustered *ERF* genes other than *ERF189* and *ERF199* is believed to be limited. Experimental evidence contradicting this view, however, exists (De Boer et al. 2011b; Sears et al. 2014). Further work is required to functionally differentiate related and mostly clustered *ERF* genes in the regulation of nicotine and other pathways.

Clusters of transcription factor genes homologous to tobacco *ERF*s are present in the genomes of other species (Cárdenas et al. 2016; Thagun et al. 2016; Paul et al. 2017). For instance, *JRE4* (also known as *GAME9*) from tomato (Fig. 9.3) and potato (*Solanum tuberosum*) (Cárdenas et al. 2016; Thagun et al. 2016, Nakayasu et al. 2018) and *ORCA*s from *Catharanthus roseus* (van der Fits and Memelink 2000; Paul et al. 2017) regulate the jasmonate-induced production of specialized metabolites, suggesting that these related factors are functionally similar (Shoji and Hashimoto 2013b). Consistent with such notion, cell type-specific and jasmonate-inducible expression of a promoter reporter of tobacco *QPT2* regulated by ERF189 was found to be mediated by JRE4 in tomato (Shoji and Hashimoto 2019).

In plants, jasmonate signals are perceived by a co-receptor complex comprised of coronatine-insensitive 1 (COI1) and jasmonate ZIM-domain (JAZ) proteins. This triggers proteasome-dependent degradation of JAZs and subsequent transcriptional activation mediated by a basic helix-loop-helix (bHLH)-family MYC2 transcription factor liberated from repression by JAZs (Wasternack and Hause 2013) (Fig. 9.2). Jasmonate-induced formation of nicotine depends on COI1, JAZ, and MYC2 proteins in tobacco (Paschold et al. 2007; Shoji et al. 2008; Todd et al. 2010; De Boer et al. 2011b; Shoji and

Hashimoto 2011c; Zhang et al. 2012) (Fig. 9.2). In cooperation with ERF factors, tobacco MYC2 factors regulate a series of nicotine pathway genes directly by recognizing G box elements in their promoters (Shoji and Hashimoto 2011c; Zhang et al. 2012). These factors also act indirectly through transcriptional activation of the *ERF* genes (Shoji and Hashimoto 2011c) (Fig. 9.2). Similar schemes linking MYC2, a central player in general jasmonate signaling, to downstream ERFs and defense metabolism have been demonstrated in other plants (Zhang et al. 2011; Cárdenas et al. 2016; Paul et al. 2017). In addition to transcriptional regulation, ERF proteins are postulated to be regulated by protein phosphorylation in *N. tabacum* (De Boer et al. 2011b; Paul et al. 2017). In tobacco, a protein phosphorylation cascade involving a mitogen-activated protein kinase kinase (MAPKK), jasmonate factor-stimulating MAPKK (JAM1), has been proposed to stimulate nicotine biosynthesis via the ERF and MYC2 transcription factors (De Boer et al. 2011b) (Fig. 9.2).

A long non-coding RNA acting as an endogenous target mimicry and its corresponding microRNA predicted to target *QPT2* gene were identified and found to be involved in topping-triggered induction of nicotine accumulation (Li et al. 2015a, b), presenting an example of regulation of specialized metabolism by a module consisting of non-coding RNAs.

9.5 Perspectives

Molecular and genomic studies have greatly advanced the understanding of the nicotine biosynthesis pathway in tobacco. Complementing gene cloning efforts based mainly on homologies and expression profiles, the genome sequences of tobacco (Sierro et al. 2014) and two wild *Nicotiana* species (Xu et al. 2017) have revealed properties of the entire suite of genes involved in nicotine biosynthesis and regulation (Kajikawa et al. 2017b; Xu et al. 2017). With several structural and regulatory genes now known, it is easier to genetically manipulate the biosynthesis of the toxic alkaloids (Sato et al. 2001; Lewis et al. 2010; De Boer et al. 2011a). Recent advances have identified a critical role for transcription factors in nicotine regulation, demonstrating conserved regulatory circuits centering on jasmonate-responsive ERF transcription factors. Molecular studies of the nicotine biosynthesis pathway have elucidated how such coordinated systems were established during plant evolution (Kajikawa et al. 2017b; Shoji 2019).

Acknowledgements Thanks to Dr. Takashi Hashimoto of Nara Institute of Science and Technology for his collaboration and support. Research in the author's group was supported in part by grants from the Japan Society for the Promotion of Science (Grant-in-Aid for Scientific Research No. 17K07447) to TS.

References

Baldwin IT (1989) Mechanism of damage-induced alkaloid production in wild tobacco. J Chem Ecol 25:3–30

Bush LP, Cui M, Shi H, Burton HR (2001) Formation of tobacco specific nitrosamines in air-cured tobacco. Rec Adv Tob Sci 27:23–46

Cárdenas PD, Sonawane PD, Pollier J et al (2016) GAME9 regulates the biosynthesis of steroidal alkaloids and upstream isoprenoids in the plant mevalonate pathway. Nat Commun 7:10654

Chintapakorn Y, Hamill JD (1990) Antisense-mediated reduction in ADC activity causes minor alterations in the alkaloid profile of cultured hairy roots and regenerated transgenic plants of *Nicotiana tabacum*. Phytochemistry 68:2465–2479

Dawson RF (1942) Accumulation of nicotine in reciprocal grafts of tomato and tobacco. Am J Bot 29:66–71

De Boer K, Lye JC, Aitken CD, Su AK, Hamill JD (2009) The *A622* gene in *Nicotiana glauca* (tree tobacco): evidence for a functional role in pyridine alkaloid synthesis. Plant Mol Biol 69:299–312

De Boer K, Dalton HL, Edward FJ, Hamill JD (2011a) RNAi-mediated down-regulation of ornithine decarboxylase (ODC) leads to reduced nicotine and increased anatabine levels in transgenic *Nicotiana tabacum* L. Phytochemistry 72:344–355

De Boer K, Tileman S, Pauwels L et al (2011b) APETALA2/ethylene response factor and basic helix-loop-helix transcription factors cooperatively mediate jasmonate-elicited nicotine biosynthesis. Plant J 66:1053–1065

Dewey RE, Xie J (2013) Molecular genetics of alkaloid biosynthesis in *Nicotiana tabacum*. Phytochemistry 94:10–27

Gavilano LB, Siminszky B (2007) Isolation and characterization of the cytochrome P450 gene *CYP82E5v2* that mediates nicotine to nornicotine conversion in the green leaves of tobacco. Plant Cell Physiol 48:1567–1574

Griffith RB, Valleau WD, Stokes GW (1955) Determination and inheritance of nicotine to nornicotine conversion in tobacco. Science 121:343–344

Heim WG, Sykes KA, Hildreth SB et al (2007) Cloning and characterization of a *Nicotiana tabacum* methylputrescine oxidase transcript. Phytochemistry 68:454–463

Hibi N, Higashiguchi S, Hashimoto T, Yamada Y (1994) Gene expression in tobacco low-nicotine mutants. Plant Cell 6:723–735

Hildreth SB, Gehman EA, Yang H et al (2011) Tobacco nicotine uptake permease 1 (NUP1) affects alkaloid metabolism. Proc Natl Acad Sci USA 108:18179–18184

Imanishi S, Hashizume K, Nakakita M et al (1998) Differential induction of methyl jasmonate of genes encoding ornithine decarboxylase and other enzymes involved in nicotine biosynthesis in tobacco cell culture. Plant Mol Biol 38:1101–1111

Junker A, Fischer J, Sichhart Y, Brandt W, Dráger B (2013) Evolution of the key alkaloid enzyme putrescine N-methyltransferase from spermidine synthase. Front Plant Sci 4:260

Kajikawa M, Hirai N, Hashimoto T (2009) A PIP-family protein is required for biosynthesis of tobacco alkaloids. Plant Mol Biol 69:287–298

Kajikawa M, Shoji T, Katoh A, Hashimoto T (2011) Vacuole-localized berberine bridge enzyme-like proteins are required for a late step of nicotine biosynthesis in tobacco. Plant Physiol 155:2010–2022

Kajikawa M, Sierro N, Hashimoto T, Shoji T (2017a) A model for evolution and regulation of nicotine biosynthesis regulon in tobacco. Plant Signal Behav 12(6):e1338225

Kajikawa M, Sierro N, Kawaguchi H et al (2017b) Genomic insights into the evolution of the nicotine biosynthesis pathway in tobacco. Plant Physiol 174:999–1011

Kato K, Shitan N, Shoji T, Hashimoto T (2015) Tobacco NUP1 transports both tobacco alkaloids and vitamin B6. Phytochemistry 113:33–40

Kato K, Shoji T, Hashimoto T (2014) Tobacco nicotine uptake permease regulates the expression of a key transcription factor gene in the nicotine biosynthesis pathway. Plant Physiol 166:2195–2204

Katoh A, Shoji T, Hashimoto T (2007) Molecular cloning of *N*-methylputrescine oxidase from tobacco. Plant Cell Physiol 48:550–554

Katoh A, Uenohara K, Akita M, Hashimoto T (2006) Early steps in the biosynthesis of NAD in Arabidopsis start with aspartate and occur in the plastid. Plant Physiol 141:851–857

Legg PG, Collins GB (1971) Inheritance of percent total alkaloids in *Nicotiana tabacum* L. II. Genetic effects of two loci in Burley21 × LA Burley21 population. Can J Genet Cytol 13:287–291

Lewis RS, Bowen SW, Keogh MR, Dewey RE (2010) Three nicotine demethylase genes mediate nornicotine biosynthesis in *Nicotiana tabacum* L.: functional characterization of the *CYP82E10* gene. Phytochemistry 71:1988–1998

Li F, Wang W, Zhao N et al (2015a) Regulation of nicotine biosynthesis by an endogenous target mimicry of microRNA in tobacco. Plant Physiol 169:1062–1071

Li W, Zhang F, Chang Y et al (2015b) Nicotine *O*-glucosylation is an evolutionally metabolic trait important for seed germination under stress conditions in *Arabidopsis thaliana*. Plant Cell 27:1907–1924

Li W, Zhang F, Wu R et al (2017) A novel *N*-methyltransferase in Arabidopsis appears to feed a conserved pathway for nicotinate for nicotinate detoxification among land plants and is associated with lignin biosynthesis. Plant Physiol 174(3):1492–1504

Morita M, Shitan N, Sawada K et al (2009) Vacuolar transport of nicotine is mediated by a multidrug and toxic compound extrusion (MATE) transporter in *Nicotiana tabacum*. Proc Natl Acad Sci USA 106:2447–2452

Naconsie M, Kato K, Shoji T, Hashimoto T (2014) Molecular evolution of *N*-methylputrescine oxidase in tobacco. Plant Cell Physiol 55:436–444

Nakayasu M, Shioya N, Shikata M et al (2018) JRE4 is a master transcriptional regulator of defense-related steroidal glycoalkaloids in tomato. Plant J 94:975–990

Noctor G, Queval G, Gakiere B (2006) NAD(P) synthesis and pyridine nucleotide cycling in plants and their potential importance in stress conditions. J Exp Bot 57:1603–1620

Pakdeechanuan P, Shoji T, Hashimoto T (2012) Root-to-shoot translocation of alkaloids is dominantly suppressed in *Nicotiana alata*. Plant Cell Physiol 53:1247–1254

Paschold A, Hailtschke R, Baldwin IT (2007) Co(i)-ordinating defenses: NaCOI1 mediates herbivore-induced resistance in *Nicotiana attenuata* and reveals the role of herbivore movement in avoiding defenses. Plant J 51:79–91

Paul P, Singh SK, Patra B et al (2017) A differentially regulated AP2/ERF transcription factor gene cluster acts downstream of a MAK kinase cascade to modulate terpenoid indole alkaloid biosynthesis in *Catharanthus roseus*. New Phytol 213:1107–1123

Saitoh K, Noma M, Kawashima N (1985) The alkaloid contents of sixty *Nicotiana* species. Phytochemistry 24:477–480

Sato F, Hashimoto T, Hachiya A et al (2001) Metabolic engineering of plant alkaloid biosynthesis. Proc Natl Acad Sci USA 98:367–372

Sears MT, Zhang H, Rushton PJ et al (2014) NtERF32: a non-*NIC2* locus AP2/ERF transcription factor required in jasmonate-inducible nicotine biosynthesis in tobacco. Plant Mol Biol 84:49–66

Shitan N, Kato K, Shoji T (2014a) Alkaloid transporters in plants. Plant Biotechnol 31:453–463

Shitan N, Minami S, Morita M et al (2014b) Involvement of the leaf-specific multidrug and toxic compound extrusion (MATE) transporter Nt-JAT2 in vacuolar sequestration of nicotine in *Nicotiana tabacum*. PLoS ONE 9:e108789

Shoji T (2014) ATP-binding cassette and multidrug and toxic compound extrusion transporters in plants: a common theme among diverse detoxification mechanisms. Int Rev Cell Mol Biol 309:308–346

Shoji T (2016) Alkaloid biosynthesis and regulation in plants. In: Arimura G, Maffei M (eds) Plant specialized metabolism: genomics, biochemistry, and biological functions. CRC Press, Boca Raton, pp 85–118

Shoji T (2019) The recuirtment model of metabolic evolution: jasmonate-responsive transcription factors and a conceptual model for the evolution of metabolic pathways. Front Plant Sci 10:e560

Shoji T, Hashimoto T (2011a) Nicotine biosynthesis. In: Ashihara H, Crozier A, Komamine A (eds) Plant metabolism and biotechnology, John Wiley & Sons, New York, pp 191–216

Shoji T, Hashimoto T (2011b) Recruitment of a duplicated primary metabolism gene into the nicotine biosynthesis regulon in tobacco. Plant J 67:949–959

Shoji T, Hashimoto T (2011c) Tobacco MYC2 regulates jasmonate-inducible nicotine biosynthesis genes directly and by a way of the *NIC2*-locus *ERF* genes. Plant Cell Physiol 52:1117–1130

Shoji T, Hashimoto T (2013a) Smoking out the masters: transcriptional regulators for nicotine biosynthesis in tobacco. Plant Biotechnol 30:217–224

Shoji T, Hashimoto T (2013b) Jasmonate-responsive transcription factors: new tools for metabolic engineering and gene discovery. In: Chandra S, Lata H, Varma A (eds) Biotechnology for medicinal plants: micropropagation and improvement. Springer Publishing, New York, pp 345–357

Shoji T, Hashimoto T (2015) Stress-induced expression of *NICOTINE2*-locus genes and their homologs encoding Ethylene Response Factor transcription factors in tobacco. Phytochemistry 113:41–49

Shoji T, Hashimoto T (2019) Expression of a tobacco nicotine biosynthesis gene depends on the JRE4 transcription factor in heterogonous tomato. J Plant Res 132:173–180

Shoji T, Inai K, Yazaki Y et al (2009) Multidrug and toxic compound extrusion-type transporters implicated in vacuolar sequestration of nicotine in tobacco roots. Plant Physiol 149:708–718

Shoji T, Kajikawa M, Hashimoto T (2010) Clustered transcription factor genes regulate nicotine biosynthesis in tobacco. Plant Cell 22:3390–3409

Shoji T, Mishima M, Hashimoto T (2013) Divergent DNA-binding specificities of a group of ETHYLENE RESPONSE FACTOR transcription factors involved in plant defense. Plant Physiol 162:977–990

Shoji T, Ogawa T, Hashimoto T (2008) Jasmonate-induced nicotine formation in tobacco is mediated by tobacco *COI1* and *JAZ1* genes. Plant Cell Physiol 49:1003–1012

Shoji T, Winz R, Iwase T et al (2002) Expression patterns of two tobacco isoflavone reductase-like genes and their possible roles in secondary metabolism in tobacco. Plant Mol Biol 50:427–440

Shoji T, Yamada Y, Hashimoto T (2000) Jasmonate induction of putrescine *N*-methyltransferase genes in the root of *Nicotiana sylvetris*. Plant Cell Physiol 41:831–839

Sierro N, Battey JN, Ouadi S et al (2014) The tobacco genome sequence and its comparison with those of tomato and potato. Nat Commun 5:3833

Siminszky B, Gavilano L, Bowen SW, Dewey RE (2005) Conversion of nicotine to nornicotine in *Nicotiana tabacum* is mediated by CYP82E4, a cytochrome P450 monooxygenase. Proc Natl Acad Sci USA 102:14919–14924

Sinclair SJ, Murphy KJ, Birch CD, Hamill JD (2000) Molecular characterization of qunolinate phosphoribosyltransferase (QPRtase) in *Nicotiana*. Plant Mol Biol 44:603–617

Steppuhn A, Gase K, Krock B, Halitschke R, Baldwin IT (2004) Nicotine's defensive function in nature. PLoS Biol 2:1074–1080

Thagun C, Imanishi S, Kudo T et al (2016) Jasmonate-responsive ERF transcription factors regulate steroidal glycoalkaloid biosynthesis in tomato. Plant Cell Physiol 57:961–975

Todd AT, Liu E, Polvi SL, Pammett RT, Page JE (2010) A functional genomics screen identifies diverse transcription factors that regulate alkaloid biosynthesis in *Nicotiana benthamiana*. Plant J 62:589–600

van der Fits L, Memelink J (2000) ORCA3, a jamonate-responsive transcriptional regulator of plant primary and secondary metabolism. Science 289:295–297

Wasternack C, Hause B (2013) Jasmonates: biosynthesis, perpection, signal transdaction and action in plant stress response, growth and development: an update to

the 2007 review in Annals of Botany. Ann Bot 11:1021–1058
Xu S, Brockmöller T, Navarro-Quezada A et al (2017) Wild tobacco genomes reveal the evolution of nicotine biosynthesis. Proc Natl Acad Sci USA 114:6133–6138
Zhang H, Hedhili S, Montiel G et al (2011) The basic helix-loop-helix transcription factor CrMYC2 controls the jasmonate-responsive expression of the *ORCA* genes regulating alkaloid biosynthesis in *Catharanthus roseus*. Plant J 67:61–71
Zhang HB, Bokowiec MT, Rushton PJ, Han SC, Timko MP (2012) Tobacco transcription factors NtMYC2a and NtMYC2b form nuclear complexes with the NtJAZ1 repressor and regulate multiple jasmonate-inducible steps in nicotine biosynthesis. Mol Plant 5:73–84

10 Impact of Genetics and Production Practices on Tobacco-Specific Nitrosamine Formation

Dongmei Xu, Marcos F. Lusso and James A. Strickland

Abstract

Tobacco (*Nicotiana tabacum* L.) plants biochemically synthesize multiple alkaloid compounds. The predominant compound, nicotine, and other alkaloids in the plant can be nitrosated to form tobacco-specific N-nitrosamines (TSNAs). TSNAs are found in cured tobacco leaf and are created primarily during the curing process. Some TSNAs have been classified by regulatory authorities as human carcinogens. Therefore, the mechanisms associated with TSNA formation in tobacco have long attracted research and industry interest. This chapter describes TSNA formation and factors affecting TSNA levels in tobacco, including alkaloid biosynthesis, nitrite compounds, tobacco type, TSNA formation propensity, environmental conditions, and production practices. It further expounds on the genetic aspects related to TSNA formation and TSNA-mitigation methods, such as the use of low-alkaloid tobacco lines, alteration of gene expression in alkaloid biosynthesis genes and regulatory genes, and nicotine transport mechanisms. Discovery of nicotine demethylase and reductions in N-nitrosonornicotine through down-regulation or knockout of genes encoding nicotine demethylase are described. Research and industry efforts to mitigate TSNA formation in the cured leaf and tobacco products are also discussed.

D. Xu (✉) · M. F. Lusso · J. A. Strickland
Altria Client Services, Research Development & Sciences, Richmond, VA, USA
e-mail: Dongmei.Xu@altria.com

10.1 Introduction

Nitrosamines are organic compounds characterized by an attached nitrous oxide moiety. N-Nitrosamines are formed by a chemical reaction between a nitrosating agent and a secondary or tertiary amine (Scanlan 1983) and are found in tobacco products, cured meats, cooked bacon, beer, some cheeses, nonfat dry milk, and sometimes in fish (Scanlan 1983). In foods, the nitrosating agent is typically nitrous anhydride, which is formed from nitrite in an aqueous acidic solution (Scanlan 1983). N-nitrosodimethylamine, N-nitrosopyrrolidine, N-nitrosodiethylamine, and N-nitrosoproline are some nitrosamines that can be found in foods (Scanlan 1983).

Tobacco-specific N-nitrosamines (TSNAs) are a class of nitrosamine compounds found in cured unburned tobacco and in tobacco smoke (Appleton et al. 2013; Gunduz et al. 2016; Hoffmann et al. 1994; Lusso et al. 2017). Researchers have focused on these compounds because of their potential toxicological relevance and because this specific class of nitrosamines is

© Springer Nature Switzerland AG 2020
N. V. Ivanov et al. (eds.), *The Tobacco Plant Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-030-29493-9_10

found only in tobacco (Hecht et al. 1978a, 1980, 1984). The International Agency for Research on Cancer (IARC) Working Group has evaluated the available scientific evidence for four TSNAs, 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK), N′-nitrosonornicotine (NNN), N′-nitrosoanabasine (NAB), and N′-nitrosoanatabine (NAT), in an effort to determine the carcinogenic potential of these compounds (IARC Working Group on the Evaluation of Carcinogenic Risk to Humans 2007). In 2007, IARC classified NNK and NNN as Group 1 carcinogens (carcinogenic to humans) and NAB and NAT as Group 3 compounds (not classifiable as to their carcinogenicity to humans). IARC's classification of NNN and NNK was based on carcinogenicity evidence of the various purified compounds tested primarily in experimental animals. IARC reported inadequate evidence of carcinogenicity by NAB and NAT in humans. The United States National Toxicology Program has declared NNN and NNK "to be reasonably anticipated to be a human carcinogen" based on sufficient evidence of carcinogenicity from studies in experimental animals (National Institute of Environmental Health Sciences 2011). Subsequently, the United States Food and Drug Administration included NNN and NNK on its list of Harmful and Potentially Harmful Constituents in tobacco products and tobacco smoke (US Food and Drug Administration 2018).

10.2 TSNA Formation in *Nicotiana tabacum*

Precursors leading to the formation of TSNAs are the naturally occurring alkaloids found in tobacco (Fig. 10.1) and certain nitrogen compounds present or generated during leaf curing. Alkaloids are a structurally diverse class of nitrogen-containing ring compounds that are produced by many organisms as a defense mechanism against insects and pathogens. Four major alkaloids are found in tobacco: nicotine, nornicotine, anabasine, and anatabine. Nicotine is the predominant alkaloid,

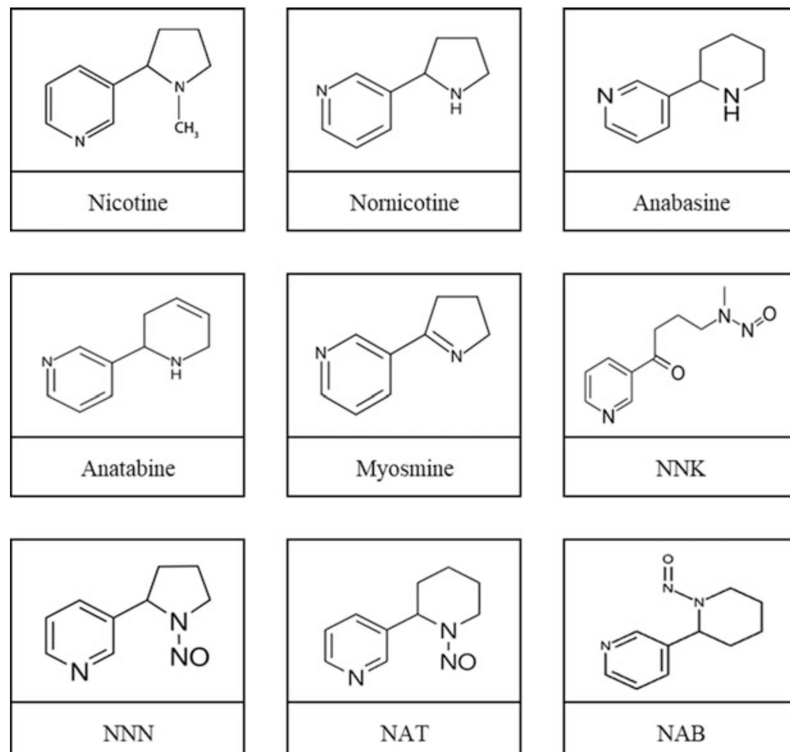

Fig. 10.1 Chemical structures of alkaloids and tobacco-specific N-nitrosamines. NAB, N′-nitrosoanabasine; NAT, N′-nitrosoanatabine; NNK, 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone; NNN, N′-nitrosonornicotine

Table 10.1 Percentage of nicotine and minor alkaloids in Burley and bright tobacco types

	Nicotine	Nornicotine	Anabasine	Anatabine	Myosmine
Total alkaloids (%)	88–92	2–4	0.5–1	3–4	0.2–0.4
Burley (%)	91	4	0.5	3	0.4
Bright (%)	92	1.5	1	4	0.2

Modified from Piade and Hoffmann (1980)

typically accounting for more than 90–95% of the total alkaloids in commercial tobacco cultivars (Table 10.1). It is hypothesized that nicotine, a tertiary alkaloid, must first undergo oxidation of the 1′,2′ bond of the pyrrolidine ring to form pseudo-oxynicotine, followed by nitrosation to produce NNK (Caldwell et al. 1991; Hecht et al. 1978b). In contrast, the remaining secondary alkaloids, nornicotine, anatabine, and anabasine, can react directly with nitrite or any of the nitrogen oxide (NO_x) species present during the curing process to form NNN, NAT, and NAB, respectively (Burton et al. 1994).

TSNAs are not present at significant levels in growing tobacco plants or fresh-cut tobacco (green tobacco) but are formed during the curing process (Wiernik et al. 1995). Bacterial populations that reside on and within tobacco leaves facilitate the reduction of nitrate to nitrite during the curing process and affect the direct catalysis of secondary amines at the physiological pH of the plant tissues. When tobacco alkaloids, which by definition have secondary amines, are nitrosated, TSNAs are formed. There is no evidence that TSNAs can revert to the non-nitrosated alkaloid state within tobacco once formed. Nitrosation of other secondary amines can also occur, resulting in the formation of other compounds, such as nitrosated amino acids and proteins. Indiscriminate nitrosation of these types of secondary amine compounds can occur, although their toxicological relevance has not been established.

Burton et al. (1989) studied changes in TSNA content in burley tobacco leaves at various maturity stages under two air-curing environments. Under normal temperature and humidity curing conditions (24 °C, relative humidity 70%), an increase in TSNA concentration occurred during the curing period from 2 weeks after housing to leaf browning. Leaf lamina and midrib TSNA levels were initially low and increased substantially during weeks 4–7 of the curing process. During air curing, water loss and cell exudates increased, indicating that cellular membrane integrity was impaired. Breakdown of the tobacco cellular content provided a food source for microbial growth, and bacterial populations capable of reducing nitrate to nitrite increased, leading to an increase in the concentration of nitrite and subsequent TSNA formation. Law et al. (2016) analyzed air-cured burley tobacco TN90H samples by 16S rRNA gene amplicon sequencing to investigate the role of leaf microbial communities in the formation of TSNAs. They found that, although proteobacteria predominated at the phylum level (90% of the operational taxonomic units), distinct changes in TSNA content were mainly attributable to increases in *Firmicutes* and *Actinobacteria*.

During nitrogen assimilation, the plant modulates nitrite levels because of its inherent cellular toxicity. To mitigate toxicity, nitrite is degraded to ammonia and then incorporated into amino acids. Accordingly, nitrite levels in green tobacco tissues are low. Nitrate can also be converted to nitric oxide rather than to nitrite by nitrate-reducing bacteria. However, nitrite levels increase during the curing of tobacco leaves because organic nitrogen, ammonia, and nitrate are converted by nitrate-reducing bacteria to nitrite or NO_x (e.g., nitrite esters and *S*-nitroso amino acids) resulting in TSNA formation. Lu et al. (2016) found that a constitutively active nitrate reductase variant in burley tobaccos dramatically decreased the nitrate content of the leaves, and this phenotype was accompanied by substantial reductions in the TSNA content of the cured leaf.

10.3 TSNA Distribution Within Tobacco Plants

TSNA levels in cured leaves are distributed differently within the plant and from the tip to the base of the leaf. Burton et al. (1992) reported that TSNAs were highly correlated with nitrite distribution within the leaf of air-cured tobacco, and that the base of the leaf had the highest TSNA levels. The top-to-bottom distribution of TSNAs in the plant was not consistent but generally followed the nitrite gradient levels within the tissue (Burton et al. 1992; Djordjevic et al. 1989).

Total alkaloid content is, in general, 10 times higher in the lamina than in the tobacco stem. Accordingly, TSNA levels in the stem can be very low (de Roton et al. 2005). However, some studies have shown that TSNA levels in the stem can increase because of its high nitrogen levels. These high nitrogen levels can increase the stoichiometric levels of nitrite, resulting in higher TSNA levels over time in the stem relative to the corresponding laminae (Burton et al. 1994). Brunnemann et al. (1983) studied the role of tobacco stems in the formation of TSNA in tobacco and mainstream and side-stream cigarette smoke, using cigarettes without filters prepared with 70% of a defined blend of laminae and 30% of various types of stems. The resulting nitrate contents ranged between 0.68 and 2.65% by dry weight. NNN (1.02–6.35 parts per million [ppm]) and the sum of total TSNAs (1.95–8.57 ppm) were strongly correlated with the nitrate content of tobacco (Pearson correlation coefficients $R^2 = 0.978$ and 0.948, respectively). The highest nitrite content was found in stems.

10.4 Variations in TSNA Levels Among Tobacco Types and Varieties

Typical TSNA levels vary between tobacco types (dark, burley, and flue-cured) because of tobacco genetics, agronomic practices, nitrogen content, and curing methods. For example, burley tobacco typically contains much higher levels of nitrate and 8–10 times higher levels of total TSNAs than flue-cured tobacco. Nitrate content has also been found to correlate positively with TSNA content between samples from the same variety and growing area (Shi et al. 2013).

The generally higher level of TSNAs in burley tobacco is partially due to its genetic properties. Burley tobacco, unlike other types, is homozygous for the recessive mutant alleles *yb1* and *yb2*, which are associated with reduced nitrogen utilization efficiency and increased total nitrogen-containing compounds in the plants. This increase in the total plant nitrogen pool leads to higher leaf NO_x during curing, and thus, likely contributes to higher TSNA levels (Edwards et al. 2017; Lewis et al. 2012). Burton et al. (1994) showed that when flue-cured tobaccos were grown and air-cured as is done with burley tobacco, TSNA levels in the flue-cured-type leaves remained lower than the TSNA levels in the corresponding burley tobacco leaves cured under normal conditions. Therefore, the genetics of the burley tobacco type are involved in the propensity to facilitate TSNA formation.

10.5 Factors Affecting TSNA Formation

A variety of factors affect the formation of TSNAs in tobacco, including alkaloid levels as a function of tobacco genotype, environment, fertilization, agronomic practices, curing procedures, and storage conditions. The main contributors are alkaloid and nitrate/nitrite levels in the plants and environmental conditions during curing and storage.

Alkaloid biosynthesis occurs predominantly in tobacco roots. Tobacco plants then transport nicotine through the vascular bundle to the leaves where nicotine is stored in the vacuoles. Alkaloid levels in tobacco are affected by its genetic makeup, the environment in which the plant is grown, harvested, and cured, and cultural practices (Burton et al. 1994; de Roton et al. 2005). Although TSNA levels are not always correlated with nicotine levels in cured tobacco leaves,

plants with low nicotine levels generally exhibit correspondingly lower TSNA content (Djordjevic et al. 1989; MacKown et al. 1984).

Nitrate content and nitrate/nitrite conversion are the major factors that affect TSNA levels in cured tobacco. Nitrite is formed mainly by microbial reduction of nitrate during the browning and stem-drying stages of the air-curing process and during the yellowing stage of the fire-curing process. Any conditions that influence nitrite formation will lead to higher TSNAs in the cured tobacco. There is also evidence that TSNAs may continue to be formed in cured tobacco when the tobacco is kept hanging in a barn after the end of curing and during storage (de Roton et al. 2005; Shi et al. 2013).

Nitrogen that accumulates in the tobacco leaf during plant growth directly affects TSNA levels after curing. For example, increased fertilization while growing dark fire-cured tobacco varieties can result in incremental increased levels of TSNAs in cured leaves (Fisher et al. 2012).

10.6 Genetic Control of TSNAs Through Alkaloid Biosynthesis

Alkaloid biosynthesis in tobacco plants is largely understood, and alteration of associated gene expression levels is known to lead to alterations in TSNA content in cured leaves. Briefly, alkaloid biosynthesis generally involves diversion of amino acid precursors from primary to secondary metabolism. The biosynthesis pathway involves a series of decarboxylation reactions by various enzymes. Exposure of plants to biotic and/or abiotic stress conditions can enhance the activities that lead to variations in alkaloid levels (Shoji and Hashimoto 2014).

Biosynthesis of alkaloids in tobacco involves a number of critical genes, including genes encoding putrescine N-methyltransferase (PMT), quinolinic acid phosphoribosyltransferase (QPT), N-methylputrescine oxidase (MPO), berberine bridge enzyme-like (BBL), and isoflavone reductase-like protein (A622). In the first committed step in alkaloid biosynthesis, PMT catalyzes the formation of N-methylputrescine (Hibi et al. 1992), which is then oxidized by a diamine oxidase and cyclized spontaneously to form the 1-methyl-Δ′-pyrrolinium cation that is condensed with nicotinic acid or its derivative by an unknown enzyme to form nicotine (Hashimoto and Yamada 1993). QPT serves as the initial enzyme in the pyridine nucleotide cycle that supplies nicotinic acid (Wagner and Wagner 1985). Nornicotine is the demethylation product of nicotine (Fannin and Bush 1992). Anabasine is formed by coordinated decarboxylation and condensation of one pyridine ring shared with nicotine formation (Solt et al. 1960) and a piperidine ring that originated from lysine (DeBoer et al. 2013; Lee and Cho 2001; Leete 1980). For anatabine, both rings originate from the pyridine nucleotide pathway; 3,6-dihydronicotinic acid is decarboxylated to form 2,5-dihydropyridine or 1,2-dihydropyridine and subsequently hydrogenated (Dewey and Xie 2013; Leete and Mueller 1982; Leete and Slattery 1976).

Tobacco lines where some of these genes are down-regulated have been reported to exhibit altered alkaloid levels. For example, down-regulation of PMT genes by RNA interference resulted in tobacco lines with substantially reduced nicotine content (Kudithipudi et al. 2016). PMT down-regulation in combination with the *Nic1–Nic2* regulatory loci has been predicted to lead to further reductions in alkaloids and TSNAs. Other genes involved in the nicotine biosynthetic pathway, including those encoding QPT, MPO, and BBL, have been down-regulated using antisense and/or RNA interference to result in low-nicotine lines (Kajikawa et al. 2011; Lewis et al. 2015; Shoji and Hashimoto 2011; Xie et al. 2004). Down-regulation of the genes encoding nicotine biosynthetic enzymes and ornithine decarboxylase also resulted in reduced nicotine levels and a concomitant increase in anatabine content (DeBoer et al. 2011; Wang et al. 2009). Overall, these studies showed that significant reductions in nicotine levels can lead to notable reductions in TSNAs, mainly NNN and NNK.

10.7 TSNA Reduction Through Existing Low-Alkaloid Lines

Great efforts have been made to reduce TSNA levels in cured tobacco. Examples include genetic breeding, control of nitrogen fertilization, implementation of modified curing conditions, and addition of nitrite or alkaloid scavengers. Krauss et al. (2003) proposed increasing the levels of antioxidants in the plant prior to harvest as a means of reducing TSNA content in cured tobacco. Specifically, they suggested root pruning and/or severing the xylem tissue of the tobacco plant prior to harvesting, and administering antioxidants and/or chemicals to increase antioxidants to the tobacco plant after harvesting. Simpler and more uniform, economical, and non-labor-intensive methods are desirable for reducing TSNA levels in cured tobacco leaves. Reduction of alkaloids is one such effective means to reduce TSNA levels.

Low-alkaloid traits initially found in strains of Cuban cigar tobacco varieties were introduced into burley tobacco varieties through a series of backcrosses (Valleau 1949). Subsequently, low-alkaloid tobacco germplasm was registered in the Burley 21 genetic background (Legg and Collins 1971). Genetic studies of low-alkaloid Burley 21 (LA BU21) lines indicated that two unlinked loci (*Nic1* and *Nic2*) contributed to nicotine levels in the tobacco leaf. The *Nic1* and *Nic2* mutations in LA BU21 are semi-dominant; they show dose-dependent effects on nicotine levels, with *Nic1* exerting effects that are approximately 2.4 times stronger than those of *Nic2*. Molecular characterization of the *Nic2* locus showed that the *Nic2* mutation involved a deleted cluster of transcription factor genes from the ethylene response factor (ERF) family (Shoji and Hashimoto 2012).

The *Nic1* locus has been characterized recently by Altria Client Services through tobacco whole-genome sequencing, resequencing of low-alkaloid isogenic lines, and bacterial artificial chromosome library-based genome walking (Adams et al. 2016). *Nic1* was shown to have 15 open reading frames and one further downstream small deletion. Further characterization of deleted genes within the *Nic1* locus is under investigation. Genetic control of alkaloid formation through *Nic1* and *Nic2* deletions led to a low-TSNA phenotype in cured leaves (Table 10.2), thus suggesting that deletion of key genes within the *Nic1* and *Nic2* loci may also result in low TSNAs because of reduced alkaloid biosynthesis.

Other regulatory genes are also shown to control nicotine formation. A MYC family of transacting factors and ERF gene promoters from the *Nic2* locus have been shown to bind to the PMT promoter (Shoji and Hashimoto 2011). Down-regulating the nicotine regulatory *MYC2* (Shoji and Hashimoto 2011) and *ERF189* (Shoji and Hashimoto 2012) genes can also lead to low-nicotine tobacco (Dewey and Xie 2013) and potentially result in TSNA reduction. Kajikawa et al. (2017) characterized a suite of structural genes involved in nicotine biosynthesis by phylogenetic and expression analyses of the tobacco genome sequence. This series of genes, which included *PMT*, *MPO*, *A622*, *BBL*, and *MATE* (multidrug and toxic compound extrusion transporters), formed a regulon that was under the control of jasmonate-responsive ERF transcription factors. (Jasmonates are phytohormones that

Table 10.2 Percentage of total alkaloids in burley tobacco types with differing genetic control of alkaloid biosynthesis

Tobacco type		Total alkaloids (%)	Tobacco-specific N-nitrosamines (ppm)			
			NNN	NNK	NAB	NAT
Burley	Normal alkaloid	6.60	7.74	2.04	0.33	6.44
	Low alkaloid	0.14	0.37	0.04	0.01	0.11

Abbreviations: NAB = N′-nitrosoanabasine; NAT = N′-nitrosoanatabine; NNK = 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone; NNN = N′-nitrosonornicotine
Authors, 2016, unpublished data

have a central signaling role in nicotine formation in *N. tabacum.*) Cognate cis-elements occurred frequently in the promoter regions of the nicotine pathway genes regulated by ERF189 and MYC2 transcription factors, which supported the hypothesis that the transcription of these genes was regulated by ERF and MYC2.

10.8 Leaf Quality of Low-Alkaloid Lines

Although the existing low-alkaloid isogenic lines derived from *Nic1*/*Nic2* deletions are not substantially different from their normal-alkaloid parents in days to flower, number of leaves, leaf size, and plant height, they produce lower quality leaves and are more susceptible to insect damage, probably because of the low nicotine content (Legg et al. 1977; Legg and Collins 1988). A study of isogenic flue-cured tobacco lines with low total alkaloid levels compared with their "normal" recurring parents reported that yield, grade index, total nitrate/nitrite, and reducing sugar content were lower in the low-alkaloid lines than in normal flue-cured cultivars (Chaplin and Weeks 1976). In general, low-alkaloid lines produce chlorophyll-rich leaves that, when cured, result in sensorial characteristics that are inferior to those of their normal alkaloid counterparts.

The low-alkaloid lines produced through down-regulation or deletion of the biosynthesis genes may provide better quality leaves than plants containing the *Nic1*/*Nic2* deletions (Kudithipudi et al. 2016). Overall, production of low-alkaloid lines with leaf quality acceptable to manufacturers remains challenging. Nonetheless, low-nicotine lines typically have lower TSNA levels; the use of these lines is one strategy to reduce TSNA levels in tobacco and tobacco products.

10.9 TSNA Reduction Through Nicotine Transport

Nicotine and related pyridine alkaloids accumulate in the leaves of tobacco plants, although no substantial synthesis of alkaloids occurs in the leaf part of the plant. Root-to-leaf translocation of tobacco alkaloids was first demonstrated unambiguously in interspecific grafting experiments between tomato (*Solanum lycopersicum*) and tobacco (Dawson 1942). Nicotine accumulated in tomato leaves when tomato scions were grafted onto tobacco stocks, but not in the reverse graft combination. Nicotine is translocated from the root to the leaf through the xylem, and various transporters may be involved. A number of transporter genes have been studied for their impact on nicotine levels in tobacco (Frederick et al. 2015; Hildreth et al. 2011; Morita et al. 2009; Shoji et al. 2009). Down-regulation of the gene encoding nicotine uptake permease (NUP1) affected alkaloid metabolism and reduced alkaloid levels (Hildreth et al. 2011; Kato et al. 2014). This reduction in nicotine levels may have led to low TSNA content as well because of the limited availability of alkaloids for nitrosation during curing.

10.10 Impact of Nornicotine on NNN Formation

Nornicotine is the demethylation product of nicotine and the precursor of NNN in tobacco. Commercial tobaccos normally contain around 2–4% of total secondary alkaloids, including nornicotine, which is synthesized by demethylation of nicotine in the leaf and root. In most tobacco varieties, nicotine accumulates as the predominant alkaloid (Bush et al. 1999), but under normal aging and curing conditions, individual plants from a natural population will accumulate nornicotine rather than nicotine. This alteration in a specific plant is termed conversion, and the nornicotine-accumulating plant is said to display a converted phenotype. The conversion process is often associated with the onset of senescence in the leaf tissue (Bush et al. 1999), but conversion can also be induced by exogenous application of ethylene, a senescence-inducing hormone (Bush et al. 1999; Xu et al. 2007). As much as 90% of the nicotine can be converted to nornicotine under these circumstances and various observations have confirmed that converter plants are likely to

be found within populations of essentially all commercial tobacco varieties (Mann et al. 1964). However, certain genotypes have a much greater propensity for producing converter progeny. The regulation factors that govern this unstable occurrence have not been well studied.

One of the most important steps in reducing NNN levels was the development and implementation of low-converter (LC) tobacco seed varieties. Initial research by Bush et al. (2001) indicated that environmental factors might be involved in inducing nornicotine production and that its biosynthesis might be enzymatically derived. Later, Miller et al. (2004) and Jack et al. (2004) discovered that levels of nornicotine in a given burley tobacco variety could be reduced by 70–80% by identifying and eliminating individual plants that converted nicotine to nornicotine prior to seed collection. This "roguing" process was accomplished by chemically analyzing seed plants grown in the field and destroying the plants that had >3% nornicotine relative to nicotine content prior to seed harvest. This screening process produced seeds with an LC phenotype, commonly termed the "LC variety." While LC varieties have diminished nornicotine levels, they are identical to the original variety in terms of yield, quality, agronomic character, and disease resistance.

In 2005, the Kentucky–Tennessee burley tobacco breeding program re-released existing burley varieties as LC varieties and began to use this designation for easier identification. In that burley tobacco season, Philip Morris USA was one of the first major cigarette manufacturers to require contracted growers to use LC varieties in an effort to address public health concerns related to TSNAs and lower levels of NNN in its tobacco products. Subsequently, seed companies discontinued production of non-LC varieties and withdrew existing seed inventories from the market. Currently, the use of LC varieties for burley tobacco production is the standard in the USA and many other parts of the world.

10.11 The Role of Nicotine Demethylase in NNN Formation

Nicotine demethylase is the enzyme that converts nicotine to nornicotine in tobacco. The genes encoding nicotine demethylase have been identified by two independent groups (Siminszky et al. 2005; Xu et al. 2007) that separately identified and characterized nicotine demethylase as a cytochrome P450 enzyme within the *CYP82E* subfamily. Xu et al. (2007) biochemically characterized nicotine demethylase and confirmed its *CYP82E4* function through in vitro assays of a recombinant enzyme produced in yeast microsomes. Two additional nicotine demethylase genes were discovered as encoding cytochrome P450 isoforms that shared a high degree of sequence identity with *CYP82E4*. *CYP82E5v2* was identified by searching for similar genes within a cytochrome P450 sequence library (Gavilano and Siminszky 2007), while *CYP82E10* was discovered by aligning expressed sequence tags from root tissue and identifying cytochrome P450 isoforms. Function was evidenced by yeast expression and identification of mutation lines (Lewis et al. 2010). These three nicotine demethylase genes are hypothesized to have originated from *N. sylvestris* and *N. tomentosiformis*, the two ancestral genomes of *N. tabacum*, during evolution: *CYP82E4* and *CYP82E5v2* from *N. tomentosiformis,* and *CYP82E10* from *N. sylvestris* (Gavilano and Siminszky 2007; Lewis et al. 2010). Both progenitors of *N. tabacum* have been reported as having nornicotine as the main alkaloid (Saitoh et al. 1985). Each of these genes has differentiated expression patterns in various tissues, with the senescence-inducible *CYP82E4* dominating and being responsible for approximately 80% of total nicotine conversion. *CYP82E5v2* expression is higher in green tobacco leaves than in other tissues and is also senescence-inducible, but its role in nicotine

conversion is minor. *CYP82E10* is expressed solely in roots.

Cai et al. (2012) studied the nicotine and nornicotine enantiomers in tobacco and found that different nicotine demethylase enzymes have selectivity for either the S or R forms of nicotine. Most of the nicotine in tobacco is (S)-nicotine, while (R)-nicotine is only a small portion of the total nicotine pool. Cai et al. (2013) also studied the accumulation of nicotine and nornicotine enantiomers in different tobacco plant parts and at different growth and post-harvest stages, confirming the enantioselectivity and unique properties of the nicotine demethylases in tobacco.

Overall, these studies suggested that the three nicotine demethylase genes (*CYP82E4*, *CYP82E5v2*, and *CYP82E10*) are regulated differently and together regulate nornicotine production in green and senescing tobacco leaves. Silencing these genes reduces nornicotine levels in tobacco leaves, leading to reductions in NNN.

The factors related to the initiation of nicotine conversion have yet to be resolved; the unstable phenotype of non-converter to converter status under natural conditions has not been characterized. Identifying global factors associated with regulation of nicotine demethylase gene expression could lead to permanent control of nornicotine formation that would result in very low NNN levels. Triple-mutant plants, where protein expression of all three demethylases has been silenced, still contain some nornicotine implying either direct synthesis, microbial generation, or other genetic mechanisms of nornicotine production. Therefore, further study of nornicotine biosynthesis is needed and may lead to lines with even lower nornicotine and NNN levels.

10.12 NNN Reduction in Low-Nornicotine Tobacco Varieties and Low-Nornicotine Cigarette Prototypes

The identification of three nicotine demethylase genes responsible for nornicotine biosynthesis in *N. tabacum* led to the use of mutant lines in an effort to develop tobacco varieties with reduced nornicotine and NNN levels. Breeding programs at the University of North Carolina, University of Kentucky, and Altria Client Services developed mutants of *CYP82E4*, *CYP82E5v2*, and *CYP82E10* in burley, flue-cured, and dark tobacco varieties. Triple-mutant tobacco varieties were named stable reduced converter (SRC) varieties and the ZYVERT® mark was adopted to identify these varieties with lower nornicotine content and to distinguish them from LC varieties. Experimental SRC varieties with the ZYVERT® technology showed a reduction in nicotine conversion and increased stability compared with standard LC varieties (Lusso et al. 2017).

Measurements in cured leaves of SRC and LC dark tobacco varieties grown under experimental conditions in three locations over a 3-year period showed an average NNN reduction of about 74% (Fig. 10.2) in SRC varieties compared with LC varieties. Similar reductions in leaf NNN content were also observed in experimental SRC burley tobacco varieties.

To determine the effect of SRC tobaccos on NNN levels in combustible products, cigarette prototypes were constructed with SRC burley tobaccos containing ZYVERT® technology and evaluated for selective NNN reduction in tobacco filler and mainstream smoke (Lusso et al. 2017). Control cigarette prototypes were constructed from LC burley tobaccos. Burley single-component cigarette prototypes with the ZYVERT® technology showed an NNN reduction of about 73% in the tobacco filler and 75% in mainstream smoke compared with NNN levels for burley single-component cigarette prototypes without the ZYVERT® technology (Lusso et al. 2017).

The effect of SRC tobaccos on NNN levels in combustible products was also evaluated in prototypes designed as typical American-blended cigarettes (Lusso et al. 2017); the burley component was accounted for 23% of the blend, from which inclusions of 0, 8, 15, and 23% SRC tobacco were tested. NNN decreased as SRC burley tobacco with the ZYVERT® technology inclusions increased (Lusso et al. 2017).

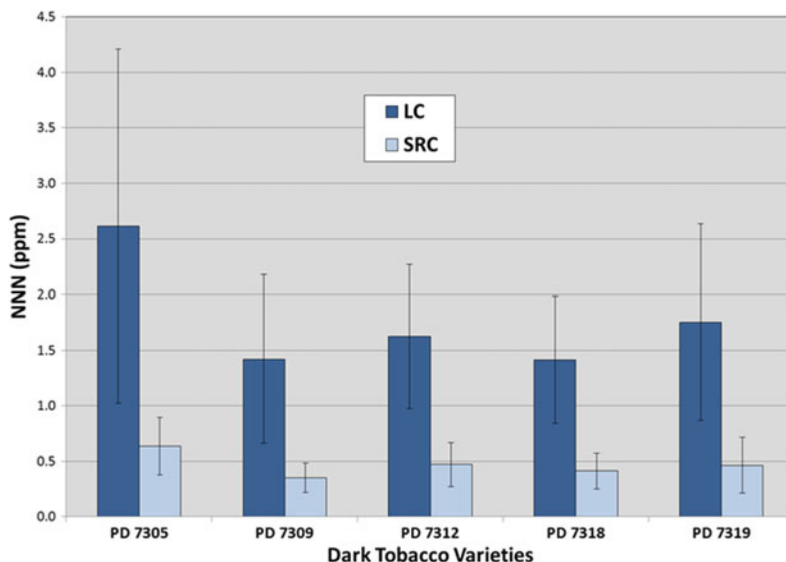

Fig. 10.2 N′-nitrosonornicotine (NNN) levels in cured leaves of stable reduced converter (SRC) and low-converter (LC) dark tobacco varieties. *Source* Lion et al. (2015)

10.13 TSNA Formation Prevention Through Microbial Mitigation and Control

During production of smokeless tobacco products, nitrate-reducing bacteria present in the cured leaves can generate nitrite and subsequently form TSNAs. Efforts to prevent TSNA formation during product processing through the end of retail shelf life have involved sanitizing equipment to control the levels of nitrate-reducing bacteria and monitoring TSNA levels in each production batch to prevent the TSNA increases relative to levels at the beginning of the process (Fisher et al. 2012).

Some bacteria can reduce nitrite to ammonia or nitrogen gas and prevent TSNA formation during curing, while other bacteria degrade TSNAs after they are formed. Wei et al. (2014) attempted to lower TSNA content in tobacco leaves during the air-curing process using a high nitrite-reductive *Bacillus amyloliquefaciens* DA9 strain with low nitrate reduction capability. A suspension of DA9 cells was applied to burley tobacco leaves during the air-curing process, and the same volume of distilled water was also sprayed onto the leaves as a control. After the air-curing process, the nitrite content of the control increased from 10.20 to 17.77 µg/g, while the maximum nitrite content in the treated sample reached 12.06 µg/g, which was 38% lower than the control. The total TSNA concentration in the burley air-cured control reached 56.94 µg/g at 20 days, and then the content decreased to 32.19 µg/g at 45 days. After treatment with DA9, the total TSNA content of the tobacco leaves was much lower than that of the control at each stage. The final TSNA content at 45 days reached 17.03 µg/g, which was 47% lower than the control. Specifically, NNN, NNK, and NAT levels decreased by 48%, 12%, and 35%, respectively (Wei et al. 2014).

Chemical reagents having microbial reduction properties can also be applied to the curing leaves to reduce TSNA formation. Thomas et al. (2010) developed a method of reducing the formation of nitrosamines in a cured tobacco by applying an aqueous solution containing a ferulic acid compound to a whole tobacco leaf. Cui et al. (2011) invented a method for reducing formation of TSNA by applying an effective amount of

chlorate in solid form to reduce bacterial and/or fungal populations on or in the tobacco leaves during the period from one day before to one day after harvest. All these findings may suggest strategies for decreasing TSNA levels during the curing process and potentially in final products.

10.14 Impact of Production Practices on TSNA Formation

Nitrogen fertilization during plant growth can impact TSNA formation. There is a direct correlation between the levels of nitrogen applied to a tobacco crop and TSNA levels in the leaf. The correlation is better when very high levels of nitrogen are applied. This is attributable to the increased alkaloid production and increased levels of nitrate in the leaf, which can lead to additional TSNA accumulation when the tobacco is stored under high-moisture conditions for prolonged periods. High alkaloid levels leading to higher TSNA accumulation in the leaf are observed when the tobacco is topped early, and low-alkaloid levels are observed when the tobacco is harvested at a late maturity date (Lewis et al. 2012). The impact of topping also has a strong correlation to TSNA formation potential.

