

Chapter 10

Designing Tobacco Genomes for Resistance to Biotic Stresses



K. Sarala, C. Nanda, K. Baghyalakshmi, Reza Darvishzadeh,
K. Prabhakara Rao, U. Sreedhar, and K. Gangadhara

Abstract Tobacco is one of the important commercial crops in the world and is cultivated in more than 120 countries. Various biotic stresses viz. pests, diseases and parasitic weeds infect tobacco from seedling stage to leaf harvest and during post-harvest leaf management there-by severely affecting its leaf yield and quality. Development and deployment of host plant resistance is a sustainable option to minimize these losses. A number of varieties resistant to major biotic stresses viz. TMV, back shank, brown spot, blue mold, powdery mildew, root-knot nematodes, caterpillar, aphid etc. infecting tobacco were developed through conventional breeding. However, lack of reliable sources of resistance, narrow genetic variability, natural barriers of crossing among existing species, longer period required for developing stable homogeneous lines, undesirable associations between the resistant gene and yield and quality contributing characters either due to pleiotropic effects of the resistance gene per se, or due to linkage drag effects caused by the presence of deleterious genes linked to resistant gene, laborious process of screening/phenotyping segregating generations, etc. are slowing down the progress in developing tobacco varieties resistant to biotic stresses through traditional breeding. These limitations can be successfully overcome through molecular breeding and genome designing

K. Sarala (✉) · K. Prabhakara Rao · U. Sreedhar
ICAR-Central Tobacco Research Institute, Rajahmundry, Andhra Pradesh, India
e-mail: sarala.ctri@gmail.com; K.Sarala@icar.gov.in

C. Nanda
ICAR-Central Tobacco Research Institute, Research Station, Hunsur, Karnataka, India
e-mail: C.Nanda@icar.gov.in

K. Baghyalakshmi
ICAR-Central Institute for Cotton Research, Research Station, Coimbatore, India
e-mail: kauverik@gmail.com

R. Darvishzadeh
Department of Plant Breeding and Biotechnology, Faculty of Agriculture, Urmia University,
Urmia, Iran
e-mail: r.darvishzadeh2016@gmail.com

K. Gangadhara
ICAR-Central Tobacco Research Institute, Research Station, Kandukur, Andhra Pradesh, India
e-mail: gangadhara.k1@icar.gov.in

strategies. In this chapter, the current knowledge about genetic resources, the status of utilization of molecular markers in germplasm characterization and development of molecular maps, identified linked markers and quantitative trait loci (QTLs) to various biotic stresses, omics resources characterized, resistant genes cloned, accessible genomic resource databases etc. were summarized for their effective utilization in designing tobacco genome for higher yields and biotic stress resistance. In addition, advances in Marker-assisted selection (MAS) strategies, gene editing technologies and other genome designing strategies, and their possible utilization in designing tobacco genotypes for biotic stress resistance were also discussed.

Keywords Tobacco · *Nicotiana* · Diseases · Pests · Stress · Resistance · Genetic maps · Diversity · Genome sequencing · Markers · Genomic resources · QTL · Genome designing · MAS · Gene editing · Cloning · Databases

10.1 Introduction

10.1.1 Economic Importance of the Crop

Tobacco (*Nicotiana* spp.), also termed as “golden leaf”, is an important commercial crop in the world and is cultivated in more than 120 countries (FAO 2019). China, India, Brazil, United Republic of Tanzania, Indonesia, Zimbabwe, Malawi, USA, Zambia, Mozambique, Turkey, Democratic People’s Republic of Korea, Bangladesh, Argentina and Pakistan, are some of the major countries growing tobacco. It is grown on less than one percent of the world’s agricultural land, and on a wide variety of soils and climate. Currently, tobacco is being grown in an area of 3.62 million hectares with the global tobacco production of 6.69 million tons (FAO 2019). China is the largest tobacco producer in the world with 2.61 million tons/year and India comes second with 0.8 million tons of yearly production. China and India together produce more than 50% of World’s total. Dried tobacco leaves are mainly used for smoking in the form of cigarettes, cigars, pipe tobacco, and flavored shisha tobacco. Leaf is also consumed as snuff, chewing tobacco, dipping tobacco and snus. It is generating huge revenue to the National governments in addition to providing employment to billions of people worldwide.

10.1.2 Reduction in Yield and Quality due to Biotic Stresses

Diseases and pests infect tobacco from seedling until leaf harvest (and even during the curing process) there-by affecting yield and quality of leaf which is the economic product. However, the extent of losses due to pests and diseases varies from year to year and location to location depending upon the weather and percentage of pest infection (Seebold et al. 2007).

According to a rough estimate, the world annual loss due to tobacco diseases is in excess of 5.7 hundred million dollars (Hossain 2020). Loss in yield due to leaf blight and black shank varies from 2 to 10% annually depending on weather conditions. Frog eye disease infection lowers nicotine (46.7%) and reducing sugars (24.3%) contents in tobacco leaves and affect leaf quality (Patel et al. 2001). In severe cases frog eye spots coalesce to become bigger spots leading to drying of leaves (Lucas 1975). Losses to the crop caused by hallow stack range from 5 to 30% (Roy et al. 2012). Tobacco brown spot caused by *Alternaria* fungal species is one of the most damaging diseases, and results in significant yield losses (Luo et al. 2009a; Mo et al. 2012). *Alternaria* produces a variety of toxic secondary metabolites that damage plant tissues resulting in necrotic lesions and premature leaf aging there-by affecting leaf quality (Melton and Shew 2000). Tobacco blue mold is one of the major foliar diseases in the United States and Canada that causes an annual loss of more than \$200 million (Schiltz 1981; Heist et al. 2002). Annual losses exceeding \$200 million recorded in North America due to blue mold epidemics (Nesmith 1984; Heist et al. 2002). Blue mold caused severe losses in Cuba between 1978 and 1980 (Pérez et al. 2003). *Fusarium* wilt adversely affects the tobacco crop in view of its seed transmissibility and its ability to survive in soil without a host plant for years (LaMondia 2015). Wilt in addition to causing severe disease, it may render heavily infested fields not suitable from tobacco production. The losses due to bacterial or granville wilt exceed up to 7% of the crop in South Carolina in 1998 (Fortnum and Martin 1998).

Tobacco Mosaic Virus (TMV) was estimated to cause a loss of one million dollars each year between 1960 and 1965 in North Carolina's flue-cured tobacco (Gooding 1969). Damage caused by TMV infection depends on the stage of infection and genetic structure of the variety and affects both yield and quality. TMV infection of younger plants results in a greater yield loss than infection of older plants. Early infection can result in up to 60% reduction in crop value (Valleau and Johnson 1927). When plants were inoculated with virus at transplanting, 20–31% (Wolf and Moss 1933) loss in yield and 24% in value was reported while late infections result in reduced yield losses of 13–17% at topping (Wolf and Moss 1933; Johnson et al. 1983; Hossain 2020). In severe leaf curl infection, yield losses vary from 60 to 70% as infected leaves do not cure well. While in mild infections, the yield loss is very marginal (Valand and Muniyappa 1992).

Plant parasitic nematodes cause severe damage to tobacco crop. Worldwide loss in tobacco yield due to nematodes was estimated to the tune of 14.7% (Sasser and Freckman 1987). Estimation of losses in production of transplants and cured leaf yield of bidi tobacco in India due to root-knot nematodes reported to be around 50% (Markose and Patel 1977; Shah et al. 1983; Patel et al. 1986). The symptoms may get aggravated under drought conditions as nematodes directly impact the plants' ability to uptake water and nutrients along with predisposition to secondary pathogens (i.e. *Fusarium* wilt).

Crop yield losses due to root parasite, *Orobanche* spp. reported to vary from 20 to 75% depending on the time of infection and the availability of soil moisture (CABI

2021a). Annual yield decline due to *Orobanche* is estimated about \$1.3 to \$2.6 billion in Middle East (Aly 2007).

Severe incidence of ground beetles reduces crop stand up to 50–60% (Sitaramaiah et al. 1999) necessitating replanting which not only adds to the cost of cultivation, but reduce the yield and quality of tobacco due to variation in crop growth. The tobacco caterpillar (*Spodoptera*) incidence causes up to 80–100% loss of transplantable seedlings in the nursery (ICAR-CTRI 2021a) and 33–71% yield loss in planted crop of tobacco (Sitaramaiah et al. 1994). The white fly infestation causes stunted plant growth leading to considerable yield reduction. Tobacco aphid causes not only loss of cured leaf and bright leaf yields (125 kg and 70 kg/ha, respectively) but also leads to deterioration of leaf quality due to development of sooty mold (Sitaramaiah et al. 1994). Further, aphid and white fly transmit viral diseases causing additional yield losses. In FCV tobacco, early incidence of capsule borer, recorded up to 2891 and 426 kg/ha loss in green leaf and cured leaf yields, respectively (Sreedhar et al. 2005). The average estimated seed loss due to capsule borer was around 89 kg/ha in chewing tobacco (Chari et al. 1983). In the years of severe incidence of capsule borer, seed production may be severely affected resulting in huge losses.

Thus, based on the percent infestation and measures taken for control in field condition, losses in tobacco due to biotic stresses may range from minimum to 100% in nursery and up 80% in planted crop. These stresses not only affect yield, but also cause reduction in leaf quality and hence, require suitable control measures to contain the losses.

10.1.3 Growing Importance in the Face of Climate Change and Increasing Population

Climate change is affecting every country on every continent. It is disrupting national economies and affecting lives. Weather patterns are altering, sea levels are rising, and weather events are becoming extreme. Year 2019 was the second warmest year and the end of the warmest decade (2010–2019) ever recorded. Carbon dioxide (CO₂) levels and other greenhouse gases in the atmosphere rose to new records in 2019 (United Nations 2019).

Climate change is likely to alter the balance between pests, their natural enemies and their hosts. Climate change is an important factor driving the spread of pests and diseases. It affect the population size, survival rate and geographical distribution of pests as well as the intensity, development and geographical distribution of diseases (Doody 2020). Temperature and rainfall are the big drivers of shifts in how and where pests and diseases spread. Since 1960, crop pests and diseases have been found to move at an average of 3 km a year in the direction of the earth's north and south poles as temperatures increase (Bebber et al. 2013). Higher temperatures and precipitation levels can slow the growth and reproduction of some pest species and destroy their

eggs and larvae by washing off from the host plant. Climate change also impacts the ecology and biology of insect pest. Increase in temperature would cause migration of insect species towards higher latitudes. Studies using computer prediction models on crop yield losses show that the crop yield loss worldwide is likely to increase by 10–25% due to global warming.

Insect pests of crop plants are highly affected by global climate change. The changing climate could result in insect outbreaks, migration, change in biodiversity, species extinction, change in host shift, and emergence of new pests or biotypes (Kumar and Singh 2016). Insect pest and disease problems in tobacco have shown a shift in the recent past due to climate, ecosystem, and technological changes. There has been an overall decline of budworm and a rise in incidence of tobacco caterpillar, aphid, white fly, stem borer, ground beetle, *Cucumber Mosaic Virus* (CMV) and *Orobanche* (Sreedhar 2016). The damage due to certain pests like ground beetles is usually noticed more in drought years and during prolonged hot spells immediately after planting of tobacco (Sitaramaiah et al. 1999). Climate change also impacts disease incidence and their severities. In case of higher latent infection, the entire tobacco crop can be wiped out due to hallow stack in the event of increased rainfall and water logging (ICAR-CTRI 2005, 2006). Tobacco plants aged 35–60 days were found to be highly susceptible to bacterial wilt infection in the field following high temperature (25–30 °C) and rainfall. Thus, climate change is likely to change the pests and disease composition and their severities in tobacco, requiring preparedness to contain the possible crop losses.

10.1.4 Limitations of Traditional Breeding and Rational of Genome Designing

One of the strategies for effectively managing the menace of biotic stresses in tobacco is through host plant resistance. Conventional breeding played a significant role in developing tobacco genotypes for higher yields and biotic stress resistance in the past and continues to be important for tobacco improvement. However, it has its limitations in terms of availability of sources of resistance and other desirable traits, narrow genetic variability, barriers for natural crossing among existing species, undesirable associations between the presence of a resistant gene and yield and quality contributing characters either due to pleiotropic effects of the resistance gene per se, or due to linkage drag effects caused by the presence of deleterious genes linked to gene of interest (Chaplin et al. 1966; Chaplin and Mann 1978; Legg et al. 1981; Zeven et al. 1983; Friebe et al. 1996; Brown 2002). Also, suppressed recombination within introgressed chromatin (Paterson et al. 1990; Liharska et al. 1996) can make it difficult to alleviate linkage drag effects through back crossing (Stam and Zeven 1981; Young and Tanksely 1989). This can even complicate the efforts to discriminate between pleiotropic and linkage drag effects (Purrlington 2000; Brown 2002). Other limitations in conventional breeding are the time required (relatively longer)

to combine different genes and laborious process of screening/phenotyping segregating generations in developing a variety. Traditional breeding depends on phenotypic screening for biotic stress resistance. Phenotypic screening for biotic stress resistance under natural conditions is not reliable as the stress incidence depend on various environmental factors which are highly variable under field conditions. Screening under artificial conditions is dependent on crop stage and creation of favorable conditions for infection development. Therefore, screening at early crop stages may not be possible and hence, large undesirable plants have to be handled up to full crop stage before rejection. Such screening related issues are acting as limitations in achieving progress in resistance breeding.

The recent technological advancements makes it possible to design plant genomes with desirable phenotypes with high yield by accumulating favorable alleles and eliminating deleterious and undesirable alleles. Genome manipulation strategies alleviate the limitation faced in conventional breeding as they depend on precise modification or editing of genome regions responsible for a phenotypic trait. The complete knowledge about target trait in terms of genomic areas and genes responsible, gene expression patterns and regulation, linked markers/quantitative trait loci (QTLs) etc. aids in targeted manipulation of desirable traits. Such modifications overcome the linkage drag effects and undesirable gene associations and reduce the time required for resistance gene transfer. Molecular marker-assisted breeding for foreground and background selection helps in saving time required in classical breeding and assists in selection of characters in precise manner in segregating populations from early stages itself without artificial inoculations and environmental influence. The current advancements in gene editing technologies are making it possible to modify and edit target genes to yield desirable phenotypes avoiding nontarget effects observed when mutation breeding employed. In genome-assisted breeding (GAB), selection of genotypes based on their breeding values that are estimated considering the all available information on markers will pave the way for the development of genotypes with all the favorable alleles for achieving maximum attainable yields along with desirable quality and resistance to biotic stresses. The time taken for developing genotypes with desirable genes can be reduced through precise manipulation of genomes without requiring generations of selfing to achieve homozygosity. Transgenic and cisgenic transfer of desirable genes from unrelated and distantly related species will enlarge the sources of resistance and other desirable traits. Genome edited crops may be considered as non-GMO (Genetically Modified Organisms) plants and hence, may avoid GMO regulations. Thus, genome designing offers greater scope for accelerated varietal development and overcomes the limitations of conventional breeding.

10.2 Description on Different Biotic Stresses

10.2.1 Taxonomy of Diseases and Insects Infecting Tobacco

Several species of insect pests, diseases and root parasite pose serious threat to the tobacco crop adversely affecting the leaf yield and quality. Some pests also transmit virus pathogens that cause viral diseases (Lucas 1975; Gopalachari 1984; Doroszewska et al. 2013).

