Characterizing the Genome of *Nicotiana tabacum*

4

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

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^{TM}) 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