TSNA formation has been well-characterized during the curing processes of flue-cured, burley, dark air-cured, and dark fire-cured tobaccos. Formation of TSNAs in flue-cured tobaccos occurs predominantly through the reaction of leaf alkaloids with NO_x species in the combustion gases that are formed during the curing process. In 1999, tobacco farm organizations, leaf purchasers, and land-grant university scientists established the Tobacco Industry Leadership Group to address the formation of TSNAs in flue-cured tobacco. The group recommended a retrofit of all direct-fired flue-cured barns with indirect-fired heating, and retrofitting curing barns with heat exchangers minimized combustion gas exposure, notably reducing the formation of TSNAs in flue-cured tobaccos. Data from the 2000 curing season showed that converting from direct- to indirect-fired flue-curing could result in cured leaf TSNA levels below 0.1 ppm in a research setting (Sumner 2007). Today, TSNA levels in flue-cured tobacco are generally 1.0 ppm or lower because of this curing improvement.

Burley and dark air-cured tobaccos are typically cured under ambient conditions for 2–3 months. During this time, rainfall, relative humidity, and temperature can impact TSNA values in the cured leaf. This extended curing period results in slow cell degradation and, under conditions favoring increased microbial populations such as high humidity and nitrite levels, can lead to TSNA increases. Surveys of US-grown burley crops found significant variability in total TSNA values between crop years, ranging on average from 3.42 ppm in 2010 to 10.81 ppm in 2017 (Fig. 10.3; Modified from Lion et al. 2015).

To assess the impact of ambient conditions on TSNA formation in burley tobacco during curing, Fig. 10.4 plots the average relative humidity in the production regions during curing against mean TSNA values for crop years 2010–2016. A correlation ($R^2 = 0.6303$) was observed between average relative humidity and mean TSNA values (Modified from Lion et al. 2015).

In dark fire-cured tobacco, the curing process allows for greater influence on curing conditions and is less impacted by changes in ambient conditions, although high temperatures during curing, NO_x gases generated by burning wood, and nitrite accumulation in the leaf through nitrate microbial reduction lead to a higher TSNA formation potential.

Analyses similar to those described above for US burley tobacco crops have been conducted for dark fire-cured tobacco. Figure 10.5 shows significant variability in TSNA values between crop years, as was observed for burley crops. TSNA mean values ranged from 4.6 ppm in 2011 to 13.4 ppm in 2005 (Modified from Lusso et al. 2012).

Figure 10.6 plots the average relative humidity in the production regions during curing against mean TSNA values in each dark fire-cured tobacco crop year. A correlation ($R^2 = 0.7217$) was observed between average relative humidity and mean TSNA values during the curing process (Modified from Lusso et al. 2012).

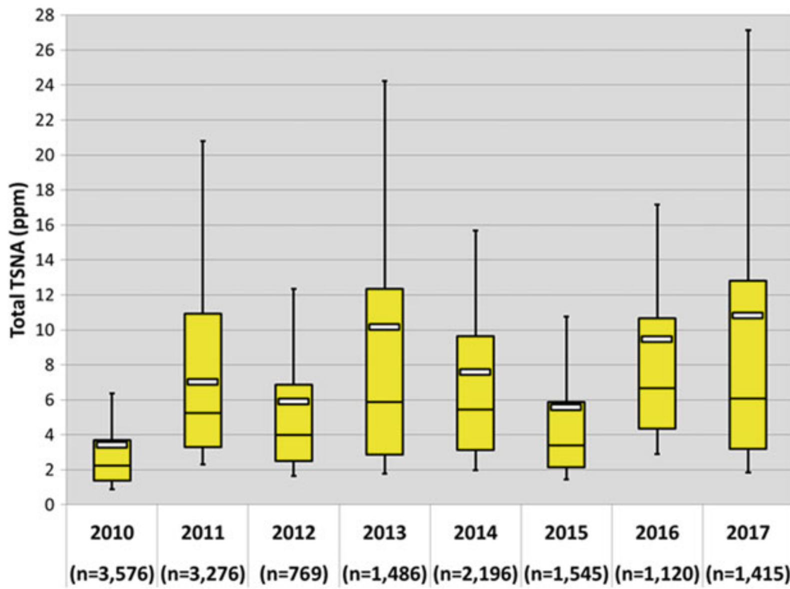

Fig. 10.3 Box-and-whisker plots for tobacco-specific N-nitrosamine (TSNA) values in US burley tobacco crops by year. [The outer whiskers of the box plot represent the 10th and 90th percentiles; the box represents the 25th to 75th percentiles; the median is represented by the black line across the middle of the box, and the annual mean is represented by the white bars. Numbers in parentheses indicate the number of samples collected for each year (Modified from Lion et al. 2015)]

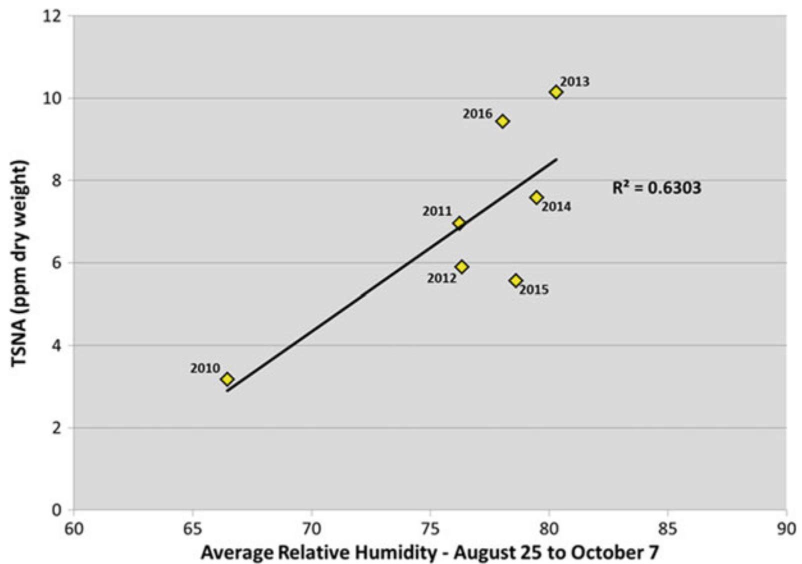

Fig. 10.4 Correlation between relative humidity and tobacco-specific N-nitrosamine (TSNA) values in US burley tobacco crops 2010–2016. *Source* Modified from Lion et al. (2015)

The effect of temperature on TSNA formation during the curing process has also been examined to understand the factors influencing TSNA formation in dark fire-cured tobacco. Figure 10.7 shows TSNA values in dark tobacco cured under both the recommended curing temperature (not exceeding 54.5 °C) (University of Kentucky et al. 2017) and temperatures higher than recommended levels (stressed conditions) (Lusso et al. 2012). A 52% increase in total TSNA

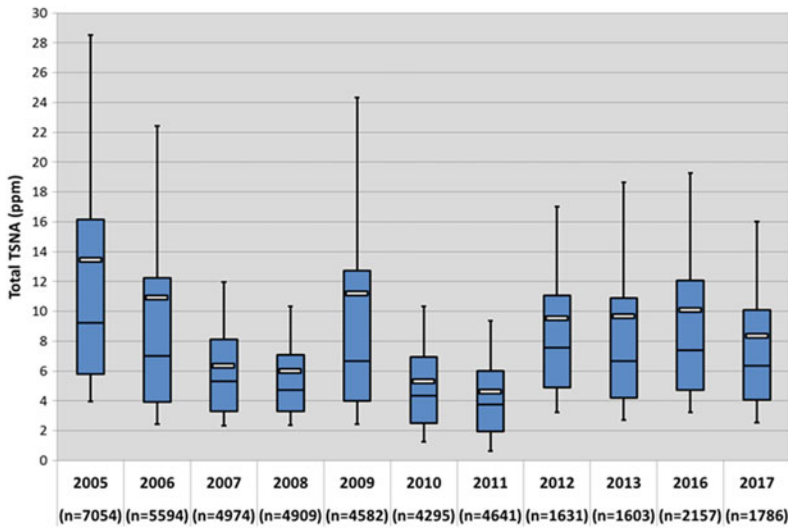

Fig. 10.5 Box-and-whisker plots for tobacco-specific N-nitrosamine (TSNA) values in US-grown dark fire-cured tobacco crops [Assessments were not conducted for 2014 and 2015 The outer whiskers of the box plot represent the 10th and 90th percentiles; the box represents the 25th to 75th percentiles; the median is represented by the black line across the middle of the box, and the annual mean is represented by the white bars. Numbers in parentheses indicate the number of samples collected for each year. Modified from Lusso et al. (2012)]

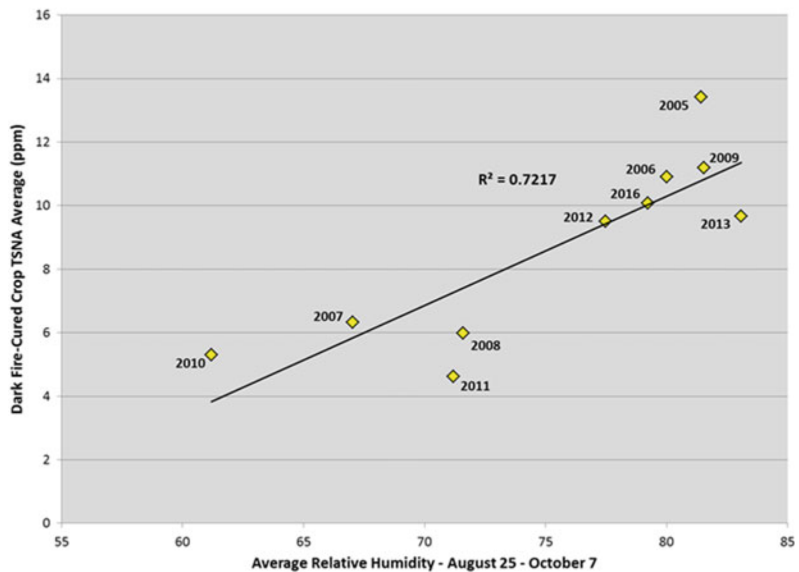

Fig. 10.6 Correlation between relative humidity and tobacco-specific N-nitrosamine (TSNA) values in US-grown dark fire-cured tobacco crops. [Assessments were not conducted for 2014 and 2015 (Modified from Lusso et al. 2012)]

values were observed in dark tobacco cured at temperatures exceeding the recommended levels. The highest temperature in the stressed barn was approximately 79 °C, whereas, in the normal curing barn, the maximum was approximately 58 °C. Of note, fire curing relies on the use of hardwood and sawdust, which vary by tree species and environmental conditions, increasing the

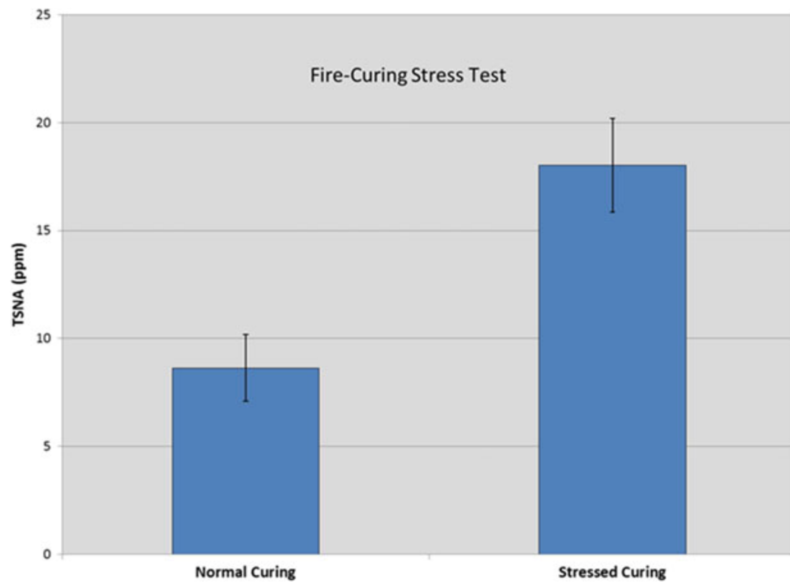

Fig. 10.7 Tobacco-specific N-nitrosamine (TSNA) levels in US-grown dark fire-cured tobacco cured under recommended temperatures (normal curing) and temperatures that exceed recommendations (stress curing). [*Source* Authors, 2017 (unpublished data). Error bars represent standard deviation of the mean]

challenges associated with temperature management during the curing process (University of Kentucky et al. 2017).

As mentioned above, the storage period following curing can also increase TSNA values depending on storage conditions and tobacco type (Jackish and Rovedder 2007; Verrier et al. 2008). de Roton and Girard (2004) reported that in ground samples of burley tobacco stored in an ambient environment for 6 months (but not in refrigerated samples), TSNA concentrations increased from the initial 1.3–1.4 µg/g to 8.1–9.5 µg/g. The mechanism of TSNA formation during storage was not elucidated. Saito et al. (2006) reported that as storage temperature increased, TSNA and nitrite contents increased; as burley tobacco has higher levels of nitrate than flue-cured tobacco, increases in TSNA levels in burley tobacco were more significant.

During storage, unstable nitrate compounds form nitrogen oxides that can easily react with alkaloids to form TSNAs by nitrosation. This may be the cause of TSNA increases during storage under relatively high temperatures. To test this hypothesis, Shi et al. (2013) designed a series of experiments assessing changes in TSNA content over 1 year of storage and examining the effect of temperature and nitrate on TSNA formation. It was found that TSNA levels increased during storage and under conditions of higher temperature, higher nitrate, and higher moisture levels (Shi et al. 2013; Wang et al. 2017). Overall, to prevent TSNA formation during leaf storage, low temperature and low humidity are recommended.

10.15 Conclusions

The formation of TSNAs in tobacco is somewhat simple in that it results from the chemical linkage of nitrite and/or NO_x gases with the secondary amine of relatively abundant alkaloids found in tobacco plants. The genetic and environmental control of TSNA formation, however, is complex. Gene expression, environmental conditions during curing, the plant microbial complement, and nitrate and nitrite pools within the plant during the curing process are all factors in total TSNA content. While efforts have been made to mitigate TSNA formation, it remains a challenge, particularly given the environmental and genetic heterogeneity of tobacco crops. Mitigation of TSNA formation has been adequately addressed in flue-cured tobacco through the use of indirect-heat exchangers; however, TSNA formation in

the air-cured tobacco types (burley and dark tobacco) and in dark fire-cured tobacco can be reduced but not eliminated through the use of LC varieties. The formation of TSNAs, specifically NNN, can be prevented substantially with wide-scale adoption of tobacco varieties containing ZYVERT® technology. Overall, production of tobacco leaf with extremely low TSNA content remains challenging.

References

Adams A, Lusso M, Pramod S, Xu D (2016) Compositions and methods for producing tobacco plants and products having altered alkaloid levels. US Patent No. 20160374387/ WO2016210303A1

Appleton S, Olegario R, Lipowicz P (2013) TSNA levels in machine-generated mainstream cigarette smoke: 35 years of data. Regul Toxicol Pharmacol 66(2):197–207. https://doi.org/10.1016/j.yrtph.2013.03.013

Brunnemann KD, Masaryk J, Hoffmann D (1983) Role of tobacco stems in the formation of N-nitrosamines in tobacco and cigarette mainstream and sidestream smoke. J Agric Food Chem 31(6):1221–1224. https://doi.org/10.1021/jf00120a020

Burton HR, Bush LP, Djordjevic MV (1989) Influence of temperature and humidity on the accumulation of tobacco-specific nitrosamines in stored burley tobacco. J Agric Food Chem 37(5):1372–1377. https://doi.org/10.1021/jf00089a036

Burton HR, Dye NK, Bush, LP (1992) Distribution of tobacco constituents in tobacco leaf tissue. 1. Tobacco-specific nitrosamines, nitrate, nitrite, and alkaloids. J Agric Food Chem 40(6):1050–1055. https://doi.org/10.1021/jf00018a028

Burton HR, Dye NK, Bush LP (1994) Relationship between tobacco-specific nitrosamines and nitrite from different air-cured tobacco varieties. J Agric Food Chem 42 (9):2007–2011. https://doi.org/10.1021/jf00045a033

Bush L, Cui M, Shi H, Burton H, Fannin F, Lei L, Dye N (2001) Formation of tobacco-specific nitrosamines in air-cured tobacco. Rec Adv Tob Sci 27:23–46

Bush L, Hempfling WP, Burton H (1999) Chapter 2: Biosynthesis of nicotine and related compounds. In: Gorrod JW, Jacob P (eds) Analytical determination of nicotine and related compounds and their metabolites. Elsevier Science, Amsterdam, pp 13–44

Cai B, Jack A, Lewis R, Dewey R, Bush L (2013) (R)-nicotine biosynthesis, metabolism and translocation in tobacco as determined by nicotine demethylase mutants. Phytochemistry 95:188–196. https://doi.org/10.1016/j.phytochem.2013.06.012

Cai B, Siminszky B, Chappell J, Dewey R, Bush LP (2012) Enantioselective demethylation of nicotine as a mechanism for variable nornicotine composition in tobacco leaf. J Biol Chem 287(51):42804–42811. https://doi.org/10.1074/jbc.M112.413807

Caldwell W, Greene J, Plowchalk D, deBethizy J (1991) The nitrosation of nicotine: a kinetic study. Chem Res Toxicol 4(5):513–516

Chaplin JF, Weeks W (1976) Association between percent total alkaloids and other traits in flue-cured tobacco. Crop Sci 16(3):416–418

Cui M, Nielsen M, Hart R, Overbey M, Watson D, Chipley J (2011) Use of chlorate, sulfur or ozone to reduce tobacco specific nitrosamines. US Patent No. 7,992,575 B2

Dawson RF (1942) Accumulation of nicotine in reciprocal grafts of tomato and tobacco. Am J Bot 29(1):66–71. https://doi.org/10.2307/2436544

de Roton C, Girard C (2004) Potential changes of TSNA composition in stored tobacco powder: Consequences for sample preparation and ground storage. Paper presented at the CORESTA Congress, Kyoto, Japan, 3–7 Oct 2004

de Roton C, Wiernik A, Wahlberg I, Vidal B (2005) Factors influencing the formation of tobacco-specific nitrosamines in French air-cured tobaccos in trials and at the farm level. Beiträge zur Tabakforschung International 21:305–319. https://doi.org/10.2478/cttr-2013-0797

DeBoer K, Dalton H, Edward F, Hamill J (2011) RNAi-mediated down-regulation of ornithine decarboxylase (ODC) leads to reduced nicotine and increased anatabine levels in transgenic Nicotiana tabacum L. Phytochemistry 72(4–5):344–355. https://doi.org/10.1016/j.phytochem.2010.12.012

DeBoer K, Dalton H, Edward F, Ryan S, Hamill J (2013) RNAi-mediated down-regulation of ornithine decarboxylase (ODC) impedes wound-stress stimulation of anabasine synthesis in Nicotiana glauca. Phytochemistry 86:21–28. https://doi.org/10.1016/j.phytochem.2012.10.016

Dewey R, Xie J (2013) Molecular genetics of alkaloid biosynthesis in Nicotiana tabacum. Phytochemistry 94:10–27. https://doi.org/10.1016/j.phytochem.2013.06.002

Djordjevic MV, Gay SL, Bush LP, Chaplin JF (1989) Tobacco-specific nitrosamine accumulation and distribution in flue-cured tobacco alkaloid isolines. J Agric Food Chem 37(3):752–756. https://doi.org/10.1021/jf00087a040

Edwards KD, Fernandez-Pozo N, Drake-Stowe K, Humphry M, Evans AD, Bombarely A, Mueller L (2017) A reference genome for Nicotiana tabacum enables map-based cloning of homeologous loci implicated in nitrogen utilization efficiency. BMC Genom 18(1):448. https://doi.org/10.1186/s12864-017-3791-6

Fannin F, Bush L (1992) Nicotine demethylation in Nicotiana. Med Sci Res 20:867–868
Fisher M, Bennett C, Hayes A, Kargalioglu Y, Knox B, Xu D, Gaworski C (2012) Sources of and technical approaches for the abatement of tobacco specific nitrosamine formation in moist smokeless tobacco products. Food Chem Toxicol 50(3–4):942–948. https://doi.org/10.1016/j.fct.2011.11.035
Frederick J, Shen Y, Xu D, Warek U, Strickland J (2015) Tobacco plants having altered amounts of one or more alkaloids in leaf and methods of using such plants. US Patent No. 20150173319/ WO/2015/085299
Gavilano L, Siminszky B (2007) Isolation and characterization of the cytochrome P450 gene CYP82E5v2 that mediates nicotine to nornicotine conversion in the green leaves of tobacco. Plant Cell Physiol 48 (11):1567–1574. https://doi.org/10.1093/pcp/pcm128
Gunduz I, Kondylis A, Jaccard G, Renaud J, Hofer R, Ruffieux L, Gadani F (2016) Tobacco-specific N-nitrosamines NNN and NNK levels in cigarette brands between 2000 and 2014. Regul Toxicol Pharmacol 76:113–120. https://doi.org/10.1016/j.yrtph.2016.01.012
Hashimoto T, Yamada Y (1993) Nicotine and tropane alkaloids. In: Lea PJ (ed) Methods in plant biochemistry, vol 9. Academic Press, London, pp 369–379
Hecht S, Castonguay A, Chung F Hoffmann D (1984) Carcinogenicity and metabolic activation of tobacco-specific nitrosamines: current status and future prospects. IARC Sci Publ (57):763–778
Hecht S, Chen C, Hirota N, Ornaf R, Tso T, Hoffmann D (1978a) Tobacco-specific nitrosamines: formation from nicotine in vitro and during tobacco curing and carcinogenicity in strain A mice. J Natl Cancer Inst 60 (4):819–824
Hecht S, Chen C, Ornaf R, Jacobs E, Adams J, Hoffmann D (1978b) Reaction of nicotine and sodium nitrite: formation of nitrosamines and fragmentation of the pyrrolidine ring. J Org Chem 43(1):72–76
Hecht S, Chen C, Ohmori T, Hoffmann D (1980) Comparative carcinogenicity in F344 rats of the tobacco-specific nitrosamines, N'-nitrosonornicotine and 4-(N-methyl-N-nitrosamino)-1-(3-pyridyl)-1-butanone. Cancer Res 40(2):298–302
Hibi N, Fujita T, Hatano M, Hashimoto T, Yamada Y (1992) Putrescine N-methyltransferase in cultured roots of hyoscyamus albus: n-butylamine as a potent inhibitor of the transferase both in vitro and in vivo. Plant Physiol 100(2):826–835
Hildreth S, Gehman E, Yang H, Lu R, Ritesh K, Harich K, Jelesko J (2011) Tobacco nicotine uptake permease (NUP1) affects alkaloid metabolism. Proc Natl Acad Sci USA 108(44):18179–18184. https://doi.org/10.1073/pnas.1108620108
Hoffmann D, Brunnemann K, Prokopczyk B, Djordjevic M (1994) Tobacco-specific N-nitrosamines and Areca-derived N-nitrosamines: chemistry, biochemistry, carcinogenicity, and relevance to humans. J Toxicol Environ Health 41(1):1–52. https://doi.org/10.1080/15287399409531825
IARC Working Group on the Evaluation of Carcinogenic Risk to Humans (2007) Smokeless tobacco and some tobacco-specific N-nitrosamines. (IARC monographs on the evaluation of carcinogenic risks to humans, No. 89.) https://www.ncbi.nlm.nih.gov/books/NBK326497/. Lyon, France: International Agency for Research on Cancer
Jack A, Bush L, Fannin N, Miller R (2004) Proceedings of the 41st workers conference, Nashville, TN
Jackish R, Rovedder J (2007) Burley tobacco post-curing management and its effect in the nitrosamine amount. Paper presented at the CORESTA Joint Study Group Meeting, Krakow, Poland, October 2007
Kajikawa M, Shoji T, Kato A, Hashimoto T (2011) Vacuole-localized berberine bridge enzyme-like proteins are required for a late step of nicotine biosynthesis in tobacco. Plant Physiol 155(4):2010–2022. https://doi.org/10.1104/pp.110.170878
Kajikawa M, Sierro N, Kawaguchi H, Bakaher N, Ivanov N, Hashimoto T, Shoji T (2017) Genomic insights into the evolution of the nicotine biosynthesis pathway in tobacco. Plant Physiol 174(2):999–1011. https://doi.org/10.1104/pp.17.00070
Kato K, Shoji T, Hashimoto T (2014) Tobacco nicotine uptake permease regulates the expression of a key transcription factor gene in the nicotine biosynthesis pathway. Plant Physiol 166(4):2195–2204. https://doi.org/10.1104/pp.114.251645
Krauss M, Li Q, Walter P (2003) A method for the reduction of tobacco-specific nitrosamines by increasing antioxidants in tobacco. US Patent No. WO2003022081A1
Kudithipudi C, Morris J, Lusso M (2016) Ultra-low nicotine tobacco lines with improved leaf quality. Paper presented at the 2016 CORESTA Congress, Berlin, Germany
Law A, Fisher C, Jack A, Moe L (2016) Tobacco, microbes, and carcinogens: correlation between tobacco cure conditions, tobacco-specific nitrosamine content, and cured leaf microbial community. Microb Ecol 72(1):120–129. https://doi.org/10.1007/s00248-016-0754-4
Lee Y, Cho Y (2001) Identification of essential active-site residues in ornithine decarboxylase of Nicotiana glutinosa decarboxylating both L-ornithine and L-lysine. Biochem J 360(Pt 3):657–665
Leete E (1980) Alkaloids derived from ornithine, lysine, and nicotinic acid. In: Bell EA, Charlwood BV (eds) Encyclopedia of plant psysiology, vol 8. Springer-Verlag, Berlin, Germany
Leete E, Mueller ME (1982) Biomimetic synthesis of anatabine from 2,5-dihydropyridine produced by the oxidative decarboxylation of baikiain. J Am Chem Soc 104(23):6440–6444. https://doi.org/10.1021/ja00387a048
Leete E, Slattery S (1976) Incorporation of [2-14C]-and [6-14C]nicotinic acid into the tobacco alkaloids. Biosynthesis of anatabine and alpha,beta-dipyridyl. J Am Chem Soc 98(20):6326–6330
Legg P, Chaplin J, Williamson R (1977) Genetic diversity in burley and flue-cured tobacco. Crop Sci 17:943–947

Legg P, Collins G (1971) Inheritance of per cent total alkaloids in Nicotiana tabacum L. II. Genetic effects of two loci in Burley 21 × LA Burley 21 populations. Can J Genet Cytol 13(2):287–291 https://doi.org/10.1139/g71-047

Legg P, Collins G (1988) Registration of HI and LI Burley 21 tobacco germplasm. Crop Science 28 (1):206–207

Lewis R, Bowen S, Keogh M, Dewey R (2010) Three nicotine demethylase genes mediate nornicotine biosynthesis in Nicotiana tabacum L.: functional characterization of the CYP82E10 gene. Phytochemistry 71(17–18):1988–1998 https://doi.org/10.1016/j.phytochem.2010.09.011

Lewis R, Lopez H, Bowen S, Andres K, Steede W, Dewey R (2015) Transgenic and mutation-based suppression of a berberine bridge enzyme-like (BBL) gene family reduces alkaloid content in field-grown tobacco. PloS one 10(2):e0117273. https://doi.org/10.1371/journal.pone.0117273

Lewis R, Parker R, Danehower D, Andres K, Jack A, Whitley D, Bush L (2012) Impact of alleles at the Yellow Burley (Yb) loci and nitrogen fertilization rate on nitrogen utilization efficiency and tobacco-specific nitrosamine (TSNA) formation in air-cured tobacco. J Agric Food Chem 60(25):6454–6461. https://doi.org/10.1021/jf2053614

Lion K, Lusso M, Adams A, Morris W, Hart F (2015) The relationship of tobacco specific nitrosamine (TSNA) levels with relative humidity during curing in U.S. burly. Paper presented at the 2015 CORESTA meeting

Lu J, Zhang L, Lewis R, Bovet L, Goepfert S, Jack AM, Dewey R (2016) Expression of a constitutively active nitrate reductase variant in tobacco reduces tobacco-specific nitrosamine accumulation in cured leaves and cigarette smoke. Plant Biotechnol J 14(7):1500–1510. https://doi.org/10.1111/pbi.12510

Lusso M, Gunduz I, Kondylis A, Jaccard G, Ruffieux L, Gadani F, Strickland J (2017) Novel approach for selective reduction of NNN in cigarette tobacco filler and mainstream smoke. Regul Toxicol Pharmacol 89:101–111. https://doi.org/10.1016/j.yrtph.2017.07.019

Lusso M, Hayes A, Lion K, Hart F, Morris W (2012) Impact of curing management practices on leaf TSNA content in dark fire-cured tobacco. Paper presented at the CORESTA Congress, Sapporo, 2012, Agronomy/Phytopathology Groups, AP 10

MacKown C, Eivazi F, Sims J, Bush L (1984) Tobacco-specific N-nitrosamines: effect of burley alkaloid isolines and nitrogen fertility management. J Agric Food Chem 32:1269–1272

Mann T, Weybrew J, Matzinger D, Hall J (1964) Inheritance of the conversion of nicotine to nornicotine in varieties of Nicotiana tabacum L. and related amphidiploids. Crop Science 4:349–353

Miller R, Jack A, Bush L (2004). Screening commercial burley tobacco cultivars for nornicotine conversion. Paper presented at the 41st Tobacco Workers Conference, Nashville, TN

Morita M, Shitan N, Sawada K, Van Montagu M, Inze D, Rischer H, Yazaki K (2009) Vacuolar transport of nicotine is mediated by a multidrug and toxic compound extrusion (MATE) transporter in Nicotiana tabacum. Proc Natl Acad Sci U S A 106(7):2447–2452. https://doi.org/10.1073/pnas.0812512106

National Institute of Environmental Health Sciences (2011) National Toxicology Program. NTP 12th Report on Carcinogens. Rep Carcinog, 12, iii-499

Piade JJ, Hoffmann D (1980) Chemical studies on tobacco smoke LXVII. Quantitative determination of alkaloids in tobacco by liquid chromatography. J Liquid Chromatogr 3(10):1505–1515. https://doi.org/10.1080/01483918008062792

Saito H, Miyazaki M, Miki J (2006) Role of nitrogen oxides in tobacco-specific nitrosamine formation in burley tobacco. Paper presented at the CORESTA Congress, Paris, France, 15–20 Oct 2006

Saitoh F, Noma M, Kawashima N (1985) The alkaloid contents of sixty Nicotiana species. Phytochemistry 24 (3):477–480. https://doi.org/10.1016/S0031-9422(00)80751-7

Scanlan R (1983) Formation and occurrence of nitrosamines in food. Cancer Res 43(5 Suppl):2435s–2440s

Shi H, Wang R, Bush LP, Zhou J, Yang H, Fannin N, Bai R (2013) Changes in TSNA contents during tobacco storage and the effect of temperature and nitrate level on TSNA formation. J Agric Food Chem 61(47):11588–11594. https://doi.org/10.1021/jf404813m

Shoji T, Hashimoto T (2011) Tobacco MYC2 regulates jasmonate-inducible nicotine biosynthesis genes directly and by way of the NIC2-locus ERF genes. Plant Cell Physiol 52(6):1117–1130. https://doi.org/10.1093/pcp/pcr063

Shoji T, Hashimoto T (2012) DNA-binding and transcriptional activation properties of tobacco NIC2-locus ERF189 and related transcription factors. Plant Biotechnology 29:35–42. https://doi.org/10.5511/plantbiotechnology.11.1216a

Shoji T, Hashimoto T (2014) Biosynthesis and regulation of tobacco alkaloids. In: Wallner F (ed) Herbacious plants: cultivation methods, grazing and environmental impacts. Nova Biomedical, pp 37–67

Shoji T, Inai K, Yazaki Y, Sato Y, Takase H, Shitan N, Hashimoto T (2009) Multidrug and toxic compound extrusion-type transporters implicated in vacuolar sequestration of nicotine in tobacco roots. Plant Physiol 149(2):708–718. https://doi.org/10.1104/pp.108.132811

Siminszky B, Gavilano L, Bowen S, Dewey R (2005) Conversion of nicotine to nornicotine in Nicotiana tabacum is mediated by CYP82E4, a cytochrome P450 monooxygenase. Proc Natl Acad Sci USA 102 (41):14919–14924. https://doi.org/10.1073/pnas.0506581102

Solt M, Dawson R, Christman D (1960) Biosynthesis of anabasine and of nicotine by excised root cultures of nicotiana glauca. Plant Physiol 35(6):887–894

Sumner P (2007) Tobacco barn retrofit information: retrofitting tobacco curing barns. Accessed from http://cpes.peachnet.edu/tobacco/retrofitinfo.htm

Thomas T, Brandon J, Bailey W, Losty T (2010) Method for reducing nitrosamines. US Patent No. 7,757,697 B2

US Food and Drug Administration (2018) Harmful and potentially harmful constituents (HPHCs). Accessed from https://www.fda.gov/tobaccoproducts/labeling/productsingredientscomponents/ucm20035927.htm

University of Kentucky, University of Tennessee, Virginia Polytechnic Institute and State University, & North Carolina State University – Cooperative Effort (2017) Burley and dark tobacco production guide. Knoxville, TN: Good Agricultural Practices Connection

Valleau W (1949) Breeding low-nicotine tobacco. Agric Res 78:171–181

Verrier J, Wiernik A, Staaf M, Cadilhac J, Onillon M, Vidal B (2008) The influence of post-curing of Burley tobacco and dark air cured tobacco on TSNA and nitrite levels. Paper presented at the CORESTA Congress, Shanghai, China, 2–7 Nov 2008

Wagner R, Wagner K (1985) The pyridine-nucleotide cycle in tobacco Enzyme activities for the de-novo synthesis of NAD. Planta 165(4):532–537. https://doi.org/10.1007/bf00398100

Wang J, Yang H, Shi H, Zhou J, Bai R, Zhang M, Jin T (2017) Nitrate and nitrite promote formation of tobacco-specific nitrosamines via nitrogen oxides intermediates during postcured storage under warm temperature. J Chem 2017:11. https://doi.org/10.1155/2017/6135215

Wang P, Zeng J, Liang Z, Miao Z, Sun X, Tang K (2009) Silencing of PMT expression caused a surge of anatabine accumulation in tobacco. Mol Biol Rep 36(8):2285–2289. https://doi.org/10.1007/s11033-009-9446-1

Wei X, Deng X, Cai D, Ji Z, Wang C, Yu J, Chen S (2014) Decreased tobacco-specific nitrosamines by microbial treatment with Bacillus amyloliquefaciens DA9 during the air-curing process of burley tobacco. J Agric Food Chem 62(52):12701–12706. https://doi.org/10.1021/jf504084z

Wiernik A, Christakopoulos A, Johansson L, Wahlberg I (1995) Effect of air-curing on chemical composition of tobacco. Rec Adv Tob Sci 21:39–80

Xie J, Song W, Maksymowicz W (2004) Biotechnology: a tool for reduced risk tobacco products—the nicotine experience from test tube to cigarette pack. Rec Adv Tob Sci 30:17–37

Xu D, Shen Y, Chappell J, Cui M, Nielsen M (2007) Biochemical and molecular characterizations of nicotine demethylase in tobacco. Physiol Plant 129(2):307–319. https://doi.org/10.1111/j.1399-3054.2006.00811.x

Check for updates

11 The Genetic Basis of Flower Color Differences in *Nicotiana tabacum*

Elizabeth W. McCarthy, Jacob B. Landis, Amelda Kurti, Amber J. Lawhorn and Amy Litt

Abstract

Nicotiana tabacum L accessions vary in flower color from light pink to magenta. The differences in flower color are attributable to differences in anthocyanin content. To determine the genetic basis of flower color differences, we generated transcriptomes and quantified transcript levels of flavonoid biosynthetic genes in four *N. tabacum* accessions and their diploid progenitors. High expression ratios of the flavonol synthase (*FLS*) gene to dihydroflavonol 4-reductase (*DFR*) gene are found in light-pink flowers, suggesting that competition between the FLS and DFR enzymes for the same substrates may drive the flux of the flavonoid biosynthetic pathway toward producing flavonols at the expense of anthocyanins, resulting in light-pink flowers. The high *FLS*:*DFR* expression ratio appears to be attributable to *DFR* activation later in development in light-pink flowers.

E. W. McCarthy · J. B. Landis · A. Kurti · A. J. Lawhorn · A. Litt (✉)
Department of Botany and Plant Sciences, University of California, Riverside, Riverside, CA 92521, USA
e-mail: amy.litt@ucr.edu

E. W. McCarthy
Department of Biological Sciences, SUNY Cortland, Cortland, NY 13045, USA

J. B. Landis
School of Integrative Plant Science, Section of Plant Biology and the L.H. Bailey Hortorium, Cornell University, Ithaca, NY 14853, USA

11.1 Introduction

The genus *Nicotiana* (Solanaceae) consists of 76 species, which have flowers with tubular corollas that display diverse colors, including white, pink, magenta, purple, red, green, yellow, and white with ultraviolet light reflectance (Goodspeed 1954; McCarthy et al. 2015). Approximately half of *Nicotiana* species are allopolyploids that arose from both whole-genome duplication and interspecific hybridization (Chase et al. 2003; Knapp et al. 2004; Clarkson et al. 2004, 2010). These hybrid species may be predicted to display intermediate phenotypes between their diploid progenitors; however, *Nicotiana* allotetraploids can show transgressive floral colors (which fall outside the range of the diploid progenitors), and related polyploid accessions from the same origin can display floral color variation (McCarthy et al. 2015). We are interested in the genetic basis of these floral color differences.

Nicotiana tabacum L originated approximately 0.6 million years ago from maternal *N.*

© Springer Nature Switzerland AG 2020

N. V. Ivanov et al. (eds.), *The Tobacco Plant Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-030-29493-9_11

sylvestris and paternal *N. tomentosiformis* progenitors (Clarkson et al. 2004, 2010, 2017). First-generation allotetraploid lines that share the same parentage as natural *N. tabacum* were created synthetically by K. Y. Lim, Queen Mary University of London (London, UK). Floral color in accessions of natural *N. tabacum* ranges from light pink to magenta, and independent synthetic lines also display floral color differences, although the range is limited from light pink to dark pink (McCarthy et al. 2015, 2017). To determine the gene expression differences underlying this variation in floral color, we focused on two natural *N. tabacum* accessions (*N. tabacum* 095-55, magenta flowers, termed 095-55, and *N. tabacum* "Chulumani", light-pink flowers, termed Chulumani), two first-generation synthetic lines (QM24, light-pink flowers, and QM25, dark-pink flowers), and their diploid progenitors, *N. sylvestris* (white flowers) and *N. tomentosiformis* (dark-pink flowers) (Fig. 11.1).

In many plant species, anthocyanin pigments are responsible for the red, pink, purple, and blue floral colors. Equally important for pollination are flavonols, which are not visible to humans, but absorb ultraviolet light and thereby provide visual cues to birds and bees. In a previous study, we quantified anthocyanidins (anthocyanins without attached sugars) and flavonols in *N. sylvestris*, *N. tomentosiformis*, 095-55, Chulumani, QM24, and QM25 to determine the biochemical basis of floral color differences. The predominant anthocyanidin in all samples was cyanidin, which provides a pink color. The amount of cyanidin was consistent with flower

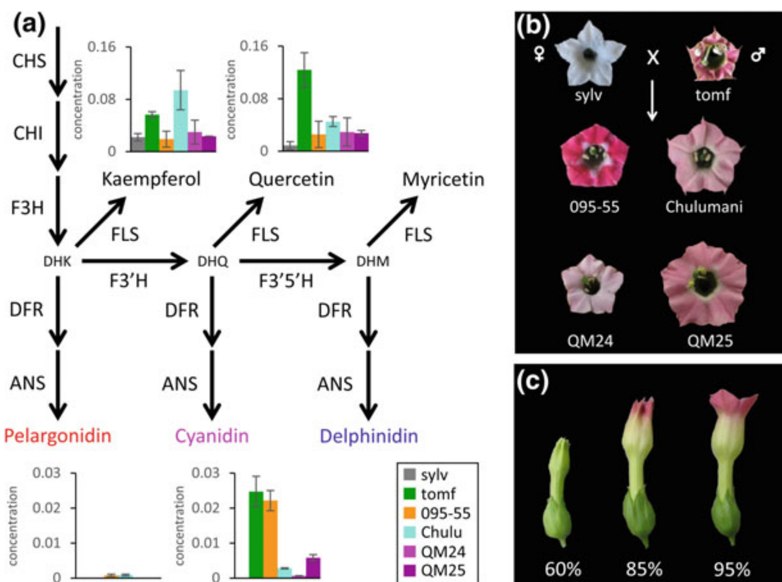

Fig. 11.1 Flavonoid biosynthetic pathway and *Nicotiana* flower colors and pigments. **a** Flavonoid biosynthetic pathway with structural enzymes in large font next to arrows; reproduced from McCarthy et al. (2017). CHS, chalcone synthase; CHI, chalcone isomerase; F3H, flavanone 3-hydroxylase; F3′H, flavonoid 3′-hydroxylase; F3′5′H, flavonoid 3′,5′-hydroxylase; FLS, flavonol synthase; DFR, dihydroflavonol 4-reductase; ANS, anthocyanidin synthase. Shared substrates appear in small font (DHK, dihydrokaempferol; DHQ, dihydroquercetin; DHM, dihydromyricetin), and resulting flavonol (black text) and anthocyanin (colored text) pigments. Bar graphs represent the pigment concentration (mg/g fresh tissue) of kaempferol, quercetin, pelargonidin, and cyanidin in the *Nicotiana* accessions examined in this chapter. **b** Floral limb photographs of the *Nicotiana* accessions used. *Nicotiana sylvestris* (white) and *N. tomentosiformis* (dark pink) are the maternal and paternal progenitors, respectively, of *N. tabacum*. Natural tobacco accessions: *N. tabacum* 095-55 (dark pink) and *N. tabacum* "Chulumani" (light pink). Synthetic lines QM24 (light pink) and QM25 (dark pink). **c** Floral bud photographs of *N. tabacum* 095-55 at 60, 85, and 95% of anthesis length, the developmental time points used in this study. 095-55, *Nicotiana tabacum* 095-55; Chulu, *N. tabacum* "Chulumani"; sylv, *N. sylvestris*; tomf, *N. tomentosiformis*. QM24 and QM25 are first-generation synthetic lines

color: low levels in the light-pink flowers of Chulumani and QM24, none in the white flowers of *N. sylvestris*, and high levels in the dark-pink flowers of *N. tomentosiformis* and QM25 and the magenta flowers of 095-55 (McCarthy et al. 2017; Fig. 11.1). In addition to cyanidin, the natural *N. tabacum* accessions contain a small amount of transgressive pelargonidin, which is not present in either progenitor or in the synthetic lines. Floral tissue in all accessions contains the flavonols kaempferol and quercetin. Kaempferol levels are low in *N. sylvestris*, 095-55, and QM25, high in *N. tomentosiformis* and Chulumani, and intermediate in QM24. Quercetin levels are low in *N. sylvestris*, high in *N. tomentosiformis*, and intermediate in natural and synthetic *N. tabacum* accessions (McCarthy et al. 2017; Fig. 11.1).

Anthocyanin and flavonol pigments are produced by the flavonoid biosynthetic pathway, a branched pathway that yields multiple pigment types (Grotewold 2006; Fig. 11.1). The enzymes chalcone synthase (CHS), chalcone isomerase (CHI), and flavanone 3-hydroxylase (F3H) act early in the pathway and are necessary for the production of both flavonols and anthocyanins. The enzymes flavonoid 3′-hydroxylase (F3′H) and flavonoid 3′,5′-hydroxylase (F3′5′H) shift the pathway along different branches toward different pigments. Flavonol synthase (FLS) produces flavonols from the intermediate substrates dihydrokaempferol, dihydroquercetin, and dihydromyricetin, whereas dihydroflavonol 4-reductase (DFR) and anthocyanin synthase (ANS) convert these substrates into anthocyanidins, which can then be converted to anthocyanins (Grotewold 2006). The transcription factors anthocyanin 1 (AN1, basic helix-loop-helix), anthocyanin 2 (AN2, R2R3 MYB), and WD40 act together as a complex to induce anthocyanin production (Grotewold 2006).

Many evolutionary changes in floral color have been attributed to mutations in genes of the flavonoid biosynthetic pathway. Floral color shifts from blue to red are the result of inactivation, down-regulation, or deletion of *F3′5′H* or *F3′H* genes, resulting in production of red pelargonidin pigmentation instead of blue cyanidin or delphinidin (Zufall and Rausher 2004; Streisfeld and Rausher 2009; Des Marais and Rausher 2010; Smith and Rausher 2011; Wessinger and Rausher 2015). Several studies of the genetic basis of shifts from pigmented to white flowers have identified the cause as inactivation of the R2R3 MYB gene (orthologous to *AN2* in *Nicotiana*) that regulates the pathway (Quattrocchio et al. 1999; Schwinn et al. 2006; Hoballah et al. 2007). Some studies, however, have found that partial reduction or complete abolition of anthocyanin pigment is driven by repression of the pathway by an R3 MYB transcription factor (Yuan et al. 2013; Gates et al. 2017). Our questions focus on the genetic basis of the variety of floral colors observed in *N. tabacum* accessions and are as follows: (1) Do the expression levels of genes involved in anthocyanin synthesis (*DFR*, *ANS*, *AN1*, and *AN2*) predict cyanidin concentration, i.e., is the expression of these genes up-regulated in dark-pink flowers? (2) Are light-pink flowers lighter than dark-pink flowers because *DFR* and *ANS* are activated later in development, resulting in lower cyanidin accumulation? (3) In pairwise comparisons of dark-pink and light-pink flowers, are the same genes involved in the flavonoid biosynthetic pathway differentially expressed?

11.2 Materials and Methods

11.2.1 Plant Growth and Material

We grew plant material in greenhouses with natural sunlight and temperatures between 10 and 30 °C. We used the following plant material: *N. sylvestris* A04750326 (Radboud University, Nijmegen, The Netherlands), *N. tomentosiformis* BRNO 4103 (acquired from A. Kovařík, Brno, Czech Republic), *N. tabacum* 095-55 (IPK Gatersleben, Gatersleben, Germany), *N. tabacum* “Chulumani” (collected in the field in Bolivia by S. Knapp), and two first-generation synthetic lines, QM24 and QM25 (created by K. Y. Lim at Queen Mary University of London by crossing 4× autotetraploid *N. sylvestris* and 4× autotetraploid *N. tomentosiformis*). Because we used multiple accessions of *N. tabacum*, we will refer to all six plant lines as accessions.

Table 11.1 Primer sequences used for digital droplet PCR

Gene	Sequence (5′→3′)
DFR F	CGTCACTGGAGCAGCTGGAT
DFR R	TCTCAGGATCACGAACAGTAGCG
FLS F	TGAAGGGTATGGTACTTCTTTGCAGA
FLS R	CGATAATTGATGGCAGAAGGAGGC
GAPDH F	TGACAGATTTGGAATTGTTGAGGGTCT
GAPDH R	CTTCCACCCCTCCAGTCCTTG
EF1α F	CTTGGTGGTATTGACAAGCGTGT
EF1α R	TGTCAAGCACCCAGGCATACT
L25 F	GGACAAAAGTTACATTCCACCGACC
L25 R	AGTTTGTTCCTTCCAGGTGCACT

11.2.2 Developmental Series, RNA Extraction, and cDNA Synthesis

We examined floral development in each accession to identify stages for use in transcriptome analyses by timing of the appearance of floral pigment. We measured corolla tube length, including the floral limb (the portion of the flower that opens at anthesis), at anthesis in 5–10 flowers per accession by dissecting the corolla tube, pinning it flat, photographing it, and measuring length using ImageJ (version 1.51 k; Rasband 1997). We used the mean tube length at anthesis to calculate the percentage of anthesis length at which pigment appeared for each accession. At 60% of anthesis length, none of the accessions displayed floral pigmentation, but different accessions began to show pigmentation by 85% or 95% of anthesis length. We therefore selected 60, 85, and 95% of anthesis corolla length as stages for transcriptome analysis. Floral buds were collected using the corolla length values ±1 mm. We dissected the corolla tissue, cut the stamens from the corolla, and placed the tissue in liquid nitrogen. We collected three biological replicates from different plants, where available, for each accession at each developmental time point. Each independent synthetic accession consisted of a single first-generation individual, so three replicate flowers were collected from each plant for these lines. Because of the scarcity of material, we collected only two biological replicates for *N. tomentosiformis* at 85% of anthesis length and two flowers for the synthetic line QM24 at 95% of anthesis length.

We extracted RNA using the RNeasy Mini Plant Kit (Qiagen, Hilden, Germany), DNase-treated using the Turbo-DNase Kit (Ambion, Thermo Fisher Scientific, Waltham, MA, USA), and synthesized cDNA from 1 μg of RNA using the SuperScript III Reverse Transcriptase Kit (Invitrogen, Carlsbad, CA, USA) with the provided oligo-dT primer.

11.2.3 Digital Droplet PCR (ddPCR)

To quantify *FLS* and *DFR* transcript accumulation, we performed ddPCR on a QX200 ddPCR system (Bio-Rad Laboratories, Hercules, CA, USA). We used QX200 ddPCR EvaGreen SuperMix and 2 pmol of each primer with the following program: 95 °C for 5 min, followed by 40 cycles of 95 °C for 30 s and 60 °C for 60 s, followed by 4 °C for 5 min and 90 °C for 5 min, using a ramp rate of 2 °C per second. We normalized transcript accumulation using the geometric mean of three reference genes: elongation factor 1 alpha (*EF1α*), glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*), and ribosomal protein gene *L25*. Table 11.1 lists the primer sequences. We analyzed the ddPCR data using QuantaSoft Analysis Pro, version 1.0 (Bio-Rad Laboratories).

11.2.4 Illumina Library Preparation and Sequencing

To prepare cDNA libraries for 60, 85, and 95% of anthesis length samples from all six accessions, strand-specific libraries were constructed from mRNA as previously described (Zhong et al. 2011) and sequenced on a HiSeq 2500 system (Illumina, San Diego, CA) with 1 × 85 bp reads. Sequencing data from a smaller, preliminary experiment was also included to create our reference assembly (see below in the "Transcriptome analyses" section for details), using corolla tissue at anthesis from three biological replicates each of *N. tabacum* 095-55 and *N. tabacum* 51789 (this accession was not used in the larger analysis). Libraries were prepared with the NEBNext Ultra Directional RNA Library Prep Kit for Illumina (New England Biolabs, Ipswich, MA, USA) according to the manufacturer's protocols. Library quality was assessed using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA), and libraries were sequenced on an Illumina NextSeq 500 platform at the University of California, Riverside Institute for Integrated Genome Biology Genomics Core Facility with 1 × 150 bp reads.

11.2.5 Transcriptome Analyses

We processed raw sequencing reads on the University of California, Riverside Biocluster by trimming adaptors using Cutadapt (version 1.15; Martin 2011), followed by filtering with a quality score of 30 and a minimum length of 60 bp using Sickle (version 1.33; Joshi and Fass 2011). We initially performed analyses of cleaned reads using a genome-guided approach in TopHat (version 2.1.1), Cufflinks (version 2.2.1; Trapnell et al. 2012), and cummerbund (version 2.20; Goff et al. 2013) with the available genomes of *N. sylvestris* (Sierro et al. 2013), *N. tabacum* (Edwards et al. 2017), and *N. tomentosiformis* (Sierro et al. 2013). Although *N. sylvestris* and *N. tomentosiformis* are the progenitors of the allotetraploid *N. tabacum*, the percentage of reads mapped varied widely (20–70%) both within and among accessions and depending on which genomes were used as the reference. We therefore adopted a de novo approach using the Trinity pipeline (version 2.4; Grabherr et al. 2011; Haas et al. 2013).

We generated a single reference assembly using all sequencing reads (60, 85, and 95% of anthesis length from the four *N. tabacum* accessions, *N. sylvestris*, and *N. tomentosiformis* and at anthesis from *N. tabacum* 095-55 and *N. tabacum* 51789), and normalized them in silico to 50× to improve the efficiency of the assembly by decreasing the number of reads included (Haas et al. 2013). We mapped sequencing reads back to the reference assembly using RNA-Seq by Expectation Maximization implemented with Bowtie2 (version 2.2.9; Langmead and Salzberg 2012). We annotated the reference assembly using the Transdecoder (version 3.0) plugin in Trinity by performing BLASTx and BLASTp searches against the Swiss-Prot (The UniProt Consortium 2017) and Pfam (Finn et al. 2016) databases. Because of short read lengths, it was difficult to assemble full gene transcripts; therefore, any given gene locus may have several corresponding contigs in the de novo assembly. Additionally, likely due to short read lengths and nonuniformity of polymorphisms between the *N. sylvestris* and *N. tomentosiformis* genomes, there were difficulties isolating homeologous sequences in the de novo assembly; therefore, some contigs correspond to *N. sylvestris* copies, some contigs correspond to *N. tomentosiformis* copies, and some correspond to both homeologs.

We performed differential expression analyses with the bioconductor package limma/voom (version 3.6; Ritchie et al. 2015). We ran two sets of comparisons, across accessions within developmental time points and across developmental time points within accessions. To ensure robust differences, we classified differentially expressed contigs as those with p-value < 0.05, false discovery rate < 0.05, and log2 fold change > 2 (absolute value).

11.2.6 Identifying Flavonoid Biosynthesis Transcripts in the de novo Assembly

To identify contigs in the de novo assembly that correspond to genes in the flavonoid biosynthetic pathway, we performed BLAST searches using previously identified flavonoid biosynthetic genes from *Iochroma* and *Petunia* species (both in family Solanaceae) to find putative orthologs from *N. sylvestris* and *N. tomentosiformis* in GenBank. A single hit was found for each gene in the flavonoid biosynthetic pathway for *N. sylvestris* and *N. tomentosiformis*, apart from *CHS* for which two genes were found. We then executed custom BLAST searches in Geneious (version 10, Auckland, New Zealand) with the detected *N. sylvestris* and *N. tomentosiformis* orthologs against the de novo reference assembly to identify the contigs of genes likely involved in flavonoid biosynthesis in *Nicotiana*. We extracted these flavonoid biosynthesis contigs from read count data for each transcriptome and from the differential expression analyses for use in additional investigations.