10.2.1.1 Insect Pests Infesting Tobacco Crop

Number of insects attack tobacco and cause damage to various degrees. The taxonomic details of insect pests that infest tobacco crop in majority of its growing areas are given at Table 10.1. Insects belonging to various families of Lepidoptera, Homoptera, Coleoptera and Hemiptera orders found to attack tobacco crop. Among

Table 10.1 Taxonomic details of insect pests infesting tobacco crop

Insect pest	Scientific name	Order	Family
Tobacco caterpillar	<i>Spodoptera litura</i> (Fabricius)	Lepidoptera	Noctuidae
Beet armyworm	<i>Spodoptera exigua</i> (Boisduval)	Lepidoptera	Noctuidae
Tobacco horn worm	<i>Manduca Sexta</i> (Linnaeus)	Lepidoptera	Sphingidae
Stem borer	<i>Scrobipalpa heliopa</i> (Lower)	Lepidoptera	Gelechiidae
Whitefly	<i>Bemisia tabaci</i> (Gennadius)	Homoptera	Aleurodidae
Tobacco Flea Beetle	<i>Epitrix hirtipennis</i> (Meisheimer)	Coleoptera	Chrysomelidae
Ground beetles	<i>Mesomorphus villiger</i> (Blanchard), <i>Seleron latipes</i> Guer, <i>Opatroides frater</i> Farum	Coleoptera	Tenebrionidae
Wireworms	<i>Conoderus vespertinus</i> (Fabricius)	Coleoptera	Elateridae
Stink bug	<i>Nezara viridula</i> (Linnaeus), <i>Euschistus servus</i> (Say)	Hemiptera	Pentatomidae
Tobacco aphid	<i>Myzus persicae nicotianae</i> (Blackman)	Homoptera	Aphididae
Tobacco budworm	<i>Heliothis virescens</i> (Fabricius)	Lepidoptera	Noctuidae
Budworm/Capsule borer	<i>Helicoverpa armigera</i> (Hubner)	Lepidoptera	Noctuidae
Cotton Mealybug	<i>Phenacoccus solenopsis</i> (Tinsley)	Hemiptera	Pseudococcidae

them, tobacco caterpillar, *Spodoptera litura* (nursery and main field), stem borer, *Scrobipalpa heliopa* (nursery and main field), whitefly, *Bemisia tabaci* (nursery and main field) vector of *Leaf Curl Virus* disease, tobacco aphid, *Myzus persicae nicotiana*, bud worm and seed capsule borer, *Helicoverpa armigera* (main field) are the major pests, whereas ground beetles, *Mesomorphus villiger* and *Spodoptera exigua* are minor and sporadic in nature. Others pests cause insignificant damage to the crop.

10.2.1.2 Microorganisms that Cause Diseases in Tobacco

Number of fungal, bacterial and viral pathogens infect tobacco causing economic losses (Lucas 1975; Gopalachari 1984; Shew and Lucas 1991)). The taxonomic details of various organisms causing disease in tobacco areas detailed at Table 10.2. Major fungal diseases are damping off, leaf blight and black shank, brown spot and blue mold; Bacterial diseases are leaf spots, wilt and hallow stack; viral diseases are TMV, leaf curl, CMV etc. Mixed infections of viruses are quite frequently observed in tobacco (Blancard et al. 1999).

10.2.1.3 Nematodes Infecting Tobacco

Two types of nematodes mostly affect tobacco causing malformation on tobacco root. They are root-knot nematodes (*Meloidogyne* spp.) and cyst or gall nematodes (*Globodera* spp.). Both the nematodes belongs to Kingdom Animalia, Phylum Nematoda, Class Secernentea, Subclasses Tylenchia and family Heteroderidae. Out of several *Meloidogyne* spp., *M. incognita* (Kofoid and White) and *M. javanica* (Treub) are the widespread and the most damaging ones on tobacco Two other species, *M. arenaria* and *M. hapla* (Chitwood 1949) have limited distribution and cause less damage. Several subspecies of *Globodera* viz. *G. tabacum*; *G. tabacum solanacearum*, *G. solanacearum*; *G. tabacum virginiae*, *G. virginianae* etc. have been reported on tobacco in the world (CABI 2021b). Other nematodes that cause damage in few areas are reniform (*Rotylenchulus reniformis*) and stunt nematodes (*Tylenchorhynchus vulgaris*) (Bairwa and Patel 2016).

10.2.1.4 Broomrape (*Orobanche* spp.)

Orobanche generally known as broomrape is a root parasites on tobacco and belongs to Orobanchaceae family (Gevezova et al. 2012). This genus is divided into four sections: Gymnocaulis Nutt., Myzorrhiza (Phil.) Beck, Trionychon Wallr., and Orobanche (syn. Osproleon Wallr.) (Greuter et al. 2000). The most important species (from an agronomic perspective) are found in the sections Trionychon and Orobanche. Section Trionychon includes *O. ramosa* L. and *O. aegyptiaca* (Paran et al. 1997). *O. cernua*, *O. ramosa* and *O. aegyptiaca* found to cause damage in tobacco.

Table 10.2 Taxonomic details of microorganisms that cause diseases in tobacco

Disease	Causal organism	Phylum	Order	Family
<i>Fungal diseases</i>				
Damping off	<i>Pythium aphanidarmatum</i> (Edson) Fitzp	Oomycota	Peronosporales	Pythiaceae
Leaf blight and Black shank	<i>Phytophthora parasitica</i> f. spp. <i>nicotianae</i> (van Breda de Hann)	Oomycota	Peronosporales	Peronosporaceae
Brown spot	<i>Alternaria alternata</i> (Fries) Keissler	Ascomycota	Pleosporales	Pleiosporaceae
Blue mold	<i>Peronospora hyoscyami</i> f. spp. <i>tabacina</i>	Oomycota	Peronosporales	Peronosporaceae
Fusarium wilt	<i>Fusarium oxysporum</i> Schlechtend.; Fr. f. spp. <i>nicotianae</i> (J. Johnson) W.C. Snyder & H.N. Hans	Ascomycota	Hypocreales	Nectriaceae
Frog eye leaf spot	<i>Cercospora nicotianae</i> Ell. and Ev	Ascomycota	Capnodiales	Mycosphaerellaceae
Curvularia leaf spot	<i>Curvularia verrucosa</i> Sivan	Ascomycota	Pleosporales	Pleiosporaceae
Powdery mildew	<i>Erysiphe cichoracearum</i> var. <i>nicotianae</i> (D.C.) V.P. Heluta, <i>Golovinomyces cichoracearum</i> var. <i>cichoracearum</i>	Ascomycota	Erysiphales	Erysiphaceae
Black root rot	<i>Thielaviopsis basicola</i> Syn. <i>Chalara elegans</i>	Ascomycota	Microascales	Ceratocystidiaceae
<i>Bacterial diseases</i>				
Angular leaf spot	<i>Pseudomonas syringae</i> pv. <i>angulata</i> Ellis and Everh	Proteobacteria	Pseudomonadales	Pseudomonadaceae
Bacterial leaf spot	<i>Dickeya dadantii</i>	Proteobacteria	Enterobacterales	Pectobacteriaceae
Bacterial wilt	<i>Ralstonia solanacearum</i>	Proteobacteria	Burkholderiales	Urkholderiaceae
Hollow stalk	<i>Erwinia carotovora</i> sub spp. <i>carotovora</i>	Proteobacteria	Enterobacterales	Pectobacteriaceae
<i>Viral diseases</i>				
Tobacco Mosaic Virus	<i>Tobacco Mosaic Virus</i> (TMV)	Kitimoviricota	Martellivirales	Virgaviridae
Potato Virus Y	<i>Potato Virus Y</i> (PVY)	Pisuviricota	Patatavirales	Potyviridae

(continued)

Table 10.2 (continued)

Disease	Causal organism	Phylum	Order	Family
Tobacco Leaf Curl Virus	<i>Tobacco Leaf Curl Virus (TLCV)</i>	Cressdnaviricota	Geplafuvirales	Geminiviridae
Cucumber Mosaic Virus	<i>Cucumber Mosaic Virus (CMV)</i>	Kitrinoviricota	Martellivirales	Bromoviridae
Tobacco Distorting Virus	<i>Tobacco Distorting Virus</i>	Pisuviricota	Sobelivirales	Luteoviridae
Tobacco Etch Virus	<i>Tobacco Etch Virus</i>	Pisuviricota	Patativirales	Potyviridae

10.2.2 Races, Isolates, Biotypes

Race/isolates/biotypes that are distinguishable based on host differentials exist in causal organisms of few biotic stresses of tobacco. *B. tabaci* is documented as a complex of cryptic species with two most important biotypes, MEAM1 (Middle East-Asia Minor 1; biotype B) and MED (Mediterranean; biotype Q) (Yao et al. 2017). Two races, race 0 and race 1 found to exist in *Phytophthora parasitica f. sp. Nicotianae* (Woodend and Mudzengerere 1992). Likewise, two races viz. race 0 and race 1 are reported in *Pseudomonas syringae* pv. *tabaci* causing wildfire. Two isolates, PVY^{NTN} and PVY^{NW} in the PVY^N strain group that produce necrotic symptoms on “VaVa” plants (BURLEY 21, K 326, NC 95) found to infect tobacco (Verrier and Doroszewska 2018). CMV constitutes two subgroups, I and II based on severity of symptoms and virulence (Blume et al. 2017). TMV has mutated into many strains and strains such as TMV-O, TMV-C and TMV-N can infect most members of Solanaceae (Holmes 1946).

10.2.3 Stages and Extent of Damage

Biotic stresses infest and cause damage to tobacco at different stages of crop growth from seedling stage in the nursery to seed collection stage in the main field. The damage caused by some major insect pests, diseases, nematodes and *Orobanche* in tobacco are discussed here.

10.2.3.1 Insect Pests

Tobacco caterpillar (*S. litura*): The tobacco caterpillar is one of the most destructive polyphagous pests worldwide (Xue et al. 2010). The young caterpillars feed on the leaf tissues in both nursery and main field. Larvae feed voraciously leaving only veins and petioles and also cut the stems of small and tender seedlings, hence, known as cut worms. In severe cases, larva feed the entire lamina leaving only veins and petioles leading to leaf skeletonisation and heavy defoliation.

Tobacco hornworm (*M. sexta*): The tobacco hornworm is a common pest of plants in the family Solanaceae, that includes tobacco, eggplant, tomato, pepper, various ornamentals and weeds (del Campo and Renwick 1999). They are voracious feeders and may completely defoliate plants if not controlled. *M. sexta* has a large distribution throughout the New World, occurring as far south as Chile.

Stem borer (*S. heliopa*): The larva bore inside the stem and midribs in nursery as well as in the transplanted crop and feed on internal tissues. As a result, swelling appears where the borer stays. Borer infested seedlings when planted in the field remain stunted and sometimes unusual branching of the plant is seen.

Whitefly (*B. tabaci*): Whitefly is a complex cryptic species and destructive insect pest, reported to attack and damage about 600 plant species (Nombela and Muniz 2010). Whiteflies are small fly like insects seen on the underside of leaves. Both adults and nymphs suck the sap from the leaves and transmit the tobacco leaf curl virus disease to the healthy seedlings/plants. The virus infected plants are stunted and twisted; leaves are puckered and thickened with abnormally prominent veins. “B” biotype was found to infest tobacco and transmit TLCV.

Ground beetles (*M. villiger*): These beetles damage newly transplanted tobacco plants by gnawing/cutting the tender stem, resulting in death of the seedlings causing up to 50–60% gaps in the field (Sitaramaiah et al. 1999).

Tobacco aphid (*M. persicae nicotianae*): In case of heavy infestation, hundreds of aphid can be seen on the underside of the leaf. By constantly sucking the sap from the leaf they debilitate the plant and there-by retard the growth. They secrete sugary juice known as ‘honey dew’ on the leaf due to which sooty mold develops rendering the leaf unfit for curing. In addition, they also transmit virus diseases eg. CMV, *Rosette* or *Bushy Top Virus* etc.

Tobacco budworm (*H. virescens*): It is principally a field crop pest, attacking such crops as tobacco, alfalfa, clover, cotton and soybean. The budworm larvae make holes in shoots and flower buds. Sometimes larvae can be found on the growing tips, the leaf petioles and the stems. In the absence of reproductive tissue, the larvae would feed on leaf material.

Budworm/Capsule borer (*H. armigera*): It is a polyphagous pest. During the vegetative phase (30–50 days), it feeds on the terminal bud and surrounding young leaves causing loss. Generally, one larva is seen on terminal buds in earlier stages and during flowering phase more than one borer per plant is seen. During flowering and capsule formation stage, larvae feed on flower buds and bore the capsules to feed on the developing seeds.

10.2.3.2 Fungal Diseases

Damping off (*Pythium* spp.): Damping off is one of the most important diseases of tobacco nurseries and is responsible for poor stand of seedlings or complete loss of nursery beds. It is caused by several soil inhabiting fungi predominant being *P. aphanidermatum* (Edson) Fitz; *P. debatyanum* Hesse, *Phytophthora* spp. and sometimes *Rhizoctonia solani* are also involved. The disease attacks the root or stem region near the soil surface. The disease may appear at any stage of the seedlings but maximum damage is observed 5–6 weeks after sowing causing rotting of the tiny seedlings. Older seedlings show shriveling and dark brown discoloration of stem at the base and ultimately collapse and topple over. The wet rotting and collapse of seedlings start in circular patches and may extend to the entire bed, if unchecked.

Leaf blight and Black shank (*P. parasitica* f. sp. *nicotianae*): Disease occur both in nursery and field crop. Young tiny seedlings in the nursery rot and die suddenly. Seedlings show blackening of roots and stem at ground level. Under continuous wet weather conditions, large circular to irregular water-soaked patches appear on the leaf

surface causing leaf blight. Symptoms of black-shank on the transplanted tobacco are seen in the form of blackening of roots and stalk. Blackening of the stalk starts at the base near the soil gradually extending upwards up to 30 cm or more above the ground level. The leaves turn yellow and the whole plants wilt and die.

Brown spot (*A. alternate*): It is a disease of senescence. The symptoms on older leaves appear as small water-soaked lesions which enlarge quickly. Once the spots enlarge, the center of the spots die and become brown, leaving a clear demarcation between diseased and healthy tissue. Circular brown spot lesions with concentric rings appear on lower leaves. In severe infection spots enlarge, coalesce and damage large areas making leaf dark brown, aged and worthless.

Blue mold of tobacco (*P. hyoscyamif. sp. tabacina*): Blue mold is one of the most important foliar diseases of tobacco that causes significant losses in the America, south-eastern Europe and the Middle East. It is highly destructive to tobacco seed beds, transplants and production fields. Single or groups of yellow lesions appear on the older and shaded leaves. The spots, often, grow together to form light brown necrotic areas. Leaves become puckered, distorted, large portions disintegrated and may lead to fall apart of the entire leaf. Blue mold can destroy all leaves at any growth stage in the event of continuous favorable weather conditions. Lesions may also appear on buds, flowers, and capsules.

***Fusarium* wilt (*F. oxysporum* f. sp. *nicotianae*):** *Fusarium* infection causes chlorosis, wilting and necrosis of tobacco leaves that leads to stunted growth and death. The symptoms often appear vertically on one side of the plant or even one side of the leaf midvein. Diagnostic chocolate-brown discoloration of the vascular tissue develops up to top of the plant. In due course of time, the discoloration on the exterior of the green stalk becomes visible.

Frog eye leaf spot (*C. nicotianae*): Generally, this disease is seen 4–5 weeks after germination and 30 days after transplanting and on the harvested crop. Round shaped brown spots akin to frog-eye form appear on the lower leaves of the seedlings. Spots appear initially on lower leaves and spread gradually to upper leaves. Under hot, dry weather frog eye lesions may be pin point in size and would not be recognized.

Powdery mildew (*E. cichoracearum* var. *nicotianae*): Powdery mildew, also known as white mold or ash disease, occurs in all types of tobacco in several Asian countries, Oceania, the Mediterranean, Africa and Canada. The disease causes severe damage in flue-cured tobacco compared to other types of tobacco. Under favorable conditions of low temperatures (16–23 °C) and high humidity, white patches spread to upper leaves, enlarge and cover the entire surface of the leaf. Mildew affected leaves get scorched on curing and show brown patches or blemishes rendering them unfit for marketing or reduce their commercial value.

Black root rot (*T. basicola*): Infected roots appear dark brown or black due to the presence of large numbers of black spores and rotting of root tissues greatly reduces the number of roots. Large scars may be seen on the main tap root. The above-ground symptoms show temporary nutrient deficiency symptoms, stunting and irregular growth of the tobacco.

10.2.3.3 Bacterial Diseases

Angular leaf spot (*P. syringae* pv. *angulate*): *Pseudomonas syringae* pv. *angulate* causes angular dark brown to black colored spots surrounded by yellow halo. Lesions are restricted between veins and leaves appear puckered and tears easily. It was regarded as a mutant of *Pseudomonas syringae* pv. *tabaci*, which does not produce tobtoxinine (Braun 1955). *P. syringae* pv. *tabaci* causes a severe type of angular leaf spot well known as wild fire in Tamil Nadu, India (Gnanamanickam et al. 1977).

Bacterial leaf spot (*D. dadantii*): It affect *bidi* and late planted FCV tobacco in Karnataka, India. Plant develops circular water soaked yellow lesions with a minute brown centre, which expand with a translucent border and wide chlorotic halo (Wolf and Foster 1917). Causes vascular discoloration of the stems, wilting and stunting (Johnson 1923; Komatsu et al. 2002).

Bacterial wilt (*R. solanacearum*): Also known as ‘Granville wilt’, the disease affects both nursery and field crop. In the field, the first symptom of the disease is drooping of 1–2 leaves during day and their recover during evening. One half of the affected leaves become flaccid, a characteristic symptom of bacterial wilt of tobacco. On slow progression of the disease, the affected leaves turn light green and may gradually turn yellow, midribs and veins get flaccid and large leaves may droop in an umbrella like fashion.

Hollow stalk (*E. carotovora* sub sp. *carotovora*): Hollow stalk has been reported from U.S.A, Canada, India and China (Roy et al. 2008). The disease may appear at any time due to stem injury but it is commonly observed 35–40 days after topping operations. Pith undergoes rapid browning and hollowing due to soft rot and tissue collapse eventually. Initially, the top leaves wilt and the infection slowly spreads downward. Black leg phase of the disease is characterized by the formation of black stripes or bands girdling the stalk and cured leaves.

10.2.3.4 Viral Diseases

Tobacco Mosaic Virus (TMV): TMV is worldwide in distribution and reduces cured leaf yield, quality and price. It is sap transmissible and spreads mainly through mechanical means. Characteristic symptoms include irregular mosaic pattern of dark and light green areas on leaf, leaf malformation and stunted plant growth. Infected young leaves are often malformed and show puckering or wrinkling. “Mosaic burn” with large, irregular, burned or necrotic areas appear on affected mature leaves causing extensive loss to the crop.

Potato Virus Y (PVY): This virus is aphid transmitted and affects tobacco crop worldwide. The virus is transmitted by aphids. This virus causes tobacco veinal necrosis or vein banding *i.e.* appearance of dark green bands along the brown necrotic veins. Necrosis extends to the vascular region and plants die out of pith necrosis.

Tobacco Leaf Curl Virus (TLCV): Caused by a virus of the genus begomovirus belonging to Geminiviridae. TLCV is transmitted by the insect vector, whitefly, *B. tabaci*. Diseased leaves become brittle, puckered, exhibit downward curling of

margins with enations or leafy outgrowths on the under surface of leaves. Leaves show vein clearing, abnormal vein thickening and twisting of petioles. Internodes get shortened resulting in dwarfing of the plant. Late infected plants show mild symptoms with small sized inwardly curled top leaves.

Cucumber Mosaic Virus (CMV): CMV enters into the leaf through wounds, principally those made by aphid. Infected plants show typical mottling and mosaic patterns, narrowing and distortion of the leaves, sometimes accompanied by plant stunting. Under severe infection, vein banding, inter veinal yellowing are observed accompanied with leaf blisters, shriveled, chlorotic and necrotic lines causing filiform leaves. Mild CMV strains would cause a faint mottling of the leaves.

Tobacco Distorting Virus: This virus causes abnormal suckering, stunting of the plant. Leaves get mottled, puckered, distorted with rat-tails. Only long midrib is seen without lamina at times.

Tobacco Etch Virus: It is transmitted by sap and also by aphids in a non-persistent manner. It infests solanaceous plants, and causes rat-tailing. On lower leaves vein clearing is seen along with some necrotic lines or etching. Mottling is seen with chlorotic and necrotic spots.

10.2.3.5 Nematode Diseases and Broomrape

Meloidogyne spp.: In most countries *Meloidogyne* spp. (root-knot nematodes) are considered as major factors limiting the production of tobacco crops. The nematode infection causes root galls and plant stunting. As nematodes damage plants at root level, above ground symptoms may not be prominent. At moderate to high infection, plants may appear wilted, yellowed and stunted.

Globodera spp.: *Globodera* presence on the roots is evident by the numerous cysts scattered over young or old roots with resultant reduced growth. The cysts of the species vary in sizes. Cyst nematodes cause lesser damage on tobacco compared to *Meloidogyne* spp.

Broomrape (*Orobanche* spp.): Broomrape affects the growth, development and morphogenesis of tobacco in the main field (Krishnamurthy 1994).

10.2.4 Control Methods

10.2.4.1 Cultural Methods of Control

Selection of disease free nursery site, deep summer ploughing, raising the seed beds, rabbing seed beds, use of optimum seed rate and regulation of watering to avoid dampness etc. controls fungal diseases in the nursery.

Deep summer ploughing, Soil fumigation, crop rotation, intercropping, companion/boarder crops (maize, sorghum, Tagetes etc.), trap cropping (castor as ovipositional trap for insects; sesame, jowar, black gram, green gram etc. for

Orobanche), application of fermented farm yard manure (avoid *Orobanche* seed) etc. have been highly effective in the management of insect, nematode, viral diseases and *Orobanche* weed (Chari et al. 1999; Sreedhar et al. 2007).

The phyto-sanitary measures that are adopted for controlling insect and diseases in tobacco include adopting right time of planting, transplanting of healthy seedlings, removal and destruction of diseased/virus infected plants and alternate hosts from the field, use of recommended fertilizer doses (including potash and avoiding excess nitrogen), collecting and destroying the crop debris after harvest etc.

Planning inter-culture operations in infected fields at the end, disinfection of implements before entering healthy fields, washing hands with soap water before and after entering infected fields, prevention of smoking and use of other tobacco products, etc. effectively control viral diseases and *Orobanche* (Lucas 1975). Physical removal of *Orobanche* before its flowering and burning reduce *Orobanche* incidence in the next season (Krishnamurthy 1994).

10.2.4.2 Chemical Methods of Control

Chemicals recommended and their doses for the control of insects and diseases vary from country to country. Guidance Residue Levels (GRLs) have been developed by the CORESTA (Cooperation Centre for Scientific Research Relative to Tobacco) Agro-Chemical Advisory Committee (ACAC) for guiding tobacco growers and those in the tobacco industry interested in Crop Protection Agents (CPAs) application and implementation of Good Agricultural Practice (GAP) in tobacco production (CORESTA 2020). The importing countries buy tobaccos when the pesticide residues in the cured leaf are below established GRLs and Maximum residue level (MRL) fixed by them. In view of the increasing stringency of regulations controlling the registration and use of CPAs resulted in an ever decreasing list of insecticides registered and recommended for use on tobacco (CORESTA 2020). Hence, new selective and low active ingredient (ai) insecticides, bio-pesticides and novel molecules are being evaluated and recommended to cope up with pesticide residue problem (Sreedhar 2020).

10.2.4.3 Biocontrol Methods with Natural Products and Biotic Agents

In view of the increasing concern about pesticide residues in tobacco, biopesticides viz. *Bacillus thuringiensis* var. kurstaki based biopesticides, Neem seed kernel suspension (NSKS), Nuclear Polyhedrosis Virus (NPV), Fungal pathogens viz. *Nomuraea rileyi*, *Beauveria bassiana*, *Verticillium lecani*, *Trichoderma harzianum*, *T. viride* etc. and bacterium, *Pseudomonas aeruginosa*, VA mycorrhiza *Glomus fasciculatum*, *Azotobacter* etc. have emerged as a strong component of integrated pest management strategies in tobacco (Chari et al. 1996; Sreeramulu et al. 1998; Ramaprasad et al. 2000; Dam et al. 2010; Sreedhar et al. 2014).

10.2.4.4 Integrated Management of Pests and Disease (IPM/IDM)

Integration of host plant resistance, cultural, chemical and biological management are always necessary for effectively controlling various diseases, nematodes and *Orobanche* without environmental pollution and for minimum residues of Crop Protection Agents (CPA), which is an issue of concern in exported tobacco. Bio-intensive IPM modules with genetic, cultural, biological and habitat management techniques as major components along with need based use of selective insecticides were highly effective in reducing pest damage, enhancement of natural enemy activity and helps in increasing the production of residue free tobaccos with favorable economics (Rao et al. 1994; Sitaramaiah et al. 2002; Sreedhar and Subbarao 2014).

10.2.5 Traditional Breeding Methods

Tobacco is a self-pollinated crop with very low natural out-crossing and all the breeding methods that are applicable to self-pollinated crops, such as introduction, mass selection, pure line breeding, pedigree method, back cross breeding, hybridization etc., are being utilized in tobacco (Bowman and Sisson 2000; Sarala et al. 2012). Introduction has played an important role in the initial years till 1960s in introducing tobacco cultivation into European, Australian and Asian countries. Pure line selection, hybridization of selected parents followed by selection in the segregating populations through pedigree, back cross and recurrent selection methods was the most pre-dominant methods adopted in breeding high yielding tobacco varieties worldwide. Back-cross breeding was used for the efficient transfer of a number of resistant factors viz. TMV (from Vamorr-50), Powdery mildew (Kofun), Black shank (Beinhart 1000-1), Caterpillar (DWFC), *Fusarium* wilt (Speight G33) etc. into tobacco varieties, from other tobacco types and wild species (Sarala et al. 2012). Backcrossing to the common tobacco variety was practiced to eliminate the genetic drag due to undesirable alleles and to recover the plant type and quality characteristics of the adapted variety. Distant hybridization/Interspecific hybridization was used to introgress useful genes like resistance to TMV, wild fire, black shank, brown spot, black root rot, blue mold, aphid, tobacco caterpillar, root knot nematode, powdery mildew, *Tomato Spotted Wilt Virus* (TSWV), PVY, cyst nematode etc. (Milla et al. 2005; Sarala et al. 2012). Incompatibility barriers in interspecific hybridization were overcome through utilization of hormones, bridge cross technique and in vitro rescue methods (Ramavarma et al. 1980). Mutation breeding has also played an important role in creating variability and developing high yielding improved tobacco varieties and parental materials for use in breeding programs (Sarala et al. 2012).

10.2.6 Use of Morphological Markers

The characters which can be readily detected by phenotype and useful to identify and characterize plants is referred to as morphological marker. In tobacco, these markers are mostly related to variation in plant (shape, habit, height, intermodal length etc.), leaf (number of economic leaves, size, color, shape, margin, tip shape, maturity interval period, etc.) floral characters (time of 50% flowering, size, color, development of stamens, height of pistil relative to stamens, etc.), capsule (shape and size), and seed (shape, color and surface characters) (Sarala et al. 2018). Morphological markers generally require neither sophisticated equipment nor preparatory procedures for scoring. Monogenic or oligogenic morphological traits are generally simple, rapid, and inexpensive to score, even from preserved specimens (Bretting and Widrechner 1995).

Morphological, karyotypical and physiological characters have been used to study the genetic background of tobacco (Goodspeed 1954; Zhang 1994; Lu 1997). Morphological markers played an important role in breeding improved tobacco varieties till date and continue to be effective in future also. As leaf is the economical product, improvement in number of harvestable leaves and leaf weight is important in realizing higher yields in tobacco along with desirable chemical quality characters. Sarala et al. (2005) observed continuous improvements for plant height, total leaves, harvestable leaves, days to flowering, leaf area, leaf growth rate, specific leaf weight and carotene from old varieties to recently released tobacco varieties. They further suggested that improvement in leaf number, leaf area and specific leaf weight in future cultivars can results in obtaining higher yields.

Conventional plant-breeding approaches in tobacco, as in other crops, rely on morphological markers representing desirable agronomic and product characteristics (that is, phenotypes including resistance to biotic stresses) for the selection of the parents, for creation of variation through their crossing or mutagenesis from the tobacco germplasm and for identification of targeted genotypes in segregating generations. Resistance related morphological characters are only expressed when there is an incidence of pests/diseases and such characterization may not be possible under natural conditions. These morphological traits are limited in number, influenced by the plant growth stages and various environmental factors (Eagles et al. 2001) because of which phenotypic identification could be misleading due to complex genotype and environment interaction that governs the trait of interest. Consequently, development of resistant varieties requires more than twelve years in case of recessive traits. As these markers are generally expressed late into the development of tobacco plant and are highly influenced by the environmental factors or growing conditions, their detection is dependent on the development stage of the crop. Further, they are less polymorphic, and exhibit dominance, pleiotropy and epistasis. In view of these limitations, genotypic markers that can be identified in early stages without environmental influence are highly useful in accelerating the tobacco varietal improvement.

10.2.7 Limitations and Prospect of Genomic Designing

Conventional biotic resistance breeding approaches in tobacco are phenotype-based, time-taking and resource-intensive. Hence, the progress thus obtained through traditional tobacco breeding is slow and is hampered by linkage drag and lack of easily transferrable sources of resistance. The genome designing strategies overcome this as they depend on the knowledge of the genome composition of the plants and gene sequence information. Targeted modification or designing of plant genome including addition of alien genes will accelerate the tobacco varietal developmental process through precise manipulation of gene functions for higher yields and stress resistance. Transfer of resistance from distance wild relatives and other unrelated sources to cultivated tobacco can also be successfully made using genome designing through trans- and cis-genesis approaches involving various processes viz. gene mapping, identification, gene transfer, gene editing etc. Marker-assisted and genome assisted breeding are going to be the order of the day with rapidly evolving technological advancements.

Availability of draft genome of *N. tabacum* and few wild species and various data sharing and analysis platforms (databases) in recent times, making it possible to understand genes, their sequences and linked molecular markers for target traits with the involvement of innovative bioinformatics tools and comparative genomic studies. The information thus available can effectively utilized to edit the genome sequences with rapidly evolving relatively precise gene editing technologies viz. meganucleases, zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), homing endonucleases, clustered regularly interspaced short palindromic repeat (CRISPR)/CRISPR-associated protein (Cas) 9 etc. However, gene editing technologies yet times do suffer from lower specificity due to their off-targets side effects (Khan 2019). Availability of high density of molecular maps and genome information further helpful in providing the knowledge of linked molecular markers and QTLs that are either closely linked or with in the target gene (s) and also allow map-based cloning of target traits. Linked markers and QTLs identified in tobacco for various biotic stresses is going to pave the way for marker-assisted breeding for resistant traits. The available information can be effectively utilized for estimating the breeding value of individuals in genome assisted breeding and accordingly selections can be made.

However, genetic engineering tools have certain limitations, including time-consuming and complicated protocols, potential tissue damage, DNA incorporation in the host genome, and low transformation efficiency. Unlike tradition breeding strategies, genome designing technology is resource intensive and require technology expertise for handling the processes.

10.3 Genetic Resources of Resistance/Tolerance Genes

Availability of stable and heritable sources of resistance to biotic stresses is essential in any crop while breeding resistant varieties. The gene pool consists of various easily crossable tobacco lines and *Nicotiana* species are to be explored for available variability and resistance factors in developing tobacco cultivars resistant. In case of non-availability in any of these gene pools it needs to be created through mutations or incorporated through genome designing.

Currently, large number of *Nicotiana* species and cultivated tobacco varieties are available (Lewis 2011). Around 83 species are available within the genus *Nicotiana* (Berbec and Doroszewska 2020). Currently there are 92 *Nicotiana* species and varieties listed at the Taxonomy Browser of National Center for Biotechnology Information (NCBI) (<https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi>) and 307 records of *Nicotiana* species and varieties available in The Plant List database (<http://www.theplantlist.org/tp1.1/search?q=Nicotiana>). International Plant Name Index (<http://www.ipni.org/ipni/plantnamesearchpage.do>) contains 450 records on the keyword '*Nicotiana*'. Large number of these species are reliable sources of resistance to biotic stresses that infest tobacco crop (Lewis 2011). Another important advantage 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. Various available sources of resistance in tobacco genetic resources are discussed below.

10.3.1 Primary Gene Pool

The primary gene pool consists of tobacco genotypes that can easily be crossed to produce fertile offspring with the cultivated tobacco. They may be cultivated ones and in wild gene pools. The cultivated gene pool covers commercial varieties of the crop, as well as landraces. While the wild gene pool comprises putative ancestors and closely related species that show a fair degree of fertile relationships with the domesticated tobacco. Large number of varieties are developed by breeders in different countries and fairly large collections of germplasm are available in *N. tabacum* and *N. rustica* that are resistant to various pests and diseases (Table 10.3). As gene transfers from such sources is easy, the first priority of the breeders is to explore them for the availability of resistance to target biotic stresses and transfer them to cultivated varieties through either traditional method of breeding or genetic engineering.