11.2.7 Comparative Quantitative Expression of Flavonoid Biosynthetic Genes

To determine whether increase in *DFR* and *ANS* expression accounts for the higher cyanidin concentration observed in dark-pink-flowered species, we generated strip charts in R (version 3.4.2) using the ggplot2 package (version 2.2.1; Wickham 2009). We plotted the log2 fold change values of differentially expressed flavonoid biosynthetic genes for pairwise comparisons between accessions within each developmental time point. To establish whether the magnitude and direction of expression differences in these transcripts reflected those observed in the transcriptome overall, we overlaid these strip charts onto violin plots (which display the distribution of the data in a manner similar to that of a histogram) of the differentially expressed genes (DEGs) from the whole transcriptome using ggplot2. As our findings suggested that some floral color differences involved more complex dynamics than simply increased levels of *DFR* and *ANS*, we also examined relative expression dynamics across the pathway to identify patterns that involved more than a single gene. To do this, we summed the transcript levels for each gene (e.g., *CHS*) and produced heatmaps in R using the mixOmics package (version 6.3.1; Cao et al. 2016). To further investigate whether competition for the same substrates may affect pigment concentrations, we calculated the *FLS*:*DFR* ratio using both transcriptome and ddPCR data and compared it to floral phenotype and total anthocyanidin concentration.

11.2.8 Comparative Quantitative Expression of Flavonoid Biosynthetic Genes in Light-Pink and Dark-Pink Flowers

To determine whether accessions with the same floral color display similar relative transcript levels of flavonoid biosynthetic genes, we plotted Venn diagrams using the R packages Venn diagram (version 1.6.18; Chen and Boutros 2011) and gplots (version 3.0.1; Warnes et al. 2009). We compared DEGs between *N. tabacum* accessions and each of their progenitors to identify those shared by light-pink or dark-pink flowers that differ from progenitor expression levels. We also examined overlap in pairwise comparisons between *N. tabacum* accessions with light-pink and dark-pink flowers (095-55 versus Chulumani, 095-55 versus QM24, and Chulumani versus QM25).

11.2.9 Comparative Quantitative Expression of Flavonoid Biosynthetic Genes Across Development

To establish whether differences in developmental timing of *DFR* and *ANS* activation

underlie changes in floral color, we examined patterns of expression of flavonoid biosynthetic genes across development within each accession by generating strip charts overlaid on violin plots and heatmaps, as described in the Sect. 11.2.7.

11.3 Results

11.3.1 Patterns of Differential Gene Expression Among Accessions and Across Development

To elucidate the genetic basis of floral color variation among white, light-pink, and dark-pink *Nicotiana* flowers, we used transcriptome analyses and ddPCR to quantify the expression of flavonoid biosynthetic pathway genes from two natural and two synthetic accessions of *N. tabacum* and their diploid progenitors. Because of difficulty in mapping the transcriptomes to the publicly available *Nicotiana* genomes, we created a de novo assembly for sequence analysis. Sequencing coverage was low for one biological replicate of QM24 at 60% of anthesis length; therefore, we excluded this replicate from further analyses. The number of reads mapped from each replicate ranged from 3.66 to 9.42 million (Table 11.2).

In pairwise comparisons of accession transcriptomes at a given developmental time point, we found the highest number of DEGs between progenitor species and the smallest number between synthetic lines (Table 11.3). There was also a higher number of DEGs between *N. tabacum* accessions and their progenitors than between *N. tabacum* accessions (Table 11.3). In comparisons across developmental time points within each accession, the highest number of DEGs was found between 60 and 95% of anthesis length, followed by 60 and 85% and between 85 and 95% (Table 11.4). Not surprisingly, far more genes were differentially expressed between accessions than across development within an accession.

11.3.2 Higher Levels of Anthocyanin-Specific Biosynthetic Transcripts Are Correlated with Pink Versus White Phenotype Comparisons, but Variation Between Pink Phenotypes Is More Complex

To determine whether relatively higher expression levels of *DFR* and *ANS* in dark-pink flowers might underlie differences in floral color, we compared DEGs involved in the flavonoid biosynthetic pathway between pairs of accessions at each developmental time point (Fig. 11.2). In comparisons between white (*N. sylvestris*) and pink (all other accessions) flowers, most of the differentially expressed flavonoid biosynthetic transcripts were higher in pink flowers. These included genes that produce anthocyanins (*DFR*, *ANS*, *AN1*, and *AN2*) as we hypothesized, but also other genes in the pathway (*CHS*, *CHI*, *F3H*, *F3′H*, and *FLS*). In comparisons between light-pink and dark-pink flowers, some of the DEGs were up-regulated in lighter flowers, whereas others were up-regulated in darker flowers. This contrasted with the findings for comparisons involving *N. sylvestris* (Fig. 11.2), suggesting that the production of dark- versus light-pink flowers is not simply the result of higher expression of anthocyanin biosynthetic genes, but that more complex expression dynamics in the pathway are involved.

To determine whether relative expression levels across the pathway as a whole are consistent with a role in determining different color phenotypes, we generated heatmaps for the summed contig counts for each flavonoid biosynthetic gene across accessions. *CHS* expression levels were higher than expression levels in the other genes, so to facilitate interpretation of expression patterns among the genes present at lower levels, we also generated heatmaps without *CHS* (Fig. 11.3). Some patterns were consistent with white/light-pink (*N. sylvestris*, Chulumani, and QM24) versus dark-pink

Table 11.2 Number of reads mapped to the de novo reference assembly

Sample	Reads mapped
sylv a 60	5,674,537
sylv b 60	6,280,955
sylv c 60	5,309,214
sylv a 85	4,224,300
sylv b 85	8,116,442
sylv c 85	8,450,752
sylv a 95	5,207,567
sylv b 95	8,235,247
sylv c 95	4,501,618
tomf a 60	6,540,640
tomf b 60	4,660,980
tomf c 60	6,062,876
tomf a 85	6,782,341
tomf b 85	8,585,868
tomf a 95	7,746,705
tomf b 95	6,238,237
tomf c 95	6,602,617
095-55 a 60	6,439,524
095-55 b 60	4,198,419
095-55 c 60	8,176,370
095-55 a 85	5,331,654
095-55 b 85	6,718,258
095-55 c 85	7,064,007
095-55 a 95	6,138,186
095-55 b 95	7,087,995
095-55 c 95	7,136,716
Chulu a 60	4,296,617
Chulu b 60	5,880,425
Chulu c 60	5,626,046
Chulu a 85	6,004,335
Chulu b 85	6,861,648
Chulu c 85	9,198,169
Chulu a 95	5,297,129
Chulu b 95	8,899,176
Chulu c 95	5,928,235
QM24a 60	4,957,023
QM24c 60	9,365,190
QM24a 85	9,415,003
QM24b 85	8,618,508
QM24c 85	4,650,289
QM24a 95	6,162,663
QM24b 95	6,400,574

(continued)

Table 11.2 (continued)

Sample	Reads mapped
QM25a 60	5,161,262
QM25b 60	6,854,781
QM25c 60	4,610,085
QM25a 85	4,661,285
QM25b 85	3,854,284
QM25c 85	4,946,783
QM25a 95	5,768,444
QM25b 95	3,660,452
QM25c 95	5,885,153

095-55, *Nicotiana tabacum* 095-55; Chulu, *N. tabacum* “Chulumani”; sylv, *N. sylvestris*; tomf, *N. tomentosiformis*. QM24 and QM25 are first-generation synthetic lines

(*N. tomentosiformis*, 095-55, and QM25) floral color categories. Early in development (60% of anthesis length), *DFR* levels were lower in white/light-pink flowers. At 60 and 85% of anthesis length, *F3H* levels were higher in white/light-pink flowers. We also saw, however, patterns that did not correspond to floral color categories. The low *DFR* and *F3H* levels were not maintained in light-pink flowers at later developmental stages. In addition, at 60 and 85% of anthesis length, *FLS* expression was higher in *N. tomentosiformis*, Chulumani, and both synthetic lines than in *N. sylvestris* and 095-55, corresponding to lower concentration of flavonols in the white *N. sylvestris* and dark-pink 095-55 flowers. At 95% of anthesis length, *FLS* expression was similar across accessions. Expression of the regulatory genes *AN1* (at all

Table 11.3 Differentially expressed contigs across the transcriptomes of *Nicotiana* accessions

Comparison	60% of anthesis length	85% of anthesis length	95% of anthesis length
sylv versus tomf	35,517	32,920	35,638
095-55 versus sylv	16,037	15,930	18,661
Chulu versus sylv	16,352	17,260	17,171
QM24 versus sylv	10,771	11,323	11,895
QM25 versus sylv	12,080	14,726	14,316
095-55 versus tomf	18,643	17,869	20,353
Chulu versus tomf	17,976	18,896	18,918
QM24 versus tomf	13,615	14,420	11,268
QM25 versus tomf	13,430	14,778	14,562
QM24 versus 095-55	7,478	7,308	5,487
QM24 versus Chulu	6,212	7,573	4,012
QM25 versus 095-55	6,207	7,592	7,257
QM25 versus Chulu	3,688	7,024	5,238
095-55 versus Chulu	510	297	35
QM24 versus QM25	0	0	0

095-55, *Nicotiana tabacum* 095-55; Chulu, *N. tabacum* “Chulumani”; sylv, *N. sylvestris*; tomf, *N. tomentosiformis*. QM24 and QM25 are first-generation synthetic lines

Table 11.4 Differentially expressed contigs across developmental time points for each *Nicotiana* accession

Comparison	*N. sylvestris*	*N. tomentosiformis*	*N. tabacum* 095-55	*N. tabacum* "Chulumani"	QM24	QM25
60% versus 85% of anthesis length	52	234	851	2,092	64	291
60% versus 95% of anthesis length	538	1,900	4,198	4,858	3,511	1,736
85% versus 95% of anthesis length	5	0	545	0	86	180

QM24 and QM25 are first-generation synthetic lines

developmental time points) and *AN2* (at 60 and 85% of anthesis length) was higher in *N. tomentosiformis*, 095-55, and Chulumani than in the other accessions (Fig. 11.3).

Because *FLS* and *DFR* encode proteins that function in different branches of the pathway and because their expression levels vary across accessions at 60% of anthesis length, we determined whether differences in the expression of these genes correlated with differences in floral color. Consistent with floral color differences, *DFR* levels at 60% of anthesis length were significantly higher in dark-pink flowers than in light-pink or white flowers (with the exception of 095-55 versus Chulumani). However, *DFR* levels were similar in white versus light-pink, light-pink versus light-pink, and dark-pink versus dark-pink comparisons (Fig. 11.2a). In contrast, *FLS* expression differences at 60% of anthesis length were not consistent with differences between flower colors; for example, in light-pink versus dark-pink comparisons, *FLS* expression was significantly higher in QM24 versus 095-55 but similar in Chulumani versus QM25 (Fig. 11.2a). Because FLS and DFR act on the same substrates and may compete for them, we calculated *FLS*:*DFR* expression ratios to establish whether this ratio was consistent with a role in determining floral color. In both the transcriptome and ddPCR data, the *FLS*:*DFR* ratio was high (>35 fold change) at 60% of anthesis length in light-pink flowers (Chulumani and QM24), whereas other accessions (with the exception of *N. sylvestris* in the ddPCR data) at 60% and all accessions at 85 and 95% of anthesis length had much lower *FLS*:*DFR* ratios (Fig. 11.4a, b). To determine whether the high ratios in light-pink (Chulumani and QM24) and white (*N. sylvestris*) flowers were consistent with lower levels of anthocyanidins, suggesting diversion of the pathway to flavonols by FLS, we plotted total anthocyanidin concentration against *FLS*:*DFR* ratio at 60% of anthesis length (Fig. 11.4c). Although the relationship was not linear, the accessions with high *FLS*:*DFR* ratios tended to have lower anthocyanidin concentration, suggesting that a high *FLS*:*DFR* ratio early in floral development may play a role in preventing the accumulation of anthocyanidins.

11.3.3 Different Pathway Modifications Underlie Shifts Between Light-Pink and Dark-Pink Flowers

To determine whether expression differences of specific genes in the flavonoid biosynthetic pathway at specific developmental time points underlie the differences between light-pink and dark-pink flowers, we used pairwise DEG comparisons between accessions. We generated Venn diagrams of all four *N. tabacum* accessions versus each of their diploid progenitors within developmental time points (Fig. 11.5a). We then extracted the identity of the contigs from the sectors of the Venn diagram that were shared between light-pink and between dark-pink accessions. Dark-pink accessions (095-55 and

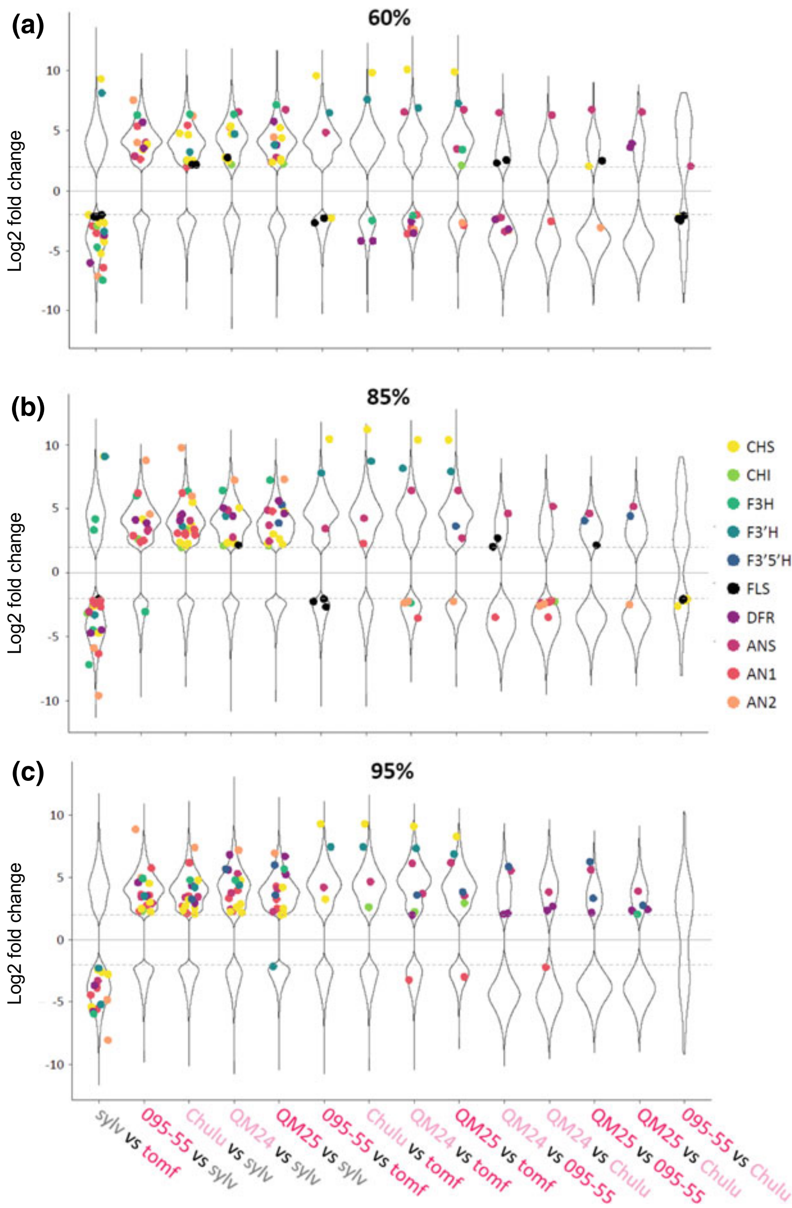

Fig. 11.2 Differential gene expression across *Nicotiana* accessions at **a** 60%, **b** 85%, and **c** 95% of anthesis length. The violin plots represent transcriptome-wide differentially expressed contigs (log2 fold change > 2 (absolute value)). The points represent differentially expressed contigs from the flavonoid biosynthetic pathway, colored for each gene. Dashed lines at 2 and −2 represent the cutoff for differentially expressed contigs. Pairwise comparisons are shown along the x-axis, and the accessions are color-coded according to flower color: gray = white; light pink = light pink; magenta = darkpink. 095-55, *Nicotiana tabacum* 095-55; Chulu, *N. tabacum* "Chulumani"; sylv, *N. sylvestris*; tomf, *N. tomentosiformis*. QM24 and QM25 are first-generation synthetic lines. Differentially expressed genes that are up-regulated in the first accession of the pair are in the positive (upper) section of each graph and those that are up-regulated in the second accession of the pair are in the negative (lower) section of each graph. CHS, chalcone synthase; CHI, chalcone isomerase; F3H, flavanone 3-hydroxylase; F3′H, flavonoid 3′-hydroxylase; F3′5′H, flavonoid 3′,5′-hydroxylase; FLS, flavonol synthase; DFR, dihydroflavonol 4-reductase; ANS, anthocyanidin synthase; AN1, anthocyanin 1; AN2, anthocyanin 2

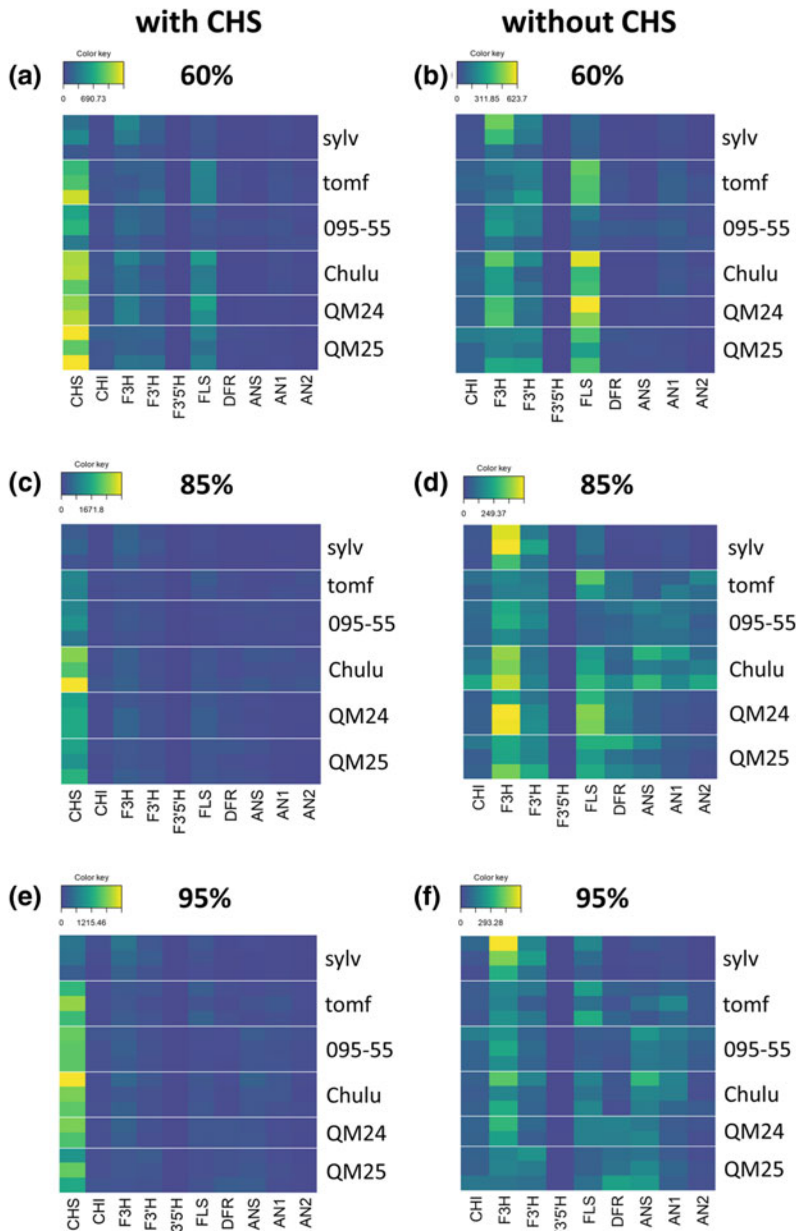

Fig. 11.3 Heatmaps of summed transcript levels for flavonoid biosynthetic genes in *Nicotiana* accessions. Transcript levels were compared at 60% (**a**, **b**), 85% (**c**, **d**), and 95% (**e**, **f**) of anthesis length, with (**a**, **c**, **e**) and without (**b**, **d**, **f**) *CHS*, which encodes chalcone synthase. CHI, chalcone isomerase; F3H, flavanone 3-hydroxylase; F3′H, flavonoid 3′-hydroxylase; F3′5′H, flavonoid 3′,5′-hydroxylase; FLS, flavonol synthase; DFR, dihydroflavonol 4-reductase; ANS, anthocyanidin synthase; AN1, anthocyanin 1; AN2, anthocyanin 2. Accessions: 095-55, *Nicotiana tabacum* 095-55; Chulu, *N. tabacum* "Chulumani"; sylv, *N. sylvestris*; tomf, *N. tomentosiformis*. QM24 and QM25 are first-generation synthetic lines

QM25) had higher *DFR* and *ANS* expression levels than *N. sylvestris* (Fig. 11.5b), which was expected because of the presence of cyanidin in the dark-pink flowers (Fig. 11.1a). Similar differences were expected for light-pink versus white comparisons, perhaps to a lesser

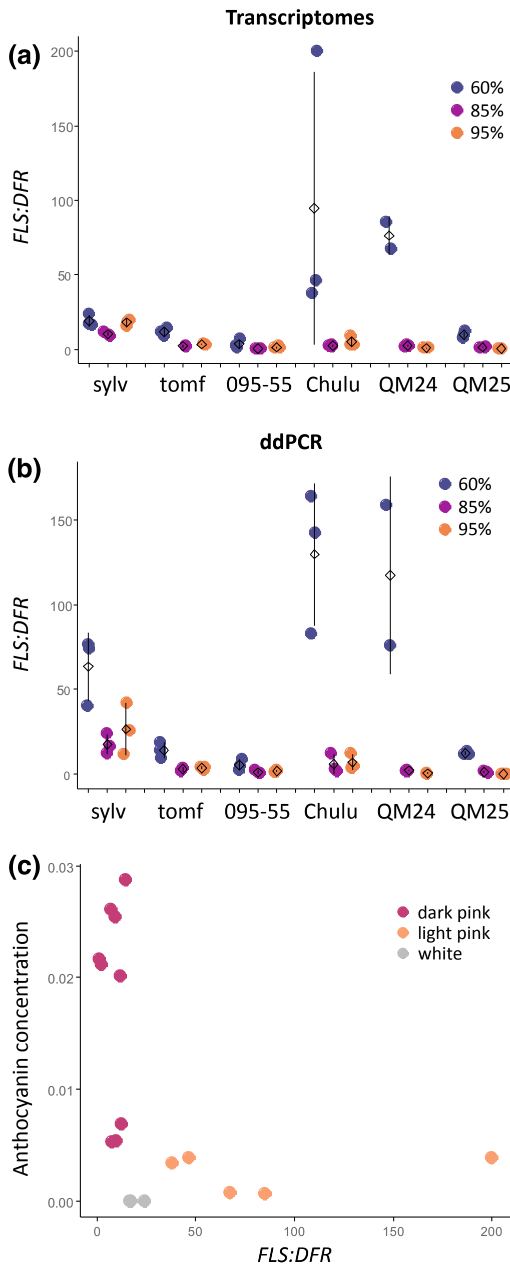

Fig. 11.4 *FLS*:*DFR* expression ratio from transcriptome (**a**) and digital droplet PCR (ddPCR) (**b**) data at 60%, 85%, and 95% of anthesis length in *Nicotiana* accessions. *FLS* encodes flavonol synthase and *DFR* encodes dihydroflavonol 4-reductase. Black diamonds and lines represent the mean and standard deviation, respectively. **c** Total anthocyanidin concentration versus transcriptome *FLS*:*DFR* ratio with points representing flower color. 095-55, *Nicotiana tabacum* 095-55; Chulu, *N. tabacum* "Chulumani"; sylv, *N. sylvestris*; tomf, *N. tomentosiformis*. QM24 and QM25 are first-generation synthetic lines

magnitude, but this was not observed. Light-pink accessions (Chulumani and QM24) had higher *FLS* expression levels than *N. sylvestris*, but dark-pink accessions (095-55 and QM25) did not (Fig. 11.5b). The higher *FLS* expression levels exclusively in light-pink flowers were not expected. This suggests that in the development of light-pink flowers, the flux of the flavonoid biosynthetic pathway may be shunted toward the production of flavonols, whereas in dark-pink flowers it may be directed toward anthocyanins. In the comparisons of the *N. tabacum* accessions to *N. tomentosiformis*, light-pink (but not dark-pink) accessions had lower *F3′H* and *DFR* expression levels, whereas dark-pink (but not light-pink) accessions had higher *ANS* expression levels (Fig. 11.5b). These findings were consistent with lower cyanidin production in light-pink flowers but the higher *ANS* expression in dark-pink *N. tabacum* accessions does not seem to result in higher cyanidin production compared to *N. tomentosiformis*, perhaps because there is no accompanying increase in *DFR* expression levels.

To ascertain whether the same genes were differentially expressed in all comparisons of light-pink and dark-pink flowers, we generated Venn diagrams to compare differential expression in pairwise comparisons between light-pink and dark-pink *N. tabacum* accessions (095-55 versus Chulumani, 095-55 versus QM24, and Chulumani versus QM25). No DEGs were shared among all three comparison sets (Fig. 11.5c, d), suggesting that similar colors were achieved in different ways. In comparisons between 095-55 and both light-pink accessions at 60 and 85% of anthesis length, however, *FLS* expression levels were higher in the light-pink flowers. In contrast, light-pink flowers had lower *DFR* levels in 095-55 versus QM24 and QM25 versus Chulumani comparisons at 60% of anthesis length. Either high *FLS* or low *DFR* expression levels can result in a high *FLS*:*DFR* ratio, suggesting that the factor driving differences between light-pink and dark-pink flowers may be the ratio of *FLS* to *DFR* expression. No other consistent differences were detected between light-pink and dark-pink flowers, but

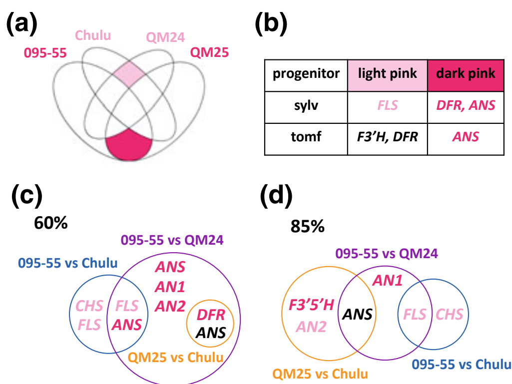

Fig. 11.5 Shared differentially expressed genes from the flavonoid biosynthetic pathway in comparisons between *Nicotiana* accessions. **a** Example Venn diagram of differentially expressed genes from pairwise comparisons of each *N. tabacum* accession with one diploid progenitor; analyses were performed for comparisons with both *N. sylvestris* and *N. tomentosiformis*, but only an example diagram is shown here. Light-pink and dark-pink sections highlight genes that are differentially expressed exclusively in light-pink accessions versus progenitor or dark-pink accessions versus progenitor, respectively. **b** Summary of the Venn diagram data for *N. tabacum* accessions versus progenitor at 60% of anthesis length. The genes in the table are those found in the light-pink and dark-pink sectors of (**a**) for comparisons between the *N. tabacum* accessions and *N. sylvestris* or *N. tomentosiformis*. These represent differentially expressed genes exclusive to light-pink or dark-pink accessions versus progenitor comparisons; pink text indicates higher expression in the *N. tabacum* accession, and black text indicates higher expression in the progenitor. (**c**, **d**) Venn diagrams of comparisons between light-pink and dark-pink *N. tabacum* accessions at 60% (**c**) and 85% (**d**) of anthesis length. Light-pink text represents genes that are up-regulated in light flowers; dark-pink text represents genes that are up-regulated in dark flowers; black text represents genes for which the expression levels do not correlate with flower color. Blue, purple, and orange labels and circles represent separate light-pink versus dark-pink comparisons. CHS, chalcone synthase; F3′H, flavonoid 3′-hydroxylase; F3′5′H, flavonoid 3′,5′-hydroxylase; FLS, flavonol synthase; DFR, dihydroflavonol 4-reductase; ANS, anthocyanidin synthase; AN1, anthocyanin 1; AN2, anthocyanin 2. Accessions: 095-55, *Nicotiana tabacum* 095-55; Chulu, *N. tabacum* "Chulumani"; sylv, *N. sylvestris*; tomf, *N. tomentosiformis*. QM24 and QM25 are first-generation synthetic lines

lower levels of the regulatory genes *AN1* (at 60 and 85% of anthesis length) and *AN2* (at 60% of anthesis length) were observed in QM24 than in dark-pink 095-55, although not in other light-pink versus dark-pink comparisons. In addition, *CHS* was found at higher levels in the light-pink flowers of Chulumani versus 095-55 at 60 and 85% of anthesis length, but not in any other comparisons. This variation in results suggests that the mechanism behind the differences between light-pink and dark-pink floral color may be a combination of several factors.

11.3.4 DFR Up-Regulation Occurs Later in Development in Light-Pink Flowers

Gene expression comparisons between light-pink and dark-pink flowers found lower *DFR* expression in light-pink flowers at 60% of anthesis length; we therefore assessed whether light-pink floral color may be the result of activation of *DFR* later in development. *DFR* expression was up-regulated at 85% of anthesis length compared to 60% of anthesis length and remained at high levels at 95% in *N. tomentosiformis*, Chulumani, and both synthetic lines (Fig. 11.6a–f). In heatmaps showing relative transcript levels across development within each accession, very low levels of *DFR* expression were observed in *N. sylvestris*, Chulumani, and QM24 at 60% of anthesis length (Fig. 11.6g–l). To compare expression across development in all accessions, we generated strip plots of *DFR* expression levels from the six transcriptomes (Fig. 11.7). *N. sylvestris* displayed low levels of *DFR* throughout development, which is consistent with its lack of anthocyanidin pigmentation. *DFR* levels in *N. tomentosiformis*, 095-55, and QM25 were significantly higher than those of *N. sylvestris* at 60% of anthesis length (Fig. 11.2a), indicating that *DFR* had already been activated in these accessions at this time point. In contrast, light-pink-flowered accessions (Chulumani and QM24) at 60% of anthesis length had *DFR* expression levels similar to those of *N. sylvestris*, suggesting that activation was delayed in light-pink flowers (after 60% but before 85% of anthesis length). This delay in *DFR* activation

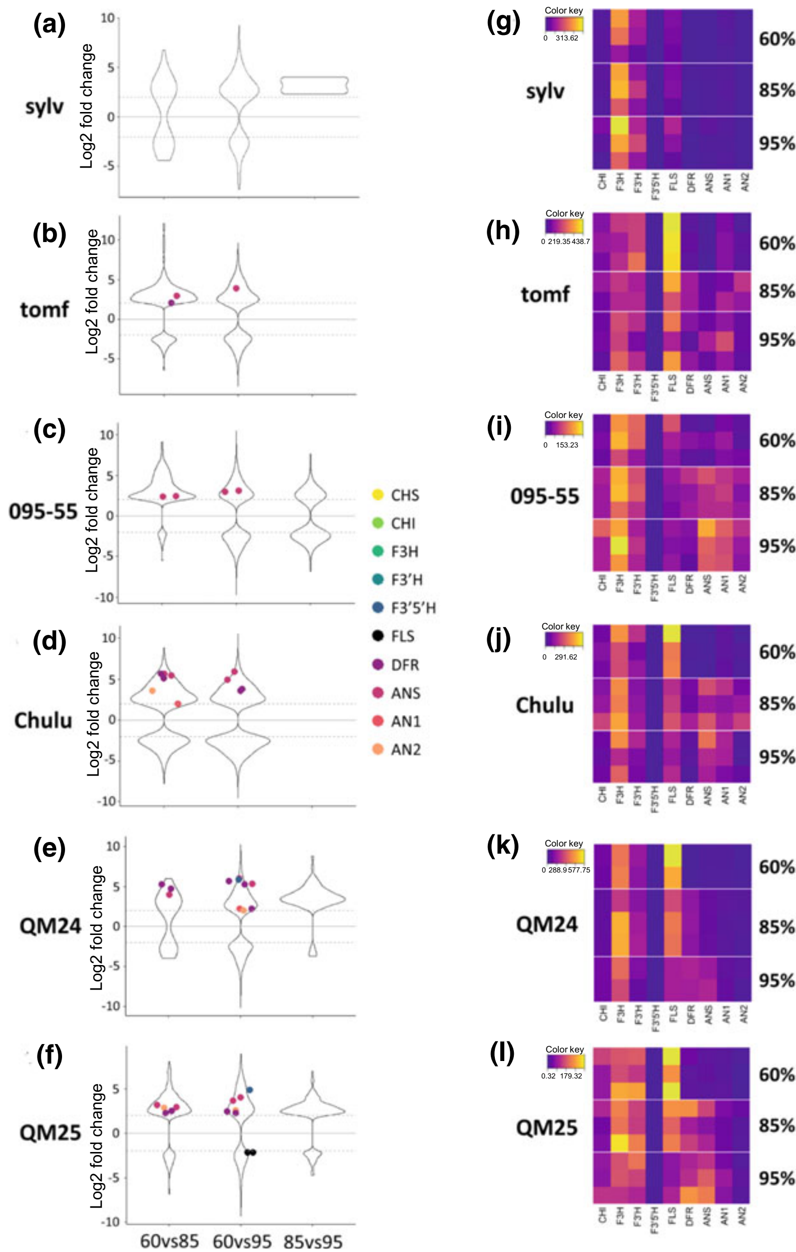

Fig. 11.6 Gene expression dynamics across developmental time points in *Nicotiana* accessions. Differentially expressed genes between different time points in **a** *N. sylvestris* (sylv), **b** *N. tomentosiformis* (tomf), **c** *N. tabacum* 095-55 (095-55), **d** Chulumani (Chulu), **e** the synthetic line QM24, and **f** the synthetic line QM25. The violin plots represent transcriptome-wide differentially expressed genes (log2 fold change > 2 (absolute value)). The points represent differentially expressed flavonoid biosynthesis genes, colored for each gene. Dashed lines at 2 and −2 represent the cutoff for differentially expressed genes. The positive (upper) portion of the graph represents genes that are up-regulated at the earlier developmental time point;the negative (lower) portion represents genes up-regulated at the later developmental time point. Heatmaps of summed transcript levels for flavonoid biosynthetic genes across development for **g** *N. sylvestris*, **h** *N. tomentosiformis*, **i** 095-55, **j** Chulumani, **k** QM24, and **l** QM25 without *CHS*, which encodes chalcone synthase. CHI, chalcone isomerase; F3H, flavanone 3-hydroxylase; F3′H, flavonoid 3′-hydroxylase; F3′5′H, flavonoid 3′,5′-hydroxylase; FLS, flavonol synthase; DFR, dihydroflavonol 4-reductase; ANS, anthocyanidin synthase; AN1, anthocyanin 1; AN2, anthocyanin 2

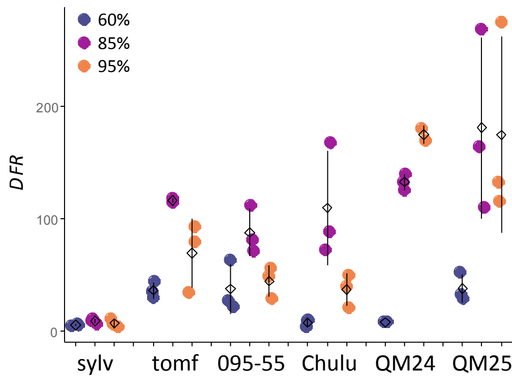

Fig. 11.7 *DFR* expression levels for each *Nicotiana* accession at 60, 85, and 95% of anthesis length. Black diamonds and lines represent the mean and standard deviation, respectively. 095-55, *Nicotiana tabacum* 095-55; Chulu, *N. tabacum* "Chulumani"; sylv, *N. sylvestris*; tomf, *N. tomentosiformis*. QM24 and QM25 are first-generation synthetic lines

may underlie the increased *FLS*:*DFR* ratio observed in these accessions, which may play a role in the decrease of anthocyanidin accumulation.

11.4 Discussion

11.4.1 Higher Expression Levels of Anthocyanin-Specific Genes Differentiate White from Pink Flowers

In comparisons between white-flowered *N. sylvestris* and each pink accession, expression of some or all anthocyanin-specific genes (*DFR*, *ANS*, *AN1*, and *AN2*) and of other genes in the pathway was higher in pink flowers. These results suggest that the difference between white and pink flowers arises from changes in pathway regulation and/or from expression of the genes that produce anthocyanins, consistent with reports that overexpression of either *AN1* or *AN2* increases anthocyanin concentration (Pattanaik et al. 2010; Bai et al. 2011). However, no *AN2* ortholog was detected in the *N. sylvestris* genome. Thus, it is not surprising that *AN2* expression was higher in all pink-flowered accessions than in *N. sylvestris* at all developmental time points examined. *AN2* activates *DFR* and *ANS* expression in *Nicotiana* (Pattanaik et al. 2010); therefore, the absence of the *AN2* gene from *N. sylvestris* is consistent with the lower levels of *DFR* and *ANS* observed in this species. Silencing or inactivation of *AN2* orthologs has been shown to yield a white-flowered phenotype in *N. tabacum* (Pattanaik et al. 2010), *Petunia* (Quattrocchio et al. 1999; Hoballah et al. 2007), and *Antirrhinum* (Schwinn et al. 2006). Taken together, these results suggest that lack of *AN2* expression due to loss of the gene from the genome may be responsible for the white floral phenotype in *N. sylvestris*.

11.4.2 Competition Between FLS and DFR May Underlie the Light-Pink Floral Phenotype

Variation in pink floral phenotypes appears to be driven by complex dynamics across the flavonoid biosynthetic pathway, consistent with differences in the degree of pigmentation rather than presence or absence. In comparisons between light-pink and dark-pink flowers, the latter exhibit increased expression of *DFR*, *ANS*, *AN1*, and/or *AN2*, whereas light-pink flowers have higher levels of *CHS* and/or *FLS*. These contrasting patterns suggest that the flux of the pathway may be directed toward producing anthocyanins in dark-pink flowers, but shunted toward flavonols in light-pink flowers. Light-pink accessions displayed high *FLS*:*DFR* ratios at 60% of anthesis length, whereas dark-pink accessions did not. In pairwise comparisons between light-pink and dark-pink accessions, this high *FLS*:*DFR* ratio in light-pink flowers was attributable to lower expression levels of *DFR*, higher expression levels of *FLS*, or both. Across development, *FLS* expression levels were high in *N. tomentosiformis*, Chulumani, and both synthetic lines at 60% of anthesis, but these accessions display both light- and dark-pink phenotypes, suggesting that high *FLS* expression alone is not sufficient to produce a light-pink floral phenotype. *DFR*

expression levels at 60% of anthesis length in the light-pink accessions Chulumani and QM24 were similar to those observed for *N. sylvestris* throughout development, suggesting that *DFR* activation is developmentally delayed until 85% of anthesis length in these accessions. In contrast, *N. tomentosiformis*, 095-55, and QM25 had higher levels of *DFR* expression than *N. sylvestris* at 60% of anthesis length, indicating that *DFR* activation had already occurred by 60% of anthesis length in these accessions. These results suggest that both high expression levels of *FLS* and low expression levels of *DFR* early in development, most likely due to delayed activation of *DFR*, may be required to generate the light-pink floral phenotype; future studies should test this hypothesis.

FLS and DFR are in competition for the same substrates, the dihydroflavonols dihydrokaempferol, dihydroquercetin, and dihydromyricetin, to produce flavonols and anthocyanins, respectively. This competition, however, involves more than only enzyme affinity for substrates. Both FLS and DFR have been shown to interact with CHS in vivo, and FLS may interfere with the DFR and CHS interaction (Crosby et al. 2011). In addition, a study in grape (*Vitis vinifera*) reported that flavonols can bind to DFR and inhibit its activity (Trabelsi et al. 2008). Several studies have shown that up-regulation or silencing of *FLS* can alter the visible floral color phenotype, despite its regulation of the production of flavonols, which are not visible to the human eye. Overexpression of *FLS* yields white or light-pink floral phenotypes in *Mimulus* (Yuan et al. 2016), *Petunia* (Sheehan et al. 2016), and *N. tabacum* (Luo et al. 2015). Similarly, silencing *FLS* increases anthocyanin content in *Arabidopsis* (Stracke et al. 2009), pink-flowered *Petunia* and *N. tabacum* (Holton et al. 1993), and white-flowered *Petunia*, although, in this study, anthocyanin pigmentation was only observed in floral buds and was not maintained at anthesis (Davies et al. 2003). This evidence supports our conclusion that differences between light- and dark-pink flower phenotypes may be attributable to competition between FLS and DFR, although several questions remain. Does FLS have a higher affinity for the shared substrates than DFR? If so, why does FLS not always outcompete DFR? How high does the *FLS*:*DFR* ratio need to be to tip the balance toward producing flavonols at the expense of anthocyanins? Further studies are required to address these questions.

Acknowledgements We thank James Giovannoni and Yimin Xu for transcriptome sequencing, Zhangjun Fei for help with sequence analysis, Loyal Goff for assistance with cummerbund, and Christopher Fiscus, Dinusha Maheepala, Glen Morrison, and Alex Rajewski for help with R programming.

Data Availability

Raw sequencing reads used in this study have been submitted to the Sequence Read Archive (https://www.ncbi.nlm.nih.gov/sra) with the following accession numbers: *Nicotiana sylvestris* (SRR6434939, SRR6434941–SRR6434948), *N. tabacum* 095-55 (SRR6434933–SRR6434937, SRR6434953, SRR6434966, SRR6434967, SRR6434696, SRR6434698, SRR6434699), *N. tabacum* 51789 (SRR6434694, SRR6434695, SRR6434697), *N. tabacum* “Chulumani” (SRR6434929–SRR6434932, SRR6434960–SRR6434964), synthetic *N. tabacum* QM24 (SRR6434921, SRR6434922, SRR6434949, SRR6434950, SRR6434965, SRR6434968, SRR6434969), synthetic *N. tabacum* QM25 (SRR653419, SRR6434920, SRR6434923–SRR6434928, SRR6434951), and *N. tomentosiformis* (SRR6434940, SRR6434952, SRR6434954–SRR6434959). Analysis code is available on GitHub (https://github.com/jblandis/Nicotiana_transcriptomics).

References

Bai Y, Pattanaik S, Patra B et al (2011) Flavonoid-related basic helix-loop-helix regulators, *NtAn1a* and *NtAn1b*,

of tobacco have originated from two ancestors and are functionally active. Planta 234:363–375

Cao L, Rohart F, Gonzalez I et al (2016) mixOmics: Omics data integration project. R package version 6 (1):1

Chase MW, Knapp S, Cox AV et al (2003) Molecular systematics, GISH and the origin of hybrid taxa in *Nicotiana* (Solanaceae). Ann Bot 92:107–127

Chen H, Boutros PC (2011) VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. BMC Bioinform 12:35

Clarkson JJ, Dodsworth S, Chase MW (2017) Time-calibrated phylogenetic trees establish a lag between polyploidisation and diversification in *Nicotiana* (Solanaceae). Plant Syst Evol 303:1001–1012

Clarkson JJ, Kelly LJ, Leitch AR et al (2010) Nuclear glutamine synthetase evolution in *Nicotiana*: phylogenetics and the origins of allotetraploid and homoploid (diploid) hybrids. Mol Phylogenet Evol 55:99–112

Clarkson JJ, Knapp S, Garcia VF et al (2004) Phylogenetic relationships in *Nicotiana* (Solanaceae) inferred from multiple plastid DNA regions. Mol Phylogenet Evol 33:75–90

Crosby KC, Pietraszewska-Bogiel A, Gadella TWJ Jr, Winkel BSJ (2011) Förster resonance energy transfer demonstrates a flavonoid metabolon in living plant cells that displays competitive interactions between enzymes. FEBS Lett 585:2193–2198

Davies KM, Schwinn KE, Deroles SC et al (2003) Enhancing anthocyanin production by altering competition for substrate between flavonol synthase and dihydroflavonol 4-reductase. Euphytica 131:259–268

Des Marais DL, Rausher MD (2010) Parallel evolution at multiple levels in the origin of hummingbird pollinated flowers in *Ipomoea*. Evolution 64:2044–2054

Edwards KD, Fernandez-Pozo N, Drake-Stowe K et al (2017) A reference genome for *Nicotiana tabacum* enables map-based cloning of homeologous loci implicated in nitrogen utilization efficiency. BMC Genom 18:448

Finn RD, Coggill P, Eberhardt RY et al (2016) The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res 44:D279–85

Gates DJ, Olson BJSC, Clemente TE, Smith SD (2017) A novel R3 MYB transcriptional repressor associated with the loss of floral pigmentation in *Iochroma*. New Phytol. https://doi.org/10.1111/nph.14830

Goff L, Trapnell C, Kelley D (2013) cummeRbund: analysis, exploration, manipulation, and visualization of Cufflinks high-throughput sequencing data. R package version

Goodspeed TH (1954) The genus *Nicotiana*. Chronica Botanica, Waltham, Massachusetts, USA

Grabherr MG, Haas BJ, Yassour M et al (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29:644–652

Grotewold E (2006) The genetics and biochemistry of floral pigments. Annu Rev Plant Biol 57:761–780

Haas BJ, Papanicolaou A, Yassour M et al (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc 8:1494–1512

Hoballah ME, Gübitz T, Stuurman J et al (2007) Single gene-mediated shift in pollinator attraction in *Petunia*. Plant Cell 19:779–790

Holton T, Brugliera F, Tanaka Y (1993) Cloning and expression of flavonol synthase in *Petunia hybrida*. Plant J 4:1003–1010

Joshi NA, Fass JN (2011) Sickle: a sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33) [Software]

Knapp S, Chase MW, Clarkson JJ (2004) Nomenclatural changes and a new sectional classification in *Nicotiana* (Solanaceae). Taxon 53:73–82

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359

Luo P, Ning G, Wang Z et al (2015) Disequilibrium of flavonol synthase and dihydroflavonol-4-reductase expression associated tightly to white vs. red color flower formation in plants. Front Plant Sci 6:1257

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. journal 17:10–12

McCarthy EW, Arnold SEJ, Chittka L et al (2015) The effect of polyploidy and hybridization on the evolution of floral colour in *Nicotiana* (Solanaceae). Ann Bot 115:1117–1131

McCarthy EW, Berardi AE, Smith SD, Litt A (2017) Related allopolyploids display distinct floral pigment profiles and transgressive pigments. Am J Bot 104:92–101

Pattanaik S, Kong Q, Zaitlin D et al (2010) Isolation and functional characterization of a floral tissue-specific R2R3 MYB regulator from tobacco. Planta 231:1061–1076

Quattrocchio F, Wing J, van der Woude K et al (1999) Molecular analysis of the *anthocyanin2* gene of Petunia and its role in the evolution of flower color. Plant Cell 11:1433–1444

Rasband WS (1997) ImageJ. US National Institutes of Health, Bethesda, MD

Ritchie ME, Phipson B, Wu D et al (2015) *limma* powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 43:e47

Schwinn K, Venail J, Shang Y et al (2006) A small family of *MYB*-regulatory genes controls floral pigmentation intensity and patterning in the genus *Antirrhinum*. Plant Cell 18:831–851

Sheehan H, Moser M, Klahre U et al (2016) *MYB-FL* controls gain and loss of floral UV absorbance, a key trait affecting pollinator preference and reproductive isolation. Nat Genet 48:159–166

Sierro N, Battey JND, Ouadi S et al (2013) Reference genomes and transcriptomes of *Nicotiana sylvestris* and *Nicotiana tomentosiformis*. Genome Biol 14:R60

Smith SD, Rausher MD (2011) Gene loss and parallel evolution contribute to species difference in flower color. Mol Biol Evol 28:2799–2810

Stracke R, De Vos RCH, Bartelniewoehner L et al (2009) Metabolomic and genetic analyses of flavonol synthesis in *Arabidopsis thaliana* support the in vivo involvement of leucoanthocyanidin dioxygenase. Planta 229:427–445

Streisfeld MA, Rausher MD (2009) Genetic changes contributing to the parallel evolution of red floral pigmentation among *Ipomoea* species. New Phytol 183:751–763

The UniProt Consortium (2017) UniProt: the universal protein knowledgebase. Nucleic Acids Res 45:D158–D169

Trabelsi N, Petit P, Manigand C et al (2008) Structural evidence for the inhibition of grape dihydroflavonol 4-reductase by flavonols. Acta Crystallogr D Biol Crystallogr D64:883–891

Trapnell C, Roberts A, Goff L et al (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc 7:562–578

Warnes GR, Bolker B, Bonebakker L, Gentleman R (2009) gplots: Various R programming tools for plotting data. R package version

Wessinger CA, Rausher MD (2015) Ecological transition predictably associated with gene degeneration. Mol Biol Evol 32:347–354

Wickham H (2009) ggplot2: Elegant Graphics for Data Analysis Springer-Verlag. New York

Yuan Y-W, Rebocho AB, Sagawa JM et al (2016) Competition between anthocyanin and flavonol biosynthesis produces spatial pattern variation of floral pigments between *Mimulus* species. Proc Natl Acad Sci U S A 113:2448–2453

Yuan Y-W, Sagawa JM, Young RC et al (2013) Genetic dissection of a major anthocyanin QTL contributing to pollinator-mediated reproductive isolation between sister species of *Mimulus*. Genetics 194:255–263

Zhong S, Joung J-G, Zheng Y et al (2011) High-throughput illumina strand-specific RNA sequencing library preparation. Cold Spring Harb Protoc 2011:940–949

Zufall RA, Rausher MD (2004) Genetic changes associated with floral adaptation restrict future evolutionary potential. Nature 428:847–850

Check for updates

Natural *Agrobacterium*-Mediated Transformation in the Genus *Nicotiana*

12

Léon Otten

Abstract

Agrobacterium can transfer genetic information to plants, transforming the plants naturally. An *Agrobacterium* plasmid fragment is transferred to the plant via bacterial infection and stably integrated into the plant's nuclear DNA. This plasmid fragment (termed transferred DNA or T-DNA) contains several genes that are inserted into the plant's chromosomes. Some natural plant species contain *Agrobacterium* T-DNA-like sequences, which have been shown to result from natural transformation. These sequences are called cellular T-DNAs or cT-DNAs. Multiple *Nicotiana* species have been shown to contain cT-DNA sequences and to express cT-DNA genes, qualifying these species as natural transformants. The composition and organization of the T-DNA sequences vary considerably. Sequencing the genome of the *Tomentosae* and *Noctiflorae* sections of the genus *Nicotiana* has identified seven cT-DNA sequences that are similar to sequences in *A. rhizogenes*, *A. tumefaciens*, and *A. vitis*. As some cT-DNA genes show strong growth effects when expressed in other species, they may influence the growth of the natural transformants as well. The precise mechanisms by which these genes alter growth patterns and their regulation by promoters and by plant transcription factors remain to be elucidated.

L. Otten (✉)
Institut de Biologie Moléculaire des Plantes, University of Strasbourg, Strasbourg, France
e-mail: leon.otten@ibmp-cnrs.unistra.fr

12.1 Introduction

Plant genetic engineering is now in large-scale application in both fundamental and applied researches (Kado 2014). The stable transfer of genes to plants depends on a great extent on the use of special *Agrobacterium* strains. This bacterium carries a unique, natural gene-transfer mechanism that has been adapted for use in the laboratory. Although *Agrobacterium* is generally believed to use its T-DNA transfer system to induce tumors, this review summarizes data showing that *Agrobacterium* infection can also result in transgenic plants under natural conditions. These plants will be termed "natural transformants" in this chapter, in contrast with man-made transformants obtained through genetic engineering.

Members of several plant genera have been found to be naturally transformed, including *Nicotiana* (White et al. 1983), *Linaria* (Matveeva et al. 2012), and *Ipomoea* (Kyndt et al. 2015). This chapter focuses on the *Nicotiana* genus and begins with a short historical overview of important *Agrobacterium* discoveries, to provide a background for discoveries on natural transformation. The chapter closes with perspectives for future research.

© Springer Nature Switzerland AG 2020

N. V. Ivanov et al. (eds.), *The Tobacco Plant Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-030-29493-9_12

12.2 Historical Overview of *Agrobacterium* Research

Agrobacterium tumefaciens is well known for its capacity to transfer genetic information to plants and has been used for over 30 years in the genetic modification of plants. The bacterium was first isolated in the early 1900s (Smith and Townsend 1907) and found to induce plant tumors. Around 1930, a related species, *A. rhizogenes*, was isolated from plants with hairy root disease (Riker et al. 1930). This bacterium induces profuse root growth in host plants (De Cleene and De Ley 1981). Other agrobacteria, such as *A. rubi* (Hildebrand 1940) and *A. vitis* (Ophel and Kerr 1990; Burr and Otten 1999), have been described.

Originally, it was believed that agrobacteria produced hormones and thereby stimulated plant growth. Detailed studies of the properties of the *Agrobacterium* tumors showed that tumors remained hormone-independent in the absence of bacteria and could be cultivated in vitro for indefinite periods, suggesting that some permanent genetic change had occurred (White and Braun 1941), perhaps through the transfer of a (hypothetical) tumor-inducing principle from the *Agrobacterium* to the plant.