Table 10.3 Few genetic resources of resistance/tolerance available in the primary gene pool of tobacco

S. No	Disease	Resistance sources	Reference
1	Leaf blight and Black shank	Beinhart 1000-1, Coker 411, F 180, F 210, MC 1610, Reams 64 and Va 770 resistant to all the races of <i>P. parasitica</i> var. <i>nicotianae</i> Coker 371, KT 200, Beinhart Coker 371-Gold, Oxford 1, NC 95, Speight NF3 16 Cigar wrapper tobacco lines <i>Jati</i> tobacco: Vaishali Special, PT-76 and Gandak Bahar Banket A1 KY 15, KT 17 PB19	Abdul Wajid et al. (1987) Martinez et al. (2011) Bowman and Sisson (2000) Monga and Dobhal (1987) Shenoi et al. (2002) Lapham (1976) Martinez et al. (2011) Bai et al. (1995)
2	Brown spot	Colombian variety 'Ambalema' 8 FCV, 10 Burley, 10 Air cured, 3 rustica, 2 each of hookah and cigar and 1 each of <i>natu</i> and snuff types Banket A1 TN 90, Burley 21, Kentucky 56 TMVRR-1, TMVRR-2 and TMVRR-3	Nolla and Roque (1933) Reddy and Nagarajan (1981) Lapham (1976) Sierro et al. (2014) Sastri et al. (1981)
3	Black root rot	<i>Bidi</i> cultivars: MRGTH 1, GT 9 2001 Cheroot variety: S 5 FCV cultivars: VT-1158, Jayasri-MR, Godavari special, CTRI special -MR, CTRI Sulakshana and FCR-15	https://ctri.icar.gov.in/files/varieties.pdf
4	<i>Tobacco Mosaic Virus</i>	Coker 176, SC 71 and SC 72, NC 567 DG 3	Bowman and Sisson (2000) Gopalachari (1984)
5	<i>Tobacco Distorting Virus</i>		

(continued)

Table 10.3 (continued)

S. No	Disease	Resistance sources	Reference
6	Poty Virus	TN 86	Miller (1987)
7	Tobacco Vein Mottling Virus (TVMV) and Tobacco Etch Virus (TEV)	TI 1406 (Virgin A Mutant) TN 86	Noguchi et al. (1999) Miller (1987)
8	Bacterial wilt	Oxford 26	Bowman and Sisson (2000)
9	Granville wilt	NC 95, Speight NF3	Bowman and Sisson (2000)
10	Wild fire	NC 567	Bowman and Sisson (2000)
		Banket A1	Lapham (1976)
11	Hollow stalk	<i>N. rustica</i> germplasm accessions, White Pathar and Bengthuli	ICAR-CTRI (2010)
12	Budworm	CU 263 I 514, TI 1687, TI 536, TI 675, TI 1068, TI 1656, CU 1097	Bowman and Sisson (2000) Sreedhar (2011)
13	Aphid	TI 1687, TI 675, TI 1068, TI 1656, CU 1097, KDH 959	Sreedhar (2011)
14	Whitefly/Leaf Curl	I 514, CU 1097 TI 1687, TI 675, TI 1068, TI 536, KDH 959	Sreedhar (2011)

(continued)

Table 10.3 (continued)

S. No	Disease	Resistance sources	Reference
15	Tobacco caterpillar	DWFC, Abhirami (CR) I 514, CU 263, TI 536, TI 1687, CU 1097, KDH 959 and TN 86	Joshi et al. (1978) Sreedhar (2011)
16	Root-knot nematodes	Florida 22, Speight NF3 Bhavya	Bowman and Sisson (2000) https://etri.icar.gov.in/files/varieties.pdf
17	Cyst nematodes	NC 567	Bowman and Sisson (2000)

10.3.2 Secondary Gene Pool

The secondary gene pool referred to wild relatives of tobacco that are discrete from the cultivated species and still closely related in such a way that are crossable with cultivated species to at least certain extent to produce few fertile offspring. Genetic resources collected from gene centers as closely related species, primitive cultivars, old land races evolved and adopted to different environments are valuable source for resistance to biotic stresses. Majority of *Nicotiana* species (58 No.) would hybridize either with *N. tabacum* or at least with one other sister *Nicotiana* species. Some (*N. mutabilis*, *N. petunioides*, *N. attenuata*, *N. corymbosa*, *N. linearis*, *N. burbridgeae*, *N. thyrsoiflora*, and *N. wigandoides*) hybridize with at least one other *Nicotiana* but not with *N. tabacum* (Berbec and Doroszewska 2020). *N. tabacum* found to produce inviable hybrids with *N. africana* ($2n = 46$), *N. excelsior* ($2n = 38$), *N. goodspeedii* ($2n = 40$), *N. gossei* ($2n = 36$), *N. maritima* ($2n = 32$), *N. megalosiphon* ($2n = 40$) and *N. velutina* ($2n = 32$) after crossing at 28 °C (Tezuka et al. 2010). However, Type II hybrid lethality with the characteristic symptoms of browning of hypocotyl and roots observed in these crosses was suppressed at elevated temperatures (34 or 36 °C). Utilization of genes from these materials in breeding program is tedious due to incompatibility and undesirable linkages. Modern molecular and biotechnological tools can be an aid in overcoming such difficulties.

10.3.3 Tertiary Gene Pool

This pool made up of even more distantly related crop wild relative species. Fourteen species (*N. azambujae*, *N. acaulis*, *N. ameghinoi*, *N. paa*, *N. cutleri*, *N. longibracteata*, *N. spegazzini*, *N. faucicola*, *N. fatuhivensis*, *N. heterantha*, *N. stenocarpa*, *N. truncata*, *N. monoschizocarpa*, and *N. symonii*) had no hybridization records (Berbec and Doroszewska 2020) with *N. tabacum*. Genes present in both the S and T subgenomes of *N. tabacum* appear to be responsible for hybrid lethality in crosses with incompatible *Nicotiana* species (Tezuka and Marubashi 2012). Use of specific breeding techniques, such as bridge crosses, various sorts of treatments of male and female flower parts, and in vitro techniques like ovary/ovule culture and embryo rescue, chromosome and genome manipulation, such as increasing (polyploidization) or decreasing (haploidization) the number of genomes, different methods of partial genome transfer that include chromosome addition and substitution lines, mutagenesis and cell fusion, translocation breeding, exchange of nuclear and cytoplasmic genomes (mitochondrial and/or chloroplastic), grafting, marker-assisted breeding (MAB), in vitro tissue culture etc. and genetic engineering are needed to transfer genes from such pools (Weil et al. 2010).

10.3.4 Artificially Induced/Incorporated Traits/Genes

When source of resistance is not available in any of the above pools, creation of resistance through mutations (physical and chemical mutagens), genetic engineering for transfer of alien genes, gene manipulation and genome editing technologies are to be adopted in developing resistant cultivars.

Some source of resistance for biotic stresses available in primary gene pool of cultivated tobacco are provided at Table 10.3. Various resistant factors introgressed to *N. tabacum* from *Nicotiana* Species (secondary and tertiary pools) are listed at Table 10.4.

10.4 Glimpses of Classical Genetics and Traditional Breeding

10.4.1 Classical Mapping Efforts

In tobacco, initial gene mapping efforts were made by Anderson and De Winton (1931), East (1932), and Brieger (1935) with the identification of linkage between a pollen color factor and the sterility factors. Brieger estimated that S and br are about 20 crossover units apart in the first linkage group. Brieger (1935) has summarized the data on linkages in *N. langsdorfii* and *N. sanderae*, and established with some degree of certainty the first two linkage groups (1) Self-sterility allele (S) and lethality (I), (2) C is the basic gene for anthocyanin color (C) and a recessive gene causing a peculiar type of growth cr (crassa). Smith (1937) confirmed the existence of linkage between self-sterility and pollen anthocyanin color. The first accounts of monosomics in *N. tabacum* was reported by Clausen and Goodspeed (1926) and they established two types of monosomics in which haplo-C (called “corrugated”) found to be associated with the chromosome where the basic color factor, Wh, is positioned. Later, Clausen and Cameron (1944) developed complete set of 24 monosomics and studied the association in transmission between monosomics and mendelian characters which led to location of 18 genes on nine chromosomes. Though inheritance of number of traits studied and the linkage of few of them established with other traits, genetic linkage maps are not fully developed in tobacco due to its allopolyploid nature (Suen et al. 1997; Narayanan et al. 2003).

Table 10.4 Biotic stress resistance traits introgressions to *N. tabacum* from wild *Nicotiana* species

Resistance trait	Source	Reference
Tobacco Mosaic Virus	<i>N. glutinosa</i>	Holmes (1938), Tbbnovsky (1941), Valleau (1952), Apple et al. (1963)
Wild fire and angular leaf spot	<i>N. longiflora</i> (against <i>P. syringae</i> pv <i>tabaci</i> race 0)	Clayton (1947)
	<i>N. rustica</i> (against <i>P. syringae</i> pv <i>tabacirace</i> 0 and 1)	Stavelly and Skoog (1976), Woodend and Mudzengerere (1992)
Black shank (<i>P. parasitica</i> var <i>nicotianae</i>) (race 0)	<i>N. longiflora</i>	Valleau et al. (1960)
	<i>N. plumbaginifolia</i>	Chaplin (1962)
	<i>N. rustica</i>	Woodend and Mudzengerere (1992)
Brown spot	Immune- <i>N. debneyi</i> ; Resistant- <i>N. exigua</i> , <i>N. glutinosa</i> , <i>N. longifolia</i> , <i>N. nesophila</i> and <i>N. plumibaginifolia</i>	Reddy et al. (1976)
Black root rot	<i>N. debneyi</i>	Clayton (1969)
Blue mold	<i>N. velutina</i>	Clayton (1967), Clayton et al. (1967), Lea (1963)
	<i>N. debneyi</i>	Clayton et al. (1967)
	<i>N. goodspeedii</i>	Wark (1970)
	<i>N. goodspeedii</i> and <i>N. exigua</i>	Gillham et al. (1977)
Aphid	<i>N. gossei</i> , <i>N. repanda</i> , <i>N. trignophylla</i> , <i>N. umbratica</i> and <i>N. excelsior</i>	Thurstun (1961)
	<i>N. gossei</i>	Sarala et al. (2012)
Tobacco caterpillar	<i>N. gossei</i> , <i>N. benthamiana</i> , <i>N. repanda</i> , <i>N. nesophila</i> and <i>N. gossei</i>	Joshi and Sitaramaiah (1975), Sarala et al. (2012)
Root knot nematode (<i>M. javanica</i>)	<i>N. longiflora</i>	Ternouth et al. (1986), Sarala et al. (2012)
	<i>N. amplexicaulis</i>	Sarala et al. (2012)
	<i>N. repanda</i>	Ternouth et al. (1986)
	<i>N. tomentosa</i>	Clayton et al. (1958), Yi et al. (1998b)
Powdery mildew	<i>N. debneyi</i> , <i>N. glutinosa</i> , <i>N. tomentosoformis</i>	Smeeton and Ternouth (1992)
Tomato Spotted Wilt Virus	<i>N. alata</i>	Gajos (1988)
Potato Virus Y	<i>N. africana</i>	Lewis (2005)
	<i>N. tomentosiformis</i>	Legg and Smeeton (1999)

(continued)

Table 10.4 (continued)

Resistance trait	Source	Reference
Tobacco cyst nematode (<i>G. tabacum</i>)	<i>N. plumbaginifolia</i>	Chaplin (1962), Johnson et al. (2002)

10.4.2 Limitations of Classical Endeavors and Utility of Molecular Mapping

Though genes are very useful markers but they are by no means ideal for mapping studies. Mapping based on morphological markers is tedious and time taking and genes governing quantitative traits cannot be mapped (Worland et al. 1987). Limited number of visual phenotypes whose inheritance could be studied and complications arising in analyzing them due to pleiotropic effect of genes is one of the major limitations for gene (trait) mapping through classical approach. Allopolyploid nature of tobacco making it difficult to score individual gene effects and recombination events at phenotype level due to epistatic interactions between homeo alleles present on constituent genomes and their functional redundancy posing a problem in classical mapping.

For making the gene maps more comprehensive, identification of characteristics that were clearly distinctive and less complex than visual ones is necessary. But, only a fraction of the total number of genes in tobacco exist in allelic forms that can be distinguished conveniently making it difficult to construct classical maps. One of the reasons why our knowledge of the details of inheritance in tobacco was so meager, is because of the prevalingly quantitative or semi-quantitative nature of majority of character including flower color in tobacco (Clausen and Cameron 1944). Gene maps are, therefore, not very comprehensive in tobacco and a detailed map based entirely on genes is not available in tobacco.

As the plant breeding progressed, biochemical markers such as protein and isozyme markers were developed (Markert and Moller 1959). Protein and isozyme markers have been successfully applied in the detection of genetic diversity, population structure, gene flow and population subdivision in tobacco (Mateu-Andres and De Paco 2005). The isozymes and other proteins mostly have neutral effect on plant phenotype and are often expressed co-dominantly making them discriminate easily between homozygote and heterozygote. However, due to their limited availability and the requirement of a different protocol for each isozyme system, utilization of protein and isozyme markers in plant breeding programs and mapping endeavors is very limited.

Hence, there is a need for other types of marker which are abundantly available, stable over environments and can be easily classified into distinct categories. DNA based molecular markers developed in 1990 could satisfy these requirements. Because molecular markers are so abundant in a genome and can easily be detected, when mapped by linkage analysis, they fill the voids between genes of known phenotype. In mapping, DNA marker is not important in itself but, the heterozygous

site is merely a convenient reference point for marking the chromosomal locations. However, the molecular marker development and genetic map construction in tobacco have lagged behind other Solanaceae crops such as the tomato, potato, and pepper plants (Tanksley et al. 1992; Barchi et al. 2007). The molecular marker based maps can be effective anchoring points for identification of linked traits for their isolation, cloning and also for use in marker-assisted breeding.

10.4.3 Breeding Objectives

The tobacco breeding mainly aims at enhancing leaf yield potential of the cultivar in addition to maintaining leaf quality, and resistance to biotic and abiotic stresses. Tobacco leaf yield is a dependent variable and is the result of the associated yield attributing independent traits such as plant height, number of leaves, leaves length and width, days to maturity, resistance to biotic and abiotic stresses (Sarala et al. 2005). Some of these traits, in breeding improved high tobacco varieties, are to be selected in positive direction such as number of leaves, days to maturity, resistance, leaf length and width etc. while some of the traits such as plant height, leaf intermodal length, reducing sugars and alkaloids etc. in negative direction. Breeding for improved (higher level) resistance to various biotic stresses as well as finding novel or improved sources of resistance remains a central part of majority of the present flue-cured breeding programs.

Bowman and Sisson (2000) analyzed the relative change for several traits of flue-cured tobacco cultivars grown in the early 1960s compared to the late 1990s and they concluded that yield, leaves per plant and days to flower were selected in positive direction while plant height, intermodal length, reducing sugars and total alkaloids were selected in negative direction. Breeders were successful in increasing the yield despite maintaining leaf quality (both are negatively correlated). The newer cultivars were modified to have longer growing period, more leaves prior to flower bud formation, relatively lower plant height, internodal length, reducing sugars and total alkaloids. Sarala et al. (2005) observed positive selection for plant height, total leaves, harvestable leaves, days to flowering, leaf area, leaf growth rate, specific leaf weight and carotene in tobacco varieties.

10.4.4 Classical Breeding Achievements

Traditional tobacco breeding aimed at developing improved tobacco varieties with higher yield, better leaf quality, resistance to pests and diseases. Significant progress has been made over the years in enhancing the tobacco leaf yield through both varietal and hybrid development, in addition to improving disease and insect resistance without significantly sacrificing in ease of curing. The classical breeding achievements in tobacco are enumerated here.

10.4.4.1 Yield and Quality

Breeders have made progress in improving tobacco leaf yield and quality over years (Bowman and Sisson 2000; Sarala et al. 2005). A number of high yielding varieties have been released so far using conventional breeding methods (<https://crosscreekseed.com>; <https://content.ces.ncsu.edu/flue-cured-tobacco-information>; ICAR-CTRI 2021b). Release of two high yielding flue-cured tobacco cultivars viz. Coker 139 (in 1955) and K 326 (in 1981) have provided important germplasm for yield enhancement and have been used widely by breeders as parental lines since release. K 326 is the most successful tobacco cultivar in the 20th century with its cultivation in majority of the tobacco growing countries (Bowman and Sisson 2000). With the development of high yielding disease resistant lines and CMS lines, F₁ hybrids were developed in tobacco. Hybrid burley cultivars have been available since 1960s, but few flue-cured hybrids were released before the late 1990s (Bowman and Sisson 2000; Sarala et al. 2012). Key advantage of hybrids found to be the speed and ease of obtaining multiple disease resistant ones based on the selection of desirable parental lines having target traits. Several hybrids of Burley and FCV are released and are under cultivation worldwide (<https://crosscreekseed.com>; <https://content.ces.ncsu.edu/flue-cured-tobacco-information>; ICAR-CTRI 2021b). One of the main reasons for the popularity of hybrids is the repeal of the 1944 Federal Tobacco Seed Law that prohibited the sale of domestic tobacco seed in abroad. Hybrid cultivars ensure the proprietary status of a new cultivar and seed sales, and distribution can be controlled (Bowman and Sisson 2000).

10.4.4.2 Disease Resistance

Much of the early tobacco breeding carried out was concentrated in incorporation of resistance to diseases that were the primary factors limiting production (Bowman and Sisson 2000). The varieties developed in recent years found to combine disease resistance with high yield potential. Conventional breeding and interspecific hybridization were used to transfer disease resistance traits viz. TMV, wild fire, black shank, brown spot, black root rot, blue mold, root knot nematode, powdery mildew, TSWV, PVY, cyst nematode etc. (Bowman and Sisson 2000; Sarala et al. 2012). Most commercial varieties are highly susceptible to blue mold disease (Ruffy and Main 1989) and functional/partial resistance reported in *N. debneyi*, *N. velutina*, *N. goodspeedii* and *N. exigua* (Clayton 1968; Gillham et al. 1977; Wark 1970) have been transferred through interspecific hybridization (Milla et al. 2005). N gene conferring TMV resistance from *N. glutinosa* was successfully transferred to *N. tabacum* by bridge cross technique (Clausen and Goodspeed 1925; Holmes 1938). Majority of the present flue-cured tobacco cultivars possess resistance to important diseases such as TMV, black shank, Granville wilt and root knot nematodes. However, resistance to black shank and Granville wilt are variable among cultivars (Bowman and Sisson 2000). The five CMV genes from wild *Nicotiana* species were pyramided into the tobacco

line Holmes and resistance derived from Holmes for multiple virus resistance has been incorporated into many cultivars (Holmes 1960, 1961; Wan et al. 1983).

Once resistance to a single disease had been achieved, pyramiding resistance to several diseases into a single cultivar became a top priority (Bowman and Sisson 2000). This objective is being continued into the twenty-first century. Majority of the current flue-cured tobacco cultivars have resistance to Granville wilt, black shank, and root knot nematodes, even though levels of resistance is highly variable among cultivars. Many of the modern hybrids possess resistance to two or more biotic stress (<https://crosscreekseed.com>; <https://content.ces.ncsu.edu/flue-cured-tobacco-information>; ICAR-CTRI 2021b). The popular FCV cultivar, K 326 is resistant to the common root-knot nematodes and have low level of resistance to black shank and bacterial wilt (<https://crosscreekseed.com/usa-varieties/usa-flue-cured-tobacco-seed/k-326/>). While, popular burley tobacco cultivar, Banket A1 is resistant to brown spot, wildfire, and TMV (Lapham 1976).

10.4.4.3 Insect Resistance

Unlike in the case of disease resistance, incorporation of insect resistance is not met much success in tobacco. A tobacco variety, CU 263 having moderate level of resistance to tobacco budworms was released in 1995. Tobacco varieties with resistance to tobacco aphids (*Myzus nicotianae*) and tobacco caterpillar have been developed for commercial cultivation in India (Joshi and Sitaramaiah 1975; Joshi et al. 1978; Murthy et al. 2014). FCV variety, CTRI Sulakshana is tolerant to aphids and resistant to TMV. Meenakshi (CR) and Abirami (CR) are caterpillar resistance chewing tobacco varieties developed through back cross breeding (ICAR-CTRI 2021b).

10.4.5 *Limitations of Traditional Breeding and Rationale for Molecular Breeding*

Major limitation in traditional breeding in tobacco are undesirable gene associations, pleiotropic gene effects and linkage drag effects caused by the presence of deleterious genes linked to gene of interest (Legg et al. 1981; Zeven et al. 1983; Friebe et al. 1996; Brown 2002). Chaplin et al. (1966), Chaplin and Mann (1978) and suppressed recombination within introgressed chromatin (Paterson et al. 1990; Liharska et al. 1996) making it difficult to alleviate linkage drag effects through back crossing (Stam and Zeven 1981; Young and Tanksely 1989). Sources of resistance from easily crossable or that can be transferred through special techniques such as bridge crossing, embryo rescue, chemical treatments are required for developing resistance species.

With the advent of advanced molecular breeding techniques like, marker-assisted selection, plant transformation, CRISPER/CAS9 gene editing etc. only gene of interest can be modified or incorporated into the selected cultivar in short span of

time without these limitations. Molecular marker-assisted plant breeding for foreground and background selection with speed breeding strategies will help in saving time.

10.5 Brief on Diversity Analysis

The major interest to study the genetic diversity in tobacco is for the conservation of genetic resources, broadening of the genetic base and cultivar development in breeding programs. Narrow genetic diversity can lead to crop losses due to reduced flexibility of varieties to combat infestations of new strains of pests or pathogens or to adapt populations to changing environmental conditions such as increasing temperatures or salinity (Moon et al. 2009a). The study of the genetic diversity is important for identifying sources for economically important traits and diverse parental combinations to create segregating progenies with maximum genetic variability for further selection in breeding (Barrett and Kidwell 1998). Utilization of diverse parents helps in developing tobacco varieties having diverse genetic backgrounds with improvement in yield and quality along with stress resistance traits. Deploying varieties with diverse sources of resistance will reduce genetic vulnerability of cultivars to evolving races and biotype of insects and pathogens. Thus, genetic diversity is essential for continued progress in breeding varieties for adaptation to future environmental challenges.

10.5.1 *Phenotype-Based Diversity Analysis*

Phenotypic viz. morphological, karyotypical and physiological characters have been used to study the genetic diversity of tobacco germplasm (Goodspeed 1954; Zhang 1994; Lu 1997; Zhang et al. 2005). Till date several agro-morphological traits (Zhang 1994; Wenping et al. 2009; Zeba and Isbat 2011; Sarala et al. 2018; Baghyalakshmi et al. 2018), chemical and cytological traits (Tso et al. 1983; Okumus and Gulumser 2001; El-Morsy et al. 2009; Darvishzadeh et al. 2011) have been used to study the genetic variation of tobacco germplasm. Agro-morphological traits usually vary with environment and affect diversity estimates under different environments (Lu 1997). Germplasm are screened for identifying sources of resistance to various biotic stresses under artificial conditions based on the extent and type of damage caused to the crop through phenotyping.

The tolerance/resistance mechanism to various stresses is now being extensively studied through newer techniques of phenotyping namely high throughput phenotyping where the system quantifies a number of traits within a described set of plant population with automated image collection and analysis, thus effectively streamlining the plant phenomics. This technology is gaining importance due to its non-destructive sampling methods, rapid screening of larger population and automated

data analysis. To measure tolerance, visual cameras are used to capture the plant growth, architecture, chlorosis and necrosis all of which can be negatively affected by insect infestation. Photosynthetic activity of crop plants can be measured in terms of chlorophyll fluorescence through fluorescence cameras as an indication of tolerance mechanisms occurring in response to insect attack (Buschmann and Lichtenthaler 1998). The technology can have numerous applications in the measure of insect damage and plant resistance to insects (Goggin et al. 2015). Ultimately the use of high throughput phenotyping systems could possibly reduce the amount of labour and screening time to identify plants tolerant to insects.

10.5.2 Genotype-Based Diversity Analysis

Before the advent of DNA markers, morphological, karyotypical, physiological and isozyme markers are mainly used in diversity analysis of tobacco. However, morphological characters usually differ with environments. The karyotypical characters are limited in number and the study of genotypic diversity based on isozyme variation is restricted to a small number of loci that control few polymorphic enzyme systems (Lu 1997). Until now, only limited information has been available on the relationship between morphological variability and genome diversity in cultured tobacco. Later, attempts were made to examine the degree of relatedness among tobacco cultivars and diversity of germplasm based on variability at DNA level. About 77% of the total genomic DNA content is composed of repetitive sequences in tobacco and therefore, the remaining non-repetitive sequences are responsible for morphological and quality trait variability (Narayan 1987).

With the advantages such as highly polymorphic nature, codominant inheritance, easy access, fast assay, high reproducibility and easy exchange of data between laboratories, different types of molecular markers have been introduced over the last two decades, that has revolutionized the entire scenario of biological sciences including tobacco (Liu and Zhang 2008). Molecular markers provide a relatively unbiased estimation of genetic diversity in plants and are abundantly available throughout the genome. DNA-based molecular markers found to be versatile tools having applications in various fields such as characterization of genetic variability, genome fingerprinting, genome mapping, gene localization, analysis of genome evolution, population genetics, taxonomy, genome comparisons, gene mapping, quantitative trait loci analysis, marker-assisted breeding diagnostics, etc.

Restricted fragment length polymorphism (RFLP), randomly amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), microsatellites or simple sequence repeat (SSR), single-nucleotide polymorphism (SNP), inter simple sequence repeats (ISSR) etc. have been employed in studying genetic diversity, gene mapping and marker-assisted breeding of tobacco.

RFLPs were the most popular molecular markers in late eighties due to their reproducibility, locus-specificity, mendelian inheritance and codominant nature (Nadeem et al. 2018). RFLP methodology is simple and require no special equipment. Another

advantage of RFLP is that the sequence used as a probe need not be known RFLPs, being co dominant markers, can detect coupling phase of DNA molecules as DNA fragments from all homologous chromosomes are detected. RFLPs are highly reliable markers in linkage analysis and breeding as they can easily differentiate whether the linked trait is present in a homozygous or heterozygous state in an individual and such information highly desirable for recessive traits (Winter and Kahl 1995). However, the large amount of DNA required for restriction digestion and laborious Southern blotting technique hampered the utility of the assay. Further, it is time-taking and labor-intensive and only few out of numerous markers may be polymorphic making it highly inconvenient especially for crosses between closely-related species. RFLP markers were the first molecular markers used in tobacco research specially to study the function of few cloned genes (Bretting and Widrechner 1995).

Invention of polymerase chain reaction (PCR) technology and PCR-based markers such as RAPD and AFLP emerged in the beginning of nineties and later microsatellite markers were used by different workers to study genetic diversity in tobacco. Compared to RFLP, these PCR based markers are preferred because of the relative ease with which PCR assays can be carried out. RAPD technology utilizes short synthetic arbitrary oligonucleotides primers to amplify fragments. Its methodology is simple, time-saving, and requires only small amount of genomic DNA. Due to these advantages, RAPD technology has been widely used in studying genetic diversity, localization of target genes, genetic mapping and evolution genetics. However, RAPD are anonymous and the level of their reproducibility is very low due to the non-specific binding of short, random primers. AFLP technology combines the power of RFLP with the flexibility of PCR-based markers. It offers a universal, multi-locus marker tool that can be applied to complex genomes from any organism. Though the AFLPs markers are advantageous in terms of their reproducibility and sensitivity, their detection method is lengthy, laborious and not amenable to automation. Both RAPD (Xu et al. 1998; Del Piano et al. 2000; Evanno et al. 2005; Zhang et al. 2005, 2008a; Arslan and Okumus 2006; Sarala and Rao 2008; Sivaraju et al. 2008; Denduangboripant et al. 2010; D'hoop et al. 2010) and AFLP (Huang et al. 2008; Zhang et al. 2008a; Chuanyin et al. 2009; Liu et al. 2009a) were used to analyze the genetic diversity and varietal identification in tobacco. In tobacco, the RAPD technique has also been used to identify markers linked to genes for resistance to pathogens (Bai et al. 1995; Yi et al. 1998a, b; Johnson et al. 2002; Collard et al. 2005; Julio et al. 2006; Gong et al. 2020). Based on AFLP markers, five TMV-resistant tobacco accessions were found to carry a TMV resistance gene (N gene derived from *N. glutinosa* L.) on chromosome H and seven accessions were found to carry a resistance factor on an alternative chromosome (Lewis et al. 2005).

Soon after the discovery of SSR markers in late 90 s and the beginning of twenty-first century, they became markers of choice as they could able to eliminate all drawbacks of earlier DNA marker technologies (Jafar et al. 2012). SSRs were highly reproducible, polymorphic, and amenable to automation. In spite of higher working costs, SSR markers are extensively employed in plant molecular genetics and breeding. Development of PCR-based molecular markers is, in general, inefficient in tobacco in view of the very limited genetic diversity between tobacco cultivars, particularly

between those of the same type (Del Piano et al. 2000; Julio et al. 2006; Rossi et al. 2001) and that cultivated tobacco is a tetraploid species with a very large genome (Livingstone et al. 1999; Ren and Timko 2001; Doganlar et al. 2002). However, Bindler et al. (2007) for the first time reported the use of microsatellite markers in the identification of tobacco varieties. They used around 637 functional SSR markers out of which 282 were highly polymorphic and were used for variety identification. Since then, SSR markers are being used in estimating diversity in tobacco. Later, an additional set of 5119 new and functional SSR markers were developed for mapping and diversity studies (Bindler et al. 2011). Tong et al. (2012b) published a set of 1365 genomic SSRs and 3521 expressed sequence tag (EST)-SSRs, which only slightly overlapped with the set published by Bindler et al. (2007, 2011). Madhav et al. (2015) developed a new set of microsatellite markers using 70 motifs (which includes perfect and imperfect repeat) and validated their applicability in differentiating different types of tobacco and diverse cultivars of Flue Cured Virginia (FCV) tobacco, and the transferability of these markers in a wide range of *Nicotiana* species. Cai et al. (2015) used EST databases of tobacco for developing EST-SSR markers and validated them in studying the genetic differentiation between *N. rustica* and *N. tabacum*, and between oriental tobacco and other accessions of *N. tabacum*. Later, Wang et al. (2018) identified a huge number of SSRs through comparative genome wide characterization of ~20 Gb sequences from seven species viz. *N. benthamiana*, *N. sylvestris*, *N. tomentosiformis*, and *N. otophora*, and from *N. tabacum* cultivars TN90, K326, and BX representing a ~74%, ~80%, ~78%, ~81%, ~84%, and ~73% of each of the genomes, respectively. Their study resulted in the development of a total of 1,224,048 non-redundant NIX (*Nicotiana* multiple (X) genome) markers (SSRs), of which 99.98% are novel (Wang et al. 2018). SSR markers development in tobacco has led to the analysis of molecular diversity (Moon et al. 2009b; Davalieva et al. 2010; Fricano et al. 2012; Gholizadeh et al. 2012; Prabhakararao et al. 2012; Xiang et al. 2017) of genetic resources, DUS testing (Binbin et al. 2020), genetic relatedness of cultivated varieties (Moon et al. 2008), estimating the changes in diversity due to breeding interventions (Moon et al. 2009a) and also for the identification of linked markers and QTLs to various biotic stresses that were discussed later in this chapter.

Genetic diversity in tobacco varieties is also been assessed using Inter simple sequence repeats (ISSR) markers (Yang et al. 2005, 2007; Qi et al. 2006) and inter-retrotransposon amplification polymorphism (IRAP) markers (Yang et al. 2007).

In recent years, SNP markers, first discovered in human genome, proved to be universal, abundant, ubiquitous, and amenable to high- and ultra-high-throughput automation are being identified and used (Jafar et al. 2012). Although SNPs are less polymorphic than SSR markers, their biallelic nature compensate this drawback (Ghosh et al. 2002). The application of SNPs in tobacco is complicated and challenging due to its tetraploid nature and complex genetic architecture (Ganal et al. 2009). However, recent studies identified number of SNPs in tobacco (Xiao et al. 2015; Thimmegowda et al. 2018; Tong et al. 2020) that are being utilized in the identification of markers linked to important traits in tobacco entries including biotic stresses in addition to development of molecular maps, and studying genome structure and organization.

Wang et al. (2021) used genotyping-by-sequencing (GBS) technique on 113 cigar tobacco accessions for the identification of 47 core Kompetitive allele specific PCR (KASP) and 24 candidate core markers utilizing SNP data. KASP markers are able to discriminate between two alleles of a SNP using a common reverse primer paired with two forward primers, one specific to each allele. Core markers were used for varietal identification and fingerprinting in 216 cigar germplasm accessions.