Crown gall tumors were found to contain unusual compounds not found in normal plants (Lioret 1956; Biemann et al. 1960; Ménagé and Morel 1964). These compounds varied with the type of strain used to induce the tumor. The compound octopine, a conjugate of arginine and pyruvate, had earlier been identified in common octopus (*Octopus vulgaris*). Subsequently, other compounds found in crown galls were collectively termed "opines" (Dessaux et al. 1998). Opines were found to accumulate in high levels in sterile tumors (Scott 1979), and the type of opine depended on the type of bacterium used to induce the galls (Petit et al. 1970). Thus, agrobacteria could be classified as octopine-type, nopaline-type, or other opine types, confirming the notion that permanent genetic changes had occurred. Soon, it was shown that opines could be used by the agrobacteria for their growth. Thus, synthesis of opines in the galls and their use by the bacterium appeared to be intimately linked, leading to the idea that opine synthesis was beneficial to the agrobacterium and the main objective of tumor induction (Schell et al. 1979; Guyon et al. 1993).

In a major discovery, Alan Kerr found that infection of plants with mixtures of *A. radiobacter* and *A. tumefaciens* would convert the avirulent *A. radiobacter* strain into a virulent bacterium (Kerr 1969). It appeared that a genetic factor carrying the tumor-inducing principle was transferred from *A. tumefaciens* to *A. radiobacter*. These experiments became known later as "Kerr crosses." The tumor-inducing principle was subsequently found to reside on plasmids, which were therefore termed tumor-inducing and root-inducing plasmids (pTi and pRi plasmids, respectively).

A small and specific part of the plasmid was shown to be transferred to the plant during the infection process and stably integrated in the plant's nuclear DNA (Chilton et al. 1977), demonstrating that plants were indeed genetically transformed by the bacterium.

The mechanism of transfer has been studied extensively (Gelvin 2012; Gordon and Christie 2014) but will not be addressed in detail in this review. The T-DNA is flanked by a specific 25-nucleotide repeat sequence (termed left and right border) that is essential for transfer (Yadav et al. 1982; Wang et al. 1984). T-DNA transfer begins at the right repeat and proceeds up to the left repeat but can also be interrupted, leading to truncated T-DNA fragments. T-DNA sequences are transferred into the plant nucleus and insert more or less randomly into the chromosomes. One or more copies can be transferred, and they can be linked together in various ways before integration, or integrated at different chromosomal locations. Some *Agrobacterium* strains contain two T-DNAs, and these can also be linked together and then inserted at a single target site.

Once it was established that *Agrobacterium* can introduce genes into plant cells, and in an efficient, precise, and reliable manner, research was directed toward the development of vectors and the transformation of model plants and plants of agronomical interest. The vectors were

derived from pTi plasmids in which the tumor and opine genes were replaced by other genes, generally including a gene for selection of the transformed cells and a gene of interest (Hernalsteens et al. 1980; Zambryski et al. 1983; Hoekema et al. 1983). The use of these vectors led to a new field in plant biology, plant genetic engineering, and a revolution in the molecular analysis of plants. Together with technical advances in molecular biology, plant genetic engineering also led to the adoption of *Arabidopsis thaliana* as a general model for flowering plants.

These important developments caused a shift of interest from *Agrobacterium* to basic and applied plant research, and likely contributed to a loss of interest in the T-DNA genes. As this review argues, however, there may soon be a renewed interest in T-DNA genes because they are probably involved in the genesis of natural transformants. More generally, they could provide new insights into how plants control their growth, a question of great interest in agriculture.

12.3 First Discoveries of Natural Transformants

When studies of *Agrobacterium* T-DNA transfer commenced, multiple sterile plant tissues were investigated. The standard technique used at the time was Southern blotting, whereby purified plant DNA was cut with restriction enzymes, separated on agarose gels, and transferred to a membrane by blotting. The DNA fragments were fixed to the membrane and placed in contact with well-defined, single-stranded radioactively labeled sequences termed DNA probes. If sequences similar to the probes were present on the membrane, the probes would bind to those sequences by hydrogen bonding between complementary nucleotides (a process called hybridization), thereby labeling the target. Using T-DNA probes, transformed tissues could be analyzed for the presence of T-DNA sequences. By chance, the DNA of non-transformed plants of the species *Nicotiana glauca* (section *Noctiflorae*) was shown to hybridize to probes from the T-DNA of *A. rhizogenes* strain A4 (White et al. 1983). This unexpected and puzzling finding was investigated, and the research group was able to recover the relevant DNA sequences from *N. glauca* and to show that they were similar to a part of the *A. rhizogenes* T-DNA (Furner et al. 1986). This DNA was termed cellular T-DNA or cT-DNA to differentiate it from bacterial T-DNA. It was proposed that in the course of the evolution of the genus *Nicotiana*, the cT-DNA had been introduced by *A. rhizogenes*-mediated hairy root induction; some of these hairy roots then formed shoots and regenerated spontaneously into plants, leading to "natural transformants."

Additional Southern blotting experiments (Furner et al. 1986) showed that other *Nicotiana* species also contained A4-like sequences, including *N. tabacum*, *N. tomentosa*, *N. tomentosiformis*, *N. otophora,* and *N. benavidesii*; all except *N. benavidesii* (section *Paniculatae*) belong to section *Tomentosae* (this review adopts the *Nicotiana* classification by Knapp et al. 2004; Fig. 12.1).

These observations qualified several *Nicotiana* species as natural transformants. Additional *Nicotiana* species (*N. benthamiana*, *N. debneyi*, *N. langsdorfii*, *N. glutinosa*, *N. rustica*, *N. knightiana*, *N. solanifolia*, *N. cordifolia*, *N. setchellii*, *N. paniculata*, and *N. raimondii*) did not show hybridization, either did other members of the Solanaceae family, such as *Solanum americanum* or *Petunia hybrida* (for a full list of non-*Nicotiana* species, see Furner et al. 1986). The same experiments revealed some structural differences between the cT-DNA sequences of different *N. tomentosiformis* accessions, but those were not explored further.

To identify cT-DNA sequences from other members of the *Nicotiana* group, polymerase chain reaction (PCR) experiments using primers for *rolC*, *rolB*, *orf13*, and *orf14* were performed with a large number of *Nicotiana* species, combined with Southern blotting to enhance sensitivity (Intrieri and Buiatti 2001). Not only did *N. glauca*, *N. tabacum*, *N. tomentosa*, *N. tomentosiformis*, *N. otophora,* and *N. benavidesii* contain cT-DNAs, but also *N. cordifolia*,

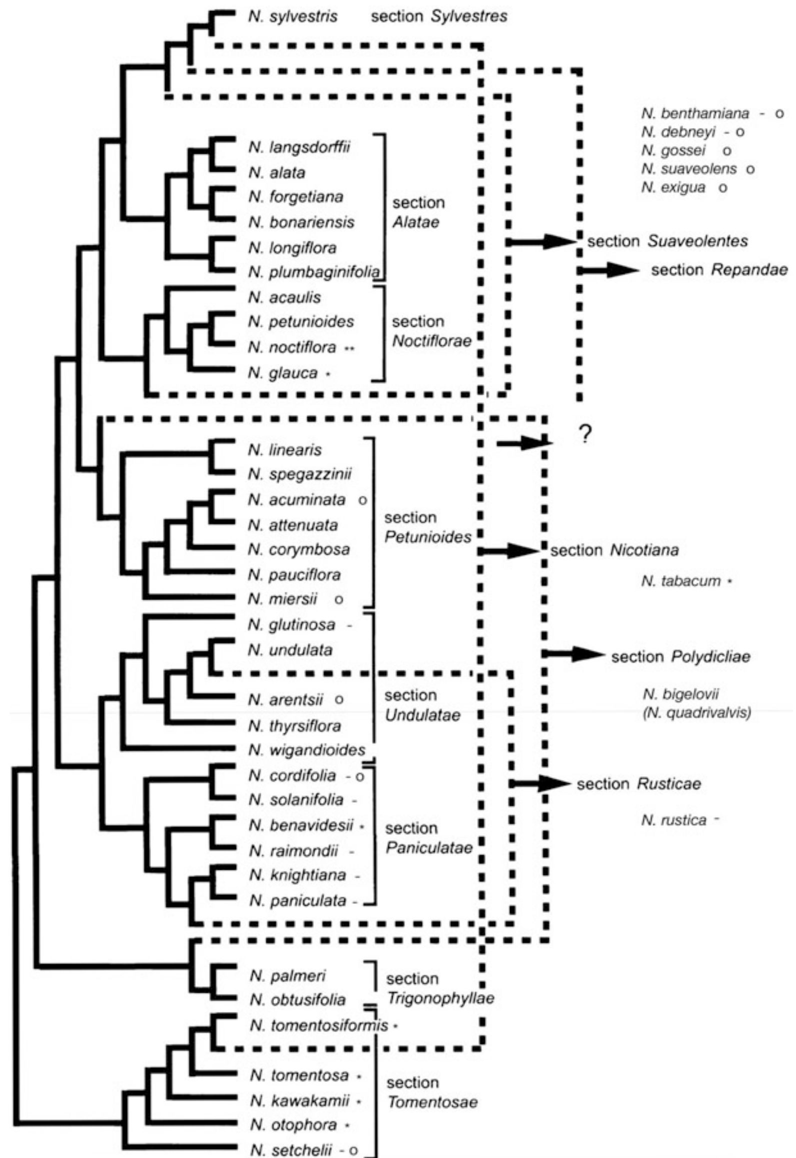

Fig. 12.1 Phylogenetic tree of *Nicotiana* species (Adapted from Knapp et al. 2004). Dashed lines indicate interspecific hybridization events, involving unknown species. Species reported to contain cellular T-DNA (cT-DNA) sequences are indicated with a symbol beside their name. * and –: According to Furner et al. (1986); **: According to transcription data from Long et al. (2016); O: According to Intrieri and Buiatti (2001)

N. setchellii, *N. arentsii*, *N. acuminata*, *N. miersi*, *N. bigelovii*, *N. debneyi*, *N. gossei*, *N. suaveolens*, and *N. exigua*. Expression of cT-DNA genes was found in *N. glauca*, *N. tabacum*, *N. tomentosiformis*, *N. cordifolia*, *N. miersi*, and *N. debneyi*. Gene sequences were reported for *rolC*, *orf13*, and *orf14* from *N. tomentosiformis*, *N. otophora*, *N. tabacum*, *N. cordifolia*, and *N. debneyi*. It appeared that natural transformants were common and widespread in the *Nicotiana* genus (Fig. 12.1), although these results require confirmation by sequencing, because they contain several puzzling inconsistencies (Chen et al. 2014).

The initially published *N. glauca* cT-DNA sequence (Furner et al. 1986) was incomplete; it was completed later (Suzuki et al. 2002) and found to consist of an inverted repeat with two arms that showed 1.5% divergence, a sequence difference that most probably occurred after the initial insertion. The sequenced cT-DNA fragment corresponded to an A4 TL-DNA fragment

with genes Ng*rolB*, Ng*rolC*, Ng*orf13*, Ng*orf14*, and Ng*mis*. All except Ng*rolB* were intact. Understanding the significance of this finding requires a summary of what is known about *Agrobacterium* T-DNA genes.

12.4 *Agrobacterium* T-DNA Genes

Early in *Agrobacterium* research, it was realized that the *Agrobacterium* group consists of various types of strains with different properties. A first classification was based on the disease phenotypes (tumors or hairy roots), leading to *A. tumefaciens* and *A. rhizogenes*; the non-pathogenic strains were grouped as *A. radiobacter*. This group was later extended to include agrobacteria such as *A. vitis* (strains infecting *Vitis vinifera*, grapevine), *A. rubi*, and *A. larrymoorei*. With the discovery of opines, another division was proposed. This led to classification of octopine and nopaline strains, with agropine, mikimopine, and several other types added later (Guyon et al. 1980). Still another, so-called "biotype" classification was proposed (Kerr and Panagopoulos 1977), based on the capacity of strains to grow on different media. Many of the characters involved in classification, however, such as oncogenicity and the capacity to induce the synthesis of specific opines, are located on mobile elements. These can be transferred by bacterial conjugation to other strains, as was seen with the Kerr cross, and therefore do not correspond to stable characteristics. Thus, the pathogenicity or opine-based classifications may have an ecological value but no phylogenetic one. Therefore, the naming and grouping of strains can be confusing. *Agrobacterium* strains carry one, two, or three T-DNA sequences on their Ti/Ri plasmids (Canaday et al. 1992).

The composition and organization of the T-DNA sequences vary considerably. They can be functionally divided into four groups. The first group contains the opine synthesis genes, including the octopine synthase (*ocs*) gene. These genes encode the enzymes that produce opines. Large quantities of purified proteins can easily be synthesized to study their enzymatic properties and determine substrate specificities, optimum pH, capacity to carry out reverse reactions, and inhibition by high substrate concentrations. Little has been done so far to relate the in vitro properties of opine enzymes to their properties in vivo, which to a large extent depend on substrate availability and on the regulatory properties of the corresponding gene promoters. There are indications that some opine enzymes (mannopine synthases Mas1' and Mas2') are membrane-bound, as implied by the predicted presence of transmembrane segments (Chen et al. 2016); these enzymes should be studied in microsomes or in other reconstituted membrane systems.

The second group encompasses the hormone–enzyme genes; these include *iaaM* and *iaaH*, coding for enzymes that catalyze indoleacetic acid production from tryptophan, and *ipt*, encoding isopentenyltransferase, which catalyzes the production of cytokinins (Schröder et al. 1984; Morris 1986). These genes are likewise easy to purify and study, and will also require activity tests in vivo.

The third group codes for a large family of proteins named the plast proteins. Plast genes are widespread on T-DNA sequences of different origins. They were originally defined on the basis of low but significant protein homology (Levesque et al. 1988); the word "plast" is derived from "morphogenetic plasticity" and refers to their capacity to induce strong phenotypes. Their origin from a common ancestor clearly suggests that their functions are related, but in spite of a large number of studies on individual members of this family, the ancestral activity has not been defined to date. Phenotypes induced by plast genes are quite specific.

Various molecular activities for individual plast genes have been proposed. Since plast proteins are relatively small, it is unlikely that they are multifunctional; some of the proposed activities are probably based on artifacts (Chen 2016). The plast genes appeared at first to be restricted to T-DNAs from *Agrobacterium*, but later (weak) homologs were found in other

organisms (Mohajjel-Shoja et al. 2011; Chen et al. 2014; Chen 2016). Nothing is known about these other plast genes.

Some T-DNA genes are not related to plast genes, opine genes, or hormone synthesis genes and may also have morphogenetic activities; examples include *rolA*, *orf3*, and gene *c*. These genes do not show any homology to genes with known activity (Chen 2016) and constitute the fourth group.

cT-DNA sequences of natural transformants are most likely derived from *A. rhizogenes*. This bacterium normally induces hairy roots, requiring at minimum the *rolA*, *rolB*, and *rolC* genes (Spena et al. 1987). In some plant species, including tobacco, hairy roots can be experimentally induced to form transformed plants. These "hairy root plants" have a typical phenotype: shorter internodes and conspicuously wrinkled leaves (Tepfer 1990; Christey 2001; Lütken et al. 2012). Natural transformants, however, do not possess these characteristics, probably because they do not carry an intact *rolB* gene (see Sect. 12.5).

12.5 The Function of CT-DNA Genes in Natural Transformants

After the initial discovery of cT-DNA sequences in *N. glauca*, the full *N. glauca* cT-DNA sequence was established. In addition to the Ng*rolC* gene (see Sect. 12.3), the *N. glauca* sequence was found to contain intact mikimopine synthase (Ng*mis*) genes. These could be expressed in *Escherichia coli* and produce an active enzyme, but no mikimopine was found in *N. glauca* (Suzuki et al. 2002). In view of the results of studies with the TB-*mas2*' gene (see Sect. 12.10), it appears nevertheless possible that some varieties of *N. glauca* produce significant quantities of mikimopine locally. The *N. glauca* cT-DNA also carries intact *orf13* and *orf14* plast genes; these were expressed in *N. tabacum*, and their effects on the plant were compared with the effects from the *A. rhizogenes* strain A4 *orf13* and *orf14* genes (A4-*orf13*, A4-*orf14*). Neither A4-*orf14* nor Ng*orf14* exert a clear effect, but A4-*orf13* and Ng*orf13* lead to a change in leaf form: while normal tobacco leaves are ovoid, the *orf13* tobacco leaves are heart-shaped. A Ng*rolB* gene with stop codons could be repaired and was shown to regain biological activity (Aoki and Syōno 1999), suggesting that it was active at the time of insertion. The *rolB* gene is the main determinant for the wrinkled-leaf phenotype of hairy root regenerants, as can be shown by

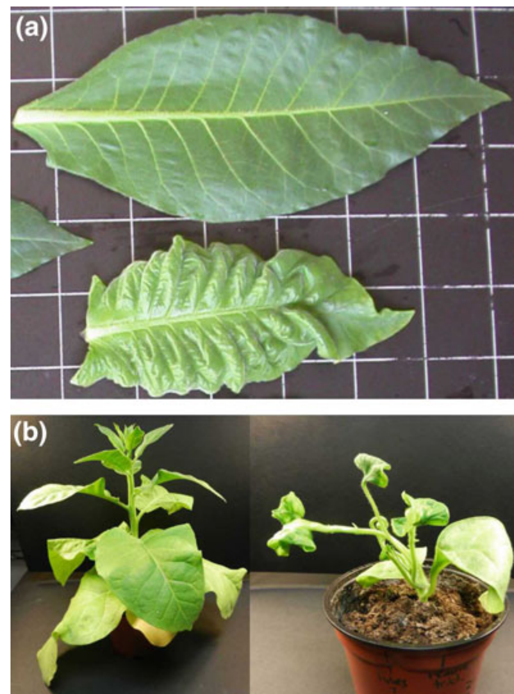

Fig. 12.2 Phenotypes of *Nicotiana tabacum* transformed with cellular T-DNA (cT-DNA) genes, *rolB* or *6b*. **a** Normal leaf from *N. tabacum* cv. Wisconsin 38 (upper leaf); wrinkled leaf from *N. tabacum* cv. Wisconsin 38 transformed with the *Agrobacterium rhizogenes* A4-*rolB* gene under tetracycline-inducible promoter control (lower leaf) (Röder et al. 1994; Mohajjel Shoja 2010). **b** Normal *N. tabacum* cv. Samsun plant (left); *N. tabacum* cv. Samsun plant transformed with *A. vitis* Tm4-*6b* gene under dexamethasone-inducible promoter control (left). The transformed plant has strongly modified leaves. For a detailed analysis of *6b*-induced changes see Chen and Otten (2016)

expression of *rolB* in tobacco (Spena et al. 1987; Mohajjel Shoja 2010, Fig. 12.2a).

The initial Southern blotting experiments showed that the common model plant tobacco also contained cT-DNA sequences (Furner et al. 1986). A few cT-DNA genes were isolated and sequenced and, as in the case of *N. glauca* cT-DNA genes, were tested in functional assays. A *torf13* (*t* for tobacco) gene (Fründt et al. 1998) and a *trolC* gene (Mohajjel-Shoja et al. 2011) were intact and had biological activity. The precise extent of the tobacco cT-DNA remained unclear. Therefore, it was not possible to establish the precise relationship between *N. glauca* and *N. tabacum* cT-DNA sequences: were they derived from one cT-DNA fragment (inserted into a common ancestor of the two species), or did they result from independent transformation events? Only the full sequence of tobacco cT-DNA finally provided a clear answer.

12.6 New Insights from the Deep-Sequencing Approach

Various strategies can be used to obtain full cT-DNA sequences from plant genomes, such as cloning of large genomic DNA fragments and inverted PCR using T-DNA primers. These methods are relatively complex and have some disadvantages. Therefore, progress in the cT-DNA field has been slow, as illustrated by the difficulty of obtaining the full *N. glauca* cT-DNA sequence, which took more than 10 years. A major technical breakthrough occurred around 2010 when deep sequencing (or next-generation sequencing) became available, allowing researchers to rapidly obtain large, redundant numbers of small genomic DNA sequences termed reads (typically 100 nucleotides) covering the entire genome. Reads can be assembled by computer programs, producing contigs. Ideally, contig assembly should lead to a reconstruction of the complete chromosomes (48 in tobacco), but because of the presence of highly repeated sequences, contigs will be much shorter. Nonetheless, collections of contigs and reads can be searched for sequences homologous to various T-DNA sequences. Once these (often small, from a few hundred to a few thousand bases) cT-DNA contigs are retrieved, further assembly is possible using other approaches. The redundancy of the deep-sequencing method theoretically guaranteed a complete analysis of a genome and complete assembly of the cT-DNA, since all genomic sequences are present in the database. This approach, however, is not as easy as was originally expected.

To obtain a full cT-DNA sequence of tobacco, the genome of the paternal ancestor of tobacco, *N. tomentosiformis*, was deep-sequenced. Four cT-DNA sequences were identified and termed TA, TB, TC, and TD (Chen et al. 2014). Their structures are shown in Fig. 12.3, along with the cT-DNA of *N. glauca*, and compared to the TL-DNA of *A. rhizogenes* strain A4.

Other *N. tomentosiformis* contigs were acquired by Philip Morris International (Sierro et al. 2014) and confirmed the analysis (Chen et al. 2014).

The presence of no less than four cT-DNA sequences in *N. tomentosiformis* was unexpected and prompted new questions about natural transformants. How could four cT-DNA sequences accumulate in one species? Did they result from one or several hairy root regeneration events? How could successive regeneration events occur, if one event is a rarity? Might the order of these events be reconstructed? Did some cT-DNA sequences appear in specific *Nicotiana* species but not in others? Was transformation perhaps not as rare as earlier believed, at least not in certain species? Could a first regeneration event favor subsequent transformation/regeneration events by the activity of cT-DNA genes?

12.7 *N. tomentosiformis* cT-DNA Sequences

The cT-DNA sequences from *N. tomentosiformis* were manually reconstructed from contigs obtained by deep sequencing (Fig. 12.3). As in *N. glauca* cT-DNA, all four consist of partial inverted repeats and are similar to the TL-DNA sequences of *A. rhizogenes* strains A4 (Slightom

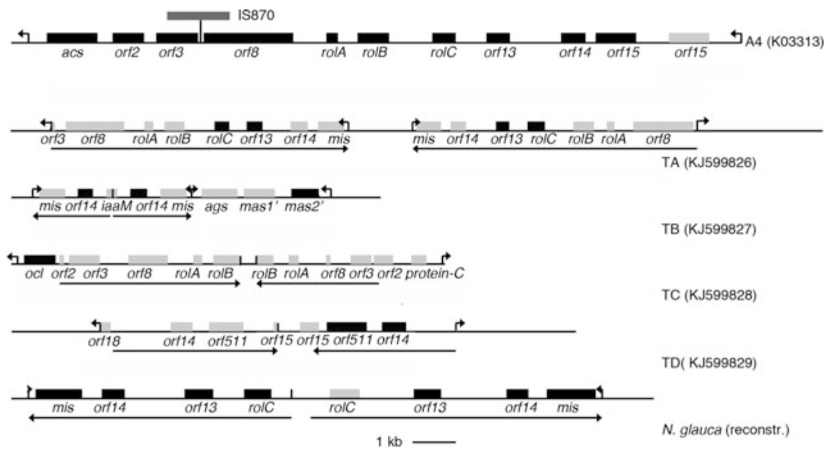

Fig. 12.3 Structures of cellular T-DNA (cT-DNA) from *Nicotiana tomentosiformis* and *N. glauca.* From the top: structure of the *Agrobacterium rhizogenes* A4 TL-DNA; structures of TA, TB, TC, and TD cT-DNAs from *N. tomentosiformis* (Chen et al. 2014); structure of the cT-DNA from *N. glauca* reconstructed from GenBank sequences AB071335.1, D16559.1, X03432.1, and AB071334.1. Intact genes are indicated in black, and partial genes are shown in gray. IS870 is a bacterial insertion sequence. Arrows below the cT-DNA sequences indicate repeats. Border sequences are indicated by arrowed hooks. *N. otophora* (Chen et al. 2018) and *N. noctiflora* also contain cT-DNA sequences, the *N. noctiflora* sequences have not yet been assembled

et al. 1986) and 1724 (Suzuki et al. 2001) but contain additional sequences that are not typically associated with *A. rhizogenes*. The A4 TL-DNA is only one of several T-DNA structures found in *A. rhizogenes*; these T-DNA sequences are evolutionary chimeras resulting from DNA recombination processes that led to mixed structures, the origin of which is difficult if not impossible to trace (van Nuenen et al. 1993; Burr and Otten 1999). This creates practically unsurmountable problems for the reconstruction of bacterial phylogeny.

The TA structure of *N. tomentosiformis* is discussed in some detail below, the structures of TB, TC, and TD are basically similar. The inverted repeat of the TA region shows a 1724-like *orf3–orf8–rolA–rolB–rolC–orf13–orf14–mis* organization, with the two right border repeats in the central part. The *orf14* gene (although at the right position) is unusual and can only be recognized because the predicted protein has low but significant (36%) homology to other Orf14 proteins. The central region separating the left and right arm is approximately 2 kb and derived from plant DNA; this fragment might have been ligated to the two partial T-DNA fragments before integration of the complete structure. The plant sequences flanking TA are 95% identical to sequences from *N. sylvestris* and show that a short eight-nucleotide sequence was probably deleted during insertion. The TA region contains the following open reading frames: TA-*rolC*-L and TA-*rolC*-R, TA-*orf13*-L, and TA-*orf13*-R.

TD appears to be a ligation of three partial fragments from the right part of an A4-like TL-DNA (Chen et al. 2014). TD also carries an unknown, large open reading frame (*orf511*) potentially coding for a 511 amino acids protein. This protein has no homologs, and its role in hairy root induction and in *N. tomentosiformis* remains to be elucidated.

12.8 cT-DNA Sequences in Other Members of the *Nicotiana* Genus

When *N. tomentosiformis* was found to carry four cT-DNA sequences, the question arose whether other members of the *Tomentosae* section would contain the same type of inserts. Primers were designed to amplify fragments

specific to parts of each cT-DNA. Another question was whether the four insertion sites were unmodified ("empty"). Since the four cT-DNA sequences contained diverged repeats, the divergence levels could be used to estimate the order of insertion: older insertion events can be expected to have higher divergence than younger events. The results suggested that the cT-DNA sequences arrived in the order TC, TB, TD, and TA. Interestingly, the TC region was deleted in *N. tabacum*, together with about 1 kb on the left and right sides. Moreover, some *N. tabacum* cultivars carried internal deletions in the TA region. Thus, intraspecific variation in cT-DNA sequences can occur. The cT-DNA combinations and repeat divergence levels were used to relate the insertion of each cT-DNA to the evolution of the *Tomentosae* section (Knapp et al. 2004). *N. setchellii* does not appear to contain cT-DNA genes, as indicated by inspection of its transcriptome (Chen et al. 2016).

The distribution of the cT-DNA sequences fully coincides with the branching pattern of the *Tomentosae* phylogenetic tree (Fig. 12.1). This leads to the question of whether the branching could be related to cT-DNA insertion, i.e., whether the new species appearing at different time points in evolution were created by the transformation of an ancestor. Since T-DNA sequences from *A. rhizogenes* lead to strong morphogenetic changes (Spena et al. 1987; Schmülling et al. 1988; Nilsson and Olsson 1997; Lemcke and Schmülling 1998), it is possible that a new cT-DNA can transform a plant in such a way that it becomes sexually incompatible with its immediate ancestor. Thus, in the *Nicotiana* genus, speciation by transformation may explain the existence of at least seven species. This concept remains to be validated by experimental removal of the cT-DNA sequences using CRISPR or gene silencing, followed by incompatibility tests. An alternate line of investigation can test whether the introduction of *A. rhizogenes* T-DNA genes, alone or in different combinations, could lead to reproductive isolation.

As *Nicotiana* species have been calculated to evolve with a divergence rate of 5.6% per 1,000,000 years (Clarkson et al. 2005), and since the first cT-DNA (TC) shows a 5.8% repeat divergence, the TC region can be estimated to have been inserted approximately 1,000,000 years ago.

Another cT-DNA region, TE, was found (Chen et al. 2014) by analysis of deep-sequencing data from *N. otophora* (Sierro et al. 2014). This region has not been found in other *Nicotiana* species.

N. otophora contains the following types of sequences: *orf2*, *orf3*, *orf8*, *rolA*, *rolB*, *ocl*, and *c* (on the TC region that is also found in *N. tomentosiformis*) and *vis*, *6b*, *iaaM*, *iaaH*, *acs*, *rolC*, *orf13*, *orf14*, *mas2'*, and *mas1'*on TE. The TE sequence has recently been assembled (Chen et al. 2018). The presence of the *vis* and *6b* genes is unexpected, since these have only been found in *A. tumefaciens* and *A. vitis*, and not in *A. rhizogenes*. The effects of the T-*6b* gene from the *A. vitis* Tm4 strain on tobacco were studied using an inducible promoter and found to lead to strong growth effects (Helfer et al. 2002; Chen and Otten 2016, Fig. 12.2b). The three *N. otophora* TE-*6b* genes are biologically active, but their effects on tobacco are different from the T-*6b* gene (Chen et al. 2018).

12.9 A New Natural Transformant Within the *Nicotiana* Genus: *N. Noctiflora*

Transcriptome data from *N. noctiflora* (Long et al. 2016) have shown that this species also contains cT-DNA sequences, including previously unknown sequences related to *rolC*, *orf13*, *orf13a*, *orf14*, *mis*, *iaaH*, and *iaaM* similar to genes from the *A. rhizogenes* TL-DNA, *N. glauca* cT-DNA, and *A. tumefaciens* C58 T-DNA (Fig. 12.4).

The *N. noctiflora* accession, however, was not indicated in the paper by Long et al. (2016), and the seeds were not made available. Initial attempts to amplify fragments from DNA of several *N. noctiflora* accessions failed. Subsequently, DNA fragments of seeds from accessions TW88 and TW89 (provided by Jessica Nifong, US Nicotiana Germplasm Collection) were amplified successfully (Fig. 12.4).

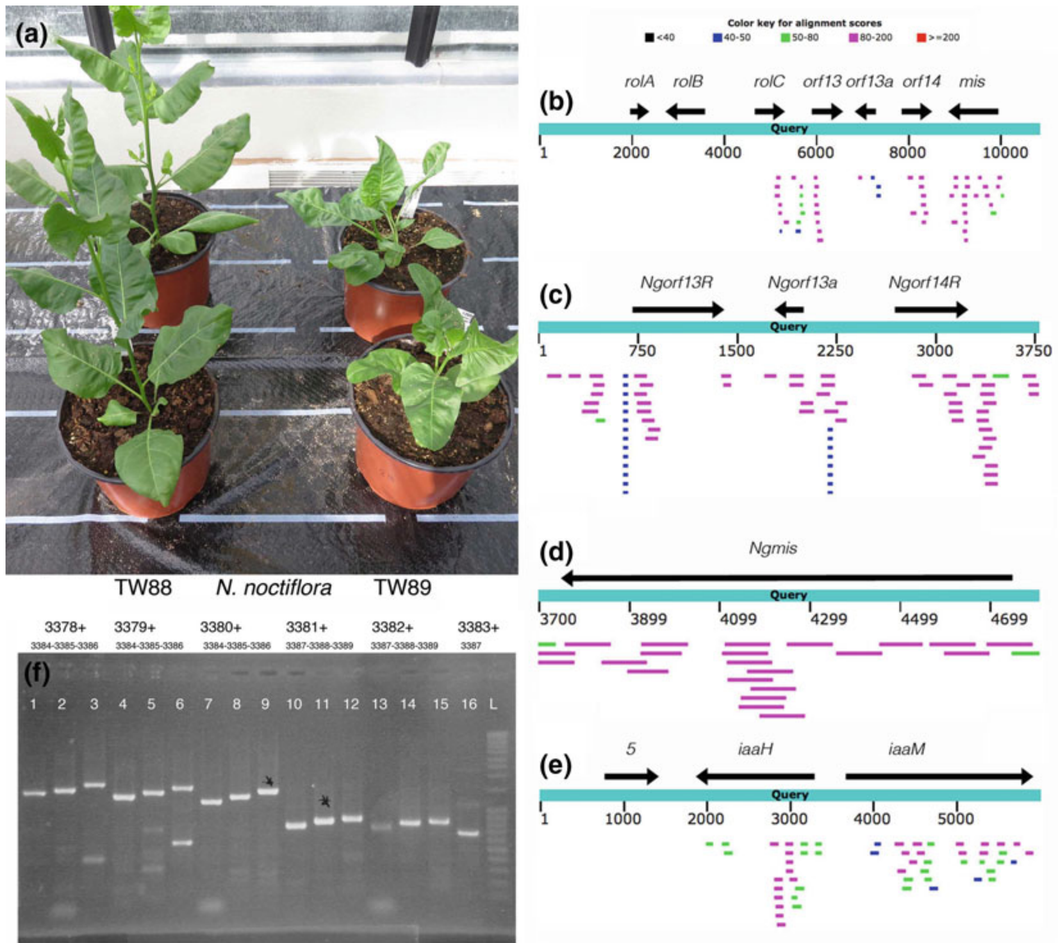

Fig. 12.4 Detection of cellular T-DNA (cT-DNA) sequences in *Nicotiana noctiflora.* **a** Two *N. noctiflora* accessions, TW88 and TW89. **b–e** Alignment of 100-bp reads from the transcriptome of *N. noctiflora* (Long et al. 2016) to different T-DNA model sequences. Colors indicate identity levels. **b** Part of *Agrobacterium rhizogenes* strain 1724 T-DNA. **c** Part of *N. glauca* cT-DNA with Ng*mis*. **d** Part of *N. glauca* cT-DNA (right arm) with Ng*orf13*, Ng*orf13a*, and Ng*orf14*. **e** Part of *A. tumefaciens* C58 T-DNA (gene *5* and genes *iaaH* and *iaaM*). **f** Amplification of cT-DNA regions from genomic DNA of *N. noctiflora* using primer combinations based on the read sequences. Forward primers: 3378–3383, reverse primers: 3384–3389. 3378: GAAACGCTGCATG-TAATTC, 3379: GCACGCAAAAGTCAGTGGTT, 3380: CCCGACGATCTTAGCCATTT, 3381: CGCGAT TGAACCCTTCGAT, 3382: GCGTCAAGAGCGC-GATGGT, 3383: CAGATCATCAAGTGCATCA, 3384: GATGTCATAAGATGTCGGAG, 3385: TAGGTGGTC-GAGCCAGAATT, 3386: TTATCCGGAACAGCCAA-GAG, 3387: CTTCTTATTGCCTACCATG, 3388: GGAACTTGTTTGCTCTCA, 3389: CGACATTGAT-GAGGAAAG. The bands have the expected sizes, and are therefore bona fide amplicons. L: DNA ladder

Once fragments are detected, the full cT-DNA sequences can be characterized by deep sequencing. *N. noctiflora* belongs to the same section (*Noctiflorae*) as *N. glauca*, and its *rolC–orf13–orf13a–orf14–mis* sequences may be part of the same fragment and inserted in the same site. Thus, both species would share this insert. The presence of *iaaM* and *iaaH* sequences in *N. noctiflora* (lacking in *N. glauca*), however, shows that there could be a second cT-DNA, resulting from either two successive insertions by two different *A. rhizogenes* strains or more likely from the transfer of two fragments from two T-DNA sequences located on the same pRi plasmid (similar to TL and TR from strain A4, see above). The new cT-DNA of *N. noctiflora*

brings the known number of cT-DNA sequences for the genus *Nicotiana* to seven.

12.10 Synthesis of Opines by Natural *Nicotiana* Transformants

The TB region of *N. tomentosiformis* contains a *mas2*' gene, TB-*mas2*'. Various *A. tumefaciens* and *A. rhizogenes* strains carry *mas2'* genes on their T-DNA sequences. These are part of an opine synthesis system comprising *mas2'*, *mas1'*, and agropine synthase (*ags*) genes (Bouchez and Tourneur 1991). These three genes encode the synthesis of agropine via two intermediates. The first step is the condensation of glutamine with fructose by Mas2', leading to desoxyfructosyl glutamine (DFG). DFG is then converted to mannopine by Mas1', and Ags converts mannopine into agropine. The TB-*mas2'* gene of *N. tomentosiformis* was introduced in *N. benthamiana* and has been shown to encode an active Mas2' enzyme (Chen et al. 2014). The TB-*mas2'* promoter leads to specific expression in root tips (Chen et al. 2016), as shown using a β-glucuronidase reporter construct (Jefferson 1987). Furthermore, while 137 of 162 tobacco cultivars were reported not to express TB-*mas2'* in roots or shoots (low-expression, LE cultivars), 25 cultivars, including Vuelta Abajo and Russian Burley, expressed high levels of TB-*mas2'* in the roots (high-expression, HE cultivars). No TB-*mas2'* sequence differences were found between these cultivars, and HE × LE crosses showed no trans effects, so it was assumed that a cis-element controlled TB-*mas2'* transcription. Paper electrophoresis of crude root-tip extracts showed that only the HE lines produced detectable amounts of DFG; this was the first study to show that natural transformants can produce opines (Chen et al. 2016). Whether the HE cultivar roots secrete DFG and whether this compound accumulates in the rhizosphere is unknown. Although HE roots develop normally, DFG production could influence the metabolism of fructose and glutamine in roots, and this will also require further study.

12.11 Natural Transformants in Plant Genera Other Than *Nicotiana*

A systematic PCR search in a large number of plants found cT-DNA sequences in the *Linaria* genus (Matveeva et al. 2012). *L. vulgaris* is a common weed, and its cT-DNA is a direct imperfect repeat that has been fully sequenced. It contains only one open reading frame, Lv*rolC*. Similar sequences were later found in other *Linaria* species, such as *L. dalmatica*, but whether these sequences are in the same chromosomal location is currently unknown. The sequences could have been transferred by interspecific hybridization, because *L. vulgaris* and *L. dalmatica* are known to hybridize in nature (Ward et al. 2009), rather than by descent from a common ancestor, as was proposed for the *Nicotiana* TA, TB, TC, and TD regions. After the discovery of the *Linaria* cT-DNA, another species with cT-DNA was identified, the food crop *Ipomoea batatas* (sweet potato) (Kyndt et al. 2015). *I. batatas* contains two cT-DNA sequences, *Ib*T-DNA1 and *Ib*T-DNA2. The first is a partial inverted repeat and contains *iaaM*, *iaaH*, *C-protein*, and *acs* (the two latter genes are unusual for *A. rhizogenes* T-DNA sequences). The second contains *rolC*, *orf13*, *orf14*, and *rolD* (typical *A. rhizogenes* T-DNA genes). The wild-type species *I. trifida* contains an *Ib*T-DNA2 *orf13* gene but no *Ib*T-DNA1 sequences. Sweet potato cultivars such as Huachano and Xu781 can carry different cT-DNA structures; the latter contains a plant-derived transposon-like sequence in the right-arm copy of the *iaaM* gene (Kyndt et al. 2015). As in the case of *Linaria*, interspecific hybridization between *I. batatas* and *I. trifida* could have transferred a cT-DNA (here *Ib*T-DNA2) from one species to the other (Roullier et al. 2013), but the *Ib*T-DNA2 insertion site in *I. trifida* is still unknown. Neither in *Linaria* nor in *Ipomoea* have the insertion events been located on the phylogenetic trees; complete cT-DNA sequences from different accessions and from related species and data on their insertion sites would be required.

12.12 Perspectives and Conclusions

Nicotiana species of the *Tomentosae* (five species) and *Noctiflorae* (two species) sections have been shown to contain *Agrobacterium* T-DNA sequences. Positive results for other *Nicotiana* species (Intrieri and Buiatti 2001) require confirmation. Variation in cT-DNA structures and function can be expected to occur, as has been shown for *N. glauca* and *N. tabacum*, and could be investigated using accessions from different geographical areas.

Although multiple *Nicotiana* cT-DNA sequences are similar to *A. rhizogenes* sequences, some appear to be derived from *A. tumefaciens* or *A. vitis*, and others are highly unusual (such as the strongly diverged *orf14* genes from TA and TD, a diverged *ags* sequence from TB, and diverged *ocl* and *c*-like sequences from TC) or even new, such as *orf511* from TD. These cT-DNA genes indicate that the *A. rhizogenes* T-DNA gene repertoire is greater than expected. Only a few *A. rhizogenes* strains have been isolated, and searching for *A. rhizogenes* in the rhizosphere of natural *Nicotiana* transformants in their native environment in South America may yield interesting findings. Present-day *Nicotiana*-associated agrobacteria may still carry T-DNA sequences that are closely related to *Nicotiana* cT-DNA sequences, and it may even be possible to find plants infected with these strains and producing hairy roots. Whether such hairy roots do indeed spontaneously regenerate into transgenic plants should then be investigated.

To reconstruct the events that led to the natural transformants, one may also attempt to regenerate plants from hairy root cultures established in the laboratory. This could provide some idea of the type of T-DNA genes that allow regeneration and of the range of transgenic plant phenotypes.

More information is also needed on species related to *Linaria* and *Ipomoea*, to reconstruct the insertion events and their relationship to the evolution of these species. With the decreasing costs of genomic sequencing, the simplest approach would be to deep sequence a large number of related species. This could readily identify more natural transformants, but obtaining their complete cT-DNA sequences will still be a complex task.

The next step in cT-DNA research will be the functional analysis of the genes: what role, if any, do they play in the growth of these plants? This can be studied using overexpression of individual genes in model plants such as tobacco or *Arabidopsis*, and using knock-out mutants of the original plant. To understand the effects of cT-DNA genes such as *rolC*, *orf13*, *6b*, and *orf14*, more efforts should be directed toward understanding the basic function of these and other plast genes, as well as their regulation by promoters and by plant transcription factors. Once these mechanisms are fully understood, these genes may become tools for analyzing plant growth mechanisms and useful in improving crops such as tobacco.

Finally, complete removal of the cT-DNA regions, or introduction of complete cT-DNA sequences in other species, could elucidate their overall effects on their host species.

Addendum

Since submission of this manuscript, it has been shown that a large number of natural plant species carry cT-DNA sequences (Matveeva and Otten 2019). Also, an important step towards the elucidation of the function of the *N. otophora 6b* gene was made by showing that the TE-6B protein targets the CIN-TCP transcription factors and thereby leads to important growth changes in the model plant Arabidopsis (Potuschak et al. 2019).

References

Aoki S, Syōno K (1999) Function of Ngrol genes in the evolution of *Nicotiana glauca*: conservation of the function of NgORF13 and NgORF14 after ancient infection by an *Agrobacterium rhizogenes*-like ancestor. Plant Cell Physiol 40:222–230. https://doi.org/10.1093/oxfordjournals.pcp.a029531

Biemann K, Lioret C, Asselineau J, Lederer E, Polonsky J (1960) Sur la structure chimique de la lysopine nouvel

acide aminé isolé de tissu de crown-gall. Bull Soc Chim Biol (Paris) 42:979–991

Bouchez D, Tourneur J (1991) Organization of the agropine synthesis region of the T-DNA of the Ri plasmid from *Agrobacterium rhizogenes*. Plasmid 25:27–39

Burr TJ, Otten L (1999) Crown gall of grape: biology and disease management. Annu Rev Phytopathol 37:53–80. https://doi.org/10.1146/annurev.phyto.37.1.53

Canaday J, Gérad JC, Crouzet P, Otten L (1992) Organization and functional analysis of three T-DNAs from the vitopine Ti plasmid pTiS4. Mol Gen Genet 235:292–303

Chen K (2016) Sequencing and functional analysis of cT-DNAs in Nicotiana. Doctoral dissertation, University of Strasbourg, France

Chen K, Dorlhac de Borne F, Julio E, Obszynski J, Pale P, Otten L (2016) Root-specific expression of opine genes and opine accumulation in some cultivars of the naturally occurring genetically modified organism *Nicotiana tabacum*. Plant J 87:258–269. https://doi.org/10.1111/tpj.13196

Chen K, Dorlhac de Borne F, Szegedi E, Otten L (2014) Deep sequencing of the ancestral tobacco species *Nicotiana tomentosiformis* reveals multiple T-DNA inserts and a complex evolutionary history of natural transformation in the genus *Nicotiana*. Plant J 80:669–682. https://doi.org/10.1111/tpj.12661

Chen K, Otten L (2016) Morphological analysis of the 6b oncogene-induced enation syndrome. Planta 243:131–148. https://doi.org/10.1007/s00425-015-2387-0

Chen K, Dorlhac de Borne F, Sierro N, Ivanov NV, Alouia M, Koechler S, Otten L (2018). Organization of the TC and TE cellular T-DNA regions in *Nicotiana otophora* and functional analysis of three diverged TE-6b genes. Plant J 94:274–287. https://doi.org/10.1111/tpj.13853

Chilton MD, Drummond MH, Merio DJ, Sciaky D, Montoya AL, Gordon MP, Nester EW (1977) Stable incorporation of plasmid DNA into higher plant cells: the molecular basis of crown gall tumorigenesis. Cell 11:263–271

Christey MC (2001) Use of ri-mediated transformation for production of transgenic plants. Vitro Cell Dev Biol Plant 37:687–700. https://doi.org/10.1007/s11627-001-0120-0

Clarkson JJ, Lim KY, Kovarik A, Chase MW, Knapp S, Leitch AR (2005) Long-term genome diploidization in allopolyploid Nicotiana section Repandae (Solanaceae). New Phytol 168:241–252. https://doi.org/10.1111/j.1469-8137.2005.01480.x

De Cleene M, De Ley J (1981) The host range of infectious hairy-root. Bot Rev 47:147–194. https://doi.org/10.1007/BF02868853

Dessaux Y, Petit A, Farrand SK, Murphy PJ (1998) Opines and opine-like molecules involved in plant-Rhizobiaceae interactions. In: Spaink HP, Kondorosi A, Hooykaas PJ (eds) The Rhizobiaceae: Molecular biology of model plant-associated bacteria. Springer, Berlin, pp 173–197

Fründt C, Meyer AD, Ichikawa T, Meins F (1998) A tobacco homologue of the Ri-plasmid orf13 gene causes cell proliferation in carrot root discs. Mol Gen Genet 259:559–568

Furner IJ, Huffman GA, Amasino RM, Garfinkel DJ, Gordon MP, Nester EW (1986) An Agrobacterium transformation in the evolution of the genus Nicotiana. Nature 319:422–427. https://doi.org/10.1038/319422a0

Gelvin SB (2012) Traversing the cell: Agrobacterium T-DNA's journey to the host genome. Front Plant Sci. https://doi.org/10.3389/fpls.2012.00052

Gordon JE, Christie PJ (2014) The Agrobacterium Ti plasmids. Microbiol Spectr. https://doi.org/10.1128/microbiolspec.PLAS-0010-2013

Guyon P, Chilton MD, Petit A, Tempé J (1980) Agropine in "null-type" crown gall tumors: evidence for generality of the opine concept. Proc Natl Acad Sci U S A 77:2693–2697

Guyon P, Petit A, Tempé J, Dessaux Y (1993) Transformed plants producing opines specifically promote growth of opine-degrading Agrobacteria. Mol Plant Microbe Interact 6:92. https://doi.org/10.1094/MPMI-6-092

Helfer A, Pien S, Otten L (2002) Functional diversity and mutational analysis of Agrobacterium 6B oncoproteins. Mol Genet Genomics 267:577–586. https://doi.org/10.1007/s00438-002-0707-0

Hernalsteens J-P, Van Vliet F, De Beuckeleer M, Depicker A, Engler G, Lemmers M, Holsters M, Van Montagu M, Schell J (1980) The *Agrobacterium tumefaciens* Ti plasmid as a host vector system for introducing foreign DNA in plant cells. Nature 287:654–656

Hildebrand EM (1940) Cane gall of brambles caused by *Phytomonas rubi* n. sp. J Agric Res 61:685–696

Hoekema A, Hirsch PR, Hooykaas PJJ, Schilperoort RA (1983) A binary plant vector strategy based on separation of vir- and T-region of the *Agrobacterium tumefaciens* Ti-plasmid. Nature 303:179–180. https://doi.org/10.1038/303179a0

Intrieri MC, Buiatti M (2001) The horizontal transfer of *Agrobacterium rhizogenes* genes and the evolution of the genus nicotiana. Mol Phylogenet Evol 20:100–110. https://doi.org/10.1006/mpev.2001.0927

Jefferson RA (1987) Assaying chimeric genes in plants: the GUS gene fusion system. Plant Mol Biol Rep 5:387–405. https://doi.org/10.1007/BF02667740

Kado CI (2014) Historical account on gaining insights on the mechanism of crown gall tumorigenesis induced by *Agrobacterium tumefaciens*. Front Microbiol. https://doi.org/10.3389/fmicb.2014.00340

Kerr A (1969) Transfer of virulence between isolates of Agrobacterium. Nature 223:1175–1176. https://doi.org/10.1038/2231175a0

Kerr A, Panagopoulos CG (1977) Biotypes of *Agrobacterium radiobacter* var. *tumefaciens* and their biological control. J Phytopathol 90:172–179. https://doi.org/10.1111/j.1439-0434.1977.tb03233.x

Knapp S, Chase MW, Clarkson JJ (2004) nomenclatural changes and a new sectional classification in Nicotiana

(Solanaceae). Taxon 53:73. https://doi.org/10.2307/4135490

Kyndt T, Quispe D, Zhai H, Jarret R, Ghislain M, Liu Q, Gheysen G, Kreuze JF (2015) The genome of cultivated sweet potato contains *Agrobacterium* T-DNAs with expressed genes: An example of a naturally transgenic food crop. Proc Natl Acad Sci U S A 112:5844–5849. https://doi.org/10.1073/pnas.1419685112

Lemcke K, Schmülling T (1998) Gain of function assays identify non-rolgenes from *Agrobacterium rhizogenes* TL-DNA that alter plant morphogenesis or hormone sensitivity. Plant J 15:423–433. https://doi.org/10.1046/j.1365-313X.1998.00223.x

Levesque H, Delepelaire P, Rouzé P, Slightom J, Tepfer D (1988) Common evolutionary origin of the central portions of the Ri TL-DNA of *Agrobacterium rhizogenes* and the Ti T-DNAs of *Agrobacterium tumefaciens*. Plant Mol Biol 11:731–744. https://doi.org/10.1007/BF00019514

Lioret C (1956) Sur la mise en évidence d'un acide aminé non-identifié particulier aux tissus de crown-gall. Bull Société Fr Physiol Végétale 2:76

Long N, Ren X, Xiang Z, Wan W, Dong Y (2016) Sequencing and characterization of leaf transcriptomes of six diploid Nicotiana species. J Biol Res Thessalon. https://doi.org/10.1186/s40709-016-0048-5

Lütken H, Clarke JL, Müller R (2012) Genetic engineering and sustainable production of ornamentals: current status and future directions. Plant Cell Rep 31:1141–1157. https://doi.org/10.1007/s00299-012-1265-5

Matveeva T, Otten L (2019) Widespread occurrence of natural genetic transformation of plants by Agrobacterium. Plant Mol Biol 101:415-437. https://doi.org/10.1007/s11103-019-00913-y

Matveeva TV, Bogomaz DI, Pavlova OA, Nester EW, Lutova LA (2012) Horizontal gene transfer from genus *Agrobacterium* to the Plant *Linaria* in nature. Mol Plant Microbe Interact 25:1542–1551. https://doi.org/10.1094/MPMI-07-12-0169-R

Ménagé A, Morel G (1964) Sur la présence d'octopine dans les tissus de crown-gall cultivés in vitro. Comptes Rendus L'Académie Sci Paris 259:4795–4796

Mohajjel Shoja H (2010) Contribution to the study of the *Agrobacterium rhizogenes* plast genes, rolB and rolC, and their homologs in *Nicotiana tabacum*. Doctoral dissertation, University of Strasbourg, France

Mohajjel-Shoja H, Clément B, Perot J, Alioua M, Otten L (2011) Biological Activity of the *Agrobacterium rhizogenes*–derived *trolC* gene of *Nicotiana tabacum* and its functional relation to other *plast* genes. Mol Plant Microbe Interact 24:44–53. https://doi.org/10.1094/MPMI-06-10-0139

Morris RO (1986) Genes specifying auxin and cytokinin biosynthesis in phytopathogens. Annu Rev Plant Physiol 37:509–538

Nilsson O, Olsson O (1997) Getting to the root: the role of the *Agrobacterium rhizogenes* rol genes in the formation of hairy roots. Physiol Plant 100:463–473. https://doi.org/10.1111/j.1399-3054.1997.tb03050.x

Ophel K, Kerr A (1990) *Agrobacterium vitis* sp. nov. for Strains of Agrobacterium biovar 3 from Grapevines. Int J Syst Bacteriol 40:236–241. https://doi.org/10.1099/00207713-40-3-236

Petit A, Delhaye S, Tempé J, Morel G (1970) Recherches sur les guanidines des tissus de crown gall. mise en evidence d'une relation biochimique specifique entre le souches d'*Agrobacterium tumefaciens* et les tumeurs qu'elles induisent. Physiol Végétale 8:205–213

Potuschak T, Palatnik J, Schommer C, Sierro N, Ivanov N, Kwon Y, Genschik P, Davière J-M, Otten L (2019) The *Agrobacterium*-derived *TE-2-6b* gene from the natural transformant *Nicotiana otophora* induces a *jaw-D* like phenotype and targets the CIN-like TCP transcription factors. The Plant J. https://doi.org/10.1111/tpj.14591

Riker A, Banfield W, Wright W, Keitt G, Sagen HE et al (1930) Studies on infectious hairy root of nursery apple trees. J Agric Res 41:507–540

Röder FT, Schmülling T, Gatz C (1994) Efficiency of the tetracycline-dependent gene expression system: complete suppression and efficient induction of the rolB phenotype in transgenic plants. Mol Gen Genet 243:32–38. https://doi.org/10.1007/BF00283873

Roullier C, Duputié A, Wennekes P, Benoit L, Fernández Bringas VM, Rossel G, Tay D, McKey D, Lebot V (2013) Disentangling the origins of cultivated Sweet Potato (*Ipomoea batatas* (L.) Lam.). PLoS One 8: e62707. https://doi.org/10.1371/journal.pone.0062707

Schell J, Montagu MV, Beuckeleer MD, Block MD, Depicker A, Wilde MD, Engler G, Genetello C, Hernalsteens JP, Holsters M, Seurinck J, Silva B, Vliet FV, Villarroel R (1979) Interactions and DNA transfer between *Agrobacterium tumefaciens*, the Ti-Plasmid and the plant host. Proc R Soc B Biol Sci 204:251–266. https://doi.org/10.1098/rspb.1979.0026

Schmülling T, Schell J, Spena A (1988) Single genes from *Agrobacterium rhizogenes* influence plant development. EMBO J 7:2621–2629

Schröder G, Waffenschmidt S, Weiler EW, Schröder J (1984) The T-region of Ti plasmids codes for an enzyme synthesizing indole-3-acetic acid. Eur J Biochem 138:387–391. https://doi.org/10.1111/j.1432-1033.1984.tb07927.x

Scott IM (1979) Opine content of unorganised and teratomatous tobacco crown gall tissues. Plant Sci Lett 16:239–248. https://doi.org/10.1016/0304-4211(79)90034-8

Sierro N, Battey JND, Ouadi S, Bakaher N, Bovet L, Willig A, Goepfert S, Peitsch MC, Ivanov NV (2014) The tobacco genome sequence and its comparison with those of tomato and potato. Nat Commun. https://doi.org/10.1038/ncomms4833

Slightom JL, Durand-Tardif M, Jouanin L, Tepfer D (1986) Nucleotide sequence analysis of TL-DNA of *Agrobacterium rhizogenes* agropine type plasmid.