10.5.3 Relationship with Other Cultivated Species and Wild Relatives

Cultivated tobacco belongs to the genus *Nicotiana*, one of the five major genera of the family Solanaceae. Comparative morphological, cytological and biochemical studies, and examinations of organellar (plastid and mitochondrial) genome organization and analysis of molecular features, such as repetitive DNA sequences and the structure of various nuclear gene family members, has been employed to study the evolution and the genetic diversity in genus *Nicotiana* (Kostoff 1943; Goodspeed 1954; Komarnitsky et al. 1998; Lim et al. 2000; Liu and Zhang 2008).

The genus, *Nicotiana* resembles the two genera, *Cestrum* (8 pairs) and *Petunia* (7 pairs of chromosomes) (Darlington and Janaki Ammal 1945) in habit and habitat. The origin, evolution and relationships among various species of the genus have been summarized in three phylogenetic arcs. In the first two arcs, the genus is envisaged as derived from a pre-generic reservoir of two related genera and evolving into three complexes, at the 12-paired level, that are hypothetical precursors of the three modern sub-genera. The third arc contains the present day species, at 12- and 24-paired chromosome level. Although 6-paired species of *Nicotiana* is not known, the predominance of 12- paired species and their compound morphological character, along with a frequency of 4–8 pairing with a mode of 6 pairs in large number of F₁ hybrids combining 12-paired species, indicates that 6 is the basic chromosome number for *Nicotiana* and both 12 and 24 are derived numbers. Additional evidence for this has been provided by secondary association and pairing in haploids of 12-paired species (Kostoff 1943; Goodspeed 1954). It is assumed that with a higher survival value, the allopolyploids have eliminated the older 6-paired types. Hybridization between the ancestral 6-paired and the present day 12-paired member was considered to be responsible for the evaluation of 9- and 10-paired species. The *N. tabacum* and *N. rustica* and other 24-paired species are modern descendants of the 12-paired progenitors entered into amphiploid origin. The aneuploid species of the section *Suaveolentes* with a chromosome number between 16 and 24 pairs are presumed to have originated as products of hybridization and segregation. Thus, among various evolutionary mechanisms, amphiploidy superimposed by amphiploidy seems to be the basic evolutionary process responsible for the 6- 12- 24-paired sequence in the genus (Goodspeed 1954). Further enlargement of the genus

took place through aneuploidy resulting from hybridization, genetic recombination and mutation.

The systematic classification of the genus was presented in detail by Goodspeed (1954) and Goodspeed and Thompson (1959) mainly based on cytogenetic studies involving chromosome morphology, behaviour in interspecific hybrids, aneuploids and amphiploids. Subsequently, additions to the classification were made by Burbridge (1960). In the revised systematics based on molecular research (Chase et al. 2003; Clarkson et al. 2004; Knapp et al. 2004), subgenera were dropped retaining the division into sections. As per the recent classification, *N. trigonophylla* Dun. was renamed *N. obtusifolia* Martens et Galeotti, *N. affinis* Hort is synonymous with *N. alata* Link et Otto, and *N. bigelovii* (Torrey) Watson with *N. quadrivalvis* Pursh. *N. sanderae* Hort. is considered to be a hybrid between *N. alata* and *N. forgetiana* Hemsl. (*Nicotiana x sanderae*) whereas *N. eastii* Kostoff is a variant of autotetraploid version of *N. suaveolens* Lehm. (Chase et al. 2003; Knapp et al. 2004).

N. tabacum and *N. rustica* are the cultivated species among the 83 wild species of *Nicotiana* genera (Lewis 2011; Berbec and Doroszewska 2020). *N. tabacum* is highly polymorphic with wide range of morphological types and diversified utilities viz. smoking, snuff, chewing, etc. Other species cultivated in smaller scale are *N. repanda* Willd ex Lehm., *N. attenuata* Torrey ex S. Watson and *N. quadrivalvis* Pursh are for smoking, *N. sylvestris* Spegazzini & Comes, *N. alata* Link and Otto, *N. langsdorffii* Weinmannm, *N. forgetiana* Hemsley, and *N. sanderae* (Hort) for ornamental and *N. glauca* Graham for industrial purpose (Lester and Hawkes 2001).

It is known that *N. tabacum* is natural amphidiploid i.e., allopolyploid ($2n = 4x = 48$) ascended by hybridization of wild progenitor species (*N. sylvestris* (S-genome) \times *N. tomentosiformis* (T-genome) and *N. rustica* L. from a cross between *N. paniculata*/*N. knightiana* (P/K-genome) \times *N. undulata* Ruiz & Pav. (U-genome) (Goodspeed 1954; Clarkson et al. 2005; Lim et al. 2005; Leitch et al. 2008; Edwards et al. 2017; Siervo et al. 2018). Comparison of whole-genomic sequences indicated that the genome of *N. sylvestris* and *N. tomentosiformis* contributed 53 and 47%, respectively, to the genome of *N. tabacum* (Siervo et al. 2014). The comparative mapping studies of the diploid ancestor of the T-genome (*N. tomentosiformis*) and a species related to the S-genome (*N. acuminata*) using conserved ortholog sequences (COS) and SSR markers revealed that the tetraploid tobacco genome has undergone a number of chromosomal rearrangements compared to these diploid genomes (Wu et al. 2009). Furthermore, in the same study, it was observed that a number of reciprocal translocations and inversions (>10) differentiate the ancestral tobacco genomes from the tomato genome. Gong et al. (2016) identified a large number of genome rearrangements occurring after the polyploidization event through mapping of ancestral genomes of *N. tomentosiformis* and *N. sylvestris* with SNP markers. Based on studying the cpSSRs and MtSSRs, S genome in *N. tabacum* was identified to originate from *N. Sylvestris* ancestor (Murad et al. 2002). The chloroplast genome of *N. otophora* revealed that *N. otophora* is a sister species to *N. tomentosiformis* within *Nicotiana* genus and *Atropa belladonna* and *Daturastramonium* are the closest relatives (Asaf et al. 2016).

Sierro et al. (2018) reported that 41% of *N. rustica* genome is originated from the paternal donor (*N. undulata*) and 59% from the maternal donor (*N. paniculata*/*N. knightiana*). Comparison of families of repetitive sequences, two from *N. paniculata* and one from *N. undulata* indicated that P- and U-genomes of *N. rustica* was similar to the putative parents, *N. paniculata* and *N. undulata*, respectively (Lim et al. 2005). Genomic in situ hybridization confirmed that *N. rustica* is originated as an allotetraploid from *N. paniculata* (maternal P-genome donor) and *N. undulata* (paternal U-genome donor) and interlocus sequence homogenization has caused the replacement of many *N. paniculata*-type intergenic spacer (IGS) of the 18-5.8-26S rDNA in *N. rustica* with an *N. undulata*-type of sequence (Matyasek et al. 2003). However, the study of nuclear and chloroplast genomes, and gene analyses showed that *N. knightiana* is closer to *N. rustica* compared to *N. paniculata*. Gene clustering revealed 14,623 ortholog groups common to other *Nicotiana* species and 207 unique to *N. rustica*. (Sierro et al. 2018).

Both *N. tabacum* and *N. rustica* shares its basic chromosome number of $n = 12$ with many other solanaceous species such as tomato, potato, pepper and eggplant (Lim et al. 2004; Clarkson et al. 2005). Microsynteny observed between the genomes of *N. tabacum* cv. TN90, K326 and BX and those of tomato and potato at the protein level (Sierro et al. 2014).

10.5.4 Relationship with Geographical Distribution

The genus, *Nicotiana*, considered to be recent origin, is presumed to have had its original habitat in and around the Andes region in South America and Central America, probably from the mild to low altitude forest margin (Goodspeed 1954). Although, occurring naturally as a perennial plant, tobacco is farmed as an annual crop. In the genus *Nicotiana*, 20 are native to Australia, one to Africa and 54 are indigenous to North /South America (Goodspeed 1954; Burbridge 1960; Clarkson et al. 2004). *N. benthamiana*, a species native to Australia, is used extensively as a model system to study various biological processes. *N. tabacum* L. (common tobacco) and *N. rustica* L, two species native to America, are the major cultivated species cultivated throughout the world.

Some 40% of *Nicotiana* species are allopolyploids, which have been generated independently in six polyploidy events several million years ago (Clarkson et al. 2004; Leitch et al. 2008). The taxonomically, some of the diploid genome donors that make up various allopolyploids are closely related and others belonging to distantly related taxonomic sections (Clarkson et al. 2004; Leitch et al. 2008).

Darvishzadeh et al. (2011) reported clustering of oriental-type tobacco genotypes based on morphological traits was in agreement with their geographical distribution and growth characteristics. However, the genetic diversity studies conducted using SSR markers in oriental (Darvishzadeh et al. 2011), and RAPD and AFLP markers in flue-cured tobaccos (Zhang et al. 2008a) could not indicate any clear pattern of their geographical origins. However, the clustering of tobacco genotypes based on

molecular diversity found to correspond to commercial classes (Flue-Cured, Burley, etc.), manufacturing trait and parentage (Sivaraju et al. 2008; Fricano et al. 2012).

10.5.5 *Extent of Genetic Diversity*

Existence of morphological diversity confirmed in various germplasm collections viz. FCV, burley, mutant lines, bidi, chewing, cheroot and cigar filler germplasm etc. (Baghyalakshmi et al. 2018; Sarala et al. 2018) maintained at tobacco genebank in India. Moon et al. (2009b) observed quite large average gene diversity among *N. tabacum* accessions from the U.S. *Nicotiana* Germplasm Collection compared FCV tobacco accessions. While, Fricano et al. (2012) reported lower SSR diversity per locus than in similar investigations carried out on TI accessions of tobacco. The genetic diversity assessment in tobacco cultivars indicated low degree of polymorphism (Xu et al. 1998; Del Piano et al. 2000; Rossi et al. 2001; Yang et al. 2005; Zhang et al. 2005; Julio et al. 2006). Richest genetic diversity reported for local group of tobacco varieties and lower diversity for introductions, and higher genetic similarity values between introductions and breeding group (Xiang et al. 2017). While, the variation among the tobacco lines were found to be higher concerning with the chemical nature of the plants (Tso et al. 1983; Darvishzadeh et al. 2011) as well as susceptibility to disease such as stem rot and powdery mildew (Darvishzadeh et al. 2010). The relatively low levels of diversity revealed by molecular markers in tobacco cultivars (Ren and Timko 2001) may be due to the utilization of only small proportion of the variability of the gene pools of the progenitor species in breeding programs (Lewis et al. 2007). The level of polymorphism among the varieties of *N. tabacum* observed to be higher compared to *N. rustica* (Sivaraju et al. 2008). However, wild tobacco accessions found to have higher genetic diversity (Chuanyin et al. 2009).

10.6 Association Mapping Studies

Association analysis, also known as linkage disequilibrium (LD) mapping or association mapping, uses the natural population as the target material. The analysis is based on linkage disequilibrium that detects the frequency of significant association between the genetic variation of markers with candidate genes and the target traits in a population (Bradbury et al. 2007; Pritchard et al. 2000). LD is defined as the non-random association of alleles at two or more loci (Fricano et al. 2012). Compared with QTL mapping, association analysis has two prominent advantages: (1) there is no need to construct population and genetic maps, and (2) uncovers (explores) elite genes from a certain number of germplasm resources in a single instance, providing

indication of genetic diversity. LD mapping has the potential to outperform traditional mapping and facilitate fine mapping in a random-mating population as only close linkage between markers and traits remains over several generations in such population.

10.6.1 Extent of Linkage Disequilibrium

In cultivated plants, extent of linkage disequilibrium is influenced by mating system, mutation rate, genetic drift, selection, recombination rate, gene conversion, and population size and structure (Flint-Garcia et al 2003). High-density genome fingerprinting could uncover long as well as short range LD. In the first case, in species with large genomes, a lower number of molecular markers can be tested (Waugh et al. 2009), although this will result in a lower mapping resolution. Conversely, short-range LD enables the fine mapping of causal polymorphisms, if large panels of markers are available (Myles et al. 2009).

The knowledge on the extent of LD is essential to determine the minimum distance between markers required for effective coverage when conducting genome-wide association studies (GWAS). Fricano et al. (2012) identified a set of 89 genotypes that captured the whole genetic diversity detected at the 49 SSR loci and evaluated LD using 422 SSR markers mapping on seven linkage groups. The pattern of intra-chromosomal LD revealed that LD in tobacco was clearly dependent on the population structure and was extended up to distances as great as 75 cM with $r^2 > 0.05$ or up to 1 cM with $r^2 > 0.2$.

10.6.2 Target Gene Based LD Studies

Thornsberry et al. (2001), for the first applied the association analysis in plants. Currently, LD has been used in association mapping to locate QTLs or major genes, based on the co-segregation of specific marker alleles and traits in tobacco (Zhu et al. 2008; Rafalski 2010). Zhang et al. (2012a, b) conducted association analysis on 13 agronomic traits in 258 flue-cured tobaccos and detected significant association of six agronomic traits with 18 sequence-related amplified polymorphism markers. Based on association analysis, Yu et al. (2014) found that the polymorphic loci of an SSR marker and six microsatellite-anchored fragment length polymorphism markers were significantly associated with the levels of tobacco-specific nitrosamines. Ren et al. (2014) found 24 SSR loci associated with aroma substances in tobacco and Basirnia et al. (2014) identified only one SSR locus from linkage group 13 that was significantly associated with low chloride accumulation rate in 70 oriental-type tobaccos. Fan et al. (2015) performed a marker–trait association analysis and obtained 11 SSR markers associated with potassium content in tobacco; five of the SSR markers were used to validate the stability of the associated markers in

130 tobacco germplasms. Through association mapping, Darvishzadeh (2016) identified the linkage of five SSR loci from linkage groups 2, 10, 11 and 18 of a tobacco reference map (Bindler et al. 2007, 2011) with gene(s) controlling *Orobanche* resistance in tobacco. Tahmasbali et al. (2021) by using mixed linear method identified a total of 16 loci to be significantly ($P < 0.05$) associated with the agronomic traits under normal (without *Orobanche*) and stress (with *Orobanche*) conditions with some common markers across normal and *Orobanche* conditions for few traits. Based on association mapping using 219 accessions for their responses to BS at two sites, Sun et al. (2018) detected six significant marker-trait associations and identified two probable candidate genes for resistance to BS (Nitab 4.5_0000264g0050.1 and Nitab 4.5_0000264g0130.1) among 31 predicted genes at locus Indel53.

SNPs, due to their abundance at genome-wide level, biallelic and reproducible nature, are considered to be the most desirable, precise and efficient tools for assessing the genetic characteristics of populations or germplasm, QTL mapping and facilitating the selection of breeding materials that bear desired genes/alleles or haplotypes in both plant genetics and breeding programs (Pace et al. 2015; Mora et al. 2016). Tong et al. (2020) made simultaneous association analysis of leaf chemistry traits in natural populations with a large amount of tobacco germplasms based on genome-wide SNP markers.

10.6.3 Genome Wide LD Studies

Association mapping studies largely depend on the genetic structure of the population. Molecular diversity in germplasm collections can be utilized to reconstruct the population structure in tobacco for association studies (Moon et al. 2009b). Ganesh et al. (2014) analyzed genetic structure of 135 FCV (Flue Cured Virginia) tobacco genotypes and observed that 25 unlinked SSR markers delineated 135 FCV genotypes revealing a total of 85 alleles with an average of 3.4 alleles per locus. However, as on date, Genome wide LD studies are not reported in tobacco.

10.6.4 Future Potential for the Application of Association Studies for Germplasm Enhancement

The population-based association study explores the availability of broader genetic variations with wider background for marker-trait correlations (i.e., many alleles evaluated simultaneously). Such studies provide higher resolution maps because of the utilization of majority recombination events from a large number of meiosis throughout the germplasm development history and exploits historically measured trait data for association without the development of expensive and tedious biparental populations in a time saving and cost-effective way (Abdurakhmonov

and Abdulkarimov 2008). Linkage disequilibrium (LD)-based association study, as a high-resolution, broader allele coverage, and cost-effective gene tagging approach provides an opportunity to dissect and exploit existing natural variations in tobacco germplasm resources for tobacco improvement. Owing to the availability of large collection of germplasm resources in tobacco over worldwide, association studies help to detect neutrally inherited markers in close proximity to the genetic causatives or genes controlling the complex quantitative target traits including resistance to biotic stresses for further germplasm enhancement and exploitation.