Identification of open reading frames. J Biol Chem 261:108–121

Smith EF, Townsend CO (1907) A plant-tumor of bacterial origin. Science 25:671–673. https://doi.org/10.1126/science.25.643.671

Spena A, Schmülling T, Koncz C, Schell JS (1987) Independent and synergistic activity of rol A, B and C loci in stimulating abnormal growth in plants. EMBO J 6:3891–3899

Suzuki K, Tanaka N, Kamada H, Yamashita I (2001) Mikimopine synthase (mis) gene on pRi1724. Gene 263:49–58. https://doi.org/10.1016/S0378-1119(00)00578-3

Suzuki K, Yamashita I, Tanaka N (2002) Tobacco plants were transformed by *Agrobacterium rhizogenes* infection during their evolution. Plant J 32:775–787. https://doi.org/10.1046/j.1365-313X.2002.01468.x

Tepfer D (1990) Genetic transformation using *Agrobacterium rhizogenes*. Physiol Plant 79:140–146. https://doi.org/10.1111/j.1399-3054.1990.tb05876.x

van Nuenen M, de Ruffray P, Otten L (1993) Rapid divergence of *Agrobacterium vitis* octopine-cucumopine Ti plasmids from a recent common ancestor. Mol Gen Genet (MGG) 240:49–57. https://doi.org/10.1007/BF00276883

Wang K, Herrera-Estrella L, Van Montagu M, Zambryski P (1984) Right 25 by terminus sequence of the nopaline t-DNA is essential for and determines direction of DNA transfer from Agrobacterium to the plant genome. Cell 38:455–462. https://doi.org/10.1016/0092-8674(84)90500-2

Ward SM, Fleischmann CE, Turner MF, Sing SE (2009) Hybridization between invasive populations of Dalmatian Toadflax (*Linaria dalmatica*) and Yellow Toadflax (*Linaria vulgaris*). Invasive Plant Sci Manag 2:369–378. https://doi.org/10.1614/IPSM-09-031.1

White FF, Garfinkel DJ, Huffman GA, Gordon MP, Nester EW (1983) Sequences homologous to *Agrobacterium rhizogenes* T-DNA in the genomes of uninfected plants. Nature 301:348–350. https://doi.org/10.1038/301348a0

White PR, Braun AC (1941) Crown gall production by bacteria-free tumor tissues. Science 94:239–241

Yadav NS, Vanderleyden J, Bennett DR, Barnes WM, Chilton MD (1982) Short direct repeats flank the T-DNA on a nopaline Ti plasmid. Proc Natl Acad Sci U S A 79:6322–6326

Zambryski P, Joos H, Genetello C, Leemans J, Montagu MV, Schell J (1983) Ti plasmid vector for the introduction of DNA into plant cells without alteration of their normal regeneration capacity. EMBO J 2:2143–2150

13 *Nicotiana attenuata* Genome Reveals Genes in the Molecular Machinery Behind Remarkable Adaptive Phenotypic Plasticity

Aura Navarro-Quezada, Klaus Gase,
Ravi K. Singh, Shree P. Pandey and
Ian T. Baldwin

Abstract

Nicotiana attenuata (coyote tobacco) has been developed into a model plant system to understand the molecular basis of traits important for the Darwinian fitness of plants in nature. Twenty years of releases of transgenic *N. attenuata* plants into a nature preserve in the plant's native habitat have greatly advanced our understanding of the traits related to herbivore resistance. We provide an account of the genome sequencing of this wild tobacco species, as well as of another sympatric diploid tobacco, *N. obtusifolia*, which has different responses to herbivore attack. *N. attenuata* is a post-fire annual with a long-lived seed bank and exhibits an impressive array of plastic responses to the environmental challenges it faces, including unpredictable herbivore and pathogen attacks, by reprogramming the expression of thousands of genes. Regulatory small RNAs, including microRNAs, along with their biogenesis and regulating machinery, are thought to tailor the much needed specificity in modulating these large-scale changes in gene expression. Here, we provide a detailed genome-informed account of the small RNA machinery and its evolution in *N. attenuata*.

A. Navarro-Quezada
Centre for Organismal Studies, Ruprecht-Karls University of Heidelberg, 69120 Heidelberg, Germany

K. Gase · S. P. Pandey · I. T. Baldwin (✉)
Department of Molecular Ecology, Max Planck Institute for Chemical Ecology, 07745 Jena, Germany
e-mail: baldwin@ice.mpg.de

R. K. Singh · S. P. Pandey
University Department of Botany, Magadh University, Bodhgaya 824234, India

13.1 Introduction

Nicotiana attenuata (coyote tobacco) has been used for more than 30 years as a model system to study the complex plant-mediated ecological interactions that occur in nature and to elucidate the genetic basis of traits important for the Darwinian fitness of plants. The choice of this diploid, largely selfing plant as a model system, is grounded in its natural history and unusual germination behavior. Dormant seeds, survivors of decades or even centuries of microbial challenges in soil, germinate synchronously into the nitrogen-rich, competition-free environments that are commonly found after fires. Dense *N. attenuata* populations then develop, in which plants are exposed not only to intense conspecific competition but also to as many as 34 herbivore taxa and numerous pathogens that colonize these post-fire populations from adjacent unburned areas.

Insect herbivores have short generation times and therefore have the capacity to rapidly evolve counter-responses to plant resistance traits. *N. attenuata* produces large quantities of nicotine

© Springer Nature Switzerland AG 2020

N. V. Ivanov et al. (eds.), *The Tobacco Plant Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-030-29493-9_13

as a defense trait (Steppuhn et al. 2004) and to optimize pollination by hummingbirds (Kessler et al. 2012). A solanaceous specialist herbivore, the tobacco hornworm (*Manduca sexta*), has evolved an unprecedented capacity to detoxify nicotine (Self et al. 1964; Wink and Theile 2002), largely by excretion but also by co-opting a small fraction for its own defense (Kumar et al. 2012). *M. sexta* larvae are found on *N. attenuata* plants because the adult moths oviposit while nectaring on the flowers. Hence, this insect plays dual roles in *N. attenuata* plants, as a devastating herbivore during the larval stages and as an important pollinator as an adult, vectoring gametes among the spatially separated plant populations that occur after fires. Attack from these nicotine-adapted larvae is specifically recognized by the plant, which displays a remarkable six-layered suite of defense, avoidance, and tolerance responses (summarized in the iBiology talk available at https://www.ibiology.org/plant-biology/studying-plants-ecological-interactions-genomics-era-story-nicotiana-attenuata/#part-2), which engages all aspects of the plant physiology. *N. attenuata* genotypes that adjust their growth and defense phenotypes most successfully to these unpredictable environmental challenges will produce the most successful future generations and attain the highest representation in the seed banks, and thus have higher Darwinian fitness.

To cope with the plethora of biotic and abiotic challenges, *N. attenuata* genotypes exhibit great adaptive biochemical and morphological phenotypic plasticity. *N. attenuata* has been developed as a model system to investigate how a native plant copes with ecologically relevant challenges, particularly insect herbivory. This research uses a well-developed molecular toolbox and a nature preserve in the plant's native range, in which genetically modified plants are released to understand gene function (Baldwin 2001; Gase and Baldwin 2012; Gaquerel et al. 2013). *N. attenuata* reconfigures its signaling and metabolic networks to address these ecological challenges (Kim et al. 2011; Gulati et al. 2013b), and these in turn are modulated by a large repertoire of small RNAs (smRNAs), which are also produced in response to insect feeding (Pandey et al. 2008; Pradhan et al. 2017). The study of the involvement of smRNAs in these defense mechanisms necessitates comprehensive knowledge of the components of the smRNA machinery in *N. attenuata*. In this chapter, we describe how the recently assembled genome of *N. attenuata* informs our understanding of some of the molecular machinery responsible for parts of this phenotypic plasticity. We review the sequencing and assembly of the *N. attenuata* and *N. obtusifolia* genomes and describe how the *N. attenuata* transcriptome was used for the genomic and evolutionary annotation of the smRNA machinery. We also present an overview of the genes related to smRNA machinery, especially those known to be directly involved in herbivore defense in *N. attenuata*.

13.2 Sequencing the Genomes of *N. attenuata* and *N. obtusifolia*

Genome sequencing was conducted on the *N. attenuata* accession UT (from Utah) because, following numerous laboratory and field studies, this is the best-studied accession to date. The UT genotype was inbred for 30 generations before being used for genome sequencing. The *N. attenuata* accession AZ (from Arizona), which had been inbred for 22 generations, was also sequenced although to a lesser extent. The UT accession originated from a 1996 collection from a native population in Washington County in Utah, USA, while the AZ accession was collected 10 km east of Flagstaff, Arizona, USA. A cross between both accessions using UT 30× (pollen donor) and AZ 22× (ovule donor) was created to generate a linkage map for the *N. attenuata* genome. This map was used to improve the assembly by analyzing the single-nucleotide polymorphisms differentiating these two accessions (Xu et al. 2017). The UT and AZ accessions are known to vary strongly with respect to important ecological traits, such as

jasmonate-mediated herbivory resistance, pollination, and volatile compound emission (Glawe et al. 2003; Steppuhn et al. 2008; Wu et al. 2008; Zhou et al. 2017).

This sequencing effort also included the genome of *N. obtusifolia*, which is closely related to *N. tomentosiformis*, the paternal ancestor of the tetraploid *N. tabacum* (Clarkson et al. 2010). *N. obtusifolia* is sympatric with the UT accession, and its seeds were collected from a native population in 2004 at the Lytle Ranch Preserve, St. George, Utah, USA. A single *N. obtusifolia* plant of the first inbred generation was sequenced. *N. attenuata* and *N. obtusifolia* have contrasting life histories (annual versus perennial) and different defense strategies against attack from herbivores (Lou and Baldwin 2003; Anssour and Baldwin 2010; Xu et al. 2015; Stanton et al. 2016), most likely related to their differing habitats (open desert for *N. attenuata* and rock crevices for *N. obtusifolia*) despite their overlapping geographic ranges. Another advantage of sequencing *N. obtusifolia* is its genome size (30% smaller than the *N. attenuata* genome), which was useful in guiding the genome assembly of the much-larger *N. attenuata* genome (2.58 Gb).

13.2.1 Assembling the Genomes of *N. attenuata* and *N. obtusifolia*—The Challenge of Repeats

The first round of sequencing the *N. attenuata* UT genome began in 2011 and provided reads with 58× coverage from 454 (Roche; Basel, Switzerland) and Illumina (San Diego, CA, USA) HiSeq 2000 technologies. The first assembly revealed that the repetitive regions were not well-assembled. Even with a scaffold N50 above 200 kb, it was not possible to recover the upstream regions of the genes involved in biosynthesis of the main defense metabolites, which appeared close to transposable elements (TEs), as is the case for nicotine biosynthetic genes (Xu et al. 2017). Thus, in the subsequent rounds of sequencing, the focus was on obtaining larger contigs. The assembly of the *N. obtusifolia* genome, which contained 20% fewer TEs and is smaller, resulted in larger contigs and was relatively good after the first round of sequencing. A contig N50 of 39.2 kb and a scaffold N50 of 134.1 kb were attained after 50× coverage from Illumina short reads (Xu et al. 2017). This helped to guide the *N. attenuata* assembly but was still not sufficient to improve it significantly and additional work was required.

The assembly was improved with PacBio sequencing (Pacific Biosciences, Menlo Park, CA, USA) to bridge gaps and with a Bionano (San Diego, CA, USA) optical map to enable fragments to be aligned (Xu et al. 2017). Both technologies required the isolation of large genomic DNA fragments in the megabase range for optical mapping. The isolation of large genomic DNA fragments from *N. attenuata* that met the quality requirements for the PacBio and Bionano technologies proved to be difficult, partly because of the large amounts of secondary metabolites that this plant produces. To overcome this problem, which, at the time, appeared specific to the PacBio sequencing and its associated chemistry, genomic DNA was obtained from plants that were transformed to silence the expression of genes that regulate large sectors of the plant's secondary metabolism. Specifically, a cross of two isogenic *N. attenuata* transgenic lines was used: an inverted-repeat allene oxide cyclase (ir-AOC) line that was impaired in jasmonic acid biosynthesis (Kallenbach et al. 2012) as the ovule donor, and a line silenced in the synthesis of the abundant 17-hydroxy-geranyllinalool diterpene glycosides (irGGPPS), described in Heiling et al. (2010), as the pollen donor. A 10× genome coverage was performed on the DNA isolated from this cross by PacBio sequencing.

After numerous rounds of optimizing genomic DNA isolation from the *N. attenuata* UT 30× inbred wild-type line, an optical map was created using the bionanotechnology, which resulted in an assembly with an N50 of 536 kb and a single-molecule N50 of 115 kb. Overall, 81.7% of the sequence assembly was mapped to the optical map and 84.7% of the optical map was aligned to the sequence assembly (Xu et al. 2017).

To further improve the assembly, the linkage map for *N. attenuata* from the UT × AZ accessions cross was used to develop an advanced intercross-recombinant inbred lines (AI-RIL) population. This population was generated by selfing an F_1 population between UT and AZ plants, followed by four generations of intercrossing 150 of the F_2 progenies, which yielded F_3 to F_6 seeds. From each F_6 plant, 96 individuals were germinated and five rounds of inbreeding (F_7 to F_{11} seeds) were performed, finally yielding 256 RILs. The genomic DNA of one F_{11} individual from each RIL was low-pass illumina sequenced, resulting in 1.2 billion paired-end clean reads (Xu et al. 2017). A 2906-cM map was obtained from this sequencing. A mashup of the sequences obtained from the different methods described above completed the *N. attenuata* contigs, which were mapped onto 12 pseudochromosomes (Fig. 13.1). A total of 2132 scaffolds representing 825.8 Mb (34.9%) of the final genome assembly, and a total of 13,076 (38.9%) of the total predicted genes were anchored to 12 pseudomolecules or pseudochromosomes. The assembled genomes are available in GenBank under the following accessions: MJEQ00000000 (*Nicotiana attenuata*, ecotype Utah), MCOF00000000 (*Nicotiana attenuata*, ecotype Arizona), and MCJB00000000 (*Nicotiana obtusifolia*).

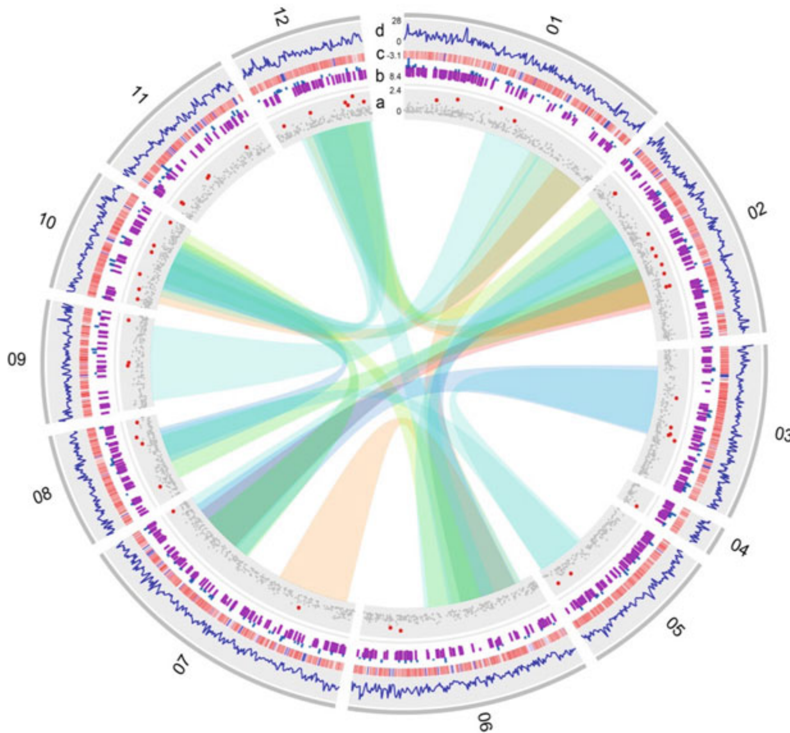

Fig. 13.1 Circos plot of the 12 assembled pseudochromosomes from *Nicotiana attenuata* (Provided by Dr. Shuqing Xu). The ribbons indicate 22 syntenic blocks found using MCSCAN (Tang 2009), each of which contain at least 20 genes. The color of the ribbons is used to visually distinguish synthetic blocks on different chromosomes. Lane a: nonsynonymous/synonymous mutation ratios (Ka/Ks) of genes between *N. attenuata* and *N. obtusifolia*; red dots indicate ratios >1. Lane b: induced gene expression changes in *N. attenuata* leaves by the oral secretions of the specialist herbivore *Manduca sexta*. Each bar indicates the log2 fold change of each gene; purple and blue bars indicate up- and down-regulated genes, respectively. Lane c: heatmap depicting the distribution of transposable elements within a 500-kb sliding window. Red and blue indicate high and low expressions of the transposable elements, respectively. Lane d: number of genes in 500-kb sliding windows

The annotation of genes and TEs is described in more detail in Xu et al. (2017). Here, we show the distribution of genes and TEs, as well as their expression, on the 12 pseudochromosomes as a Circos plot (Fig. 13.1). One of the surprising findings of the genome analysis was the relative proximity of TEs to genes, in particular long interspersed nuclear elements and miniature inverted-repeat TEs, which were preferentially found close to genes (Xu et al. 2017). The Circos plot also highlights the distribution of genes that have nonsynonymous/synonymous mutation ratios >1, and an even distribution of genes with rapid evolutionary rates; for these genes, no significant association was found with TEs. Another interesting finding in the functional analysis of the mapped genes was the capacity of the genes to respond to simulated herbivory when the oral secretion of a specialist herbivore, *M. sexta*, was applied to a mechanically wounded leaf (Wu and Baldwin 2009). The induced responses to simulated herbivory have an important adaptive value, as most of the genes analyzed from *N. attenuata* were up-regulated during this stimulus. These genes are displayed as a track shown in Fig. 13.1. The assembly of the *N. attenuata* and *N. obtusifolia* genomes sets the stage for investigating the differences between these two species with contrasting responses to herbivore attack in future studies (Lou and Baldwin 2003; Wu et al. 2006). The common responses to herbivory have been analyzed from multiple transcriptome datasets, as we discuss in the next section.

13.3 Annotating the *N. attenuata* Genome and Transcriptome in Light of Herbivory-Induced Responses

One of the main objectives of sequencing the *N. attenuata* genome was to understand the mechanisms responsible for the remarkable phenotypic plasticity of *N. attenuata*. This plant can distinguish between damage caused by feeding insects and mechanical damage by perceiving herbivore-associated elicitors (HAEs) to induce rapid early defense signaling. Research with targeted analyses of differentially expressed genes spotted onto an in-house-constructed microarray identified important components of the HAE-regulated signaling network in *N. attenuata*: protein kinases, such as wounding-induced protein kinase (Wu et al. 2007; Wu and Baldwin 2010) and calcium-dependent protein kinases (Yang et al. 2012); transcription factors, such as *NaWRKY3/6* (Skibbe et al. 2008); and ethylene biosynthesis and perception genes (*NaETR1*, *NaACO*, and *NaACS*) that modulate jasmonic acid-regulated defense responses such as nicotine biosynthesis (Kahl et al. 2000; Voelckel et al. 2001; Onkokesung et al. 2010).

An extensive co-transcriptional network analysis was needed to understand the molecular mechanisms of the HAE-induced early defense signaling network, including their genetic architecture and the identification of still-unknown components of the network, because functionally related genes tend to be transcriptionally coordinated (Stuart et al. 2003; Persson et al. 2005).

A complete *N. attenuata* transcriptome was acquired with 454 sequencing (GS FLX Titanium, 2010, GenBank GBGF01000000). For this, RNA was isolated from multiple organs (complete flowers, nectaries, anthers, pistils, developing seed capsules, seedlings, petioles, leaf midribs, and laminae) and pooled in equal amounts. A total of 1,107,883 *N. attenuata* expressed sequence tags were obtained and assembled into 43,533 contigs. The best-supported contigs, i.e., those with the highest read coverage, were used to produce a 44-k Agilent (Santa Clara, CA, USA) microarray (GEO GPL13527). The 454-derived transcriptome sequence data and the microarray probes designed from this information represented 75% of the *N. attenuata* transcriptome. This array was used extensively to study changes in the transcriptome of *N. attenuata* after ecologically relevant elicitations. Transcriptional profiles of co-regulated genes allowed the assignment of genes with unknown functions to partly elucidated biological pathways (Gulati et al. 2013b).

To identify missing factors in HAE-induced early defense signaling networks, whole-transcriptome sequencing and the design of a whole-transcriptome array for gene transcription studies were important parts of the *N. attenuata* genome sequencing project. For complete transcriptome sequencing, RNA from multiple organs of the UT 30× accession (flower buds, opening flowers, early-opening corollas, late-opening corollas, pedicels, ovary, anthers, nectaries, selfed style, outcrossed style, unpollinated style, pollen tubes, stigmas, dry seeds, water-imbibed seeds, smoke-treated imbibed seeds, stems, leaves and roots from plants treated with *M. sexta* oral secretion, and untreated controls) and from leaves of the AZ 22× accession (untreated and treated with *M. sexta* oral secretions) were harvested at various time points from plants grown under greenhouse conditions.

The UT 30× and AZ 22× transcriptome assemblies were used to support the predictions of transcripts and exons within the respective genome sequencing projects and therefore were submitted as features of the genome sequences. The UT 30× assembly (BioProject PRJNA355166) was shown to be complete with respect to all genes expressed under the selected growth conditions. A publicly accessible database (the *Nicotiana attenuata* Data Hub, http://nadh.ice.mpg.de/NaDH/) that contains information on transcript abundance for each gene in a particular organ was created (Brockmoller et al. 2017). This database includes information from two different 66-k Agilent microarrays; the first contained gene predictions based on hidden Markov models derived from solanaceous genome sequence data (GEO GPL19764), and the second contained all transcripts, including gene predictions, derived from the *N. attenuata* genome sequence (GEO GPL21138).

To obtain additional information on the roles of particular genes, experiments were performed not only with wild-type plants but also with transgenic plants in which the expression of genes involved in the HAE-elicited network was manipulated. The large datasets that were produced revealed that HAE-specific gene and metabolic networks exhibit reconfiguration dynamics that are tissue-specific as long-distance signals spread throughout the plant (Gulati et al. 2013a, 2014). Central to HAE signaling is the rewiring of phytohormone signaling, particularly jasmonate signaling, which is pivotal to the transcriptional regulation of most secondary metabolism biosynthetic genes. In *N. attenuata*, combinations of tissue-specific metabolomics and transcriptomics analyses allowed the deciphering of the molecular organization of several of the most prominent secondary metabolic pathways of this plant. In addition, candidate genes were pinpointed for metabolic branch-specific transgenic manipulations of these pathways and rigorous tests of their anti-herbivore functions. Apart from the increased synthesis of nicotine that occurs in the roots and its import to the shoot, major reconfigurations in gene expression in local and systemic leaves translate into de novo production of defensive phenolic–polyamine conjugates and 17-hydrogeranyllinalool diterpene glycosides. Importantly, rapid metabolic reconfigurations connected to the early HAE-phytohormonal reconfigurations are also elicited. These metabolic reconfigurations include shifts in the source-to-sink tissue mobilization of primary metabolites, most notably the increased rootward transport of recently fixed phytoassimilates as a resource-conserving mechanism, but also the emission of green leaf volatiles produced from fatty acid oxidation and degradation that attract the natural enemies of the plant's herbivores. More detailed descriptions of these metabolic changes have been reported (Schwachtje and Baldwin 2008; Allmann and Baldwin 2010; Gaquerel et al. 2010; Gulati et al. 2013b; Li et al. 2015; Brockmoller et al. 2017).

A crucial piece in the puzzle of the evolution of herbivory was obtained through co-expression network analysis of data from the second whole-transcriptome microarray and data from HAE-induced leaf transcriptomes from six closely related *Nicotiana* species. This analysis revealed a key gene co-expression network designated "M4 module." This network is co-activated with the HAE-induced jasmonic acid accumulations but is elicited independently of jasmonic acid. A putative leucine-rich repeat receptor kinase

was identified as a hub gene of the M4 module (Zhou et al. 2016). This study also found that multiple genes are conserved in the response to herbivore attack on the selected *Nicotiana* species. While the existence of the M4 module is inferred to have existed in the most recent common ancestor of the six closely related *Nicotiana* species used for this analysis, it is difficult to test directly, because this ancestor has been extinct for millions of years. Consequently, the co-expression network analysis should be conducted with as many accessions of *N. attenuata* as are available, so that the modern version of the HAE-elicited M4 module can be robustly tested in existing *N. attenuata* ecotypes.

The previously described *N. attenuata* transcriptomes were essential to support the *Nicotiana* genome annotation pipeline created especially for the annotation of *N. attenuata* genes. This pipeline used both protein homology with MAKER2 v. 2.28 (Cantarel et al. 2008) and hint-guided Augustus (v. 2.7) predictions (Stanke et al. 2006). To confirm the robustness of the annotations, 300 genes were curated manually (Zhou et al. 2016) using WebApollo (Lee et al. 2013). Robust assemblies of the *N. attenuata* and *N. obtusifolia* genomes and their transcriptome profiles were critical to investigating the signatures of an inferred whole-genome triplication (WGT) located at the base of Solanaceae and hence shared with all our sequenced *Nicotiana* species (Xu et al. 2017).

The annotation helped to validate the inference of a WGT event shared between *Nicotiana* and *Solanum*, and a subset of genes that were triplicated in either *Solanum* or *Nicotiana* was analyzed further. The genes that underwent a triplication event in *Solanum* were identified by their particular tree structures that consisted of one outgroup node (at least one of *Arabidopsis*, *Populus*, *Cucumis*, or *Vitis*) and two duplication events shared between tomato (*Solanum lycopersicum*) and potato (*S. tuberosum*) without gene loss in either species (Fig. 13.2). Similarly, the genes that underwent a triplication event in *Nicotiana* were identified based on their tree structure, which consisted of one outgroup node and two duplication events shared between *N. attenuata* and *N. obtusifolia*, also without gene loss in either species (Fig. 13.2). To reduce false positives, only the tree nodes with approximate Bayes branch supports of >0.9 were considered. The number of triplication events shared between *Solanum* and *Nicotiana* based on the complete

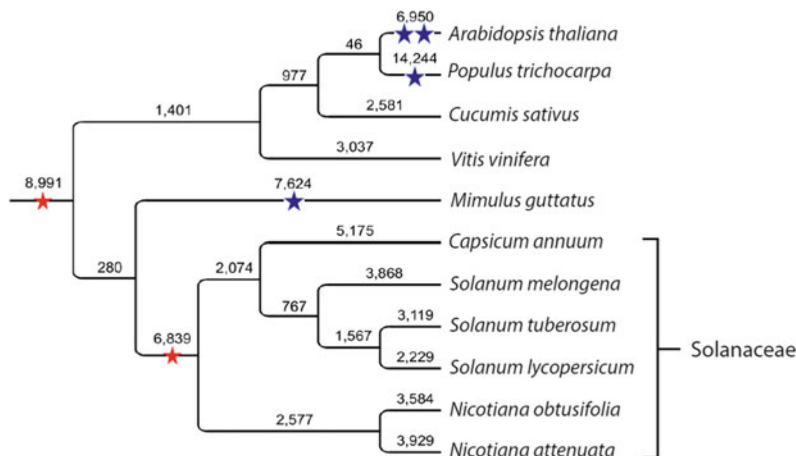

Fig. 13.2 Tree depicting gene family size evolution among 11 plant genomes (modified from a figure provided by Thomas Brockmoller). The numbers of gene duplication events are shown above the branch nodes. The duplication events were estimated from phylogenetic trees constructed from homologous groups. Only branches with approximated Bayes branch support values >0.9 are shown. Known whole-genome triplication (red stars) and duplication (blue stars) events are indicated

tree structure was then analyzed (Fig. 13.2). In total, 229 and 435 triplication events were identified in *Solanum* and *Nicotiana*, respectively. In both the *Nicotiana* and *Solanum* datasets, the majority of triplication events (79.9% and 89.4%, respectively) were shared with the other genus. These findings are consistent with the interpretation that the WGT event found in tomato is indeed shared between *Solanum* and *Nicotiana*.

This information was crucial to understanding the interplay of gene duplication and transposition in the origins of the genes involved in the biosynthesis of defense compounds in *N. attenuata*, notably those involved in the biosynthesis of the key defense compound, nicotine (Xu et al. 2017). Similar to our analysis of nicotine biosynthesis, in the following sections. we use the WGT evidence to understand in greater depth the evolution of the parts of the smRNA machinery that prior research has associated with transcriptional responses to herbivory (Pandey and Baldwin 2007; Bozorov et al. 2012b; Pradhan et al. 2017).

13.3.1 smRNAs and the smRNA Machinery of *N. attenuata*

The transcripts of the smRNA machinery genes obtained from the UT 30× transcriptome tended to occur at very low levels. Two genes (Argonaute 8 [*AGO8*] and Dicer-like 4 [*DCL4*]), which were previously described to exhibit organ-specific differential responses to herbivore elicitation (Pradhan et al. 2017), were up-regulated mainly in root and reproductive tissues after leaf wounds were treated with herbivore oral secretions (Fig. 13.3).

smRNAs are key players in the regulation of gene expression in plants and can be broadly classified into small interfering RNAs (siRNAs) and microRNAs (miRNAs). The biosynthesis and action of smRNAs consist broadly of two events: (i) the generation of double-stranded precursors and their preprocessing and export to the cytoplasm (in the case of miRNAs), and (ii) the maturation of smRNAs in the cytoplasm, where they interact with protein scaffolds and regulate gene expression by pairing with complementary regions of their target mRNAs to enforce translational repression or mRNA degradation. miRNAs have been shown to be involved in nearly all aspects of plant development and adaptation. In *N. attenuata*, miRNAs regulate the reshaping of the plant's responses to herbivory (Pandey et al. 2008; Bozorov et al. 2012a; Pradhan et al. 2017; Xu et al. 2017).

Here, we describe our annotation of genes that encode proteins of seven families that mainly modulate smRNA biogenesis and action: AGOs, DCLs, RNA-dependent RNA polymerases (RdRs), dsRNA-binding proteins (DRBs), HUA enhancer proteins (HENs), silencing defective proteins (SDEs), and suppressor of gene silencing proteins (SGSs) (Brodersen and Voinnet 2006; Borges and Martienssen 2015). The completion of the *N. attenuata* genome enabled the first in-depth analysis of these protein families in a Solanaceae genome. We compare the smRNA machinery genes found in the *N. attenuata* genome with corresponding genes in other Solanaceae genomes (*S. lycopersicum*, *S. tuberosum*, and *Capsicum annuum*). We also annotate the miRNAome and infer the functions of the annotated genes in pathogen defense from our transcriptome data.

13.3.2 The *N. attenuata* miRNAome

We identified at least 125 unique miRNAs, corresponding to more than 135 unique reads, transcribed from more than 158 genomic locations. These unique miRNAs are conserved in 34 plant species and all were identified in the *N. attenuata* genome (Pradhan et al. 2017; Xu et al. 2017; Pandey et al. 2018). Additionally, we estimated that the *N. attenuata* genome contained at least 166 putative novel miRNA candidates; thus, we consider that at least 291 miRNAs are transcribed from the *N. attenuata* genome. We will refer to these miRNAs as Na-miRNAs. Most Na-miRNA genes (88% conserved and 81% novel) were found in intergenic or intronic regions of the genome. Interestingly, we found

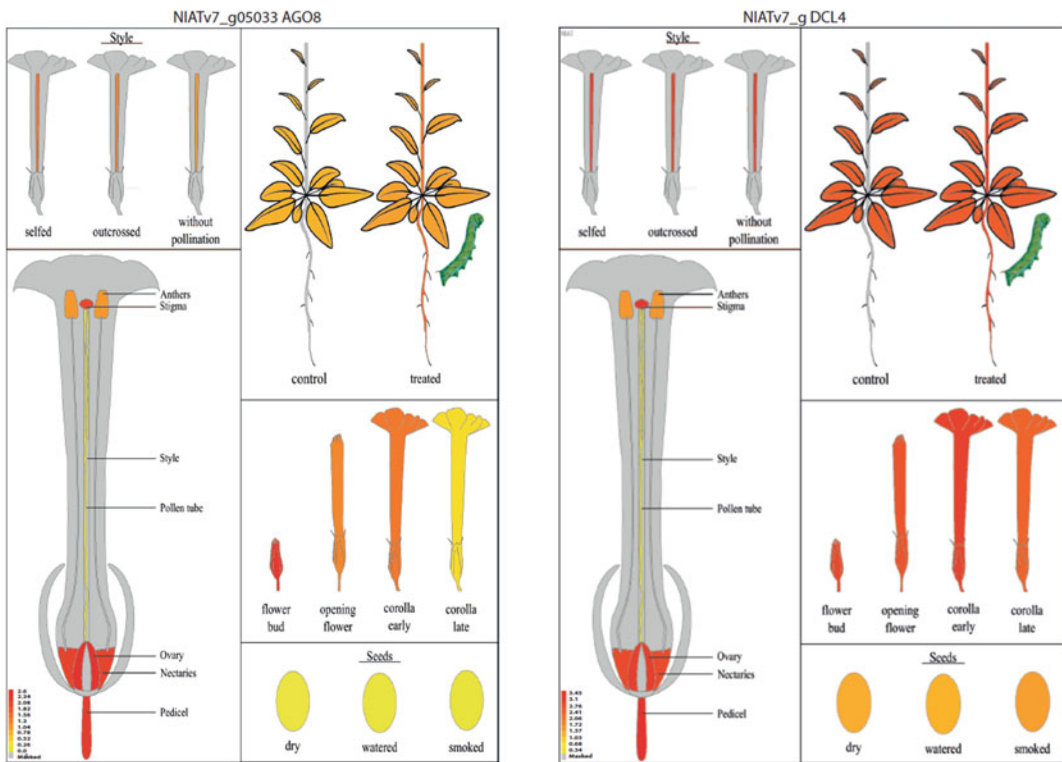

Fig. 13.3 Diagram illustrating *AGO8* and *DCL4* gene expression in *Nicotiana attenuata. AGO8* and *DCL4* are central components of the smRNA machinery. The transcript abundances of these two genes changed significantly in response to herbivory; their abundances increased in roots and stems, as well as in flowering parts after leaf-puncture wounds were treated with *Manduca sexta* oral secretions. The color scale (bottom left corner of each panel) indicates relative abundances

that 12.8% and 19.2% of the miRNAs (16 conserved and 31 novel, respectively) were derived from regions associated with exons; two conserved and two novel miRNAs were from untranslated regions, and two conserved and five novel miRNAs were from coding sequences. The remaining 12 conserved and 25 novel miRNAs were from introns. The processed Na-miRNAs ranged from 21 to 24 nucleotides (nt), and their precursor lengths varied from 45 to 111 nt, with an average length of 71 nt. Overall, most smRNA precursors had average lengths of 71 and 90 nt (50% conserved and 46% novel, respectively).

The Na-miRNAs display complex patterns of spatial-, temporal-, and signal-dependent transcription. Twelve Na-miRNAs were expressed only during the day (miR-160c-3p, miR-171b-3p, miR-398a-3p, miR-426, miR-1429-5p, miR-5069, miR-5497, miR-6020b, miR-6021, miR-6206, miR-7711-5p.4, and miR7744-5p) and 14 Na-miRNAs were expressed only during the night (miR-160-5p, miR-408-5p, miR-2609a, miR-4379, miR-5015, miR-5042-3p, miR-5255, miR-5303c, miR-5635a, miR-5741a, miR-6444, miR-7750-5p, miR-8143, and miR-8764) (Xu et al. 2017). The *N. attenuata* genome extensively expresses isomiRs, which are miRNAs that have sequence variants or different mature sequences. isomiRs of a given miRNA are thought to be expressed differently between day and night. Examples of *N. attenuata* isomiRs are miR-172c, miR-172j, miR-5303, miR-5303a, miR-6149a, miR-6161a, miR-6161b, miR-6161c, miR-6161d, mi-R6164a, miR-6164b, and miR-7997c. Herbivory-specific signals were also found to elicit large-scale reprogramming of the miRNAome, inducing the accumulation of more than 203 miRNAs (94 conserved and 109 novel miRNAs). Such

large-scale, HAE-elicited changes in the transcription of miRNAs suggest post-transcription regulation that mediates ecological processes, including plant defense (Pradhan et al. 2017).

13.3.3 *N. attenuata* smRNA Machinery is Involved in Herbivore Defense

Most of the structural and functional analysis described for Na-smRNA genes was derived from co-expression analyses and comparative genetics (Pandey and Baldwin 2008; Bozorov et al. 2012a, b; Pradhan et al. 2017). Here, we provide an overview of the genes that form the smRNA machinery, with emphasis on those *AGO*, *DCL*, and *RdR* genes that are known to be involved directly in herbivore defense in *N. attenuata* (Pandey and Baldwin 2007; Bozorov et al. 2012b; Pradhan et al. 2017).

AGO proteins. An evolutionary highly dynamic gene family (Zhang et al. 2015), AGO genes evolved from multiple duplication events in plants. They encode proteins that form the functional core of the RNA-induced silencing complex in eukaryotes, which catalytically cleaves target transcripts during post-transcriptional gene silencing or methylates key cytosines in the genome during transcriptional gene silencing. Structurally, AGOs are multi-domain proteins, characteristically comprising an N-terminal domain (which facilitates the separation of the smRNA–mRNA duplex), a PAZ domain (which harbors a characteristic hydrophilic cleft that recognizes the 2-nt overhang at the 3′ end of miRNAs and serves as an anchor), MID domains (which have a characteristic 'nucleotide specificity loop' that recognizes and binds to the 5′ phosphate of smRNAs), and a PIWI domain (which harbors the characteristic catalytic tetrad, D-E-D-H/D at the active site and has the capacity to break the phosphodiester bonds of nucleic acid). For an in-depth discussion of these domains, see Song et al. (2004) and Hutvagner and Simard (2008).

The *N. attenuata* genome contains 12 AGO genes (Singh et al. 2015), in line with the numbers found in other solanaceous genomes (11–15 AGO genes; Table 13.1). The *N. attenuata* genome contains three variants of *AGO1* (a, b, and c) and two each of *AGO4* and *AGO10*, but lacks orthologs of *AGO3* and *AGO6*, which are found in the *Arabidopsis thaliana* genome. All homologs contain the four characteristic domains: N, PAZ, MID, and PIWI. The AGO paralogs have been classified into four phylogenetic clades in plants (Singh et al. 2015). The AGOs associate with the other component of the RNA-induced silencing complex, the smRNAs, based on smRNA length and the 5′ residues of the smRNAs (Singh et al. 2015). In ongoing projects, we obtained evidence that in *N. attenuata*, AGOs fulfill different functions in the plant's environmental responses. For example, NaAGO7 may be involved in mediating responses to competition with other plants, and NaAGO8 may be involved in defense against herbivore attack (Pradhan et al. 2017).

Table 13.1 Numbers of the small RNA machinery genes in solanaceous genomes

	AGO	DCL	RdR	DRB	HEN		SDE		SGS	
					HEN1	HEN2	SDE3	SDE5	SGS1	SGS3
Nicotiana attenuata	12	4	4	7	1	4	1	1	3	1
Nicotiana obtusifolia	10	2	4	8	1	4	1	1	3	1
Nicotiana benthamiana	15	3	7	9	2	8	4	1	5	3
Solanum lycopersicum	15	7	6	6	1	4	2	2	1	2
Solanum tuberosum	11	2	4	6	1	3	1	2	1	1
Capsicum annuum	11	2	5	6	1	3	1	1	1	1

AGO Argonaute; *DCL* Dicer-like; *RdR* RNA-dependent RNA polymerase; *DRB* dsRNA-binding; *HEN* HUA enhancer; *SDE* silencing defective; *SGS* gene silencing

DCL proteins. DCLs recognize and process double-strand RNA (dsRNA) and are the most upstream proteins in the RNA interference pathway. DCLs are large multi-domain proteins encoded by multigene families. In plants, they are distributed among four phylogenetic classes that contain the homologs DCL1 to DCL4 (Mukherjee et al. 2013). The *N. attenuata* genome contains these four DCLs (Bozorov et al. 2012b), but in other solanaceous genomes, the numbers range from two to seven (Table 13.1). DCLs have various domain architectures, which are characteristic of kingdom-specific lineages (Mukherjee et al. 2013). Functional overlap or redundancy is thought to occur among the paralogs, where one DCL may functionally compensate for the loss of another (Abrouk et al. 2012).

Overall, structurally different NaDCLs are thought to have different functional targets. As a common core, they contain an N-terminal DEAD, an RNA helicase, dsRNAb, PAZ, RNaseIII, and DSRM domains. The RNaseIII catalytic domain functions as a homodimer and is ubiquitous among all the DCLs, whereas the PAZ domain binds to the end of target dsRNAs that contain a 3′ two-base overhang. The two dsRNA-binding domains, dsRNAb (dsRNA-binding) and DSRM (dsRNA-binding motif), typically bind dsRNA targets and regulate substrate specificity. The DSRM domain is the most variable domain in DCLs, and DCL2 lacks this domain entirely. The DSRM, PAZ, and RNaseIII a and b domains are responsible for recognizing and processing dsRNAs (Kurzynska-Kokorniak et al. 2015; Mickiewicz et al. 2017).

The excision of mature smRNAs from their precursors by DCLs is an essential first step in smRNA maturation. NaDCL1 is critical to the production of mature miRNAs (Bozorov et al. 2012b) and in transposon silencing by mediating DNA methylation. NaDCL2 participates in processing 22–24-nt natural antisense siRNAs, while NaDCL3 processes the 22–24-nt heterochromatin-derived siRNAs. DCL4 processes a broad range of miRNAs, trans-acting siRNAs, secondary viral smRNAs, and siRNAs from transgenes, in addition to the transitivity of silencing signals (Blevins et al. 2006; Brodersen and Voinnet 2006; Chapman and Carrington 2007).

In general, plant DCLs exhibit complex expression patterns during a plant's adaption to various environmental stresses. In *N. attenuata*, *NaDCL3* and *NaDCL4* mRNAs display expression patterns indicative of diurnal fluctuations, while *NaDCL1*, *NaDCL2*, and *NaDCL3* are suppressed and enhanced at various time points after wounding or simulated herbivore attack (Bozorov et al. 2012b). The NaDCL4-guided generation of 21-nt siRNAs is correlated with the cell-to-cell spread of transgene silencing, as well as with the long-range within-plant spread of gene silencing.

RNA-dependent RNA polymerases (RdRs). RdRs synthesize long dsRNA molecules that are typically the precursors of siRNAs in plants. RdRs are characterized by a conserved catalytic domain in the C-terminal region of the protein. RdRs have three sub-domains: palm, thumb, and fingers. The palm sub-domain contains a catalytic signature motif, D-F/L-D-G-D, at the active site that coordinates divalent metal cations, in most cases Mg^{2+}. RNA synthesis generally occurs in the $5' \rightarrow 3'$ direction using smRNAs as primers. RdR genes (previously termed RdRPs) are found in the genomes of plants, fungi, protists, lower animals, and RNA viruses (Lesburg et al. 1999; Willmann et al. 2011).

The *N. attenuata* genome sequence confirmed the previously annotated three RdR genes (Pandey and Baldwin 2007) (Table 13.1). In *N. attenuata*, NaRdR1 functions in herbivory defense and mediates changes in the smRNA population during herbivore attack. NaRdR2 largely contributes to the accumulation of 24-nt heterochromatin-derived siRNAs that direct the RNA-dependent DNA methylation pathways, and NaRdR3 (the homolog of *Arabidopsis* RdR6) regulates aspects of auxin homeostasis in the plant, which is important for competition with neighboring plants in natural habitats (Pandey and Baldwin 2007).

***Other smRNA machinery genes* (DRB, HEN, SGS, SEG)**. *N. attenuata* and *N. obtusifolia* contain the same numbers of other smRNA

machinery genes as other Solanaceae genomes analyzed to date. These gene family numbers appear to have evolved similarly (Table 13.1), with the exception of *SGS1*, which appears to have been retained in only three copies in *Nicotiana* genomes; all other Solanaceae genomes harbor only one copy. All other smRNA machinery genes are involved in gene silencing (DRBs) or in the post-transcriptional silencing of protein expression (SDEs). They are also involved in smRNA biogenesis (DRBs) and degradation (*HEN2*), methylation (*HEN1*), miRNA accumulation (DRBs), and interactions with RNAs (SGSs) (Brodersen and Voinnet 2006; Chapman and Carrington 2007). These genes display complex expression patterns, which vary between tissues and developmental stages.

Methylation and post-transcriptional gene silencing are often involved in defense against viruses, but also have developmental roles. Often, similar functions are found for closely related genes. For example, *HEN1* encodes an S-adenosyl-L-methionine-dependent RNA methyltransferase that methylates the 2′ OH termini of both strands of smRNA duplexes in a sequence-dependent manner to protect them from uridylation and degradation from the 3′ end (Chen et al. 2002; Huang et al. 2009). HEN1 contains two dsRNA-binding motifs and is involved in plant development and specification of reproductive organ identities, as well as in controlling floral determinacy by regulating the C-class floral homeotic gene *AGAMOUS*. Another related gene, *HEN2*, encodes an ATP-dependent RNA helicase that is involved in degrading various types of non-coding nuclear RNA substrates, such as small nucleolar RNAs, incompletely spliced mRNAs, miRNA precursors, and spurious transcripts that are generally produced from pseudogenes and intergenic regions. HEN2 regulates floral organ spacing and identity, either through the regulation of protein synthesis or through mRNA degradation.

Although many of these proteins have similar functions, their structures differ; for instance, SDEs and SGSs both function as helicases. Among the SGSs, SGS1 contains a helicase domain and DEAD and HRDC domains that help in strand unwinding and RNA binding, respectively, and SGS3 possesses a coiled-coil domain, which helps in oligomerization, and a zinc-finger domain (Mourrain et al. 2000; Dalmay et al. 2001; Hernandez-Pinzon et al. 2007; Bernstein et al. 2009). SGS1 has been associated with the maintenance of genome stability by suppressing spontaneous homologous recombinations in somatic cells and modulating the DNA damage response (Bernstein et al. 2009). Multiple functions have been assigned to these smRNA-related proteins, but not all mechanisms are well explored. For instance, DCLs and DRBs are known to function together in solanaceous genomes, including the *N. attenuata* genome, but the mechanism remains to be studied.

13.3.4 Evolution of the smRNA Machinery in *N. attenuata* and Other Solanaceae Genomes

The highly variable numbers of genes associated with the smRNA pathways in *N. attenuata* indicate that they may have evolved through numerous duplication and loss events that were affected mainly by the WGT event shared between *N. attenuata* and other Solanaceae genomes. The dispersed chromosomal distribution of the smRNA machinery gene homologs is consistent with this interpretation (Fig. 13.4).

To infer the phylogeny and evolutionary history of the smRNA pathway genes, we constructed phylogenetic trees for each gene family using the homologous genes from *A. thaliana*, *Chlamydomonas*, and *Physcomitrella* as outgroups. To evaluate the history of expansion of these gene families, each phylogenetic tree was reconciled with the species tree using NOTUNG (Chen et al. 2000). Figure 13.5 depicts the evolution of AGOs, DCLs, and RdRs, the families involved in defense against herbivores.

The phylogenetic analysis suggested the presence of four AGO clades in the solanaceous genomes, which is one more than has been

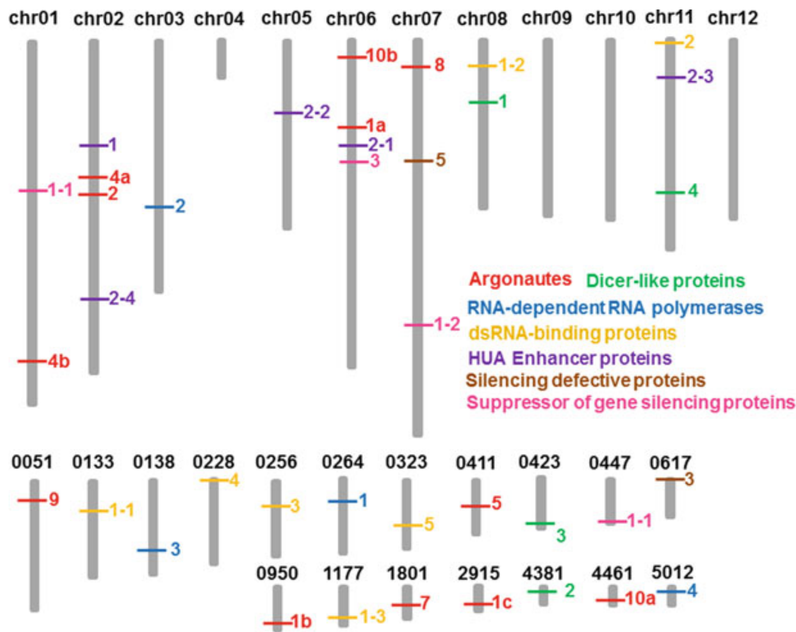

Fig. 13.4 Distribution of the small RNA machinery genes in the *Nicotiana attenuata* genome. Bars in the upper panel represent the 12 chromosomes (chrs) of *N. attenuata*. Genes in the seven families that mainly modulate smRNA biogenesis and action are fairly uniformly distributed across the chromosomal landscape. Chromosomes 4, 9, 10, and 12 are not yet known to contain any of these genes; however, several scaffolds, not yet assigned to any chromosomes (lower panel), contain one or more of the genes from the seven families

detected for other flowering plants (Zhang et al. 2015). The expansion history revealed that gene duplication events could have taken place largely in the parent genomes through whole-genome duplications/triplications before the divergence into respective species (Fig. 13.5). The common ancestor of AGOs in the solanaceous genomes may have undergone four duplications and one loss event before the bifurcation into *Nicotiana* and *Solanum*. Furthermore, the ancestors of the *Nicotiana* and *Solanum* genera may have undergone five and four AGO duplication events, respectively, before the divergence into species. In parallel, both genera experienced four and one loss events, respectively. After speciation, the genomes of each of the species experienced variable numbers of gene losses. A maximum of five loss events could have occurred in the *N. attenuata* and *S. tuberosum* genomes, only two losses occurred in the *C. annuum* genome, and one loss event occurred in *N. obtusifolia* and *S. lycopersicum*. *N. attenuata*, *N. obtusifolia*, and *N. benthamiana* lost the *SlAGO3* ortholog, while the genomes of *S. tuberosum* and *N. benthamiana* appear to have lost the *SlAGO6* ortholog.

DCL and RdR gene numbers reflect the WGT. DCL genes had four gene copies in the Solanaceae ancestor (Fig. 13.5). This common ancestor of the solanaceous genomes underwent one duplication event before bifurcation into *Nicotiana* and *Solanum* lineages. After this bifurcation, the ancestral *Nicotiana* genome may have undergone two loss events before speciation, whereas the ancestral *Solanum* genome may have undergone one duplication and one loss event before speciation. One of the DCL genes in *S. lycopersicum* duplicated tandemly twice (Fig. 13.5). The *S. tuberosum* and *C. annuum* genomes may have experienced three loss events each, and the *N. obtusifolia* genome experienced one loss event. As a result of these events, only *N. attenuata*, *N. benthamiana*, and *S. lycopersicum* still contain the *AtDCL3* ortholog. Further, *S. tuberosum* appears to have lost the *AtDCL4* ortholog.

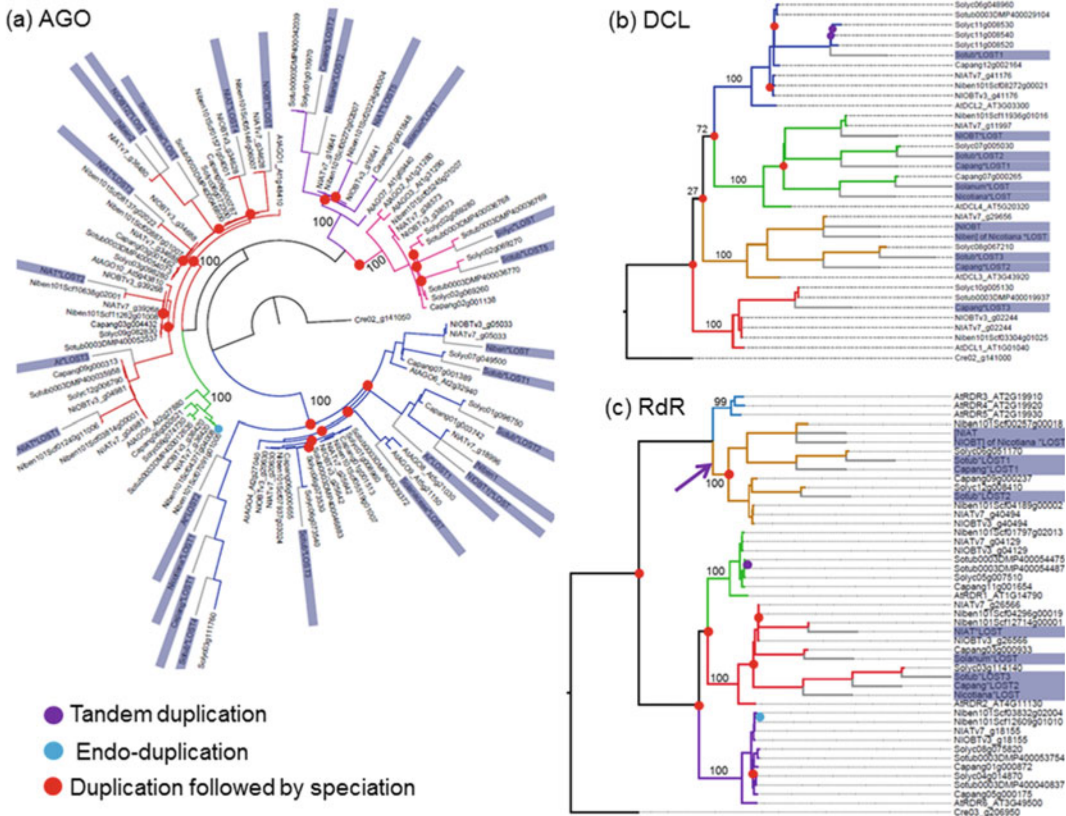

Fig. 13.5 Evolution of the main gene families (*AGO*, *DCL*, and *RdR*) that compose the smRNA machinery across solanaceous genomes. Phylogenetic relationships of **a** Argonautes (AGOs), **b** Dicer-like proteins (DCLs), and **c** RNA-dependent RNA polymerases (RdRs) among six solanaceous species reconstructed from maximum-likelihood methods implemented in RAxML (Stamatakis 2014). The phylogenies indicate that after duplication in four distinct phylogenetic clades followed by speciation, the AGO family expanded more than the DCL and RdR families. The shaded homolog names at the terminal nodes indicate the loss of a homolog in the respective species. The speciation events at the various nodes were calculated after reconciling these trees with the trees containing the following species: *Capsicum annuum*, *Solanum tuberosum*, *S. lycopersicum*, *Nicotiana attenuata*, *N. obtusifolia*, *N. benthamiana*, *Arabidopsis thaliana*, and *Chlamydomonas reinhardtii* using NOTUNG (Chen et al. 2000). The tandem and endo-duplication events were estimated using the method described in Xu et al. (2017). Homologs from *C. reinhardtii* were used as an outgroup to root the tree. The bootstrap values on the branches reflect their statistical support

RdR genes also underwent most of their duplications in the common ancestor of the solanaceous genome well before the speciation and bifurcation into *Nicotiana* and *Solanum*. After the bifurcation, the ancestor of the *Nicotiana* genus probably underwent one duplication and two loss events before speciation (Fig. 13.5). The ancestor of the *Solanum* genomes appears to have undergone one duplication and one loss event before the divergence into species (Fig. 13.5). The *AtRdR1* ortholog appears to have been tandemly duplicated in the *S. tuberosum* genome (Fig. 13.5). Three, two, and one gene-loss events may also have occurred in the genomes of *S. tuberosum*, *C. annuum*, and *N. attenuata*, respectively (Fig. 13.5). Overall, the phylogenetic patterns suggest the presence of *AtRdR1* orthologs in all six solanaceous species; *S. tuberosum* appears to have lost the *AtRdR2* ortholog.

For all other smRNA machinery gene families depicted in Table 13.1, WGT was the main event

leading to duplication, with numerous differential gene losses, as is expected for gene duplicates in plants (Rensing 2014). *Nicotiana* species appear to have retained more of these duplicated genes than other Solanaceae species, at least for three of the smRNA families: DCLs, *HEN2*, and *SGS1* (Table 13.1). This may be related to the predilection of plants in this genus to form polyploids, which could be the ancestors of the present-day genomes, and may also reflect their functional importance.

Overall, these patterns indicate that complex evolutionary scenarios have occurred in solanaceous genomes, shaping the regulatory non-coding RNAomes of these species, a complexity that highlights the importance of studies to evaluate functional specificity in mediating environmental responses.

13.3.5 Perspectives

N. attenuata can grow rapidly post-fire in the Great Basin Desert in North America. The post-fire environment, rich in nitrogen and phosphorus and with few competitors, resembles the monocultures of the slash-and-burn farming of subsistence agriculture. In these man-made agricultural systems, herbivores pose a great threat to crop yields. By identifying the genes responsible for adaptations to this primordial agricultural niche in the native tobacco (*N. attenuata*) that evolved in this habitat, it may be possible to pinpoint traits that allow plants not evolved in these habitats (i.e., all of the major crop plants) to function more efficiently and with fewer agricultural inputs when grown in intensive monocultures.

The recent availability of the *N. attenuata* genome sequence enables the identification of genes involved in ecologically relevant traits using forward-genetic approaches based on the natural variation among individuals collected from native populations (Li et al. 2015). Natural variation can also be used to identify the function of these genes (Tonsor et al. 2005). Baldwin and coworkers have collected seeds over the past three decades from more than 250 native populations (accessions) of *N. attenuata* throughout the southwestern United States. These native accessions have been used to enhance the forward-genetic tools available for the species. In addition to the available bi-parental AI-RIL population (Xu et al. 2017), a multi-parent advanced generation intercross (MAGIC) population is currently being created from 26 of the 250 accessions, which were selected to capture the greatest diversity in ecologically interesting traits. The hope is that this MAGIC population would capture the genetic variation present in the species, shown at the metabolic level (Li et al. 2015). The newly assembled genome will have a central role in the quantitative trait locus mapping of traits observed in this MAGIC population.

The *N. attenuata* genome sequence will aid in understanding the genetic basis and the role of regulatory elements that mediate the ecological adaptations of this native tobacco. The ecological function of the smRNA machinery will require a major effort to elucidate. The genome encodes for stimulus-dependent specificity in the recruitment of smRNAs as well as in the machinery that generates the smRNAs (Fig. 13.6). For instance, herbivore attack recruits *AGO5*, *AGO8*, *DCL3/4*, and *RdR1* to produce specific herbivory-induced smRNAs. Similarly, attack by pathogens recruits a machinery comprised of *RdR1/2*, *DCL3*, and *AGO4*, while stress from ultraviolet B rays is modulated by *RdR2*. *AGO7* may be involved in plant competition or in interactions with arbuscular mycorrhizal fungi, whereas *AGO9* is specifically expressed in flowers and may mediate flower-specific functions.

The availability of the *N. attenuata* genome sequence has opened opportunities to integrate functional genetics with ecological field studies within the genus *Nicotiana*. Plants in this genus occupy various habitats and have evolved specific adaptations, differences that will facilitate the testing of hypotheses related to their genetic basis. For example, the genus harbors a diversity of flower colors, shapes, and chemical constituents (McCarthy et al. 2015) that may be important for its interactions with pollinators. Furthermore, TEs may shape ecological

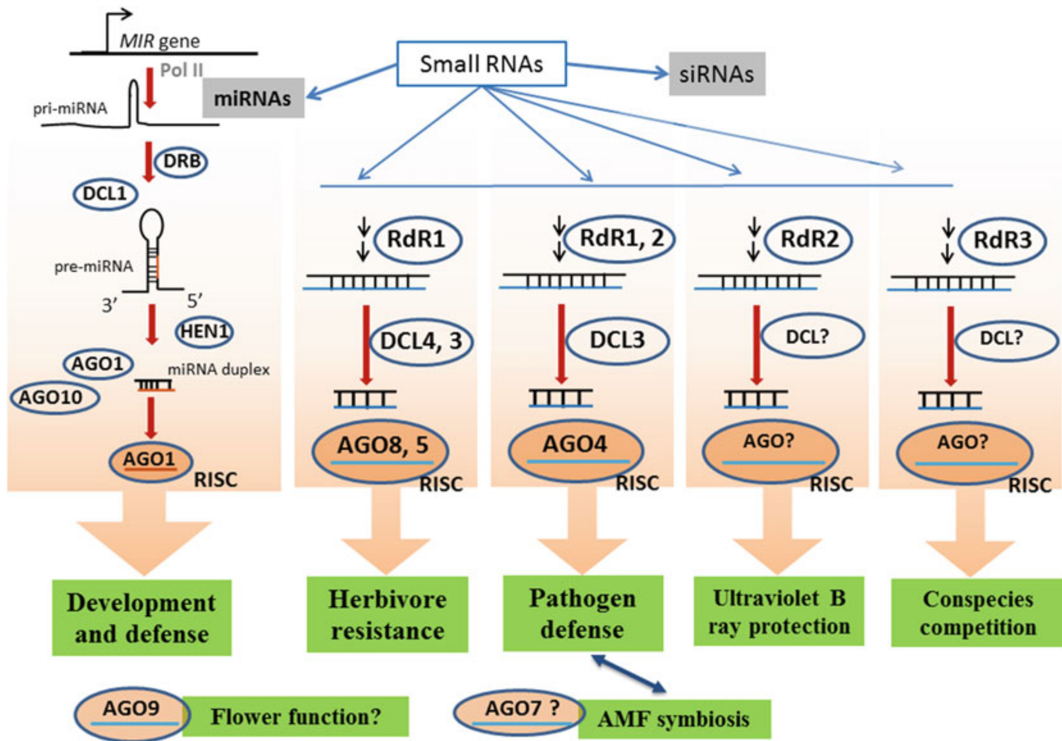

Fig. 13.6 Specialization of the smRNA machinery in *Nicotiana attenuata*. Various combinations of Argonaute (AGO), Dicer-like protein (DCL), and RNA-dependent RNA polymerase (RdR) genes define the functional specificity of molecular responses to various ecological stresses. The remarkable phenotypic plasticity of *Nicotiana attenuata* may be an evolutionary outcome of the challenges the plant faces in its natural habitat, the post-fire habitat where selection pressures from unpredictable populations of herbivores, pathogens, and competitors are particularly important. AMF, arbuscular mycorrhizal fungi; RISC, RNA-induced silencing complex

adaptations via numerous genetic mechanisms, including by modifying smRNAs (Kuang et al. 2009). The tests of these hypotheses will be facilitated by the availability of a high-quality and well-annotated genome sequence.

Acknowledgements We thank Dr. Emmanuel Gaquerel for useful comments on the text, Dr. Shuqing Xu for providing Fig. 13.1, Thomas Brockmoller for providing a raw version of Fig. 13.2, and Thomas Hahn for IT support.

References

Abrouk M, Zhang R, Murat F, Li A, Pont C, Mao L, Salse J (2012) Grass MicroRNA gene paleohistory unveils new insights into gene dosage balance in subgenome partitioning after whole-genome duplication. Plant Cell 24:1776–1792

Allmann S, Baldwin IT (2010) Insects betray themselves in nature to predators by rapid isomerization of green leaf volatiles. Science 329:1075–10758

Anssour S, Baldwin IT (2010) Variation in antiherbivore defense responses in synthetic *Nicotiana* allopolyploids correlates with changes in uniparental patterns of gene expression. Plant Physiol 153:1907–19018

Baldwin IT (2001) An ecologically motivated analysis of plant-herbivore interactions in native tobacco. Plant Physiol 127:1449–1458

Bernstein KA, Shor E, Sunjevaric I, Fumasoni M, Burgess RC, Foiani M, Branzei D, Rothstein R (2009) Sgs1 function in the repair of DNA replication intermediates is separable from its role in homologous recombinational repair. EMBO J 28:915–925

Blevins T, Rajeswaran R, Shivaprasad PV, Beknazariants D, Si-Ammour A, Park HS, Vazquez F, Robertson D, Meins F Jr, Hohn T, Pooggin MM (2006) Four plant Dicers mediate viral small RNA biogenesis and DNA virus induced silencing. Nucleic Acids Res 34:6233–6246

Borges F, Martienssen RA (2015) The expanding world of small RNAs in plants. Nat Rev Mol Cell Biol 16:727–741

Bozorov TA, Baldwin IT, Kim S-G (2012a) Identification and profiling of miRNAs during herbivory reveals jasmonate-dependent and-independent patterns of accumulation in *Nicotiana attenuata*. BMC Plant Biol 12:209

Bozorov TA, Pandey SP, Dinh ST, Kim SG, Heinrich M, Gase K, Baldwin IT (2012b) DICER-like proteins and their role in plant-herbivore interactions in *Nicotiana attenuata*. J Integr Plant Biol 54

Brockmoller T, Ling Z, Li D, Gaquerel E, Baldwin IT, Xu S (2017) *Nicotiana attenuata* Data Hub (NaDH): an integrative platform for exploring genomic, transcriptomic and metabolomic data in wild tobacco. BMC Genom 18:79

Brodersen P, Voinnet O (2006) The diversity of RNA silencing pathways in plants. Trends Genet 22:268–280

Cantarel BL, Korf I, Robb SM, Parra G, Ross E, Moore B, Holt C, Sanchez Alvarado A, Yandell M (2008) MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res 18

Chapman EJ, Carrington JC (2007) Specialization and evolution of endogenous small RNA pathways. Nat Rev Genet 8:884–896

Chen K, Durand D, Farach-Colton M (2000) NOTUNG: a program for dating gene duplications and optimizing gene family trees. J Comput Biol 7

Chen X, Liu J, Cheng Y, Jia D (2002) HEN1 functions pleiotropically in *Arabidopsis* development and acts in C function in the flower. Development 129:1085–1094

Clarkson JJ, Kelly LJ, Leitch AR, Knapp S, Chase MW (2010) Nuclear glutamine synthetase evolution in *Nicotiana*: phylogenetics and the origins of allotetraploid and homoploid (diploid) hybrids. Mol Phylogenet Evol 55:99–112

Dalmay T, Horsefield R, Braunstein TH, Baulcombe DC (2001) SDE3 encodes an RNA helicase required for post-transcriptional gene silencing in *Arabidopsis*. EMBO J 20:2069–2078

Gaquerel E, Heiling S, Schoettner M, Zurek G, Baldwin IT (2010) Development and validation of a liquid chromatography-electrospray ionization-time-of-flight mass spectrometry method for induced changes in *Nicotiana attenuata* leaves during simulated herbivory. J Agric Food Chem 58:9418–9427

Gaquerel E, Stitz M, Kallenbach M, Baldwin IT (2013) Jasmonate Signaling in the Field, Part I: Elicited Changes in Jasmonate Pools of Transgenic *Nicotiana attenuata* Populations. In: Goossens A, Pauwels L (eds) Jasmonate signaling. Humana Press, Totowa, NJ, pp 83–95

Gase K, Baldwin IT (2012) Transformational tools for next-generation plant ecology: manipulation of gene expression for the functional analysis of genes. Plant Ecol Divers 5:485–490

Glawe GA, Zavala JA, Kessler A, Van Dam NM, Baldwin IT (2003) Ecological costs and benefits correlated with trypsin protease inhibitor production in *Nicotiana attenuata*. Ecology 84:79–90

Gulati J, Baldwin IT, Gaquerel E (2013a) An integrative statistical method to explore herbivory-specific responses in plants. Plant Signal Behav 8. https://doi.org/10.4161/psb.25638

Gulati J, Kim SG, Baldwin IT, Gaquerel E (2013b) Deciphering herbivory-induced gene-to-metabolite dynamics in *Nicotiana attenuata* tissues using a multifactorial approach. Plant Physiol 162:1042–1059

Gulati J, Baldwin IT, Gaquerel E (2014) The roots of plant defenses: integrative multivariate analyses uncover dynamic behaviors of gene and metabolic networks of roots elicited by leaf herbivory. Plant J 77:880–892

Heiling S, Schuman MC, Schoettner M, Mukerjee P, Berger B, Schneider B, Jassbi AR, Baldwin IT (2010) Jasmonate and ppHsystemin regulate key Malonylation steps in the biosynthesis of 17-Hydroxygeranyllinalool Diterpene Glycosides, an abundant and effective direct defense against herbivores in *Nicotiana attenuata*. Plant Cell 22:273–292

Hernandez-Pinzon I, Yelina NE, Schwach F, Studholme DJ, Baulcombe D, Dalmay T (2007) SDE5, the putative homologue of a human mRNA export factor, is required for transgene silencing and accumulation of trans-acting endogenous siRNA. Plant J 50:140–148

Huang Y, Ji LJ, Huang QC, Vassylyev DG, Chen XM, Ma JB (2009) Structural insights into mechanisms of the small RNA methyltransferase HEN1. Nature 461:823–U86

Hutvagner G, Simard MJ (2008) Argonaute proteins: key players in RNA silencing. Nat Rev Mol Cell Biol 9:22–32

Kahl J, Siemens DH, Aerts RJ, Gabler R, Kuhnemann F, Preston CA, Baldwin IT (2000) Herbivore-induced ethylene suppresses a direct defense but not a putative indirect defense against an adapted herbivore. Planta 210:336–342

Kallenbach M, Bonaventure G, Gilardoni PA, Wissgott A, Baldwin IT (2012) *Empoasca* leafhoppers attack wild tobacco plants in a jasmonate-dependent manner and identify jasmonate mutants in natural populations. Proc Natl Acad Sci USA 109:E1548–E1557

Kessler D, Bhattacharya S, Diezel C, Rothe E, Gase K, Schottner M, Baldwin IT (2012) Unpredictability of nectar nicotine promotes outcrossing by hummingbirds in *Nicotiana attenuata*. Plant J 71:529–38

Kim SG, Yon F, Gaquerel E, Gulati J, Baldwin IT (2011) Tissue specific diurnal rhythms of metabolites and their regulation during herbivore attack in a native tobacco, *Nicotiana attenuata*. PloS One 6:e26214

Kuang H, Padmanabhan C, Li F, Kamei A, Bhaskar PB, Ouyang S, Jiang J, Buell CR, Baker B (2009) Identification of miniature inverted-repeat transposable elements (MITEs) and biogenesis of their siRNAs

in the Solanaceae: new functional implications for MITEs. Genome Res 19:42–56

Kumar P, Pandit SS, Baldwin IT (2012) Tobacco rattle virus vector: A rapid and transient means of silencing *Manduca sexta* genes by plant mediated RNA interference. PLoS ONE 7:e31347

Kurzynska-Kokorniak A, Koralewska N, Pokornowska M, Urbanowicz A, Tworak A, Mickiewicz A, Figlerowicz M (2015) The many faces of Dicer: the complexity of the mechanisms regulating Dicer gene expression and enzyme activities. Nucleic Acids Res 43:4365–4380

Lee E, Helt GA, Reese JT, Munoz-Torres MC, Childers CP, Buels RM, Stein L, Holmes IH, Elsik CG, Lewis SE (2013) Web Apollo: a web-based genomic annotation editing platform. Genome Biol 14:R93

Lesburg CA, Cable MB, Ferrari E, Hong Z, Mannarino AF, Weber PC (1999) Crystal structure of the RNA-dependent RNA polymerase from hepatitis C virus reveals a fully encircled active site. Nat Struct Biol 6:937–943

Li D, Baldwin IT, Gaquerel E (2015) Navigating natural variation in herbivory-induced secondary metabolism in coyote tobacco populations using MS/MS structural analysis. Proc Natl Acad Sci USA 112:E4147–E4155

Lou Y, Baldwin IT (2003) *Manduca sexta* recognition and resistance among allopolyploid *Nicotiana* host plants. Proc Natl Acad Sci USA 100:14581–14586

McCarthy EW, Arnold SE, Chittka L, Le Comber SC, Verity R, Dodsworth S, Knapp S, Kelly LJ, Chase MW, Baldwin IT, Kovarik A, Mhiri C, Taylor L, Leitch AR (2015) The effect of polyploidy and hybridization on the evolution of floral colour in *Nicotiana* (Solanaceae). Ann Bot 115:1117–1131

Mickiewicz A, Sarzynska J, Milostan M, Kurzynska-Kokorniak A, Rybarczyk A, Lukasiak P, Kulinski T, Figlerowicz M, Blazewicz J (2017) Modeling of the catalytic core of *Arabidopsis thaliana* Dicer-like 4 protein and its complex with double-stranded RNA. Comput Biol Chem 66:44–56

Mourrain P, Beclin C, Elmayan T, Feuerbach F, Godon C, Morel JB, Jouette D, Lacombe AM, Nikic S, Picault N, Remoue K, Sanial M, Vo TA, Vaucheret H (2000) Arabidopsis SGS2 and SGS3 genes are required for posttranscriptional gene silencing and natural virus resistance. Cell 101:533–542

Mukherjee K, Campos H, Kolaczkowski B (2013) Evolution of animal and plant dicers: early parallel duplications and recurrent adaptation of antiviral RNA binding in plants. Mol Biol Evol 30:627–641

Onkokesung N, Galis I, von Dahl CC, Matsuoka K, Saluz HP, Baldwin IT (2010) Jasmonic acid and ethylene modulate local responses to wounding and simulated herbivory in *Nicotiana attenuata* leaves. Plant Physiol 153:785–798

Pandey SP, Baldwin IT (2007) RNA-directed RNA polymerase 1 (RdR1) mediates the resistance of *Nicotiana attenuata* to herbivore attack in nature. Plant J 50:40–53

Pandey SP, Baldwin IT (2008) Silencing RNA-directed RNA polymerase 2 increases the susceptibility of *Nicotiana attenuata* to UV in the field and in the glasshouse. Plant J 54:845–862

Pandey SP, Shahi P, Gase K, Baldwin IT (2008) Herbivory-induced changes in the small-RNA transcriptome and phytohormone signaling in *Nicotiana attenuata*. Proc Natl Acad Sci USA 105:4559–4564

Pandey P, Wang M, Baldwin IT, Pandey SP, Groten K (2018) Complex regulation of microRNAs in roots of competitively-grown isogenic *Nicotiana attenuata* plants with different capacities to interact with arbuscular mycorrhizal fungi. BMC Genom 19:937

Persson S, Wei H, Milne J, Page GP, Somerville CR (2005) Identification of genes required for cellulose synthesis by regression analysis of public microarray data sets. Proc Natl Acad Sci USA 102:8633–8638

Pradhan M, Pandey P, Gase K, Sharaff M, Singh RK, Sethi A, Baldwin IT, Pandey SP (2017) Argonaute 8 (AGO8) Mediates the elicitation of direct defenses against herbivory. Plant Physiol 175:927–946

Rensing SA (2014) Gene duplication as a driver of plant morphogenetic evolution. Curr Opin Plant Biol 17:43–48

Schwachtje J, Baldwin IT (2008) Why does herbivore attack reconfigure primary metabolism? Plant Physiol 146:845–851

Self LS, Guthrie FE, Hodgson E (1964) Metabolism of nicotine by tobacco-feeding insects. Nature 204:300–301

Singh RK, Gase K, Baldwin IT, Pandey SP (2015) Molecular evolution and diversification of the Argonaute family of proteins in plants. BMC Plant Biol 15:23

Skibbe M, Qu N, Galis I, Baldwin IT (2008) Induced plant defenses in the natural environment: *Nicotiana attenuata* WRKY3 and WRKY6 coordinate responses to herbivory. Plant Cell 20:1984–2000

Song JJ, Smith SK, Hannon GJ, Joshua-Tor L (2004) Crystal structure of Argonaute and its implications for RISC slicer activity. Science 305

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313

Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B (2006) AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic Acids Res 34: W435–W439

Stanton MA, Prebetaler J, Paetz C, Boland W, Svatos A, Baldwin IT (2016) Plant-mediated pheromone emission by a hemipteran seed feeder increases the apparency of an unreliable but rewarding host. New Phytol 211:113–125

Steppuhn A, Gase K, Krock B, Halitschke R, Baldwin IT (2004) Nicotine's defensive function in nature. PLoS Biol 2:E217

Steppuhn A, Schuman MC, Baldwin IT (2008) Silencing jasmonate signalling and jasmonate-mediated defences reveals different survival strategies between two *Nicotiana attenuata* accessions. Mol Ecol 17:3717–3732

Stuart JM, Segal E, Koller D, Kim SK (2003) A gene-coexpression network for global discovery of conserved genetic modules. Science 302:249–255

Tang H (2009) Multiple collinearity scan—mcscan. In, 1.1 edn. Plant Genome Mapping Lab, University of Georgia, pp MCscan is a computer program that can simultaneously scan multiple genomes to identify homologous chromosomal regions and subsequently align these regions using genes as anchors. This is the toolset for generating the synteny correspondences in Plant Genome Duplication Database. It is intended as an easy-to-use and quick way to identify conserved gene arrays both within the same genome and across different genomes

Tonsor SJ, Alonso-Blanco C, Koornneef M (2005) Gene function beyond the single trait: natural variation, gene effects, and evolutionary ecology in *Arabidopsis thaliana*. Plant Cell Environ 28:2–20

Voelckel C, Schittko U, Baldwin IT (2001) Herbivore-induced ethylene burst reduces fitness costs of jasmonate- and oral secretion-induced defenses in *Nicotiana attenuata*. Oecologia 127:274–280

Willmann MR, Endres MW, Cook RT, Gregory BD (2011) The functions of RNA-dependent RNA polymerases in *Arabidopsis*. Arab Book 9:e0146

Wink M, Theile V (2002) Alkaloid tolerance in *Manduca sexta* and phylogenetically related sphingids (Lepidoptera: Sphingidae). Chemoecology 12:29–46

Wu J, Baldwin IT (2009) Herbivory-induced signalling in plants: perception and action. Plant Cell Environ 32:1161–1174

Wu J, Baldwin IT (2010) New insights into plant responses to the attack from insect herbivores. Annu Rev Genet 44:1–24

Wu J, Hettenhausen C, Baldwin IT (2006) Evolution of proteinase inhibitor defenses in North American allopolyploid species of *Nicotiana*. Planta 224:750–760

Wu J, Hettenhausen C, Meldau S, Baldwin IT (2007) Herbivory rapidly activates MAPK signaling in attacked and unattacked leaf regions but not between leaves of *Nicotiana attenuata*. Plant Cell 19:1096–1122

Wu J, Hettenhausen C, Schuman MC, Baldwin IT (2008) A comparison of two *Nicotiana attenuata* accessions reveals large differences in signaling induced by oral secretions of the specialist herbivore *Manduca sexta*. Plant Physiol 146:927–939

Xu S, Zhou W, Pottinger S, Baldwin IT (2015) Herbivore associated elicitor-induced defences are highly specific among closely related *Nicotiana* species. BMC Plant Biol 15:2

Xu S, Brockmoller T, Navarro-Quezada A, Kuhl H, Gase K, Ling Z, Zhou W, Kreitzer C, Stanke M, Tang H, Lyons E, Pandey P, Pandey SP, Timmermann B, Gaquerel E, Baldwin IT (2017) Wild tobacco genomes reveal the evolution of nicotine biosynthesis. Proc Natl Acad Sci USA 114:6133–6138

Yang DH, Hettenhausen C, Baldwin IT, Wu J (2012) Silencing *Nicotiana attenuata* calcium-dependent protein kinases, CDPK4 and CDPK5, strongly up-regulates wound- and herbivory-induced jasmonic acid accumulations. Plant Physiol 159:1591–1607

Zhang H, Xia R, Meyers BC, Walbot V (2015) Evolution, functions, and mysteries of plant ARGONAUTE proteins. Curr Opin Plant Biol 27:84–90

Zhou W, Brockmoller T, Ling Z, Omdahl A, Baldwin IT, Xu S (2016) Evolution of herbivore-induced early defense signaling was shaped by genome-wide duplications in *Nicotiana*. eLife 5:e19531

Zhou W, Kugler A, McGale E, Haverkamp A, Knaden M, Guo H, Beran F, Yon F, Li R, Lackus N, Kollner TG, Bing J, Schuman MC, Hansson BS, Kessler D, Baldwin IT, Xu S (2017) Tissue-specific emission of (E)-alpha-bergamotene helps resolve the dilemma when pollinators are also herbivores. Curr Biol 27:1336–1341

Check for updates

14 Nicotiana benthamiana, A Popular Model for Genome Evolution and Plant–Pathogen Interactions

Marina A. Pombo, Hernan G. Rosli, Noe Fernandez-Pozo and Aureliano Bombarely

Abstract

Nicotiana benthamiana originates from northern Australia and belongs to the Suaveolentes section. It is used extensively as a model organism for many types of research, including plant–pathogen interactions, RNA interference, and functional genomics. Recent publications that used *N. benthamiana* as a model for plant–pathogen interactions focused mainly on bacteria, viruses, oomycete, and fungi. Two different *N. benthamiana* whole genome assemblies were published in 2012. These assemblies have been improved and structurally annotated in later versions but are still incomplete. The lineage most widely used in research originates from a population that has retained a loss-of-function mutation in *Rdr1* (RNA-dependent RNA polymerase 1) that makes it highly susceptible to viruses. In this chapter, we review some of the techniques used in *N. benthamiana* to study plant–pathogen interactions, including virus-induced gene silencing, transient protein expression by agroinfiltration, stable genetic manipulation, and transcriptomics analysis, and discuss some of the results. Descriptions and links to some of the most relevant online resources for *N. benthamiana* are also provided.

14.1 Evolutionary History of *Nicotiana benthamiana*

N. benthamiana Domin is among the most popular plants used for plant pathology studies. Despite this, specific details about the origin of this species are still unknown. *N. benthamiana* is a herbaceous plant with white flowers that is native to Australia where it can be found along the north coast, the Northern Territory, and Queensland (Fig. 14.1) (Global Biodiversity Information Facility (GBIF) 2018).

It can be identified following the dichotomous key described by Burbidge. *N. benthamiana* is distinguished from *N. umbratica* specimens by: “Upper cauline leaves sessile and forming leafy

M. A. Pombo · H. G. Rosli
Instituto de Fisiología Vegetal, INFIVE, Universidad Nacional de La Plata, CONICET, La Plata, Buenos Aires, Argentina

N. Fernandez-Pozo
Faculty of Biology, University of Marburg, Karl-Von-Frisch-Str. 8, 35043 Marburg, Germany

A. Bombarely (✉)
School of Plant and Environmental Sciences, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA
e-mail: aureliano.bombarely@unimi.it

A. Bombarely
Dipartimento Di Bioscienze, Universita Degli Studi Di Milano, Milan, Italy

© Springer Nature Switzerland AG 2020

N. V. Ivanov et al. (eds.), *The Tobacco Plant Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-030-29493-9_14

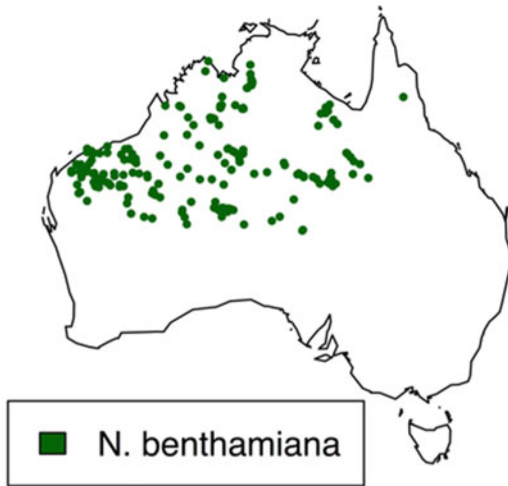

Fig. 14.1 Natural distribution of *N. benthamiana* in Australia

bracts. Laminae ovate to broad lanceolate, obtuse. Corolla lobes obtuse compared with the corolla lobes acute of the last one" (Burbidge 1960). The first available record of *N. benthamiana* can be found in the Herbarium Hookerianum at Kew Royal Botanical Gardens (http://specimens.kew.org/herbarium/K000196107), with a specimen collected by Benjamin Bynoe during the voyage of the Beagle (1837–1843) (Orchard 1999). Like other Australian *Nicotiana* species, *N. benthamiana* belongs to the Suaveolentes section, which contains 30 species and represents an important radiation of the *Nicotiana* genus in Australia (26 species), South Pacific (3 species), and Africa (1 species) (Marks et al. 2011). They are adapted to a wide range of conditions, from the high humidity coastal regions of Cairns (e.g., *N. debneyi*) to the extremely dry Great Sandy and Tanami deserts (e.g., *N. benthamiana*). All the Suaveolentes species have been described as allotetraploids with a variable number of chromosomes ranging from 2n = 4 × = 30 (*N. suaveolens*) to 2n = 4 × = 48 (*N. debneyi*).

Elucidation of the origin of the Suaveolentes section is still ongoing. An early publication by DeWolf and Goodspeed (1957) proposed that it could have formed after the hybridization of members of the Petunioides, Alatae/Sylvestres, and/or Noctiflorae sections. Later phylogenetic studies using plastid gene sequences such as *matK* indicated a Noctiflorae species as maternal donor (Aoki and Ito 2000; Clarkson et al. 2004). The use of interspacers ITS1 and ITS2, and 5.8S ribosomal genes indicated that an ancestral member of Alatae could be the possible paternal progenitor (Chase et al. 2003). A study using the *GS* (glutamine synthetase) gene, with both copies usually retained in polyploid *Nicotiana* species, indicated Sylvestres as the maternal donor and Trigonophyllae as the paternal donor (Clarkson et al. 2010). Nevertheless, the most accepted hypothesis based on several genes, such as *ADH* (alcohol dehydrogenase), *LFY/FLO* (LEAFY/FLORICAULA), *GS*, and nrITS (nuclear ribosomal interspacer), suggests a complex history where the most probable donors are Sylvestres (paternal) and Noctiflorae/Petunioides (maternal) (Kelly et al. 2013). The divergency ages between the Suaveolentes subgenomes and the corresponding proposed progenitors were estimated as 6.4 million years ago for the maternal contribution from the Noctiflorae section and 5.5 million years ago for the paternal contribution from the Sylvestres section (Clarkson et al. 2017).

14.2 *Nicotiana benthamiana* Genome Assemblies and Genetic Data

N. benthamiana was one of the first plant models positively affected by next-generation sequencing (NGS) technologies. Although several plant whole genome assemblies were built using NGS before 2012, most of them were for crops such as cucumber (Huang et al. 2009), apple (Velasco et al. 2010), and soybean (Schmutz et al. 2010). Two different *N. benthamiana* whole genome assemblies were published in 2012 (Bombarely et al. 2012; Naim et al. 2012). *N. benthamiana* is an allotetraploid plant (2n = 4 × = 38) with a large genome of 3.1 Gb, which made it difficult to assemble using the short reads obtained by NGS. Although both genomes were incomplete,

Table 14.1 Details of the two *N. benthamiana* 2012 genome assemblies

Assembly	Niben0.4.2[a]		Nbv0.3[b]	
Assembly statistics	Contigs	Scaffolds	Contigs	Scaffolds
Total assembly size (Gb)	2.46	2.63	2.44	2.44
Total assembled sequences	461,463	141,339	300,384	275,036
Longest sequence length (Kb)	208.21	615.59	307.11	447.13
Average sequence length (Kb)	5.37	18.61	8.12	8.88
N50 index (sequences)	42,984	8,897	22,438	22,068
N50 length (Kb)	16.48	89.78	31.25	31.83

[a]Bombarely et al. 2012; [b]Naim et al. 2012

they have been used extensively for plant–pathogen research. Statistics of the two *N. benthamiana* genome assemblies are summarized in Table 14.1.

Although a chromosome level assembly has still not been achieved (as October 2018), these assemblies have been improved. Statistics of the latest versions are summarized in Table 14.2. Evaluation of these genome assemblies using BUSCO (Simao et al. 2015) indicated they had similar completeness (95.4% for Niben1.0.1 and 94.4% for Nbv0.5), but Niben1.0.1 had a higher proportion of duplicated genes (46.0%) than Nbv0.5 (43.4%). These assemblies were also structurally annotated revealing 59,814 and 49,818 primary transcripts for Niben1.0.1 and Nbv0.5, respectively. Because of the ongoing diploidization process in *N. benthamiana,* its gene number is higher than the gene numbers for diploid Solanaceae species such as *Solanum lycopersicum* (Tomato Genome Consortium 2012), *Solanum tuberosum* (Potato Genome Sequencing Consortium 2011), *Capsicum annuum* (Kim et al. 2014), *N. sylvestris* and *N. tomentosiformis* (Sierro 2013), and *Petunia axillaris* (Bombarely et al. 2016), but lower than the gene number for allotetraploid *N. tabacum* (Sierro et al. 2014; Edwards et al. 2017). As expected, the BUSCO evaluation of the completeness of the annotations produced a lower value (88.1% for Niben1.0.1 and 75.2% for the transcriptome Nbv5.1 from the assembly Nbv0.5) than the evaluation of the genome assembly. Kourelis et al. (2018) used the genomes of other *Nicotiana* species to reanalyze the four genomes (Niben1.0.1, Niben0.4.2, Nbv0.5, and Nbv0.3) improving the quality of the *N. benthamiana* gene annotations.

To date, dozens of experiments using NGS technologies to investigate Solanaceae species have been published by the National Center for Biotechnology Information (NCBI). Indeed, the NCBI's Taxonomy Browser has links to 342 SRA datasets, 442 BioSamples, and 74 BioProjects (up

Table 14.2 Details of the latest publicly available *N. benthamiana* genome assemblies (up to October 2018)

Assembly	Niben1.0.1[a]		Nbv0.5[b]	
Assembly statistics	Contigs	Scaffolds	Contigs	Scaffolds
Total assembly size (Gb)	2.49	2.97	2.51	2.55
Total assembled sequences	288,736	56,094	180,357	147,949
Longest sequence length (Kb)	187.66	2,838.18	965.20	2,628.40
Average sequence length (Kb)	8.62	52.94	13.92	17.23
N50 index (sequences)	37,001	1,738	7,104	1,936
N50 length (Kb)	20.13	520.10	105.31	396.15

[a]Niben1.0.1 is available on the Sol Genomics Network (https://solgenomics.net/organism/Nicotiana_benthamiana/genome); [b]Nbv0.5 is available from Queensland University of Technology (http://benthgenome.qut.edu.au/)

to December 2018) for *N. benthamiana*. These BioProjects not only study plant–pathogen interaction using resistance gene enrichment sequencing (RenSeq) (e.g., BioProject accession PRJNA496490) and RNA sequencing (RNA-Seq) (e.g., PRJNA360110), but also study expression regulatory mechanisms involving small RNAs (e.g., PRJNA481240, PRJNA309389) and circular RNAs (e.g., PRJNA422356), transcriptomic landscapes in several organs such as nectaries (e.g., PRJNA448133), transition and floral meristems (e.g., PRJNA343677), and grafting experiments (e.g., PRJDB3306).

14.3 *Nicotiana benthamiana* as a Model for the Study of Plant–Pathogen Interactions

The use of *N. benthamiana* as a model for plant–pathogen interactions has been thoroughly reviewed by Goodin et al. 2008; therefore, we will focus mainly on the information that has been generated since then. In their review, Goodin et al. performed a PubMed search using the term "Nicotiana benthamiana" and found 1,743 publications until 2006. We conducted a similar search, which yielded 3,606 hits (as of October 2018). By curating the publications since 2008, we identified 314 papers in which *N. benthamiana* was used as a tool to study different pathosystems, not taking into account studies in which *N. benthamiana* was used merely for transient protein expression and was unrelated to plant immunity. Among the 314 publications, the focus of study included plant interactions with bacterium (Senthil-Kumar and Mysore 2010; Kim et al. 2016), virus (Pavli et al. 2011; Zhu et al. 2014), oomycete (Adachi et al. 2015; King et al. 2014), fungus (De Jonge et al. 2012; Li et al. 2015), nematode (Mantelin et al. 2011; Ali et al. 2015), aphid (Peng et al. 2016; Atamian et al. 2013), insect (Chen et al. 2014), and viroid (Adkar-Purushothama et al. 2015). Among these pathogen types, the first four were the most widely studied using *N. benthamiana* (Fig. 14.2).

N. benthamiana was adopted as a model in the plant–virus research field because of its remarkable susceptibility (Goodin et al. 2008). In particular, the lineage employed at that time, which continues to be widely used in research, has a disruptive insertion in the gene coding RNA-dependent RNA polymerase I (*Rdr1*), which enhances plant fitness but simultaneously leads to its high susceptibility to viruses (Balli et al. 2015). Other reasons why *N. benthamiana* is so extensively used include high efficiency of

Fig. 14.2 Pathogen types studied in publications involving *N. benthamiana* (2009 to October 2018). The input for WordSift (https://wordsift.org) was derived from the information from 314 publications in which *N. benthamiana* was employed in studies of plant–pathogen interaction published after 2008. Word font size indicates frequency of use

gene silencing, ease of protein transient expression using *Agrobacterium tumefaciens*, the possibility of stable genetic manipulation (Todesco and De Felippes 2016), and the availability of a draft genome (Bombarely et al. 2012; Naim et al. 2012). In addition, the ongoing development of NGS techniques may further enhance the use of *N. benthamiana* to study plant–pathogen interactions. In the following sections, we describe and discuss each of the uses and techniques mentioned in this section.

14.4 Virus-Induced Gene Silencing (VIGS)

This technique relies on the silencing machinery of plants in order to be able to target the transcripts of genes of interest (Burch-Smith et al. 2004). Tobacco rattle virus (TRV)-based vectors are commonly used in numerous plant species, particularly those belonging to the Solanaceae family (Senthil-Kumar and Mysore 2014). VIGS technology offers several advantages such as rapid reverse genetic screens, mainly because the time-consuming process of plant transformation is avoided (Velasquez et al. 2009). VIGS also allows the simultaneous knockdown of multiple genes by selecting a single fragment with enough homology to the target genes or by arranging several fragments in a single construct (Miki et al. 2005; Zhou and Zeng 2017). An important general aspect of VIGS is the choice of a control against which the target gene's silencing performance can be compared. An empty vector (TRV2::00) is not recommended as a control, rather TRV2 carrying an insert is preferred (Hartl et al. 2008; Wu et al. 2011). Inserts derived from green fluorescent protein (*GFP*) (Ryu et al. 2004), β-glucuronidase (Gonorazky et al. 2014), or an *Escherichia coli* gene (*Ec1*) (Rosli et al. 2013) have been used previously. Another key aspect of VIGS is the selection of an insert that will effectively target the gene(s) of interest while avoiding off-targets. The Sol Genomics Network (SGN) VIGS Tool allows the interactive identification of most probable targets and off-targets, thereby assisting in construct design (Fernandez-Pozo et al. 2015a, b) (see the section "Online resources for *Nicotiana benthamiana*" in this chapter for more details). In *N. benthamiana*, VIGS has been used to silence some target genes (Kang et al. 2010; Choi et al. 2011; Liebrand et al. 2012 Rosli et al. 2013; Pombo et al. 2014), as well as large sets of candidate genes in a high-throughput fashion (Chakravarthy et al. 2010; Zhu et al. 2010; Mantelin et al. 2011; Rojas et al. 2012; Senthil-Kumar and Mysore 2012; Xu et al. 2012; Du et al. 2013; Lee et al. 2013; Nakano et al. 2013). The silencing step is followed by a readout experiment that depends on the particular process under study and may involve transient protein expression and elicitation of programmed cell death (PCD) (see below), pathogen challenge (Asai et al. 2008; Tanaka et al. 2009; Senthil-Kumar and Mysore 2010; Chaparro-Garcia et al. 2011; Kiba et al. 2012; Du et al. 2013; Ohtsu et al. 2014; Adachi et al. 2015; Bruckner et al. 2017; Turnbull et al. 2017), reactive oxygen species (ROS) production (Shibata et al. 2010; Segonzac et al. 2011; Deng et al. 2016; Pfeilmeier et al. 2016; Saur et al. 2016), nitric oxide production (Zhang et al. 2010), and stomatal aperture measurement (Zhang et al. 2012, 2016). Although *N. benthamiana* is susceptible to many pathogens, it may not be a host to the pathogen under study. This has been overcome by engineering the pathogen (Wei et al. 2007) or using a related species that causes disease in *N. benthamiana* (Yu et al. 2012; Yin et al. 2013; Wang et al. 2016). Chakravarthy et al. (2010) developed an assay that can be used to test the effect of silencing a candidate gene on the pattern-triggered immunity (PTI) response. This requires the infiltration of a PTI inducer (*Pseudomonas fluorescens*) and, a few hours later, a second infiltration performed in an overlapping manner with a challenger (*Pseudomonas syringae* pv. *tomato*, Pst). The speed of PCD progression in the overlapping area is related to the functionality of PTI. The authors coupled this assay with VIGS high-throughput screening and identified genes involved in the PTI response.

Recently, Zhou and Zeng 2017 developed a novel VIGS strategy to specifically and

efficiently knockdown members of a highly homologous gene family using fragments of approximately 70 base pairs. The authors combined the SGN VIGS Tool (Fernandez-Pozo et al. 2015a, b) with a manual optimization step to select the fragments in order to analyze functional redundancy among members of a gene family.

14.5 Transient Protein Expression by Agroinfiltration

This technique uses *Agrobacterium tumefaciens* carrying an expression vector system. Usually a suspension is infiltrated into a leaf and within 1–3 days the tissue is ready for downstream analysis or treatment. The most commonly used vector cloning methods rely mainly on Gateway (Life Technologies, Carlsbad, CA, USA) (Karimi et al. 2002; Nakagawa et al. 2007) and type IIS assembly-based technologies (Golden Gate and GreenGate) (Engler et al. 2008; Lampropoulos et al. 2013). Expression vectors allow the targeted proteins to be expressed under different promoters, depending on the final purpose of the experiment. In plant–pathogen studies, transient expression in *N. benthamiana* has been driven mainly by the constitutive cauliflower mosaic virus 35S (CaMV 35S) promoter (Kang et al. 2010; Anderson et al. 2012; Stirnweis et al. 2014; Song et al. 2015; Saur et al. 2016) and in some cases under the native promoter (Sato et al. 2014; El Kasmi et al. 2017; Ramachandran et al. 2017). Alternatively, when controlled timing of protein expression was required or when prolonged protein expression could lead to detrimental effects, gene expression was modulated by inducible promoters (Stork et al. 2015; Hwang et al. 2017). To study plant immunity, transient expression in *N. benthamiana* has been followed mainly by subcellular and tissue localization (Thiel et al. 2012; Rodriguez et al. 2014; Su et al. 2015; Zhuang et al. 2016), protein–protein interactions through co-immunoprecipitation (Co-IP) (Zhao et al. 2013; Hurni et al. 2014; Kim and Hwang 2015; El Kasmi et al. 2017), bimolecular fluorescence complementation (Sun et al. 2014; Du et al. 2015; Liu et al. 2016) or luciferase complementation imaging assay (Du et al. 2013), effect on pathogen performance (Bae et al. 2011; Medina-Hernandez et al. 2013; Song et al. 2015), expression of immune-related or marker genes (Nguyen et al. 2010; Li et al. 2014; Rodriguez et al. 2014; Su et al. 2014; Chaparro-Garcia et al. 2015), electrolyte leakage (Yu et al. 2012; Teper et al. 2014; Gupta et al. 2015), and ROS production (Stork et al. 2015). Because *N. benthamiana* does not respond with a ROS burst when challenged with the microbe-associated molecular pattern (MAMP) flgII-28 from flagellin, transient overexpression of the tomato receptor-like kinase flagellin-sensing FLS3 was used to confer responsiveness to *N. benthamiana* (Hind et al. 2016). The transient expression of a mutated version of FLS3 showed that its kinase activity was required for downstream signaling associated with the flgII-28 ROS burst.

Transient protein expression in *N. benthamiana* also was employed to identify the receptor of another MAMP from the bacterial cold shock protein, the csp22 peptide (Saur et al. 2016). Under the hypothesis that this receptor should interact with the coreceptor BAK1 (that is part of several activated receptor complexes) upon csp22 challenge, BAK1 was expressed fused to a GFP tag that was used for immunoprecipitation, followed by liquid chromatography–mass spectrometry (LC-MS/MS). This strategy allowed the identification of the *N. benthamiana* receptor-like protein required for csp22 responsiveness, NbCSPR.

Attachment of fatty acids as a post-translational modification is important for the regulation of protein location and is of particular interest in the study of plant–pathogen interactions (Boyle and Martin 2015). In *N. benthamiana*, transient protein expression coupled with click chemistry has been exploited for the detection of modifications such as N-myristoylation and S-acylation of both pathogen and host proteins (Boyle et al. 2016).

Transient overexpression of the transcription factor CabZIP63 from *Capsicum annuum* (pepper) in *N. benthamiana* leaves, followed by chromatin immunoprecipitation combined with

PCR, was used to study the transcriptional regulation that CabZIP63 exerted on CaWRKY40, a transcription factor involved in the response to the bacterial pathogen *Ralstonia solanacearum* (Shen et al. 2016).

One of the most frequently used outcomes following transient protein expression is the observation of, usually macroscopic, PCD symptoms (Kang et al. 2010; Cunnac et al. 2011; Chronis et al. 2013; Mafurah et al. 2015). The large *N. benthamiana* leaves allow testing several elicitors using different concentrations or combinations. Coupled with VIGS, this approach has been used to test if the targeted candidate gene participates in PCD associated with highly divergent types of pathogens, such as bacterium, virus, nematode, and oomycete (Del Pozo et al. 2004; Oh and Martin 2011; Pombo et al. 2014). This approach also has been used to test PCD-suppressing activity of pathogen effectors by co-infiltration with inducers of PCD (Teper et al. 2014; Stork et al. 2015). The importance of using appropriate controls in these types of experiments has been highlighted by Adlung and Bonas 2017 who found that some effectors affected *Agrobacterium tumefaciens* performance in leaf tissues, which could lead to overall lower amounts of the PCD-eliciting protein than expected.

Protein expression also has been used in high-throughput approaches. A simple toothpick method (Takken et al. 2000) coupled with the observation of PCD development allowed the identification of plant proteins involved in resistance (Nasir et al. 2005; Coemans et al. 2008; Takahashi et al. 2009). High-throughput transient *in planta* expression assays were performed to study the biological activities of pathogen effector proteins (Caillaud et al. 2012; Stam et al. 2013; Petre et al. 2015). The availability of genome sequences from a variety of pathogens allows the computational prediction of candidate effector genes based on conserved host translocation motifs and their presence in well-defined genome regions (Pais et al. 2013). Cloning and transient *Agrobacterium*-mediated expression of candidate effectors can give valuable insights into the virulence activities of effector proteins, particularly regarding the suppression of host plant immunity (Pais et al. 2013). Using this approach named "effectoromics," Petre et al. 2015 selected, cloned, and expressed 20 candidate effectors in *N. benthamiana* leaf cells to determine their subcellular localizations and to identify the plant proteins they interacted with, through downstream experiments such as Co-IP and mass spectrometry. A similar approach was used for the phenotypic characterization of 84 members of a subclass of *Phytophthora capsici* effectors, which allowed the identification of one member that, when expressed *in planta*, enhanced *P. capsici* virulence in *N. benthamiana* (Stam et al. 2013). By transiently expressing 49 RxLR effector candidates (HaRxLs) of the filamentous phytopathogen *Hyaloperonospora arabidopsidis* fused to fluorescent tags in *N. benthamiana*, two major classes of HaRxLs were defined as those that accumulated in the plant cell nucleus and those that accumulated in the plant membranes. Functional analysis revealed that, in particular, a membrane-localized effector, HaRxL17, enhanced the susceptibility of *N. benthamiana* to this pathogen (Caillaud et al. 2012).

14.6 Stable Genetic Manipulation

The technique has been used frequently in studies of plant–virus interactions for stably overexpressing virus-derived transcripts by taking advantage of post-transcriptional gene silencing (PTGS) to generate *N. benthamiana* plants more resistant to a pathogen (Ling et al. 2008; Reyes et al. 2009). Plant PTGS machinery has been exploited to improve resistance by overexpressing transcripts derived from viral DNA fragments (Lin et al. 2012), double-strand RNA from viral replicase (Pavli et al. 2012), artificial microRNA (Ali et al. 2013; Wagaba et al. 2016), and interfering satellite RNA and RNA interference (RNAi) (Montes et al. 2014). Expression of a whitefly GroEL chaperonin, a protein that can bind to several viruses, produced *N. benthamiana* plants more tolerant to tomato leaf curl virus and cucumber mosaic virus (Edelbaum et al. 2009).

Broad-spectrum resistance was explored using stable expression of artificial transcript activator-like effectors, assembled based on highly conserved regions within begomovirus genomes, that conferred partial resistance to three begomoviruses tested (Cheng et al. 2015). The CRISPR/Cas9 system (clustered regulatory interspaced short palindromic repeat/CRISPR-associated DNA endonuclease 9), which revolutionized plant and animal genome editing (Samanta et al. 2016), has been used successfully in *N. benthamiana* (Nekrasov et al. 2013; Li et al. 2013). CRISPR/Cas9 *N. benthamiana* plants with an inactivated Argonaute 2 gene were used to investigate broad range resistance, which showed that the Argonaute 2 protein had antiviral activity against at least three viruses in a virus-specific manner (Ludman et al. 2017). Stable plastid protein expression (transplastomics) in *N. benthamiana* using a plastid-transformation vector and biolistic was employed to express multiple defense genes (Chen et al. 2014). The results showed that a combination of sweet potato sporamin, taro cystatin, and chitinase from *Paecilomyces javanicus* conferred broad-spectrum resistance against insects, pathogens, and abiotic stresses.

As mentioned in the "Transient protein expression by agroinfiltration" section, *N. benthamiana* has been used to study the molecular mechanisms of pathogen effectors. *Plasmopara viticola* effectors with immune-suppressing activities have been identified by combining transient and stable protein expression. In particular, the overexpression of the effector PvRxLR28 in *N. benthamiana* and grapevine produced plants with enhanced susceptibility to this oomycete (Xiang et al. 2016). Interestingly, the stable overexpression of two *Phytophthora sojae* effectors enhanced disease resistance and tolerance to salt and drought stresses in *N. benthamiana* plants (Rajput et al. 2015; Zhang et al. 2015). These results suggested the possible use of these effectors in crop breeding strategies.