10.7 Brief Account of Molecular Mapping of Resistance Genes and QTLs

10.7.1 Brief History of Mapping Efforts

Gene mapping in tobacco found to initiate with the identification of linkage between a pollen color factor and the sterility factors (Anderson and De Winton 1931; East 1932; Brieger 1935; Smith 1937). Clausen and Goodspeed (1926) developed two types of monosomics and showed that both haplo-C (then called “corrugated”) and basic color factor, *Wh* are associated with a chromosome. Later, Clausen and Cameron (1944) studied the association in transmission between 24 monosomics developed by them and mendelian characters which led to location of 18 genes in nine chromosomes. Though genes controlling various characters and their linkages with other genes are identified, detailed map based entirely on genes is not available in tobacco (Suen et al. 1997; Narayanan et al. 2003).

However, with the advent of DNA markers in early nineties, efforts were made to map tobacco genome using molecular markers. Initially, RFLP, RAPD and AFLPs markers were used to map and tag resistant genes linked to biotic stresses. In 2001, for the first time, RFLP and RAPD markers were used to map *Nicotiana* spp. (Lin et al. 2001). Later, RAPD, AFLPs and ISSR were used in construction of genetic maps (Lin et al. 2001; Nishi et al. 2003; Julio et al. 2006; Xiao et al. 2006). After the discovery of SSR markers in late 1990s, for the first time SSR based molecular map showing 24 linkage groups was developed in *N. tabacum* (Bindler et al. 2007). This SSR map was improved further with identification of more number of SSRs (Bindler et al. 2011; Tong et al. 2012b). Recently, with identification of SNPs, high density SNP based tobacco genetic map has been developed with 24 linkage groups (Tong et al. 2020). Currently, maps are available for FCV, burley tobacco and their intra type. Maps are also available few *Nicotiana* spp.

10.7.2 Evolution of Marker Types

Molecular markers allow detection of variations or polymorphisms that exist among individuals for specific regions of DNA, thus serves as useful tools in mapping of genetic material. Molecular genetic markers, such as RFLP, RAPD, AFLP, SSRs and SNPs have been used in genetic linkage mapping and QTL mapping in tobacco (Liu and Zhang 2008).

Initially, molecular genetic markers such as RFLP, RAPD and AFLP were used in mapping studies. In the beginning of nineties, PCR-based RAPD markers were used by different workers to map and tag resistant genes linked to biotic stresses due to their relative ease in spite of the reproducibility issues. Though, reproducibility and sensitivity of AFLPs markers is higher, they were used in a limited extent in mapping of resistance gene owing to their lengthy and laborious detection method and non-suitability to automation.

After the discovery of SSR markers in late 90s, SSRs and EST-SSRs became markers of choice for mapping in tobacco (Bindler et al. 2007, 2011; Tong et al. 2012b). Currently, more than 10,000 SSR markers available in tobacco for their use in QTL/gene mapping studies (Bindler et al. 2007, 2011; Tong et al. 2012b; Cai et al. 2015; Madhav et al. 2015). In addition, Wang et al. (2018) identified a huge number of about 1,200,000 non-redundant and novel NIX (*Nicotiana* multiple (X) genome) markers (SSRs) for use in tobacco.

In recent years, SNPs have been identified and used in mapping of tobacco genome. Xiao et al. (2015) developed SNPs using two different methods (with and without a reference genome) based on restriction-set associated DNA sequencing (RAD-seq). Thimmegowda et al. (2018) identified SNPs by whole-genome resequencing of 18 flue-cured Virginia (FCV) tobacco genotypes and positioned SNPs in linkage groups. Using the genome of *N. tabacum* (K326 cultivar) as a reference, Tong et al. (2020) identified and mapped 45,081 SNPs to 24 linkage groups in the tobacco genetic map.

In addition to the above markers; sequence-specific amplification polymorphism (SSAP), Sequence-related amplified polymorphism (SRAP), cleaved amplified polymorphic sequence (CAPS) and Diversity arrays technology (DArT) markers were also used in generating molecular linkage maps in tobacco.

10.7.3 Mapping Populations Used

For molecular mapping in tobacco, diverse populations viz. F₂ populations, doubled haploid (DH) lines, recombinant inbred lines (RILs), BC₁ progenies, BC₁F₁, BC₄F₃ populations etc. have been used as the mapping population (Table 10.5). Majority of the maps developed based on F₂ and DH populations and three maps developed based on NGS technologies. Practically 99–381 individuals are selected in a mapping population for higher resolution and fine mapping.

Table 10.5 Genetic linkage maps constructed in *Nicotiana* (in chronological order)

Type of tobacco	Population	Markers used	Linkage groups identified	Total length covered (cM)	References
<i>Nicotiana</i> species	99 F ₂ plants from the cross <i>N. plumbaginifolia</i> × <i>N. longiflora</i>	69 RFLP & 102 RAPD	9	1062	Lin et al. (2001)
Burley	125 DH lines, derived from F ₁ hybrids between W6 and <i>Michinoku 1</i>	117 AFLPs	10	383	Nishi et al. (2003)
Flue-cured	DH population from a cross between Speight G-28 and NC2326	11 ISSRs & 158 RAPDs	27	2094.6	Xiao et al. (2006)
Flue-cured	114 RILs	138 ISSR, AFLP & SSAP markers	18	707.6	Julio et al. (2006)
Flue-cured	186 F ₂ derived from Hicks Broad Leaf and Red Russian	293 SSRs	24	1920	Bindler et al. (2007)
Flue-cured/burley	187 F ₂ derived from the cross Taiyan 7 (FCV) and Bailei 21 (burley)	112 (92 SRAP and 10 ISSR)	26	1560.2	Ma et al. (2008)
Burley	92 DH population derived from Burley 37 and Burley 21	112 AFLP loci & 6 SRAP loci	22	1953.6 cM	Cai et al. (2009)
Wild diploid <i>Nicotiana</i> species	<i>N. tomentosiformis</i>	489 SSRs/CAPS	12	~1000 cM	Wu et al. (2010)
Wild diploid <i>Nicotiana</i> species	<i>N. acuminata</i>	308 SSRs/CAPS	12	~1000 cM	Wu et al. (2010)

(continued)

Table 10.5 (continued)

Type of tobacco	Population	Markers used	Linkage groups identified	Total length covered (cM)	References
Flue-cured	186 F2 derived from cultivar, Hicks Broad Leaf and Red Russian	2318 SSRs	24	3270	Bindler et al. (2011)
Flue-cured	207 DH population derived from FCV cultivars, HHDIY & Hicks Broad Leaf	851 (238 DArT & 613 SSRs)	24	2291	Lu et al. (2012)
<i>N. tabacum</i>	Parents [HD (Hong hua Da jin yuan) and RBST (resistance to black shank tobacco)] and their F ₁ (HD × RBST) and 193 BC ₁ progenies	Map 1: 4138 SNPs & Map 2: 2162 SNPs	24	Map 1: 1944.74 (with reference genome); Map 2: 2000.9 (de novo identification SNPs by RAD-seq)	Xiao et al. (2015)
Flue-cured tobacco	Specific locus-amplified fragment sequencing	4215 SNPs and 194 SSRs	24	2662.43	Gong et al. (2016)
Flue-cured tobacco	213 BC ₁ s derived from Y3 and K326	626 SSRs	24	1120.45	Tong et al. (2016)
Flue-cured tobacco	Parents (Y3 & K326), their F ₁ & 271 RILs	45,081 SNPs	24	3486.78 cM	Tong et al. (2020)
Flue-cured tobacco	Parents (NC82 & TT8), their F ₁ & 200 BC ₁ F ₁ s	13,273 SNPs	24	3421.80	Cheng et al. (2019)
Flue-cured tobacco	Parents (Y3 & K326), their F ₁ & 381 BC ₄ F ₃ s	24,142 SNPs	24	2885.36	Tong et al. (2021)

Note *RFLP* Restriction fragment length polymorphisms; *RAPD* Random amplified polymorphic DNA; *AFLP* Amplified fragment length polymorphism; *JSSR* Inter simple sequence repeat; *SSAP* Sequence-specific amplification polymorphisms; *SSR* Simple sequence repeats; *SRAP* Sequence-related amplified polymorphism; *CAPS* Cleaved amplified polymorphic sequences; *DArT* Diversity arrays technology; *SNP* single nucleotide polymorphisms, *DH* Double haploid lines; *RILs* Recombinant inbred lines

10.7.4 Mapping Software Used

Mapmaker program (Lander et al. 1987; Lin et al. 2001; Wu et al. 2010), JoinMap® 3.0 program (Van Ooijen and Voorrips 2001; Bindler et al. 2007, 2011) Map Manager QTXb20 (Manly et al. 2001; Bindler et al. 2011; <http://www.mapmanager.org>) and JoinMap 4.0 (Van Ooijen 2006; Lu et al. 2012; Tong et al. 2016) and LepMap3 software (Rastas 2017; Tong et al. 2020, 2021) were used in developing molecular maps in tobacco. Among various mapping softwares, JoinMap 3.0/4.0 program is the widely used one in tobacco for construction of molecular maps based on markers. LepMap3 software was used in the construction of maps using NGS data.

10.7.5 Maps of Different Generations

Genetic linkage maps are essential for studies of genetics, genomic structure, genomic evolution and for mapping essential traits. Construction of molecular genetic maps in tobacco have lagged behind other *Solanaceae* crops such as the tomato, potato, and pepper plants (Tanksley et al. 1992; Jacobs et al. 2004; Barchi et al. 2007). Till the end of twentieth century, very little information was available on genetic mapping and molecular development in tobacco (Suen et al. 1997). Construction of genetic linkage maps in tobacco was started since the beginning of twenty-first century (Lin et al. 2001). Various maps constructed are briefly discussed here.

10.7.5.1 Mapping of *Nicotiana* Species

Lin et al. (2001) constructed a genetic linkage map based on the 99 individuals of the F₂ plants derived from tobacco wild species, *Nicotiana plumbaginifolia* × *N. longiflora*. This map covers 1062 cM with the distribution of 60 RFLP and 59 RAPD loci on nine major linkage groups. Owing to the shortage of markers, the map has not coalesced into ten linkage groups, corresponding to the haploid chromosome number of *N. plumbaginifolia*. Later, another two maps for wild diploid *Nicotiana* species, *N. tomentosiformis* and *N. acuminata* with 12 linkage groups spanning ~1000 cM are generated by Wu et al. (2010). The *N. tomentosiformis* map was created with the combination of 489 SSR and CAPS markers and *N. acuminata* (closely related to *N. sylvestris*) with 308 SSR and CAPS markers (Wu et al. 2010).

10.7.5.2 Mapping of Burley Tobacco

Nishi et al. (2003) constructed a genetic linkage map of the burley tobacco containing 10 linkage groups based on DH lines, derived from F₁ hybrids between burley entries, W6 and Michinoku 1 using AFLP markers. Cai et al. (2009) constructed

currently available high density burley linkage map spanning 1953.6 cM using a double haploid (DH) population derived from a cross between Burley 37 with high nicotine content and Burley 21 with low nicotine content assembling 112 AFLP loci and six SRAP loci into 22 linkage groups (A1-A22).

10.7.5.3 Mapping of Flue-Cured Tobacco

The first linkage map of flue-cured tobacco based on a DH population from a cross between Speight G-28 and NC2326 was constructed by Xiao et al. (2006) using 169 ISSR/RAPD molecular markers covering 27 linkage groups. While Julio et al. (2006) constructed a molecular linkage map of flue-cured tobacco with 18 linkage groups covering 138 ISSR, AFLP and SSAP markers based on 114 flue-cured tobacco recombinant inbred lines.

Bindler et al. (2007) for the first time constructed an SSR based linkage map using F_2 plants derived from a cross of Hicks Broadleaf \times Red Russian. Later, this map was further improved (Bindler et al. 2011) with the identification of more number of SSR markers. This high density SSR map of flue-cured tobacco constructed with 2318 SSR markers mapping on 24 linkage groups covering a total length of 3270 cM is the most widely used map of tobacco. The length of individual linkage groups in this map varied from 86 to 199 cM, and the average genetic distance between adjacent markers was 1.4 cM. However, in spite of the high marker density, there exists some gaps spanning \sim 16 cM in the map. (Fig. 10.1; Bindler et al. 2011). Tong et al. (2012a, b) utilized a population of 207 DH lines derived from a cross between 'Honghua Dajinyuan' and 'Hicks Broad Leaf' and constructed a genetic map of flue-cured tobacco consisting of 611 SSR loci distributed on 24 tentative linkage groups and covering a total length of 1882.1 cM with a mean distance of 3.1 cM between adjacent markers. Utilizing a 213 backcross (BC1) individuals derived from an intra-type cross between two flue-cured tobacco varieties, Y3 and K326, Tong et al. (2016) further constructed a genetic map consisting of 626 SSR loci distributed across 24 linkage groups and covering a total length of about 1120 cM with an average distance of 1.79 cM between adjacent markers.

Xiao et al. (2015) constructed two SNP linkage maps of flue-cured tobacco using two different methods (with and without a reference genome) based on restriction-set associated DNA sequencing (RAD-seq) of back cross population. Overall, 4138 and 2162 SNP markers with a total length of 1944.74 and 2000.9 cM were mapped to 24 linkage groups in these genetic maps based on reference genome and without reference, respectively. An SNP-based high density genetic map, *N. tabacum* 30 k Infinium HD consensus map 2015 (Fig. 10.2) was released at SOL Genome Network to facilitate the fine mapping of different trait of interest (https://solgenomics.net/cview/map.pl?map_version_id=178). Cheng et al. (2019) constructed a high-density SNP genetic map of flue-cured tobacco using restriction site-associated Illumina DNA sequencing. In this map, a total 13,273 SNP markers were mapped on 24 high-density tobacco genetic linkage groups spanning around 3422 cM, with an average distance of 0.26 cM between adjacent markers. While, Tong et al. (2020) performed



Fig. 10.1 Part SSR map (1–6 linkage groups) constructed by Bindler et al. (2011) with 2318 microsatellite markers covering a total length of 3270 cM

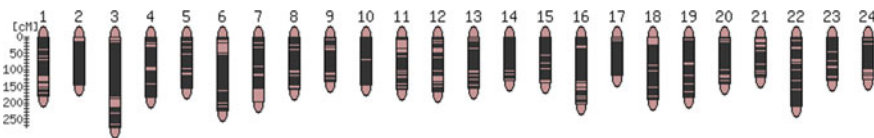


Fig. 10.2 *N. tabacum* 30 k Infinium HD consensus map 2015 (https://solgenomics.net/cview/map.pl?map_version_id=178)

whole-genome sequencing of an intraspecific RIL population, a F_1 generation and their parents and identified SNPs. Using the *N. tabacum* (K326 cultivar) genome as reference, a total of 45,081 identified SNP markers (with 7038 bin markers) were characterized to construct a high-density SNP genetic map of flue-cured tobacco spanning a genetic distance of around 3487 cM (Fig. 10.3). Tong et al. (2021) successfully constructed another high-density genetic map containing 24,142 SNP markers using a BC_4F_3 population derived from inbred of flue-cured tobacco lines Y3 (recurrent parent) and K326 (donor parent). This map included 4895 bin markers with a genetic distance of 2885.36 cM and an average genetic distance of 0.59 cM. Based on the genotype of these located markers, a binmap was also constructed using the chromosome as a unit.