A good degree of conservation of certain molecular pathways, which allows interfamily gene transfer, has been key to the use of *N. benthamiana* in plant–pathogen studies. For example, stably expressed proteins from *Arabidopsis thaliana* (Lacombe et al. 2010; Narusaka et al. 2013; Huang et al. 2014; Wang et al. 2016), tomato (Rommens et al. 1995), and cotton (Lu et al. 2013; Li et al. 2014; Xu et al. 2014) were shown to have functional roles in *N. benthamiana* immunity. A pathogen-induced nucleotide-binding (NB)-leucine-rich repeat (LRR) candidate gene from *Vitis amurensis* was stably overexpressed in *N. benthamiana* (Li et al. 2017) and the transgenic plants were more resistant not only to the oomycete *Plasmopara viticola*, but also to drought and salt stresses, suggesting that the NB-LRR protein may have immune and non-immune roles.

Some stably modified *N. benthamiana* lines have been employed as tools to study the plant immune response. Line SLJR15 expresses the reporter protein Aequorin (Knight et al. 1993), which allows cytoplasmatic Ca^{2+} dynamics to be studied through luminescence imaging (Segonzac et al. 2011; Saur et al. 2016). Line 16c, which expresses *Aequorea victoria* GFP targeted to the endoplasmic reticulum (Ruiz et al. 1998), is the most frequently used *N. benthamiana* line, in particular, to study small RNAs (Philips et al. 2017). Using NGS, the T-DNA insertion region was identified in line 16c and, surprisingly, a portion of a bacterial transposon was found to have co-integrated with this insertion, raising the concern that such events may occur in lines designed for commercial use (Philips et al. 2017). A Cas9-overexpressing (Cas9-OE) *N. benthamiana* line was developed as part of a virus-mediated genome editing system (Ali et al. 2015). In these plants, the DNA endonuclease Cas9 is stably overexpressed under the 35S promoter, and the single guide RNA (sgRNA), which determines the target sequence, is systemically delivered via tobacco rattle virus (Ali et al. 2015). This approach was used to rapidly test different sgRNAs to confer better immunity more efficiently against the DNA virus, tomato yellow leaf curl virus.

14.7 Transcriptomics Analysis of Plant–Pathogen Interactions Using *N. benthamiana*

Before a microarray derived entirely from *N. benthamiana* expressed sequence tags (EST) was developed, potato cDNA arrays were used to determine changes in gene expression in response to virus infection (Senthil et al. 2005; Dardick 2007). Subsequently, a *N. benthamiana* microarray was developed and used to comparatively analyze gene expression changes in response to the necrotrophic *Pectobacterium carotovorum* and hemibiotrophic Pst DC3000 bacteria (Kim et al. 2011). Analysis of the data showed that the transcriptomic expression profiles of *N. benthamiana* in response to *P. carotovorum* were similar to those in response to a mutated Pst DC3000 without a type III secretion system.

The development of NGS techniques and the availability of a draft genome (Bombarely et al. 2012; Naim et al. 2012) may further enhance the use of *N. benthamiana* as a model plant. So far, the Illumina RNA-Seq approach has been used to analyze changes in messenger RNA (mRNA) or small RNA levels. High-throughput small RNA sequencing was used to study the effectiveness of different RNA silencing approaches in the control of virus infections based on the expression of large virus-derived sequences (Montes et al. 2014; Zhao et al. 2015). NGS also was employed to identify and characterize microRNAs involved in the N protein-mediated immune response to tobacco mosaic virus (Yin et al. 2015). RNA-Seq was recently used for the identification of a set of stably expressed genes in *N. benthamiana* which were validated as reference genes for reverse transcription-quantitative PCR (qPCR) in plant–bacteria interaction experiments (Pombo et al. 2019).

The use of integrated omics, which included RNA-Seq analysis of healthy and Odontoglossum ringspot virus (ORSV)-infected *N. benthamiana* leaves combined with proteomics, allowed the identification of putative host proteins that interacted with ORSV capsid protein, which is important for viral long-distance movement in *N. benthamiana* (Lin et al. 2015). Recently, transcriptomic differences detected between mock-treated and *Phytophthora parasitica*-inoculated *N. benthamiana* leaves provided broad insights into *N. benthamiana* defense mechanisms against this oomycete pathogen (Shen et al. 2016).

14.8 Online Resources for *Nicotiana benthamiana*

Many online resources are available for *N. benthamiana* ranging from bioinformatics tools and sequence databases to germplasm collections. Some of the most relevant of these resources are summarized in Table 14.3.

The three main genomics and transcriptomics resources for *N. benthamiana* sequences and annotations are the Queensland University of Technology (QUT) database, the SGN, and NCBI's GenBank. The data in the QUT database are based mostly on the Nbv0.5 genome assembly and transcriptomes v5.1 and v6.1 from Naim et al. 2012. The main features and tools in this database are BLAST and keyword searches, data downloading, genome browsing, expression visualization, and a transcript lookup tool to find corresponding transcripts among the four genome assemblies (Niben1.0.1, Niben0.4.4, Nbv0.5, and Nbv0.3) of Bombarely et al. (2012) and Naim et al. (2012). The SGN contains resources and tools for Niben0.4.4 and Niben1.0.1 (Bombarely et al. 2012). The sequences, annotations, and proteomics resources from these genome versions are available for downloading and the data also can be queried in BLAST and JBrowse (genome browser) tools. The SGN also hosts the SGN VIGS Tool (see below) and SolCyc, a bioinformatics tool to visualize metabolic pathways based on genes from Solanaceae species. GenBank is a large database that contains sequences, annotations, scientific publications, and much more information for all species. Links to the most useful NCBI resources for *N. benthamiana* can be found on the *N. benthamiana* page in NCBI's Taxonomy Browser (https://www.ncbi.nlm.nih.

Table 14.3 Online resources for *N. benthamiana*

Resource	Type	URL
The Plant List	Taxonomic/Systematics	http://www.theplantlist.org/tpl1.1/record/kew-2382877
TimeTree[a]	Taxonomic/Systematic	http://timetree.org/
GBIF	Populations/Natural occurrences	https://www.gbif.org/species/3800423
Atlas of Living Australia	Populations/Natural occurrences	http://bie.ala.org.au/species
GRIN	Germplasm collections	https://npgsweb.ars-grin.gov/gringlobal/taxonomydetail.aspx?25258
IPK Gatersleben	Germplasm collections	https://gbis.ipk-gatersleben.de/
NCBI/GenBank[b]	Molecular/Genomics/Literature	https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?id=4100
QUT *N. benthamiana*[c]	Molecular/Genomics/Transcriptomics	http://benthgenome.qut.edu.au/
Sol Genomic Network[d]	Molecular/Genomics/Transcriptomics	https://solgenomics.net/organism/Nicotiana_benthamiana/genome
miRNEST[e]	MicroRNAs and predicted targets	http://rhesus.amu.edu.pl/mirnest/copy/browse.php
SolCyc[f]	Metabolic pathways	https://solgenomics.net/tools/solcyc/index.pl
CRISPR-P[g]	Biotechnology (CRISPR)	http://crispr.hzau.edu.cn/CRISPR/
CCTop[h]	Biotechnology (CRISPR)	https://crispr.cos.uni-heidelberg.de/
VIGS database[i]	Biotechnology (VIGS)	https://vigs.noble.org/
SGN VIGS Tool[j]	Biotechnology (VIGS)	http://vigs.solgenomics.net/
JOVE VIGS videos[k]	Biotechnology (VIGS)	https://www.jove.com/video/1292/virus-induced-gene-silencing-vigs-in-nicotiana-benthamiana-and-tomato
BTI *N. benthamiana*	Biotechnology	https://btiscience.org/our-research/research-facilities/research-resources/nicotiana-benthamiana/
iGEM Foundation	Biotechnology (Synthetic biology)	http://parts.igem.org/Collections/Plants#Nicotiana_benthamiana

[a]Kumar et al. (2017); [b]Benson et al. (2004); [c]Nakasugi et al. (2013); [d]Fernandez-Pozo et al. (2015a, b); [e]Szczesniak and Makalowska (2014); [f]Foerster et al. (2018); [g]Lei et al. (2014); [h]Stemmer et al. (2015); [i]Senthil-Kumar and Mysore (2014); [j]Fernandez-Pozo et al. (2015a, b); [k]Velasquez et al. 2009)

gov/Taxonomy/Browser/wwwtax.cgi?id=4100). In addition to these three web portals, the Boyce Thompson Institute (BTI) *N. benthamiana* website (https://btiscience.org/our-research/research-facilities/research-resources/nicotiana-benthamiana/) has a collection of links to bioinformatics tools and experimental protocols and resources for *N. benthamiana*.

Two bioinformatics tools, CRISPR-P and CCTop, provide support to design targets for genome editing using CRISPR with Niben0.4.4 and Niben1.0.1, respectively, as the reference genomes (see Table 14.3).

VIGS is an important and efficient tool for functional genomics in *N. benthamiana* (see the "Virus induced gene silencing (VIGS)" section). Several resources for designing and performing VIGS analysis in *N. benthamiana* are available, including, for example, the SGN VIGS Tool (Fernandez-Pozo et al. 2015), the VIGS database (Senthil-Kumar and Mysore 2014), and a video titled "Virus-induced gene silencing (VIGS) in *Nicotiana benthamiana* and tomato" (Velasquez et al. 2009). The SGN VIGS Tool assists in the design of VIGS constructs based on Niben1.0.1 or Niben0.4.4 using an interactive and intuitive

interface. This tool predicts the best target of a gene of interest, thereby allowing the design of constructs to silence multiple genes and minimizing the silencing of off-target genes. The VIGS database contains phenotypic information for a large number of genes silenced in *N. benthamiana*. Currently, the database contains about 1,300 descriptions and/or photographs of gene-silenced plants as well as sequence information of about 4,500 ESTs used for VIGS. This database also includes keyword and BLAST searches to explore all the resources.

Another resource of interest for biotechnology is the International Genetically Engineered Machine (iGem) Foundation's Registry of Standard Biological Parts for *N. benthamiana* (http://parts.igem.org/Collections/Plants#Nicotiana_benthamiana), which provides a collection of expression constructs, reporters, promoters, and other elements tested or that could be used in *N. benthamiana*.

More information about *N. benthamiana* taxonomic resources, populations and natural occurrence, and germplasm collections can be found in the links provided in Table 14.3.

References

Adachi H, Nakano T, Miyagawa N et al (2015) WRKY transcription factors phosphorylated by MAPK regulate a plant immune NADPH oxidase in *Nicotiana benthamiana*. Plant Cell 27:2645–2663

Adkar-Purushothama CR, Kasai A, Sugawara K et al (2015) RNAi mediated inhibition of viroid infection in transgenic plants expressing viroid-specific small RNAs derived from various functional domains. Sci Rep 5:17949

Adlung N, Bonas U (2017) Dissecting virulence function from recognition: cell death suppression in *Nicotiana benthamiana* by XopQ/HopQ1-family effectors relies on EDS1-dependent immunity. Plant J 91:430–442

Ali I, Amin I, Briddon RW, Mansoor S (2013) Artificial microRNA-mediated resistance against the monopartite begomovirus Cotton leaf curl Burewala virus. Virol J 10:231

Ali S, Magne M, Chen S et al (2015a) Analysis of *Globodera rostochiensis* effectors reveals conserved functions of SPRYSEC proteins in suppressing and eliciting plant immune responses. Front Plant Sci 6:623

Ali Z, Abul-Faraj A, Li L et al (2015b) Efficient virus-mediated genome editing in plants using the CRISPR/Cas9 system. Mol Plant 8:1288–1291

Ali Z, Abulfaraj A, Idris A et al (2015c) CRISPR/Cas9-mediated viral interference in plants. Genome Biol 16:238

Anderson RG, Casady MS, Fee RA et al (2012) Homologous RXLR effectors from *Hyaloperonospora arabidopsidis* and *Phytophthora sojae* suppress immunity in distantly related plants. Plant J 72:882–893

Asai S, Ohta K, Yoshioka H (2008) MAPK signaling regulates nitric oxide and NADPH oxidase-dependent oxidative bursts in *Nicotiana benthamiana*. Plant Cell 20:1390–1406

Atamian HS, Chaudhary R, Cin VD et al (2013) In planta expression or delivery of potato aphid *Macrosiphum euphorbiae* effectors Me10 and Me23 enhances aphid fecundity. Mol Plant Microbe Interact 26:67–74

Aoki S, Ito M (2000) Molecular phylogeny of Nicotiana (Solanaceae) based on the nucleotide sequence of the matK gene. Plant Biol 2:316–324

Bae H, Roberts DP, Lim HS et al (2011) Endophytic *Trichoderma* isolates from tropical environments delay disease onset and induce resistance against *Phytophthora capsici* in hot pepper using multiple mechanisms. Mol Plant Microbe Interact 24:336–351

Balli J, Nakasugi K, Jia F et al (2015) The extremophile *Nicotiana benthamiana* has traded viral defence for early vigour. Nat Plants 1:15165

Benson DA, Karsch-Mizrachi I, Lipman DJ et al (2004) GenBank: update. Nucleic Acids Res 32:D23–D26

Burbidge NT (1960) The Australian species of Nicotiana L. (Solanaceae). Aust J Botany 8(3):342–380

Bombarely A, Rosli HG, Vrebalov J et al (2012) A draft genome sequence of *Nicotiana benthamiana* to enhance molecular plant-microbe biology research. Mol Plant Microbe Interact 25:1523–1530

Bombarely A, Moser M, Amrad A et al (2016) Insight into the evolution of the Solanaceae from the parental genomes of Petunia hybrida. Nature Plants 2:16074

Boyle PC, Martin GB (2015) Greasy tactics in the plant-pathogen molecular arms race. J Exp Bot 66:1607–1616

Boyle PC, Schwizer S, Hind SR et al (2016) Detecting N-myristoylation and S-acylation of host and pathogen proteins in plants using click chemistry. Plant Methods 12:38

Bruckner FP, Xavier ADS, Cascardo RS et al (2017) Translationally controlled tumour protein (TCTP) from tomato and *Nicotiana benthamiana* is necessary for successful infection by a potyvirus. Mol Plant Pathol 18:672–683

Burch-Smith TM, Anderson JC, Martin GB et al (2004) Applications and advantages of virus-induced gene silencing for gene function studies in plants. Plant J 39:734–746

Caillaud MC, Piquerez SJ, Fabro G et al (2012) Subcellular localization of the Hpa RxLR effector repertoire

identifies a tonoplast-associated protein HaRxL17 that confers enhanced plant susceptibility. Plant J 69:252–265

Chakravarthy S, Velasquez AC, Ekengren SK et al (2010) Identification of *Nicotiana benthamiana* genes involved in pathogen-associated molecular pattern-triggered immunity. Mol Plant Microbe Interact 23:715–726

Chaparro-Garcia A, Schwizer S, Sklenar J et al (2015) Phytophthora infestans RXLR-WY Effector AVR3a Associates with Dynamin-Related Protein 2 Required for Endocytosis of the Plant Pattern Recognition Receptor FLS2. PLoS ONE 10:e0137071

Chase M, Knapp S, Cox AV et al (2003) Molecular systematics, GISH and the origin of hybrid taxa in Nicotiana (Solanaceae). Ann Bot-London 92:107–127

Chaparro-Garcia A, Wilkinson RC, Gimenez-Ibanez S et al (2011) The receptor-like kinase SERK3/BAK1 is required for basal resistance against the late blight pathogen *Phytophthora infestans* in *Nicotiana benthamiana*. PLoS ONE 6:e16608

Chen PJ, Senthilkumar R, Jane WN et al (2014) Transplastomic *Nicotiana benthamiana* plants expressing multiple defence genes encoding protease inhibitors and chitinase display broad-spectrum resistance against insects, pathogens and abiotic stresses. Plant Biotechnol J 12:503–515

Cheng X, Li F, Cai J et al (2015) Artificial TALE as a convenient protein platform for engineering broad-spectrum resistance to Begomoviruses. Viruses 7:4772–4782

Choi HW, Kim YJ, Hwang BK (2011) The hypersensitive induced reaction and leucine-rich repeat proteins regulate plant cell death associated with disease and plant immunity. Mol Plant Microbe Interact 24:68–78

Chronis D, Chen S, Lu S et al (2013) A ubiquitin carboxyl extension protein secreted from a plant-parasitic nematode *Globodera rostochiensis* is cleaved in planta to promote plant parasitism. Plant J 74:185–196

Clarkson JJ et al (2004) Phylogenetic relationships in *Nicotiana* (Solanaceae) inferred from multiple plastid DNA regions. Mol Phylogenet Evol 33:75–90

Clarkson JJ, Kelly LJ, Leitch AR et al (2010) Nuclear glutamine synthetase evolution in Nicotiana: phylogenetics and the origins of allotetraploid and homoploid (diploid) hybrids. Mol Phylogenet Evol 55:99–112

Clarkson JJ, Dodsworth S, Chase MW (2017) Time-calibrated phylogenetic trees establish a lag between polyploidisation and diversification in Nicotiana (Solanaceae). Plant Syst Evol 303:1001–1012

Coemans B, Takahashi Y, Berberich T et al (2008) High-throughput in planta expression screening identifies an ADP-ribosylation factor (ARF1) involved in non-host resistance and R gene-mediated resistance. Mol Plant Pathol 9:25–36

Cunnac S, Chakravarthy S, Kvitko BH et al (2011) Genetic disassembly and combinatorial reassembly identify a minimal functional repertoire of type III effectors in *Pseudomonas syringae*. Proc Natl Acad Sci USA 108:2975–2980

Dardick C (2007) Comparative expression profiling of *Nicotiana benthamiana* leaves systemically infected with three fruit tree viruses. Mol Plant Microbe Interact 20:1004–1017

De Jonge R, Van Esse HP, Maruthachalam K et al (2012) Tomato immune receptor Ve1 recognizes effector of multiple fungal pathogens uncovered by genome and RNA sequencing. Proc Natl Acad Sci USA 109:5110–5115

DeWolf GP, Goodspeed TH (1957) The Genus Nicotiana. Origins, Relationships and Evolution of Its Species in the Light of Their Distribution, Morphology and Cytogenetics. The Southwestern Naturalist 2, p 177

Del Pozo O, Pedley KF, Martin GB (2004) MAPKKKalpha is a positive regulator of cell death associated with both plant immunity and disease. EMBO J 23:3072–3082

Deng XG, Zhu T, Zou LJ et al (2016) Orchestration of hydrogen peroxide and nitric oxide in brassinosteroid-mediated systemic virus resistance in *Nicotiana benthamiana*. Plant J 85:478–493

Du J, Tian Z, Liu J et al (2013a) Functional analysis of potato genes involved in quantitative resistance to *Phytophthora infestans*. Mol Biol Rep 40:957–967

Du Y, Berg J, Govers F et al (2015) Immune activation mediated by the late blight resistance protein R1 requires nuclear localization of R1 and the effector AVR1. New Phytol 207:735–747

Du Y, Zhao J, Chen T et al (2013b) Type I J-domain NbMIP1 proteins are required for both *Tobacco mosaic virus* infection and plant innate immunity. PLoS Pathog 9:e1003659

Edelbaum D, Gorovits R, Sasaki S et al (2009) Expressing a whitefly GroEL protein in *Nicotiana benthamiana* plants confers tolerance to tomato yellow leaf curl virus and cucumber mosaic virus, but not to grapevine virus A or tobacco mosaic virus. Arch Virol 154:399–407

Edwards KD, Fernandez-Pozo N, Drake-Stowe K et al (2017) A reference genome for *Nicotiana tabacum* enables map-based cloning of homeologous loci implicated in nitrogen utilization efficiency. BMC Genom 18:448

El Kasmi F, Chung EH, Anderson RG et al (2017) Signaling from the plasma-membrane localized plant immune receptor RPM1 requires self-association of the full-length protein. Proc Natl Acad Sci USA 114: E7385–E7394

Engler C, Kandzia R, Marillonnet S (2008) A one pot, one step, precision cloning method with high throughput capability. PLoS One 3:e3647

Fernandez-Pozo N, Rosli HG, Martin GB, Mueller LA (2015a) The SGN VIGS tool: user-friendly software to design virus-induced gene silencing (VIGS) constructs for functional genomics. Mol Plant 8:486–488

Fernandez-Pozo N, Menda N, Edwards JD et al (2015b) The Sol genomics network (SGN)—from genotype to

phenotype to breeding. Nucleic Acids Res 43(D1): D1036–D1041

Foerster H, Bombarely A, Battey JND et al (2018) SolCyc: a database hub at the sol genomics network (SGN) for the manual curation of metabolic networks in Solanum and Nicotiana specific databases. Database, Volume 2018, 1 January 2018, bay035

Global Biodiversity Information Facility (GBIF) (2018) The species database. https://www.gbif.org/species/3800423. Accessed 19 Oct 2018

Gonorazky G, Ramirez L, Abd-El-Haliem A et al (2014) The tomato phosphatidylinositol-phospholipase C2 (SlPLC2) is required for defense gene induction by the fungal elicitor xylanase. J Plant Physiol 171:959–965

Goodin MM, Zaitlin D, Naidu RA, Lommel SA (2008) *Nicotiana benthamiana*: its history and future as a model for plant-pathogen interactions. Mol Plant Microbe Interact 21:1015–1026

Gupta MK, Nathawat R, Sinha D et al (2015) Mutations in the predicted active site of *Xanthomonas oryzae* pv. *oryzae* XopQ differentially affect virulence, suppression of host innate immunity, and induction of the HR in a nonhost plant. Mol Plant Microbe Interact 28:195–206

Hartl M, Merker H, Schmidt DD, Baldwin IT (2008) Optimized virus-induced gene silencing in *Solanum nigrum* reveals the defensive function of leucine aminopeptidase against herbivores and the shortcomings of empty vector controls. New Phytol 179:356–365

Hind SR, Strickler SR, Boyle PC et al (2016) Tomato receptor FLAGELLIN-SENSING 3 binds flgII-28 and activates the plant immune system. Nat Plants 2:16128

Huang PY, Yeh YH, Liu AC et al (2014) The Arabidopsis LecRK-VI.2 associates with the pattern-recognition receptor FLS2 and primes *Nicotiana benthamiana* pattern-triggered immunity. Plant J 79:243–255

Huang S et al (2009) The genome of the cucumber, Cucumis sativus L. Nat Genet 41:1275–1281

Hurni S, Brunner S, Stirnweis D et al (2014) The powdery mildew resistance gene Pm8 derived from rye is suppressed by its wheat ortholog Pm3. Plant J 79:904–913

Hwang IS, Brady J, Martin GB, Oh CS (2017) Ser360 and Ser364 in the kinase domain of tomato SlMAPKKKalpha are critical for programmed cell death associated with plant immunity. Plant Pathol J 33:163–169

Kang HG, Oh CS, Sato M et al (2010) Endosome-associated CRT1 functions early in resistance gene-mediated defense signaling in Arabidopsis and tobacco. Plant Cell 22:918–936

Karimi M, Inze D, Depicker A (2002) GATEWAY vectors for Agrobacterium-mediated plant transformation. Trends Plant Sci 7:193–195

Kelly LJ, Leitch AR, Clarkson JJ et al (2013) Reconstructing the complex evolutionary origin of wild allopolyploid tobaccos (Nicotiana section Suaveolentes). Evolution 67:80–94

Kiba A, Nakano M, Vincent-Pope P et al (2012) A novel Sec14 phospholipid transfer protein from *Nicotiana benthamiana* is up-regulated in response to Ralstonia solanacearum infection, pathogen associated molecular patterns and effector molecules and involved in plant immunity. J Plant Physiol 169:1017–1022

Kim HS, Park SC, Ji CY et al (2016) Molecular characterization of biotic and abiotic stress-responsive MAP kinase genes, IbMPK3 and IbMPK6, in sweetpotato. Plant Physiol Biochem 108:37–48

Kim HS, Thammarat P, Lommel SA et al (2011) *Pectobacterium carotovorum* elicits plant cell death with DspE/F but the *P. carotovorum* DspE does not suppress callose or induce expression of plant genes early in plant-microbe interactions. Mol Plant Microbe Interact 24:773–786

Kim NH, Hwang BK (2015) Pepper heat shock protein 70a interacts with the type III effector AvrBsT and triggers plant cell death and immunity. Plant Physiol 167:307–322

Kim S et al (2014) Genome sequence of the hot pepper provides insights into the evolution of pungency in Capsicum species. Nat Genet 46:270–278. https://doi.org/10.1038/ng.2877

King SR, Mclellan H, Boevink PC et al (2014) *Phytophthora infestans* RXLR effector PexRD2 interacts with host MAPKKK epsilon to suppress plant immune signaling. Plant Cell 26:1345–1359

Knight MR, Read ND, Campbell AK, Trewavas AJ (1993) Imaging calcium dynamics in living plants using semi-synthetic recombinant aequorins. J Cell Biol 121:83–90

Kourelis J, Kaschani F, GrossHolz FM et al (2018) Re-annotated *Nicotiana benthamiana* gene models for enhanced proteomics and reverse genetics. bioRxiv: e373506

Kumar S, Stecher G, Suleski M, Hedges SB (2017) Timetree: a resource for timelines, timetrees, and divergence times. Mol Biol Evol 34(7):1812–1819

Lacombe S, Rougon-Cardoso A, Sherwood E et al (2010) Interfamily transfer of a plant pattern-recognition receptor confers broad-spectrum bacterial resistance. Nat Biotechnol 28:365–369

Lampropoulos A, Sutikovic Z, Wenzl C, Maegele I, Lohmann JU, Forner J (2013) GreenGate—a novel, versatile, and efficient cloning system for plant transgenesis. PLoS One 8:e83043

Lee JH, Kim YC, Choi D, Park JM (2013) Identification of novel pepper genes involved in Bax- or INF1-mediated cell death responses by

high-throughput virus-induced gene silencing. Int J Mol Sci 14:22782–22795

Lei Y, Lu L, Liu H, Sen L, Xing F, Chen L (2014) CRISPR-P: a web tool for synthetic single-guide RNA design of CRISPR-system in plants. Mol Plant 7 (9):1494–1496

Li D, Zhang H, Song Q et al (2015) Tomato Sl3-MMP, a member of the matrix metalloproteinase family, is required for disease resistance against *Botrytis cinerea* and *Pseudomonas syringae* pv. *tomato* DC3000. BMC Plant Biol 15:143

Li JF, Norville JE, Aach J et al (2013) Multiplex and homologous recombination-mediated genome editing in *Arabidopsis* and *Nicotiana benthamiana* using guide RNA and Cas9. Nat Biotechnol 31:688–691

Li X, Zhang Y, Huang L et al (2014a) Tomato SlMKK2 and SlMKK4 contribute to disease resistance against Botrytis cinerea. BMC Plant Biol 14:166

Li X, Zhang Y, Yin L, Lu J (2017) Overexpression of pathogen-induced grapevine TIR-NB-LRR gene *VaRGA1* enhances disease resistance and drought and salt tolerance in *Nicotiana benthamiana*. Protoplasma 254:957–969

Li Y, Zhang L, Lu W, Wang X, Wu CA, Guo X (2014b) Overexpression of cotton GhMKK4 enhances disease susceptibility and affects abscisic acid, gibberellin and hydrogen peroxide signalling in transgenic *Nicotiana benthamiana*. Mol Plant Pathol 15:94–108

Liebrand TW, Smit P, Abd-El-Haliem A et al (2012) Endoplasmic reticulum-quality control chaperones facilitate the biogenesis of Cf receptor-like proteins involved in pathogen resistance of tomato. Plant Physiol 159:1819–1833

Lin CY, Tsai WS, Ku HM, Jan FJ (2012) Evaluation of DNA fragments covering the entire genome of a monopartite begomovirus for induction of viral resistance in transgenic plants via gene silencing. Transgenic Res 21:231–241

Lin PC, Hu WC, Lee SC et al (2015) Application of an integrated omics approach for identifying host proteins that interact with *Odontoglossum ringspot virus* capsid protein. Mol Plant Microbe Interact 28:711–726

Ling KS, Zhu HY, Gonsalves D (2008) Resistance to *Grapevine leafroll* associated virus-2 is conferred by post-transcriptional gene silencing in transgenic *Nicotiana benthamiana*. Transgenic Res 17:733–740

Liu C, Pedersen C, Schultz-Larsen T et al (2016) The stripe rust fungal effector PEC6 suppresses pattern-triggered immunity in a host species-independent manner and interacts with adenosine kinases. New Phytol. https://doi.org/10.1111/nph.14034

Lu W, Chu X, Li Y, Wang C, Guo X (2013) Cotton GhMKK1 induces the tolerance of salt and drought stress, and mediates defence responses to pathogen infection in transgenic *Nicotiana benthamiana*. PLoS One 8:e68503

Ludman M, Burgyan J, Fatyol K (2017) Crispr/Cas9 mediated inactivation of Argonaute 2 reveals its differential involvement in antiviral responses. Sci Rep 7:1010

Mafurah JJ, Ma H, Zhang M et al (2015) A virulence essential CRN effector of *Phytophthora capsici* suppresses host defense and induces cell death in plant nucleus. PLoS One 10:e0127965

Mantelin S, Peng HC, Li B, Atamian HS, Takken FL, Kaloshian I (2011) The receptor-like kinase SlSERK1 is required for Mi-1-mediated resistance to potato aphids in tomato. Plant J 67:459–471

Marks CE, Newbigin E, Ladiges PY (2011) Comparative morphology and phylogeny of Nicotiana section Suaveolentes (Solanaceae) in Australia and the South Pacific. Aust Syst Bot 24:61–86

Medina-Hernandez D, Rivera-Bustamante RF, Tenllado F, Holguin-Pena RJ (2013) Effects and effectiveness of two RNAi constructs for resistance to *Pepper golden mosaic virus* in *Nicotiana benthamiana* plants. Viruses 5:2931–2945

Miki D, Itoh R, Shimamoto K (2005) RNA silencing of single and multiple members in a gene family of rice. Plant Physiol 138:1903–1913

Montes C, Castro A, Barba P et al (2014) Differential RNAi responses of *Nicotiana benthamiana* individuals transformed with a hairpin-inducing construct during *Plum pox virus* challenge. Virus Genes 49:325–338

Naim F et al (2012) Advanced engineering of lipid metabolism in *Nicotiana benthamiana* using a draft genome and the V2 viral silencing-suppressor protein. PLoS One 7:e52717

Nakagawa T, Suzuki T, Murata S et al (2007) Improved Gateway binary vectors: high-performance vectors for creation of fusion constructs in transgenic analysis of plants. Biosci Biotechnol Biochem 71:2095–2100

Nakano M, Nishihara M, Yoshioka H et al (2013) Suppression of DS1 phosphatidic acid phosphatase confirms resistance to *Ralstonia solanacearum* in *Nicotiana benthamiana*. PLoS One 8:e75124

Nakasugi K, Crowhurst RN, Bally J, Wood C, Hellens RP, Waterhouse PM (2013) De novo transcriptome sequence assembly and analysis of RNA silencing genes of Nicotiana. PLoS One 8(3):e59534. https://doi.org/10.1371/journal.pone.0059534

Narusaka M, Kubo Y, Hatakeyama K et al (2013) Interfamily transfer of dual NB-LRR genes confers resistance to multiple pathogens. PLoS One 8:e55954

Nasir KH, Takahashi Y, Ito A et al (2005) High-throughput in planta expression screening identifies a class II ethylene-responsive element binding factor-like protein that regulates plant cell death and non-host resistance. Plant J 43:491–505

Nekrasov V, Staskawicz B, Weigel D, Jones JD, Kamoun S (2013) Targeted mutagenesis in the model plant *Nicotiana benthamiana* using Cas9 RNA-guided endonuclease. Nat Biotechnol 31:691–693

Nguyen HP, Chakravarthy S, Velasquez AC et al (2010) Methods to study PAMP-triggered immunity using tomato and *Nicotiana benthamiana*. Mol Plant Microbe Interact 23:991–999

Oh CS, Martin GB (2011) Tomato 14-3-3 protein TFT7 interacts with a MAP kinase kinase to regulate immunity-associated programmed cell death mediated by diverse disease resistance proteins. J Biol Chem 286:14129–14136

Ohtsu M, Shibata Y, Ojika M et al (2014) Nucleoporin 75 is involved in the ethylene-mediated production of phytoalexin for the resistance of *Nicotiana benthamiana* to *Phytophthora infestans*. Mol Plant Microbe Interact 27:1318–1330

Orchard AE (1999) A history of systematic botany in Australia. Australian Biological Resources Study/CSIRO, Canberra, pp 11–103. ISBN: 0643059652

Pais M, Win J, Yoshida K et al (2013) From pathogen genomes to host plant processes: the power of plant parasitic oomycetes. Genome Biol 14:211

Pavli OI, Kelaidi GI, Tampakaki AP, Skaracis GN (2011) The hrpZ gene of *Pseudomonas syringae* pv. *phaseolicola* enhances resistance to rhizomania disease in transgenic *Nicotiana benthamiana* and sugar beet. PLoS One 6:e17306

Pavli OI, Tampakaki AP, Skaracis GN (2012) High level resistance against rhizomania disease by simultaneously integrating two distinct defense mechanisms. PLoS One 7:e51414

Peng HC, Mantelin S, Hicks GR, Takken FL, Kaloshian I (2016) The conformation of a plasma membrane-localized somatic embryogenesis receptor kinase complex is altered by a potato aphid-derived effector. Plant Physiol 171:2211–2222

Petre B, Saunders DG, Sklenar J et al (2015) Candidate effector proteins of the rust pathogen *Melampsora larici-populina* target diverse plant cell compartments. Mol Plant Microbe Interact 28:689–700

Pfeilmeier S, Saur IM, Rathjen JP, Zipfel C, Malone JG (2016) High levels of cyclic-di-GMP in plant-associated *Pseudomonas* correlate with evasion of plant immunity. Mol Plant Pathol 17:521–531

Philips JG, Naim F, Lorenc MT, Dudley KJ, Hellens RP, Waterhouse PM (2017) The widely used *Nicotiana benthamiana* 16c line has an unusual T-DNA integration pattern including a transposon sequence. PLoS ONE 12:e0171311

Pombo MA, Zheng Y, Fernandez-Pozo N, Dunham DM, Fei Z, Martin GB (2014) Transcriptomic analysis reveals tomato genes whose expression is induced specifically during effector-triggered immunity and identifies the Epk1 protein kinase which is required for the host response to three bacterial effector proteins. Genome Biol 15:492

Pombo MA, Ramos RN, Zheng Y, Fei Z, Martin GB, Rosli HG (2019) Transcriptome-based identification and validation of reference genes for plant-bacteria interaction studies using *Nicotiana benthamiana*. Sci Rep 9(1):1632

Potato Genome Sequencing Consortium et al (2011) Genome sequence and analysis of the tuber crop potato. Nature 475:189–195

Rajput NA, Zhang M, Shen D et al (2015) Overexpression of a *Phytophthora* cytoplasmic CRN effector confers resistance to disease, salinity and drought in *Nicotiana benthamiana*. Plant Cell Physiol 56:2423–2435

Ramachandran SR, Yin C, Kud J et al (2017) Effectors from wheat rust fungi suppress multiple plant defense responses. Phytopathol 107:75–83

Reyes CA, Pena EJ, Zanek MC, Sanchez DV, Grau O, Garcia ML (2009) Differential resistance to *Citrus psorosis virus* in transgenic *Nicotiana benthamiana* plants expressing hairpin RNA derived from the coat protein and 54 K protein genes. Plant Cell Rep 28:1817–1825

Rodriguez PA, Stam R, Warbroek T, Bos JI (2014) Mp10 and Mp42 from the aphid species *Myzus persicae* trigger plant defenses in *Nicotiana benthamiana* through different activities. Mol Plant Microbe Interact 27:30–39

Rojas CM, Senthil-Kumar M, Wang K, Ryu CM, Kaundal A, Mysore KS (2012) Glycolate oxidase modulates reactive oxygen species-mediated signal transduction during nonhost resistance in *Nicotiana benthamiana* and *Arabidopsis*. Plant Cell 24:336–352

Rommens CMT, Salmeron JM, Oldroyd GED, Staskawicz BJ (1995) Intergeneric transfer and functional expression of the tomato disease resistance gene *Pto*. Plant Cell 7:1537–1544

Rosli HG, Zheng Y, Pombo MA et al (2013) Transcriptomics-based screen for genes induced by flagellin and repressed by pathogen effectors identifies a cell wall-associated kinase involved in plant immunity. Genome Biol 14:R139

Ruiz MT, Voinnet O, Baulcombe DC (1998) Initiation and maintenance of virus-induced gene silencing. Plant Cell 10:937–946

Ryu CM, Anand A, Kang L, Mysore KS (2004) Agrodrench: a novel and effective agroinoculation method for virus-induced gene silencing in roots and diverse Solanaceous species. Plant J 40:322–331

Samanta MK, Dey A, Gayen S (2016) CRISPR/Cas9: an advanced tool for editing plant genomes. Transgenic Res 25:561–573

Sato Y, Ando S, Takahashi H (2014) Role of intron-mediated enhancement on accumulation of an *Arabidopsis* NB-LRR class R-protein that confers resistance to *Cucumber mosaic virus*. PLoS One 9: e99041

Saur IM, Kadota Y, Sklenar J et al (2016) NbCSPR underlies age-dependent immune responses to bacterial cold shock protein in *Nicotiana benthamiana*. Proc Natl Acad Sci USA 113:3389–3394

Schmutz J et al (2010) Genome sequence of the palaeopolyploid soybean. Nature 463:178–183

Segonzac C, Feike D, Gimenez-Ibanez S, Hann DR, Zipfel C, Rathjen JP (2011) Hierarchy and roles of pathogen-associated molecular pattern-induced responses in *Nicotiana benthamiana*. Plant Physiol 156:687–699

Senthil G, Liu H, Puram VG, Clark A, Stromberg A, Goodin MM (2005) Specific and common changes in *Nicotiana benthamiana* gene expression in response to infection by enveloped viruses. J Gen Virol 86:2615–2625

Senthil-Kumar M, Mysore KS (2010) Assessing functional role of three water deficit stress-induced genes in nonhost disease resistance using virus-induced gene silencing in *Nicotiana benthamiana*. Plant Signal Behav 5:586–590

Senthil-Kumar M, Mysore KS (2012) Ornithine-delta-aminotransferase and proline dehydrogenase genes play a role in non-host disease resistance by regulating pyrroline-5-carboxylate metabolism-induced hypersensitive response. Plant Cell Environ 35:1329–1343

Senthil-Kumar M, Mysore KS (2014) *Tobacco rattle virus-based virus*-induced gene silencing in *Nicotiana benthamiana*. Nat Protoc 9:1549–1562

Shen D, Chai C, Ma L, Zhang M, Dou D (2016a) Comparative RNA-Seq analysis of *Nicotiana benthamiana* in response to *Phytophthora parasitica* infection. Plant Growth Regul 80:59–67

Shen L, Liu Z, Yang S et al (2016b) Pepper CabZIP63 acts as a positive regulator during *Ralstonia solanacearum* or high temperature-high humidity challenge in a positive feedback loop with CaWRKY40. J Exp Bot 67:2439–2451

Sierro N, *et al*. (2013) Reference genomes and transcriptomes of *Nicotiana sylvestris* and *Nicotiana tomentosiformis*. Genome Biol 14, R60 (2013)

Sierro N et al (2014) The tobacco genome sequence and its comparison with those of tomato and potato. Nat Commun 5:3833

Simao F et al. (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. 31, 3210–3212

Shibata Y, Kawakita K, Takemoto D (2010) Age-related resistance of *Nicotiana benthamiana* against hemibiotrophic pathogen *Phytophthora infestans* requires both ethylene- and salicylic acid-mediated signaling pathways. Mol Plant Microbe Interact 23:1130–1142

Song Q, Li D, Dai Y et al (2015) Characterization, expression patterns and functional analysis of the MAPK and MAPKK genes in watermelon (*Citrullus lanatus*). BMC Plant Biol 15:298

Stam R, Jupe J, Howden AJ et al (2013) Identification and characterisation CRN effectors in *Phytophthora capsici* shows modularity and functional diversity. PLoS One 8:e59517

Stemmer M, Thumberger T, del Sol Keyer M, Wittbrodt J, Mateo JL (2015) CCTop: an intuitive, flexible and reliable CRISPR/Cas9 target prediction tool. PLoS One. https://doi.org/10.1371/journal.pone.0124633

Stirnweis D, Milani SD, Jordan T, Keller B, Brunner S (2014) Substitutions of two amino acids in the nucleotide-binding site domain of a resistance protein enhance the hypersensitive response and enlarge the PM3F resistance spectrum in wheat. Mol Plant Microbe Interact 27:265–276

Stork W, Kim JG, Mudgett MB (2015) Functional analysis of plant defense suppression and activation by the *Xanthomonas* core type III effector XopX. Mol Plant Microbe Interact 28:180–194

Su Y, Xu L, Fu Z et al (2014) ScChi, encoding an acidic class III chitinase of sugarcane, confers positive responses to biotic and abiotic stresses in sugarcane. Int J Mol Sci 15:2738–2760

Su Y, Xu L, Wang S et al (2015) Identification, phylogeny, and transcript of chitinase family genes in sugarcane. Sci Rep 5:10708

Sun Y, Wang C, Yang B et al (2014) Identification and functional analysis of mitogen-activated protein kinase kinase kinase (MAPKKK) genes in canola (*Brassica napus* L.). J Exp Bot 65:2171–2188

Szczesniak MW, Makalowska I (2014) miRNEST 2.0: a database of plant and animal microRNAs. Nucleic Acids Res 42:D74–D77

Takahashi Y, Berberich T, Kanzaki H et al (2009) Serine palmitoyltransferase, the first step enzyme in sphingolipid biosynthesis, is involved in nonhost resistance. Mol Plant Microbe Interact 22:31–38

Takken FL, Luderer R, Gabriels SH et al (2000) A functional cloning strategy, based on a binary PVX-expression vector, to isolate HR-inducing cDNAs of plant pathogens. Plant J 24:275–283

Tanaka S, Ishihama N, Yoshioka H et al (2009) The *Colletotrichum orbiculare* SSD1 mutant enhances *Nicotiana benthamiana* basal resistance by activating a mitogen-activated protein kinase pathway. Plant Cell 21:2517–2526

Teper D, Salomon D, Sunitha S, Kim JG, Mudgett MB, Sessa G (2014) *Xanthomonas euvesicatoria* type III effector XopQ interacts with tomato and pepper 14-3-3 isoforms to suppress effector-triggered immunity. Plant J 77:297–309

Thiel H, Hleibieh K, Gilmer D, Varrelmann M (2012) The P25 pathogenicity factor of *Beet necrotic yellow vein virus* targets the sugar beet 26S proteasome involved in the induction of a hypersensitive resistance response via interaction with an F-box protein. Mol Plant Microbe Interact 25:1058–1072

Todesco M, De Felippes FF (2016) Why *Benthamiana* went viral. Trends Plant Sci 21:4–6

Tomato Genome Consortium (2012) The tomato genome sequence provides insights into fleshy fruit evolution. Nature 485:635–641

Turnbull D, Yang L, Naqvi S et al (2017) RXLR effector AVR2 up-regulates a brassinosteroid-responsive bHLH transcription factor to suppress immunity. Plant Physiol 174:356–369

Velasco R et al (2010) The genome of the domesticated apple (Malus x domestica Borkh.). Nat Genet 42:833

Velasquez AC, Chakravarthy S, Martin GB (2009) Virus-induced gene silencing (VIGS) in *Nicotiana benthamiana* and tomato. J Vis Exp 28:1292

Wagaba H, Patil BL, Mukasa S, Alicai T, Fauquet CM, Taylor NJ (2016) Artificial microRNA-derived resistance to Cassava brown streak disease. J Virol Methods 231:38–43

Wang Y, Nsibo DL, Juhar HM, Govers F, Bouwmeester K (2016) Ectopic expression of Arabidopsis L-type lectin receptor kinase genes LecRK-I.9 and LecRK-IX.1 in Nicotiana benthamiana confers Phytophthora resistance. Plant Cell Rep 35:845–855

Wei CF, Kvitko BH, Shimizu R et al (2007) A *Pseudomonas syringae* pv. *tomato* DC3000 mutant lacking the type III effector HopQ1-1 is able to cause disease in the model plant *Nicotiana benthamiana*. Plant J 51:32–46

Wu C, Jia L, Goggin F (2011) The reliability of virus-induced gene silencing experiments using tobacco rattle virus in tomato is influenced by the size of the vector control. Mol Plant Pathol 12:299–305

Xiang J, Li X, Wu J, Yin L, Zhang Y, Lu J (2016) Studying the mechanism of *Plasmopara viticola* RxLR effectors on suppressing plant immunity. Front Microbiol 7:709

Xu G, Li S, Xie K et al (2012) Plant ERD2-like proteins function as endoplasmic reticulum luminal protein receptors and participate in programmed cell death during innate immunity. Plant J 72:57–69

Xu L, Zhang W, He X et al (2014) Functional characterization of cotton genes responsive to *Verticillium dahliae* through bioinformatics and reverse genetics strategies. J Exp Bot 65:6679–6692

Yin K, Tang Y, Zhao J (2015) Genome-wide characterization of miRNAs involved in *N* gene-mediated immunity in response to *Tobacco mosaic virus* in *Nicotiana benthamiana*. Evol Bioinform Online 11:1–11

Yin W, Dong S, Zhai L, Lin Y, Zheng X, Wang Y (2013) The *Phytophthora sojae Avr1d* gene encodes an RxLR-dEER effector with presence and absence polymorphisms among pathogen strains. Mol Plant Microbe Interact 26:958–968

Yu X, Tang J, Wang Q et al (2012) The RxLR effector Avh241 from *Phytophthora sojae* requires plasma membrane localization to induce plant cell death. New Phytol 196:247–260

Zhang H, Dong S, Wang M et al (2010) The role of vacuolar processing enzyme (VPE) from *Nicotiana benthamiana* in the elicitor-triggered hypersensitive response and stomatal closure. J Exp Bot 61:3799–3812

Zhang H, Teng W, Liang J et al (2016) MADS1, a novel MADS-box protein, is involved in the response of *Nicotiana benthamiana* to bacterial harpin$_{Xoo}$. J Exp Bot 67:131–141

Zhang H, Wang M, Wang W et al (2012) Silencing of G proteins uncovers diversified plant responses when challenged by three elicitors in *Nicotiana benthamiana*. Plant Cell Environ 35:72–85

Zhang M, Ahmed Rajput N, Shen D et al (2015) A *Phytophthora sojae* cytoplasmic effector mediates disease resistance and abiotic stress tolerance in *Nicotiana benthamiana*. Sci Rep 5:10837

Zhao J, Liu Q, Zhang H, Jia Q, Hong Y, Liu Y (2013) The rubisco small subunit is involved in tobamovirus movement and Tm-2(2)-mediated extreme resistance. Plant Physiol 161:374–383

Zhao M, San Leon D, Mesel F, Garcia JA, Simon-Mateo C (2015) Assorted processing of synthetic trans-acting siRNAs and its activity in antiviral resistance. PLoS One 10:e0132281

Zhou B, Zeng L (2017) Elucidating the role of highly homologous *Nicotiana benthamiana* ubiquitin E2 gene family members in plant immunity through an improved virus-induced gene silencing approach. Plant Methods 13:59

Zhu F, Xi DH, Yuan S, Xu F, Zhang DW, Lin HH (2014) Salicylic acid and jasmonic acid are essential for systemic resistance against *Tobacco mosaic virus* in *Nicotiana benthamiana*. Mol Plant Microbe Interact 27:567–577

Zhu X, Caplan J, Mamillapalli P, Czymmek K, Dinesh-Kumar SP (2010) Function of endoplasmic reticulum calcium ATPase in innate immunity-mediated programmed cell death. EMBO J 29:1007–1018

Zhuang J, Coates CJ, Mao Q, Wu Z, Xie L (2016) The antagonistic effect of *Banana bunchy top virus* multifunctional protein B4 against *Fusarium oxysporum*. Mol Plant Pathol 17:669–679

Check for updates

15 Production of Valuable Compounds in Tobacco

Kacper Piotr Kaminski, Simon Goepfert, Nikolai V. Ivanov and Manuel C. Peitsch

Abstract

Traditional tobacco was originally considered to be sacred by Native Americans and was used for ceremonial and medicinal purposes for thousands of years. It was, in fact, what we know today as the *Nicotiana rustica* species that was considered to be "traditional" tobacco. In addition to its traditional purposes, before the arrival of Europeans, Native Americans consumed tobacco by chewing it, in a very similar manner as coca leaves. The chewing custom was gradually adopted by the new settlers, resulting in commercial production and spreading across multiple continents over time. Another indigenous way of consuming tobacco, sniffing pulverized leaves, was first brought to Europe by the Spanish in the early sixteenth century and also became popular among Europeans. The most popular method of tobacco consumption, however, is smoking cigarettes. Although it was likely nicotine that was the primary driver behind the popularity of tobacco products, there is an extremely complex biosynthetic mechanism that leads to the production of many different compounds in the tobacco plant. Without a doubt, plants are one of the most important sources of valuable compounds. Our food, paper, clothes, and numerous medicines are derived from them. Even the fossil fuels we use so much today originate from the decomposition of natural plant biomass. Common tobacco (*Nicotiana tabacum L.*), though primarily associated today with consumer goods production, is no different from other plants and efficiently produces variety of valuable organic compounds. This chapter describes tobacco biosynthetic potential, from alkaloids and phenolics to complex cembranoids, and presents in detail state-of-the-art knowledge of metabolic pathways as well as the methods and factors affecting their accumulation.

K. P. Kaminski · S. Goepfert · N. V. Ivanov (✉) · M. C. Peitsch
PMI R&D, Philip Morris Products S.A., Quai Jeanrenaud 5, Neuchâtel 2000, Switzerland
e-mail: Nikolai.Ivanov@pmi.com

15.1 Biosynthetic Potential of Tobacco

Common tobacco (*Nicotiana tabacum L.*) is an allotetraploid plant with a 4.5 Gb genome (with more than 70% repeats) and approximately 90,000 gene models (Sierro et al. 2014). This makes its genome one of the largest and most gene-rich among all common crops, roughly five times larger than potato (Potato Genome Sequencing 2011) and tomato (Tomato Genome 2012). The hot pepper genome (3.5 Gb) is closer

© Springer Nature Switzerland AG 2020

N. V. Ivanov et al. (eds.), *The Tobacco Plant Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-030-29493-9_15

in size to tobacco, yet it has only 35,000 protein-coding genes predicted (Kim et al. 2014), approximately the same number/content of genes as potato and tomato and half of that for tobacco. With its 90,000 genes, tobacco may present superior potential for producing biologically complex compounds that require an elaborated system of enzymatic reactions.

It is estimated that in all tobacco tissues, we find around 5,700 (Rodgman and Perfetti 2013) different chemical constituents (compared to only 200 discovered in 1960), including carbohydrates, nitrogenous compounds, alkaloids, pigments, isoprenoids, carboxylic acids, phenolics, sterols, and inorganic compounds (Rodgman and Perfetti 2013). These compounds can be classified in a number of different ways (e.g., as in Beilstein ["The Master Catalog"], where the "parents and derivatives" concept is followed) (Heller 1990). For the purpose of this chapter, however, compounds are classified according to their biosynthetic pathway (Fig. 15.1), as it is the most appropriate in the discussion of how their overall production can be influenced at both genetic and physiological levels.

Understanding general biochemical pathways as well as factors influencing yields of desired compounds is crucial. Figure 15.2 represents numerous physiological and elicitation methods of accumulation of target tobacco compounds.

First and foremost, one must determine if target compounds should be produced by influencing the genetic factors of the organism, as shown in Fig. 15.2. Genetically modified organism (GMO)-based methods open possibilities for direct biochemical pathway modification, as knocking out specific genes (e.g., CRISPR/Cas9) (Jiang et al. 2013), overexpression/silencing, or even the introduction of completely new sets of genes can create entirely new biochemical pathways (*Agrobacterium*-mediated) (Fuentes 2016; Farre et al. 2014).

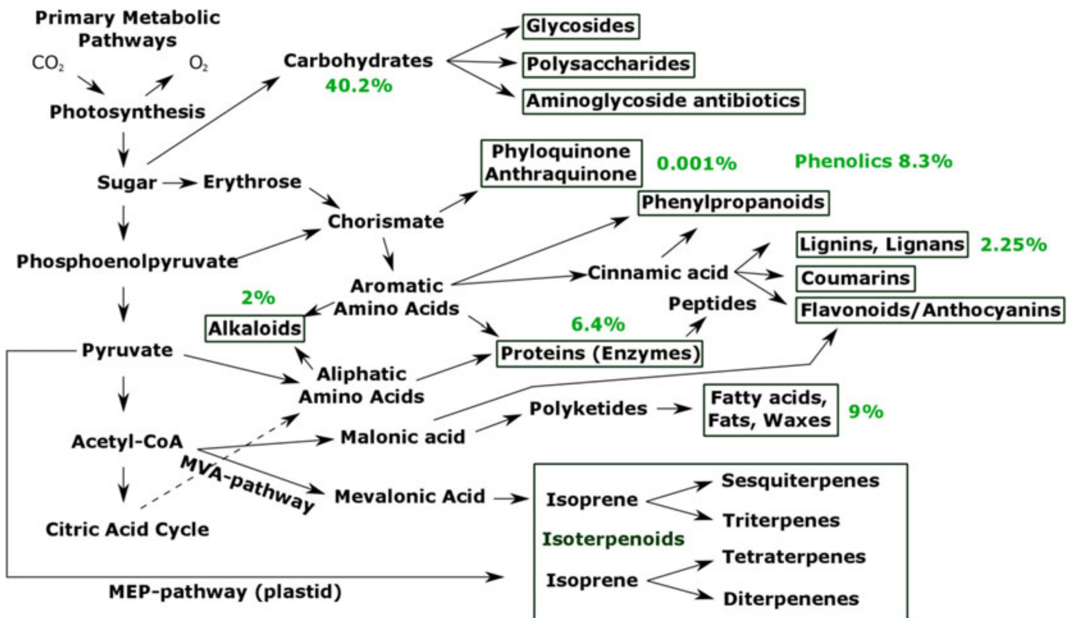

Fig. 15.1 Biosynthetic pathways leading to primary and secondary metabolites. In green, we see the approximate dry weight composition [%] of the most important compounds (Nugroho and Verpoorte 2002). Desired classes of commercially viable compounds are indicated with black rectangles. This figure was assembled from a number of different scientific sources (Nugroho and Verpoorte 2002; Lu et al. 2016; Davis and Nielsen 1999; Fernandez-Pozo 2015)

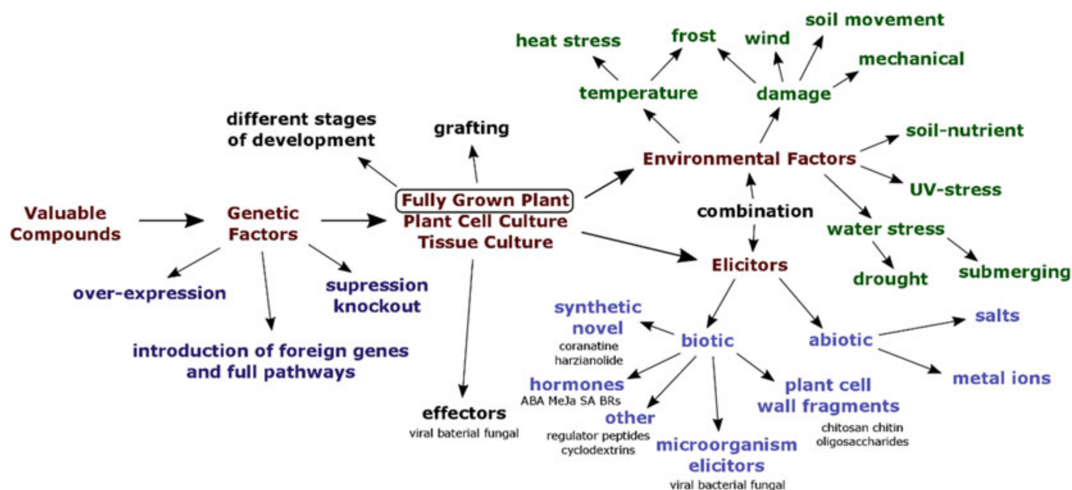

Fig. 15.2 Production methods of target tobacco compounds

Once the use of genetic factors is determined, one can also choose which cultivation form should be employed. Production of target compounds can be performed either at the whole, fully grown plant basis or in plant cell cultures or plant tissue cultures. Harvesting tobacco plants at different developmental stages and implementation of grafting techniques are also interesting and are performed regularly in the agricultural production of many different plants. Tobacco offers numerous grafting possibilities in that regard, yet there have been very few investigations so far, despite the significant molecular diversity in the worldwide tobacco collection (Fricano 2012). Notable exceptions include improving salinity resistance (Ruiz et al. 2006) or the distantly related grafting of tomato and tobacco to improve growth and leafing (Haberal et al. 2016) (grafted "Tomacco," Patent no: TR-2008-05391-B).

Industrial purposes quite often employ the breeding of tobacco varieties to improve phenotypical performance, particularly in flavor or disease resistance. This proves to be quite challenging due to the tetraploid nature of tobacco and the resulting endless recombination possibilities. In addition, breeding can be valuable in functional gene discovery.

The chemical composition of tobacco, like that of many other plant species, changes in response to a variety of environmental factors, which can include abiotic stresses such as temperature, humidity, drought, light, or carbon dioxide concentration. Factors like nitrogen supply, salt stress, or availability of a number of macro- and microelements should also be taken into consideration (Ramakrishna and Ravishankar 2011). Experiments have also been performed in tobacco and other closely related plants using elicitors. These elicitors can be of biotic (hormones, peptides, cell wall fragments, etc.) or abiotic nature (here partly overlapping with environmental factors) and can even be entirely synthetic/novel (Ramirez-Estrada et al. 2016).

In the majority of publications describing tobacco accumulation of target compounds, only a single factor influencing chemical composition is investigated (Jiang et al. 2013; Ruiz et al. 2006; Ramakrishna and Ravishankar 2011; Ramirez-Estrada et al. 2016). It is also very common that tobacco, if physiologically stressed or seriously stimulated, will almost certainly have significant growth impairment (Ramakrishna and Ravishankar 2011; Ramirez-Estrada et al. 2016). The single-factor approach to improve target compound accumulation is also prone to a number of other bottlenecks in synthesis that are unaffected by the specific stress applied. Although rarely pursued by scientists, a

combination of moderate environmental stress with slight elicitation, ideally targeting different phases in biosynthetic pathways, could be employed. This approach has potentially optimal GMO-free ability to improve compound yields, particularly if bottlenecks are eliminated.

15.2 Complex Mixture of Tobacco Alkaloids

Certain tobacco alkaloids were recently found to act effectively against major diseases affecting millions of people. Anatabine administration, for example, was shown to reduce Alzheimer's disease pathology and improve socio-behavioral deficits (Verma et al. 2015). It has also proven to be effective in decreasing thyroglobulin antibodies in patients with chronic lymphocytic autoimmune (Hashimoto's) thyroiditis. Furthermore, both anatabine and anabasine acted as effective stimulants of memory and attention (Levin et al. 2014).

Because of these applications, it is very important to have an effective production procedure for tobacco alkaloids other than nicotine, specifically anatabine and anabasine as well as other closely related compounds found naturally in low quantities. There are two major approaches to achieve this goal and, in general, to produce natural plant bioactive compounds. Traditional production involves plant cultivation and relies on alkaloid extraction procedures, as in the case of atropine from *Duboisia* hybrids or quinine from the *Chincona* genus (Schlager and Drager 2016). Recently, however, astounding progress has been made in the other production alternative: using microbial cells as a production system, particularly genetically engineered bacteria and yeast (Schlager and Drager 2016; Suastegui and Shao 2016). It is promising that immensely complicated benzylisoquinoline alkaloids like morphine, codeine, and oxycodone have been successfully produced in yeast, as recently published in two *Nature* articles (Thodey et al. 2014; DeLoache et al. 2015; Fossati et al. 2015). Particularly interesting work was performed by Fossati et al. (Fossati et al. 2015), in which all enzyme gene sequences from *Papaver somniferum* (opium poppy) were transferred to yeast. It should be noted that this approach requires knowledge about specific enzymes involved in the biosynthetic pathways of these complex compounds. For traditional plant-based production, it is therefore crucial to elucidate the specific biochemical mechanisms of synthesis, particularly for the purpose of proper cultivation and elicitation if optimal yields are to be obtained.

N. tabacum (common tobacco) is among the highest alkaloid-producing species in the *Nicotiana* genus. Another species with very promising biosynthetic yield potential, where nicotine can be accumulated nine times more than *N. tabacum* (Stanfill et al. 2015), is *N. rustica*, the genome of which was recently published (Sierro et al. 2018). Alkaloids in common tobacco constitute 2% of dry weight, with the majority being nicotine (95%) and nornicotine, anabasine, cotinine, and anatabine making up most of the rest. These are the major alkaloids, but tobacco also produces a complex mixture of other alkaloids (e.g., myosmine, 2′,3′-dehydronicotine, nicotyrine, N-methylanatabine, or N-methylanabasine (Baranska et al. 2012)) (Fig. 15.3).

Nicotine and other *N. tabacum* alkaloid nornicotine have been well studied, and their synthesis pathways are described to a high extent in the literature, as broadly summarized in the review by Dewey and Xie (Dewey and Xie 2013). Nonetheless, the final synthesis enzymes' specific action is still to be identified. Nicotine consists of two rings: a pyridine ring derived from nicotinic acid and a pyrrolidine ring derived from putrescine. Putrescine is synthesized from ornithine by ornithine decarboxylase but can also be synthesized from arginine by a series of enzymatic reactions with arginine decarboxylase, agmatine deiminase, and carboamylputrescine aminohydrolase. Putrescine is then converted to N-methyl putrescine by putrescine methyltransferase and subsequently to N-methyl aminobutanol by methyl putrescine oxidase. The latter compound is then spontaneously cyclized to N-methyl-Δ^1-pyrrolinium cation, which is directly incorporated to form nicotine. Incorporation of the second

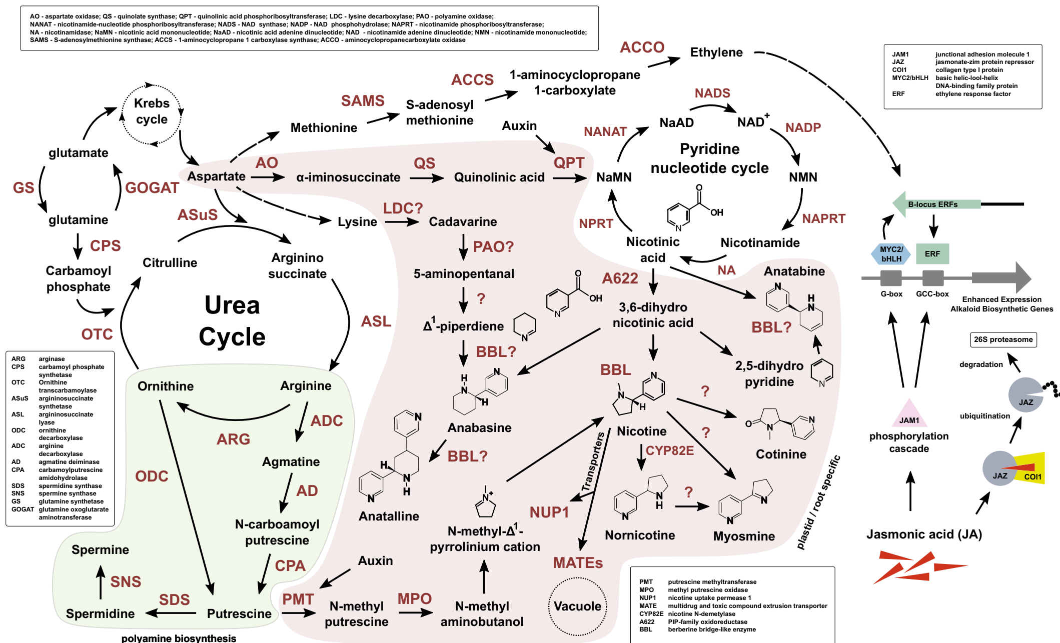

Fig. 15.3 Tobacco alkaloid biosynthetic pathways

pyridine ring into nicotine from nicotinic acid is governed by several genes. Among them is isoflavone reductase-like gene (A622), which converts nicotinic acid to 3,6-dihydro nicotinic acid, which is subsequently transferred to form nicotine by berberine bridge enzyme-like (BBL) enzymes (Lewis et al. 2015), although the specific mechanism of the final biosynthesis step is not well understood. Finally, nicotine is converted to nornicotine by nicotine N-demethylase (CYP82E). In common tobacco (*N. tabacum*), besides nicotine and nornicotine, anatabine is yet another alkaloid accumulated in significant quantities. Its synthesis differs from that of all other alkaloids, as two of its rings originate from the pyridine pathway, which is substantially increased with the inhibition of putrescine methyltransferase expression (Wang et al. 2009). One can hypothesize that overabundance of a nicotinic acid-derived ring with a reduced level of the N-methylpyrrolinium ion leads to anatabine accumulation, yet notable gaps in its biosynthetic mechanism still remain unexplored. For another tobacco alkaloid, anabasine, the current hypothesis is that the first step in its biosynthesis is the conversion of lysine to cadaverine by a yet unidentified lysine decarboxylase (Dewey and Xie 2013; Bunsupa et al. 2014; Shoji and Hashimoto 2008). Cadaverine is then deaminated by diamine oxidase to 5-aminopentanol, which is subsequently believed to be spontaneously cyclized to Δ^1-piperdine, which together with nicotinic acid forms anabasine. Furthermore, it was reported that BBL enzymes are necessary for this last step of anabasine formation, similarly to nicotine (Kajikawa et al. 2011). Due to recent discoveries, we now know that there are a number of transporter genes that were previously identified as important for alkaloid transport (Hildreth et al. 2011) and key transcription factors (Kato et al. 2014).

15.3 Phenolics: Flavonoids and Anthocyanins

Flavonoids and anthocyanins, besides being responsible for the color of flowers, fruits, and sometimes leaves, are also important secondary metabolites that protect tobacco from damage

from reactive oxygen species (ROS) (Fu 2016). They are synthesized from coumaroyl-coenzyme A (CoA) by the chalcone synthase enzyme to form naringenin chalcone, which is then converted to naringenin (C6-C3-C6 structure) by chalcone isomerase (Winkel-Shirley 2002) and subsequently to specific flavonoids and anthocyanins (Fig. 15.4). Very recently, the identification of several downstream genes was described. These genes include flavanone 3β-hydroxylase (F3H), flavonol synthase (FLS), and anthocyanidin synthase (ANS) (Wang 2018). F3H catalyzes the very first step in flavonol and anthocyanin synthesis, converting naringenin to dihydrokaempferol, which is subsequently converted to dihydroquercetin. These two compounds are then transformed to kaempferol and quercetin, respectively, by FLS for flavanol production. ANS, on the other hand, is essential in the last steps of anthocyanin synthesis, where

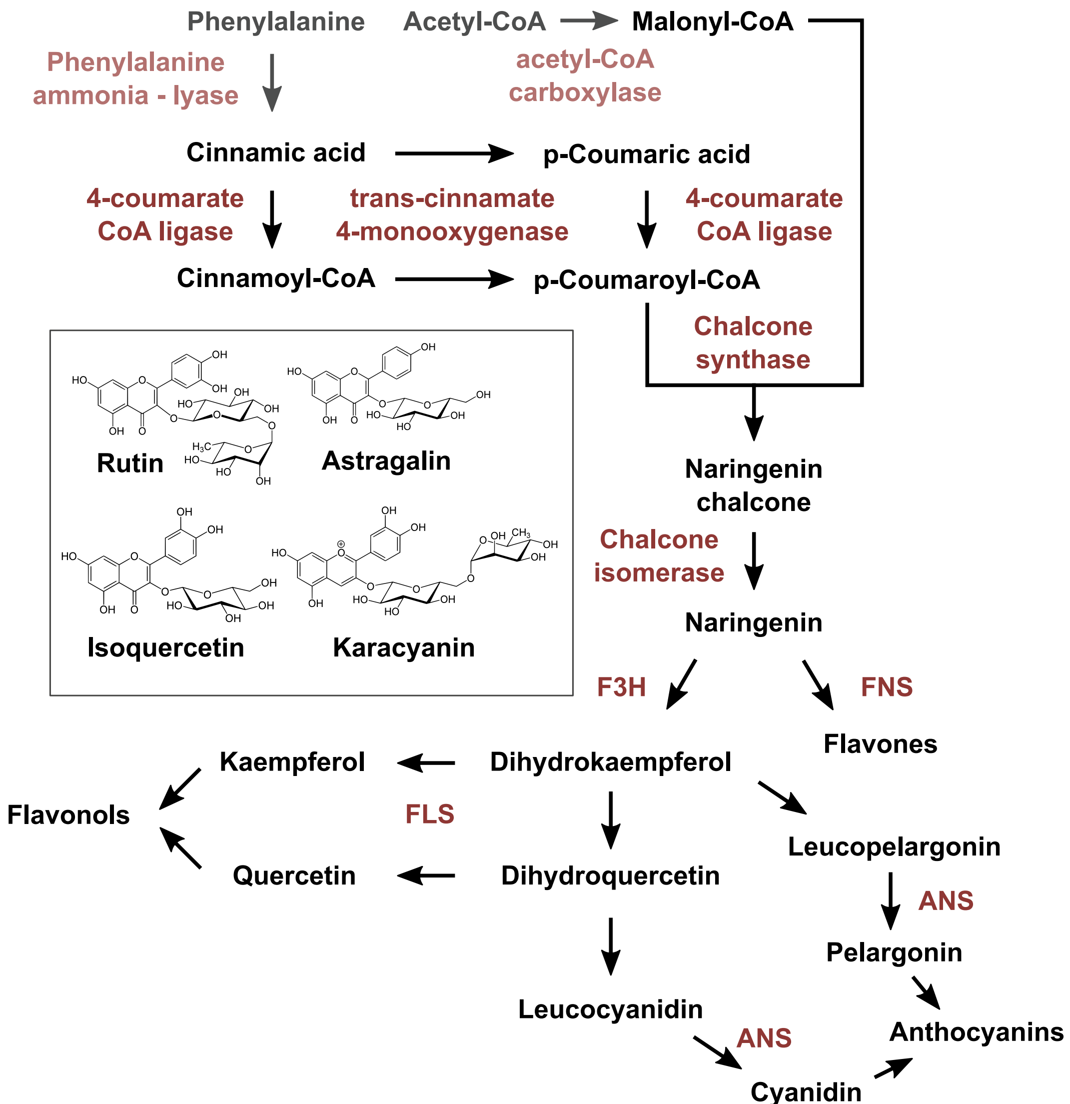

Fig. 15.4 Biosynthesis pathways of anthocyanins and flavonoids. Structures of common molecules rutin, astragalin, isoquercetin, and keracyanin are presented

leucocyanidin and leucopelargonin are converted to cyanidin and pelargonin, respectively. There are more than 4,000 known flavonoids identified in tobacco (Nugroho and Verpoorte 2002), with the most predominant in tobacco being rutin (nearly 8 mg/g), astragalin, isoquercetin, and keracyanin (Davis and Nielsen 1999). Nearly all flavonoids have potential use and are studied for their antioxidant properties, use against cancer, management of cardiovascular diseases, or antimicrobial properties (Trantas et al. 2015; Sun et al. 2012).

Flavonoids protect tobacco from ROS, which increases dramatically during UV radiation. This is one of the most obvious production improvements that one could apply. It has been shown that phytochrome B is directly involved in the regulation of flavonoid production. Production of phenolic compounds can also be enhanced by chemical elicitors, pressure, electric field, heavy metals, pH change, or temperature. It has also been shown that microorganisms, fungi, and bacteria can enhance phenolic compound accumulation (Dias et al. 2016).

15.4 Solanesol and Its Derivatives

Solanesol is a valuable secondary metabolite for which production and purification to high purity can be a very complex and expensive process. Taking into consideration that it is quite abundant in tobacco leaves, even up to 3.6% of dry weight (Yan et al. 2015), it is a valuable end product for industrial production. Solanesol is used as an intermediate for production of its

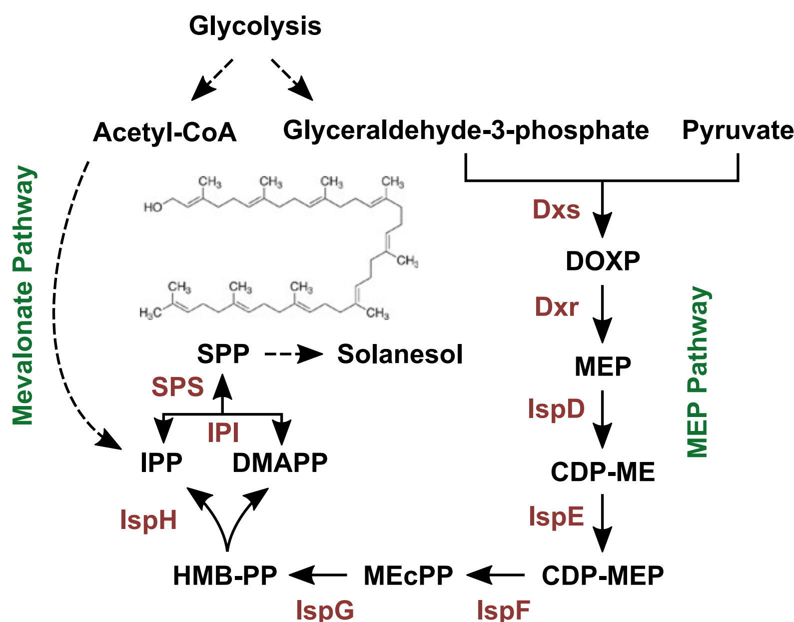

Fig. 15.5 Solanesol biosynthesis in tobacco plastids. Molecules in the solanesol biosynthetic pathway: DOXP, 1-deoxy-D-xylulose 5-phosphate; MEP, 2-C-methyl-D-erythritol-4-phosphate; CDP-ME, 4-disphosphocytidyl-2-C-methyl-D-erythritol; CDP-MEP, 4-diphosphocytidyl-2-C-methyl-D-erythritol-2-phosphate; MEcPP, 2-C-methyl-D-erythritol-2,4-cyclodiphosphate; HMB-PP, hydroxymethylbutenyl-4-diphosphate; IPP, isopentenyl diphosphate; DMAPP, dimethylallyl diphosphate; SPP, solanesyl diphosphate. Enzymes in the solanesol biosynthetic pathway: Dxs, 1-deoxy-D-xylulose-5-phosphate synthase; Dxr, 1-deoxy-D-xylulose-5-phosphate-reductoisomerase; IspD, 4-diphosphocytidyl-2C-methyl-D-erythritol synthase; IspE, 4-diphosphocytidyl-2C-methyl-D-erythritol kinase; IspF, 2-C-methyl-D-erythritol-2,4-cyclodiphosphate synthase; IspG, 1-hydroxy-2-methyl-2-(E)-butenyl-4-diphosphate synthase; IspH, 1-hydroxy-2-methyl-2-(E)-butenyl-4-diphosphate reductase; IPI, isopentenyl pyrophosphate isomerase; SPS, solanesyl diphosphate synthase. Figure adapted from (Yan et al. 2015, 2016; Kanehisa et al. 2016)

derivatives, which have very broad applications: coenzyme Q10 has applications in anticancer treatment, has antioxidant and antiaging properties, and has potential for treatment of neurodegenerative and cardiovascular diseases (Yan et al. 2015); vitamin K2 is an important agent in bone formation and has therapeutic effects on osteoporosis (Weber 2001). Solanesol itself has antimicrobial, anticancer, and anti-inflammatory properties (Li and Chase 2010; Serebryakov and Nigmatov 1990). Market demand for this compound is steadily increasing due to its wide applications in the medical industry (Fig. 15.5).

Solanesol biosynthesis occurs via the mevalonate pathway (MEP pathway) in plastid, and the key end enzyme is SPS. Accumulation of solanesol is also naturally dependent on eight other genes in the MEP pathway: DXS, DXR, IspD, IspE, IspF, IspG, IspH, and IPI (Yan et al. 2015). Changes in regulation in each of these genes have potential impact on final accumulation. Besides genetic factors, a number of environmental factors can affect solanesol production. These include infection with tobacco mosaic virus, biotic stress, or moderate drought (Bajda et al. 2009; Burton et al. 2015).

15.5 Carboxylic Acids/Fatty Acids

Carboxylic acids are one of the major constituents of tobacco biomass. Major acids include citric, malic, oxalic, and malonic acids and comprise from 5 to 18% by weight among different tobacco varieties (Davis and Nielsen 1999). Particularly, citric and malic acids have extensive use in the food industry. Their derivatives often occur as salts with nicotine or ammonia and also form anions with calcium, potassium, and sodium. Despite the overall high content of these major acids, they are only a few of the complex mixtures present in tobacco. Of particular interest in the case of biosynthesis are higher fatty acids, such as linolenic, palmitic, linoleic, oleic, stearic, or myristic acids (Fig. 15.6), which in total comprise approximately 0.35% of tobacco weight. These are

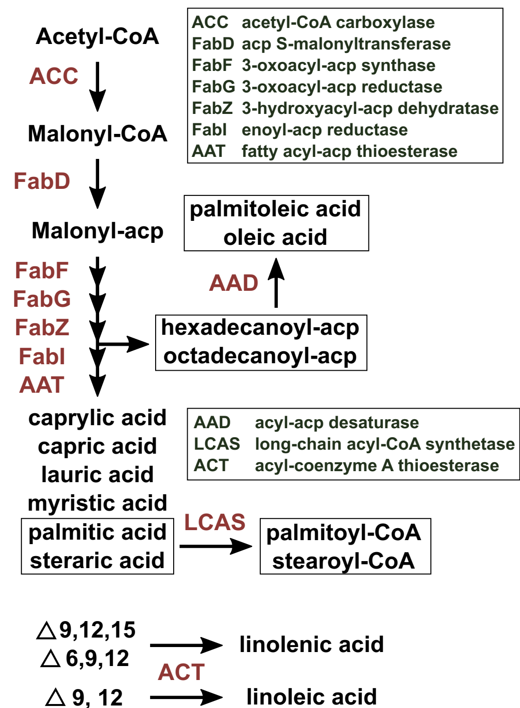

Fig. 15.6 Proposed fatty acid biosynthetic pathways

substrates for storage lipids (such as triacylglycerols), particularly in seeds.

On a molecular basis, fatty acid biosynthesis has not been extensively investigated in tobacco thus far. The majority of genes can only be predicted from the tobacco reference genome sequence. A notable exception is a recent publication exploring the role of β-ketoacyl-acyl carrier protein synthase I (KASI) in plant growth and fatty acid biosynthesis (Yang 2016). Two KASI homologs from tobacco were analyzed and designated as NtKASI-1 and NyKASI-2. In general, fatty acids are produced in plants by the action of acetyl-CoA carboxylase converting acetyl-CoA to malonyl-CoA (Gornicki and Haselkorn 1993) and the subsequent action of β-ketoacyl-ACP synthases, necessary for fatty acids chain elongation (Fig. 15.7) (Hwang et al. 2000).

Besides the high value of major fatty acids in tobacco, it is also very relevant to investigate VLCFAs (longer than 18 carbons) and their role

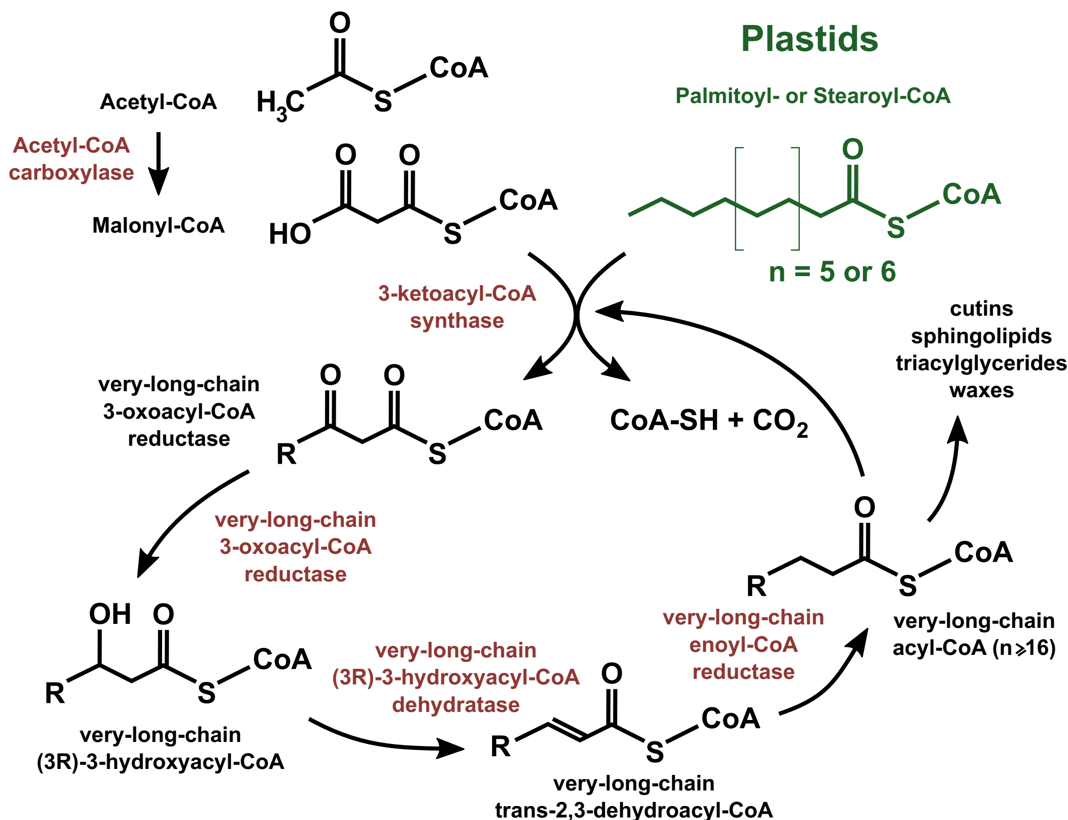

Fig. 15.7 Proposed very long chain fatty acid (VLCFA) biosynthetic pathways

in stress signaling (De Bigault Du Granrut and Cacas 2016), governing both biotic and abiotic responses. Additionally, as indicated in experiments when VLCFA levels are genetically disturbed, it leads to phenotypical differences such as severe growth and development retardation (De Bigault Du Granrut and Cacas 2016).

15.6 Isoterpenoids: Tobacco-Specific Cembranoids

To date, 87 cembranoids, nor-cembranoids, seco-cembranoids, and cyclic cembranoids have been identified in tobacco (Yan et al. 2016). The two most abundant are the diterpenes (1S,2E,4S,7E,11E)-cembra-2,7,11-triene-4,6-diol (α-CBT-ol) and (1S,2E,4R,7E,11E)-cembra-2,7,11-triene-4,6-diol (β-CBT-ol), which were first discovered more than 50 years ago (Roberts and Rowland 1962). Little was known about their potential use at that time. Tobacco cembranoids have been reported to have anticancer (Baraka et al. 2011; Ebrahim et al. 2016; Hailat et al. 2017), antimicrobial (Abuznait et al. 2011; Aqil et al. 2011; Duan 2016), and neuroprotective (Eterović et al. 2011, 2013; Eaton et al. 2004; Ferchmin et al. 2001, 2005; Vélez-Carrasco 2015) properties and can even help in treatment of Parkinson's disease (Hu 2017).

Cembranoids are produced solely in the gland of trichomes (Wagner 1991) and are relatively abundant in tobacco, considering their potential use (e.g., in Virginia varieties, they can account for 0.23% by weight) (Davis and Nielsen 1999). In other tobacco germplasms, they can be accumulated in different amounts, ranging from 0.005% to 0.896% (Yan et al. 2016).

As with other secondary metabolites, trichome secretion of total cembranoids changes due to

environmental factors. Production increase can be achieved by optimal nutrient supply and soil type, water stress, high temperature, or limited light (Yan et al. 2016).

The two most abundant cembranoids, α-CBT-ol and β-CBT-ol, are synthetized by a series of enzymatic reactions involving cembrantrien-ol synthase (Wang and Wagner 2003) and cytochrome P450 hydroxylase (Wang et al. 2001) following the MEP pathway in plastids (Yan et al. 2016). Synthesis of at least two other cembranoids, Z-abienol and labdanieol, is carried out by the enzymes Z-abienol cyclase and labdene-diol cyclase, respectively (Fig. 15.8) (Guo et al. 1994; Guo and Wagner 1995). We currently lack knowledge about the biosynthesis of other potentially important cembranoids in tobacco due to their limited quantities.

15.7 Proteins and Enzymes

Since the production of the first pharmaceutical protein by a genetically engineered organism (bacteria), insulin, in 1978 and the success story behind the Genentech company, there have been numerous efforts to produce target proteins in different organisms. Recently, companies like Medicago (Li et al. 2016), Biopharmaceutical (Olinger et al. 2012), Mapp, or Planet Biotechnology (Planet Biotechnology 2017) have used tobacco as an effective expression system for vaccines, therapeutically active proteins, or antibodies. In particular, Medicago has been working on a target to prevent viral infections by employing the Proficia® virus-like particle technology for transient protein expression in *N. benthamiana*: seasonal flu, pandemic, rotavirus, and norovirus. In parallel, they have been developing vaccines for monoclonal antibody production. It is particularly interesting that *N. benthamiana* became the most popular model for plant virology. Attention was first turned to this species because a particularly large number of plant viruses could effectively infect it. Together with the implementation of virus-induced gene silencing and transient protein expression, this makes *N. benthamiana* an ideal host for studying protein interaction as well as protein expression systems (Goodin et al. 2008). In 2015, the first in-human trials for tobacco-produced monoclonal antibodies were conducted (Ma et al. 2015), and methods for producing therapeutic proteins human apolipoprotein ApoA1 (Oishi and Peitsch 2012) and deoxyribonuclease DNaseI (Oishi and Peitsch 2012) in tobacco have recently been devised. There are a number of widely used plant enzymes of industrial importance (e.g., proteases like papain, ficain, or actinidain, or amylases from malted barley). Although the majority of industrial enzymes are still produced in prokaryotes due to folding mechanisms and post-transitional modification processes, some more complex structures can only be produced in plants. Furthermore, there have been a number of plant antimicrobial peptides discovered to date. The majority of them have very high toxicity and are rich in cysteine, resulting in the formation of many disulfide bonds. As a consequence, they have high thermal, chemical, and proteolytic stability (Tam et al. 2015), all very desirable for industrial production and application. This gives them great potential for possible future alternatives to antibiotics.

Proteins, enzymes, and peptides constitute around 6.4% of tobacco biomass (Rodgman and Perfetti 2013), making this crop one of the most amino acid-rich within industrially produced plants and a potential source of valuable food. It should, however, be taken into consideration that efficient protein synthesis coincides with secondary metabolite accumulation. Amino acids are also used for biosynthesis of alkaloids and phenolics. Furthermore, pyruvate is used for amino acid synthesis, and as a result, it could be shifted away from isoterpenoid production. In eukaryotic organisms, such as tobacco, one is able to produce numerous drugs, vaccines, and antibodies with the introduction of genes by transformation. Due to the nature of tobacco, possible production possibilities of protein-based drugs from other sources are great and exceed that of most other host organisms (Yao et al. 2015). Additionally, extraction and functional analysis of tobacco enzymes can also be

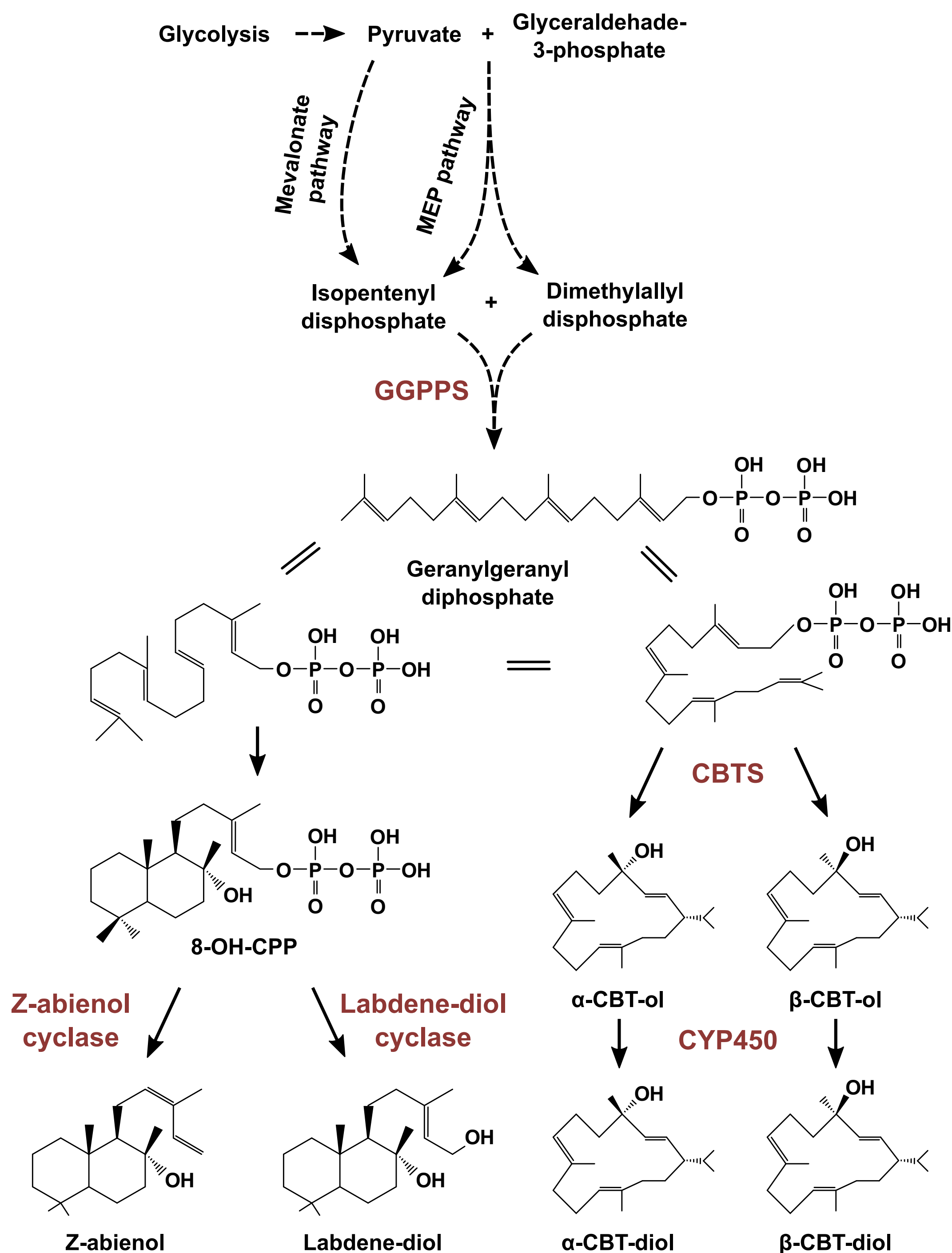

Fig. 15.8 Tobacco cembranoid biosynthetic pathways. 8-OH-CPP, 8-α-hydroxy-copalyl-pyrophosphate; α-CBT-ol, (1S,2E,4S,7E,11E)-cembra-2,7,11-triene-4,6-diol; β-CBT-ol, (1S,2E,4R,7E,11E)-cembra-2,7,11-triene-4,6-diol

performed in order to determine their conversion potential and activity. This includes, for example, lignin degradation enzymes (interesting, as tobacco has quite low lignin content compared to other crops), amylases, cellulases, and more. Antimicrobial peptides secreted by tobacco leaves upon infection can also be valuable. Few studies have been performed on this subject in plants, and to date, no studies of this kind have been performed in tobacco.

15.8 Conclusions and Future Perspectives

Valuable compounds presented in this chapter are only some of the ideas for the production of commercially valuable compounds, and this list is by no means complete. Tobacco was recently used as an effective host for the production of an important antimalarial drug, sesquiterpene lactone, artemisinin precursors (Zhang et al. 2011), and artemisinin itself (Farhi et al. 2011). Another compound of very high value as red pigment, keto-carotenoid astaxanthin, belonging to the terpene class, was produced in tobacco flowers (Mann et al. 2000). Caffeine, a valuable alkaloid of enormous industrial importance, was also successfully produced in tobacco by expressing three coffee N-methyltransferases (Uefuji et al. 2005). Very recently, the genome of the tobacco tree, *N. glauca*, was published (Usadel 2018), which has been shown as a promising source of biofuel due to its high hydrocarbon content (Mortimer et al. 2012). Sunchem SA, on the other hand, has developed a nicotine-free variety of common tobacco, "Solaris," with significant high-quality oil content (which was recently used as jet fuel to power a South African Airways Boeing 737 flight) (Project solaris. https://projectsolaris.it/ 2018). There are many different interesting studies being conducted in the tobacco community from virtually every compound class, and more will come shortly as attention is once again turned to this valuable crop.

References

Abuznait AH et al (2011) Induction of expression and functional activity of P-glycoprotein efflux transporter by bioactive plant natural products. Food Chem Toxicol 49(11):2765–2772

Aqil F et al (2011) Antimicrobial, antioxidant, and antimutagenic activities of selected marine natural products and tobacco cembranoids. Drug Chem Toxicol 34(2):167–179

Bajda A et al (2009) Role of polyisoprenoids in tobacco resistance against biotic stresses. Physiol Plant 135 (4):351–364

Baraka HN et al (2011) Bioactive natural, biocatalytic, and semisynthetic tobacco cembranoids. Planta Med 77(5):467–476

Baranska M et al (2012) Tobacco alkaloids analyzed by Raman spectroscopy and DFT calculations. J Raman Spectrosc 43(8):1065–1073

Bentinger M, Tekle M, Dallner G (2010) Coenzyme Q–biosynthesis and functions. Biochem Biophys Res Commun 396(1):74–79

Bunsupa S et al (2014) Revisiting anabasine biosynthesis in tobacco hairy roots expressing plant lysine decarboxylase gene by using ^{15}N-labeled lysine. Plant Biotechnol 31(5):511–518

Burton HR, Leggett E, Phillips RE (2015) Factors influencing the concentration of solanesol in burley tobacco, vol 14, p 313–320

Davis DL, Nielsen MT (1999) Tobacco: production chemistry and technology. Blackwell Science

De Bigault Du Granrut A, Cacas J-L (2016) How very-long-chain fatty acids could signal stressful conditions in plants? Front Plant Sci 7:1490

DeLoache WC et al (2015) An enzyme-coupled biosensor enables (S)-reticuline production in yeast from glucose. Nat Chem Biol 11(7):465–471

Dewey RE, Xie J (2013) Molecular genetics of alkaloid biosynthesis in *Nicotiana tabacum*. Phytochemistry 94:10–27

Dias MI et al (2016) Exploring plant tissue culture to improve the production of phenolic compounds: a review. Ind Crops Prod 82:9–22

Duan S et al (2016) Chemical basis of the fungicidal activity of tobacco extracts against *Valsa mali*. Molecules 21(12)

Eaton MJ et al (2004) Differential inhibition of nicotine- and acetylcholine-evoked currents through α4β2 neuronal nicotinic receptors by tobacco cembranoids in Xenopus oocytes. Neurosci Lett 366(1):97–102

Ebrahim HY et al (2016) (1S,2E,4S,7E,11E)-2,7,11-Cembratriene-4,6-diol semisynthetic analogs as novel c-Met inhibitors for the control of c-Met-dependent breast malignancies. Bioorg Med Chem 24(22):5748–5761

Eterović VA et al (2011) A cembranoid protects acute hippocampal slices against paraoxon neurotoxicity. Toxicol In Vitro 25(7):1468–1474

Eterović VA et al (2013) Protective activity of (1S,2E,4R,6R,7E,11E)-2,7,11-cembratriene-4,6-diol analogues against diisopropylfluorophosphate neurotoxicity: Preliminary structure-activity relationship and pharmacophore modeling. Bioorg Med Chem 21 (15):4678–4686

Farhi M et al (2011) Generation of the potent anti-malarial drug artemisinin in tobacco. Nat Biotechnol 29:1072

Farre G et al (2014) Engineering complex metabolic pathways in plants. Annu Rev Plant Biol 65:187–223

Ferchmin PA et al (2001) Tobacco cembranoids block behavioral sensitization to nicotine and inhibit neuronal acetylcholine receptor function. J Neurosci Res 64(1):18–25

Ferchmin PA et al (2005) Tobacco cembranoids protect the function of acute hippocampal slices against NMDA by a mechanism mediated by α4β2 nicotinic receptors. J Neurosci Res 82(5):631–641

Fernandez-Pozo N et al (2015) The Sol Genomics Network (SGN)–from genotype to phenotype to breeding. Nucleic Acids Res 43(Database issue): D1036–D1041

Fossati E et al (2015) Synthesis of morphinan alkaloids in *Saccharomyces cerevisiae*. PLoS ONE 10(4): e0124459

Fricano A et al (2012) Molecular diversity, population structure, and linkage disequilibrium in a worldwide collection of tobacco (Nicotiana tabacum L.) germplasm. BMC Genet 13:18

Fu B et al (2016) The influence of light quality on the accumulation of flavonoids in tobacco (Nicotiana tabacum L.) leaves. J Photochem Photobiol B 162:544–549

Fuentes P et al (2016) A new synthetic biology approach allows transfer of an entire metabolic pathway from a medicinal plant to a biomass crop. Elife 5

Goodin MM et al (2008) *Nicotiana benthamiana*: its history and future as a model for plant-pathogen interactions. Mol Plant Microbe Interact 21(8):1015–1026

Gornicki P, Haselkorn R (1993) Wheat acetyl-CoA carboxylase. Plant Mol Biol 22(3):547–552

Guo Z, Wagner GJ (1995) Biosynthesis of labdenediol and sclareol in cell-free extracts from trichomes of *Nicotiana glutinosa*. Planta 197(4):627–632

Guo ZH, Severson RF, Wagner GJ (1994) Biosynthesis of the diterpene cis-abienol in cell-free extracts of tobacco trichomes. Arch Biochem Biophys 308 (1):103–108

Haberal M et al (2016) Grafting tomato onto tobacco rootstocks is a practical and feasible application for higher growth and leafing in different tobacco–tomato unions. Biol Agric Hortic 32(4):248–257

Hailat MM et al (2017) The tobacco cembranoid (1S,2E,4S,7E,11E)-2,7,11-cembratriene-4,6-diol as a novel angiogenesis inhibitory lead for the control of breast malignancies. Bioorg Med Chem 25(15):3911–3921

Heller SR (1990) The Beilstein online database. In: ACS symposium series, vol. 436. American Chemical Society, p 180

Hildreth SB et al (2011) Tobacco nicotine uptake permease (NUP1) affects alkaloid metabolism. Proc Natl Acad Sci USA 108(44):18179–18184

Hu J et al (2017) 4R-cembranoid improves outcomes after 6-hydroxydopamine challenge in both in vitro and in vivo models of Parkinson's Disease. Front Neurosci 11(May)

Hwang SK, Kim KH, Hwang YS (2000) Molecular cloning and expression analysis of 3-ketoacyl-ACP synthases in the immature seeds of *Perilla frutescens*. Mol Cells 10(5):533–539

Jiang W et al (2013) Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice. Nucleic Acids Res 41(20):e188–e188

Kajikawa M et al (2011) Vacuole-localized berberine bridge enzyme-like proteins are required for a late step of nicotine biosynthesis in tobacco. Plant Physiol 155 (4):2010–2022

Kanehisa M et al (2016) KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res 44 (D1):D457–D462

Kato K, Shoji T, Hashimoto T (2014) Tobacco nicotine uptake permease regulates the expression of a key transcription factor gene in the nicotine biosynthesis pathway. Plant Physiol 166(4):2195–2204

Kim S et al (2014) Genome sequence of the hot pepper provides insights into the evolution of pungency in Capsicum species. Nat Genet 46(3):270–278

Levin ED et al (2014) Effects of tobacco smoke constituents, anabasine and anatabine, on memory and attention in female rats. J Psychopharmacol 28 (10):915–922

Lewis RS et al (2015) Transgenic and mutation-based suppression of a Berberine Bridge Enzyme-Like (BBL) Gene family reduces alkaloid content in field-grown tobacco. PLoS One 10(2):e0117273

Li J, Chase HA (2010) Development of adsorptive (non-ionic) macroporous resins and their uses in the purification of pharmacologically-active natural products from plant sources. Nat Prod Rep 27(10):1493–1510

Li J et al (2016) Multiplexed, targeted gene editing in *Nicotiana benthamiana* for glyco-engineering and monoclonal antibody production. Plant Biotechnol J 14(2):533–542

Lu X, Tang K, Li P (2016) Plant metabolic engineering strategies for the production of pharmaceutical terpenoids. Front Plant Sci 7(1647)

Ma JK et al (2015) Regulatory approval and a first-in-human phase I clinical trial of a monoclonal antibody produced in transgenic tobacco plants. Plant Biotechnol J 13(8):1106–1120

Mann V et al (2000) Metabolic engineering of astaxanthin production in tobacco flowers. Nat Biotechnol 18:888

Mortimer C, Bramley P, Fraser PD (2012) The identification and rapid extraction of hydrocarbons from

Nicotiana glauca: a potential advanced renewable biofuel source, vol 5, p 455–458
Nugroho LH, Verpoorte R (2002) Secondary metabolism in tobacco. Plant Cell Tissue Organ Cult 68(2):105–125
Oishi K, Peitsch M (2012) Method for producing apolipoprotein in plants
Oishi K, Peitsch M (2012) Method for expressing desxoyribonuclease in plants
Olinger GG et al (2012) Delayed treatment of Ebola virus infection with plant-derived monoclonal antibodies provides protection in rhesus macaques. Proc Natl Acad Sci USA 109(44):18030–18035
Parmar SS et al (2015) Coenzyme Q10 production in plants: current status and future prospects. Crit Rev Biotechnol 35(2):152–164
Planet Biotechnology (2017)
Potato Genome Sequencing C et al (2011) Genome sequence and analysis of the tuber crop potato. Nature 475(7355):189–195
Project solaris. https://projectsolaris.it/ (2018)
Ramakrishna A, Ravishankar GA (2011) Influence of abiotic stress signals on secondary metabolites in plants. Plant Signal Behav 6(11):1720–1731
Ramirez-Estrada K et al (2016) Elicitation, an effective strategy for the biotechnological production of bioactive high-added value compounds in plant cell factories. Molecules 21(2):182
Roberts DL, Rowland RL (1962) Macrocyclic diterpenes. α- and β-4,8,13-duvatriene-1,3-diols from tobacco. J Organic Chem 27(11):3989–3995
Rodgman A, Perfetti T (2013) The chemical components of tobacco and tobacco smoke, 2nd edn
Ruiz JM et al (2006) Grafting between tobacco plants to enhance salinity tolerance. J Plant Physiol 163(12): 1229–1237
Schlager S, Drager B (2016) Exploiting plant alkaloids. Curr Opin Biotechnol 37:155–164
Serebryakov ÉP, Nigmatov AG (1990) Biologically active derivatives of polyprenylacetic acids and related compounds (review). Pharm Chem J 24(2): 88–98
Shoji T, Hashimoto T (2008) Why does anatabine, but not nicotine, accumulate in jasmonate-elicited cultured tobacco BY-2 cells? Plant Cell Physiol 49(8):1209–1216
Sierro N et al (2014) The tobacco genome sequence and its comparison with those of tomato and potato. Nat Commun 5:3833
Sierro N et al (2018) The impact of genome evolution on the allotetraploid *Nicotiana rustica*—an intriguing story of enhanced alkaloid production. BMC Genom 19(1):855
Stanfill SB et al (2015) Comprehensive chemical characterization of Rapé tobacco products: nicotine, un-ionized nicotine, tobacco-specific N′-nitrosamines, polycyclic aromatic hydrocarbons, and flavor constituents. Food Chem Toxicol 82:50–58
Suastegui M, Shao Z (2016) Yeast factories for the production of aromatic compounds: from building blocks to plant secondary metabolites. J Ind Microbiol Biotechnol 43(11):1611–1624
Sun Y et al (2012) Determination of rutin in cigarette tobacco, filters, mainstream smoke and burned ash of different branded cigarettes by high performance liquid chromatography. Molecules 17(4):3751–3760
Tam JP et al (2015) Antimicrobial peptides from plants. Pharmaceuticals 8(4):711–757
Thodey K, Galanie S, Smolke CD (2014) A microbial biomanufacturing platform for natural and semisynthetic opioids. Nat Chem Biol 10(10):837–844
Tomato Genome C (2012) The tomato genome sequence provides insights into fleshy fruit evolution. Nature 485(7400):635–641
Trantas EA et al (2015) When plants produce not enough or at all: metabolic engineering of flavonoids in microbial hosts. Front Plant Sci 6:7
Uefuji H et al (2005) Caffeine production in tobacco plants by simultaneous expression of three coffee N-methyltrasferases and its potential as a pest repellant. Plant Mol Biol 59(2):221–227
Usadel B et al (2018) The genome and metabolome of the tobacco tree, *Nicotiana glauca*: a potential renewable feedstock for the bioeconomy. bioRxiv, p 351429
Vélez-Carrasco W et al (2015) Pharmacokinetics and metabolism of 4R-cembranoid. PLoS One 10(3)
Verma M et al (2015) Chronic Anatabine treatment reduces Alzheimer's Disease (AD)-like pathology and improves socio-behavioral deficits in a transgenic mouse model of AD. PLoS One 10(5):e0128224
Wagner GJ (1991) Secreting Glandular Trichomes: more than just hairs. Plant Physiol 96(3):675
Wang Z et al (2018) Evolutionary and functional analyses of the 2-oxoglutarate-dependent dioxygenase genes involved in the flavonoid biosynthesis pathway in tobacco. Planta
Wang E, Wagner GJ (2003) Elucidation of the functions of genes central to diterpene metabolism in tobacco trichomes using posttranscriptional gene silencing. Planta 216(4):686–691
Wang E et al (2001) Suppression of a P450 hydroxylase gene in plant trichome glands enhances natural-product-based aphid resistance. Nat Biotechnol 19 (4):371–374
Wang P et al (2009) Silencing of PMT expression caused a surge of anatabine accumulation in tobacco. Mol Biol Rep 36(8):2285–2289
Weber P (2001) Vitamin K and bone health. Nutrition 17 (10):880–887
Winkel-Shirley B (2002) Biosynthesis of flavonoids and effects of stress. Curr Opin Plant Biol 5(3):218–223
Yan N et al (2015) Solanesol: a review of its resources, derivatives, bioactivities, medicinal applications, and biosynthesis. Phytochem Rev 14(3):403–417
Yan N et al (2016) Chemical structures, biosynthesis, bioactivities, biocatalysis and semisynthesis of

tobacco cembranoids: an overview. Ind Crops Prod 83:66–80

Yang T et al (2016) Beta-ketoacyl-acyl Carrier Protein Synthase I (KASI) plays crucial roles in the plant growth and fatty acids synthesis in tobacco. Int J Mol Sci 17(8)

Yao J et al (2015) Plants as factories for human pharmaceuticals: applications and challenges. Int J Mol Sci 16(12):28549–65

Zhang Y et al (2011) The production of artemisinin precursors in tobacco. Plant Biotechnol J 9(4):445–454