Lu et al. (2012) developed a high-density integrated linkage map (2291 cM) of flue-cured tobacco that included 851 markers [238 DArT and 613 SSR] in 24 linkage groups. Gong et al. (2016) generated a high-density 2662.43 cM length integrated genetic map of flue-cured tobacco containing 4215 SNPs and 194 SSRs distributed

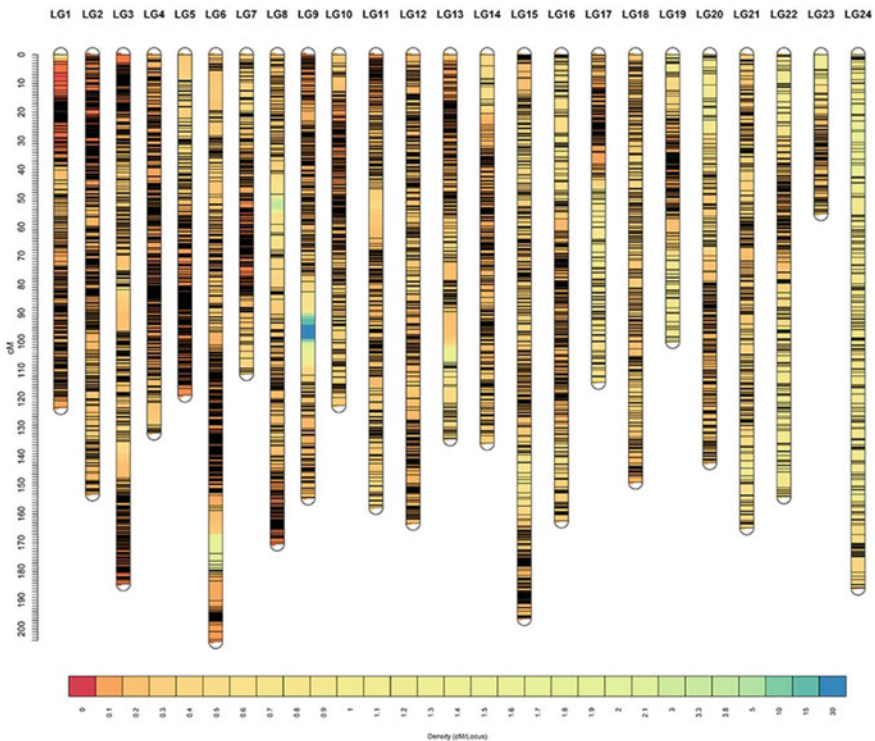


Fig. 10.3 Linkage map constructed based on the tobacco reference genome by Tong et al. (2020). These linkage maps were constructed with 45,081 SNP markers (with 7038 bin markers) covering a total length of 3486.78 cM

on 24 linkage groups (LGs) with an average distance of 0.60 cM between adjacent markers.

10.7.5.4 Intra Type Genetic Maps

Ma et al. (2008) constructed an intra type genetic map containing 26 linkage groups and 112 markers by using flue-cured and burley tobaccos based on SRAPs and ISSR markers.

Currently, the high-density maps in tobacco are constructed with SSR (Bindler et al. 2011) and SNP (Gong et al. 2016; Tong et al. 2020) markers. The widely referred SSR map of Bindler et al. (2011) was constructed with 2318 microsatellite markers covering a total length of 3270 cM while the SNP map of Tong et al. (2020) covers 3486.78 cM with 45,081 SNPs. The combination of SNPs and genetic maps, if developed, helps in designing precise breeding strategies and genomic selection in tobacco. Diverse genetic maps constructed can be effectively utilized in mapping QTLs, positional cloning, comparative genomics analysis, marker-assisted breeding and genomic selection etc. It would be necessary to further build the genetic linkage maps of tobacco in different cultivating types for their effective utilization in breeding of those types.

10.7.6 Enumeration of Mapping of Simply-Inherited Stress Related Traits

Availability of draft *Nicotiana* genome sequences (Sierro et al. 2014; Edwards et al. 2017) and high-density molecular maps in recent times are laying the foundation for trait discovery and fine mapping of trait of interest in tobacco (Yang et al. 2019). Linked molecular markers to various biotic stress namely black root rot, black shank, wildfire, blue mold, brown spot, powdery mildew, TMV, PVY, TSWV, root-knot etc. are identified in tobacco for their use in breeding programs are enumerated at Table 10.6. Majority of the markers found to be dominant ones.

10.7.7 Framework Maps and Markers for Mapping Resistance QTLs

Framework maps are being constructed mostly using SSR and SNP markers that are identified and mapped to linkage groups in tobacco (Tong et al. 2020; Lu et al. 2012). High density SSR map of Bindler et al. (2011) constructed with 2318 microsatellite markers covering a total length of 3270 cM and the SNP map of Tong et al. (2020) covering 3486.78 cM with 45,081 SNPs can be used in constructing framework maps

Table 10.6 Biotic stress resistance traits and linked markers in tobacco

Resistance trait	Gene	Marker	Linkage	Gene action	Reference
Black root rot, (ex <i>N. debneyi</i>)	–	SCAR	1 cM	Dominant	Julio et al. (2006)
Black root rot, (ex <i>N. debneyi</i>)	–	CAPS	Closely linked	Co-dominant	Qin et al. (2018)
Black shank (race 0)(ex <i>N. longiflora</i>)	<i>Phl</i>	SCAR	Closely linked	Dominant	Unpublished, personal communication
Black shank (race 0)(ex <i>N. plumbaginifolia</i>)	<i>Php</i>	RAPD	Closely linked	Dominant	Johnson et al. (2002)
Black Shank (Race 0) Coupling Phase	<i>Php</i>	SCAR	Closely linked	Dominant	Unpublished, personal communication
Black Shank (Race 0) Repulsion Phase	<i>Php</i>	SCAR	Closely linked	Dominant	Unpublished, personal communication
Wildfire (race 0)	–	RAPD	Closely linked	Dominant	Yi et al. (1998a)
Blue mold	–	SCAR	Closely linked	Dominant	Milla et al. (2005)
Blue mold	–	SCAR	Closely linked	Dominant	Julio et al. (2006)
Blue mold	–	SCAR	Closely linked	Dominant	Wu et al. (2015)
Brown spot	–	RAPD	Closely linked	Dominant	Zhang et al. (2008b)
Powdery mildew	<i>NtMLO1</i> <i>NtMLO2</i>	SSCP	Gene Specific	Co-dominant	Fujimura et al. (2015)
TMV	<i>N-gene</i>	SCAR	Genes specific	Dominant	Lewis (2005)
N-Gene (<i>N. Glutinosa</i> virus resistance)	<i>N-gene</i>	Gene specific	Gene specific	Dominant	Whitham et al. (1994)
Potyvirus susceptibility	<i>eiF4E1.S</i>	Gene specific	Gene specific	Dominant	Dluge et al. (2018)
PVY	<i>va</i>	SCAR	5.1 cM	Dominant	Julio et al. (2006)
PVY	<i>va</i>	RAPD	Closely linked	Dominant	Lewis (2005)
PVY	<i>va</i> (<i>eIF4E</i>)	Gene Specific	Gene Specific	Dominant	Julio et al. (2014)
PVY RBV	<i>eIF(iso)4E</i>	dCAPS	Mutation Specific	Co-dominant	Takakura et al. (2018)
TSWV	AFLP	SCAR	<5 cM	Dominant	Moon (2006)

(continued)

Table 10.6 (continued)

Resistance trait	Gene	Marker	Linkage	Gene action	Reference
Root-Knot, <i>M. incognita</i>	<i>Rk</i>	RAPD	Closely Linked	Dominant	Yi et al. (1998b)

Source CORESTA Guide No. 16—March 2019 (Yang et al. 2019)

while mapping various traits (Edwards et al. 2017). The release of the SNP-based high density genetic map, *N. tabacum* 30 k Infinium HD consensus map 2015 can be one of the finest resources for fine mapping any trait of interest (https://solgenomics.net/cview/map.pl?map_version_id=178).

RAPD and SSR markers are the most widely used markers in QTL mapping of biotic traits in tobacco followed by SNP and AFLP markers (Table 10.7). In few cases, RAPD markers are converted into SCAR markers for reproducibility and reliability. SCAR markers are developed after sequencing RAPD bands and designing 18–25 base PCR primers that can specifically amplify the sequenced DNA segment more reliably. CAPS, COS, Random amplified microsatellite polymorphism (RAMP), ISSRs and Target region amplification polymorphism (TRAP) are some of the other markers used. CAPS markers developed are the primers designed on the known sequence of a gene of Interest. COS primers used are universal primers based on sequence alignments of orthologs (genes that are conserved in sequence and copy number) from multiple solanaceous species. RAMP markers include SSR primers that amplify the genomic DNA in the presence or absence of RAPD primers (Liu et al. 2009b). TRAP are two PCR-based primers, one from target EST and the other is an arbitrary primer.

10.7.8 QTL Mapping Software Used

Mapmaker/Exp 3.0 is the most widely used QTL mapping software in QTL mapping various biotic traits in tobacco followed by various version of Join Map and Map Chart (Table 10.7). Mapmaker/QTL, QTL IciMapping 4.1, QTL Network 2.1, R/QTL, AYM-SS, Stat Graphics Plus 5.0 and QTL Cartographer V 2.5 are some of the other softwares used.

10.7.9 Details on Trait Wise QTLs

QTLs have been identified for various biotic traits in tobacco for their use in gene introgression and genome selection (Table 10.7). Details of QTLs developed for resistance to Bacterial wilt, Brown spot, Black Shank and CMV are briefly discussed here.

Table 10.7 Details of few trait wise QTLs identified in tobacco

Trait	Cross	Population	Marker	Condition	QTL(s)	Linkage groups	Position (cM)	Variance (R ² (%) _c)	LOD	Gene action		Software used	References	
										Additive	Dominance			
Bacterial wilt	TI448A×Enshu	F _{2,3} & F _{2,4}	SSR	Pooled	<i>qBWR-3a</i> <i>qBWR-3b</i>	3	19.6	9	4.6	0.5	0.27	Mapmaker/Exp 3.0	Qian et al. (2013)	
	TI448A×Yanyam 97	21.1					19.7	6.17	0.03	0.3				
	TI448A×Enshu	68.8					17.3	4.12	0.08	0.31				
	TI448A×Yanyam 97	43.6					17.4	3.18	0.03	0.28				
Brown spot	Jingyehuang × NC82	F ₂	SSR	Field	<i>Jimo_2014</i> <i>Jimo_2014</i> <i>Jimo_2015</i> <i>Zhuzheng_2015</i>	2a 3a 5	18	18	7	-1.6	0.5	JoinMap 4.1, MapChart 2.2	Sun et al. (2018)	
		F _{2,3}					8.6	8.6	3.5	-5.1	0.4			
		F _{2,3}					10.1	10.1	4.2	-7.6	2.2			
		F _{2,3}					9.4	9.4	3.9	-7.7	-2.0			
	Changbohuang × Jingyehuang	F ₂		SSR	-	<i>qBS 2</i> <i>qBS 3</i> <i>qBS 5</i>	2a 3a 5	8.1	14.3	4.9	8	-0.89	Mapmaker/Exp 3.0b	Tong et al. (2012a)
								8.7	8.7	5.11	8.3	-2.3		
Black Shank	Beinhart-1000 × Hicks	DH	Microsatellite	Field	<i>I</i> <i>2</i> <i>3</i> <i>4</i>	7 15 17	8	11.9	25	13		JoinMap 4	Ma et al. (2019)	
							4	40.8	20	11				
							11	5.8	6.2	4.6				
							14	4	5.5	3.3				
K326 × Beinhart 1000	Near isogenic lines (NILs), Sub NILs		Microsatellite, SNP	Field	<i>I</i> <i>2</i> <i>3</i>	7 15 17	76	35	15.9		R/QTL	Ma et al. (2020)		
							108	25	12.4					
							88	8.5	4.8					

(continued)

Table 10.7 (continued)

Trait	Cross	Population	Marker	Condition	QTL(s)	Linkage groups	Position (cM)	Variance (R ² %)	LOD	Gene action		Software used	References																																										
										Additive	Dominance																																												
	Beinhardt 1000-1 × Xiaochuangjin 1025	BC ₁ F ₂	SSR	Field	<i>qBS7</i>	7	8.9	17.4	4.98	-6.9		Join Map 4.1	Zhang et al. (2018)																																										
														BC ₁ F _{2,3}	114.26	11.6	3.12	-1.62																																					
		7																	7.9	25.6	8.04	-16.82																																	
																							17	110.26	6.9	2.46	-9.37																												
																												7	7.9	17.48	5.55	-15.14																							
																																	17	111.26	10.30	3.48	-12.34																		
																																						7	56	48.8	32	19.08	qBS7	QTL Ici Mapping 4.1											
																																													14	72.8	6.78	4.96	-7.3	qBS14					
																																																			16	114	4.05	2.9	-5.7
														24																																									
1	46		2.63	2.47	-4.6	qBS1																																																	
							7	62	16.9	8.7		Join-Map version 4.0, MapChart	Xiao et al. (2013)																																										
		24													9.5	18.6	11.7																																						
																			2	9.2	3.5																																		

(continued)

Table 10.7 (continued)

Trait	Cross	Population	Marker	Condition	QTL(s)	Linkage groups	Position (cM)	Variance (R ² (%) _c)	LOD	Gene action		Software used	References			
										Additive	Dominance					
CMV at seedling stage	TT8 × NC 82	F ₁ and BC ₁ F ₁	SNP	Green house	3	14	4.9	15.3	7.04			Join Map 4 Map Chart 2	Cheng et al. (2019)			
				Field				6	3							
				Green house				10	8.1							
				Field				–	–							
				Green house				5.6	4							
				Field				7.1	4.1							
				Green house				5.3	2.6							
				Field				4.2	2.7							
				Green house				6.9	5							
				Field				6.6	5.9							
				Green house				1.2	4.9							
				Seedling				4D5	5	40	7.70			3.39	–12.19	–
								4D18	8	88	7.20			3.24	–5.00	–

10.7.9.1 Resistance to Bacterial Wilt Disease

The resistance for bacterial wilt in tobacco is governed by either polygenes or in combination of polygenes and a major gene (Smith and Clayton 1948; Kelman 1953; Matsuda 1977; Hayward 1991). Furthermore, the influence of environment makes difficulty in screening/phenotyping, hence it is essential to identify closely linked molecular marker to developing resistant varieties through marker-assisted selection. Earlier, QTLs affecting the resistance to bacterial wilt have been reported on chromosome 6, chromosome 7, and chromosome 12 (Margin et al. 1999; Danesh et al. 1994). Nishi et al. (2003) detected a QTL for the bacterial wilt resistance of W6 explaining more than 30% of the variance. Qian et al. (2013) detected four QTLs viz. *qBWR-3a*, *qBWR-3b*, *qBWR-5a* and *qBWR-5b* in linkage group 3 and 5 strongly associated with resistance and explaining 9.00, 19.70, 17.30, and 17.40% of the variance in resistance, respectively. These loci had the close linkage with the markers PT20275 and PT3022. Lan et al. (2014) identified eight QTLs with significant main effects on chromosomes 2, 6, 12, 17 and 24 through genome wide QTL analysis. Also detected a major QTL (*qBWR17a*) on chromosome 17 which explained up to 30% of the phenotypic variation.

10.7.9.2 Resistance to Black Shank Disease

Clayton (1958) suggested that resistance to black shank disease in tobacco was simply inherited and controlled by recessive alleles. Moore and Powell (1959) reported resistance to be partially dominant and affected by modifying factors. Others have suggested that in the line Florida 301 resistance to be polygenic and additive in nature (Crews et al. 1964; Chaplin 1966). Incomplete penetrance and variable expressivity have probably complicated the interpretation of data generated to investigate the inheritance of resistance in this line. Vontimitta and Lewis (2012a, b) suggest that Beinhart 1000 and Florida 301 share a major gene affecting black shank resistance, but probably differ in allelic variability at a fair number of additional loci having smaller effects. The genetic control of black shank resistance in Florida 301 is of the classic polygenic type and that it is controlled by a combination of a few genes with large effects and a greater number of genes with small to intermediate effects. Zhang et al. (2018) identified stable QTLs (*qBS7* and *qBS17*) for resistance to black shank disease using the F₂, BC1F₂ individuals and BC1F_{2,3} lines derived from a cross between Beinhart 1000-1 and Xiaohuangjin 1025 (Fig. 10.4). QTL *qBS7* was mapped to the region between PT30174 and PT60621 and explained 17.40–25.60% of the phenotypic variance under various conditions. One major QTL, labeled as *Phn7.1*, was found to be the largest contributor to partial *P. nicotianae* resistance in the highly black shank resistant cigar tobacco cultivars, Beinhart 1000 and Florida 301 and highly resistant flue-cured tobacco cultivar, K346 (Vontimitta and Lewis 2012b; Xiao et al. 2013; Ma et al. 2019). A second QTL, *Phn15.1* with a large effect was identified on *N. tabacum* linkage group 15 in Beinhart 1000 (Ma et al. 2020). Gong et al. (2020) fine mapped the QTLs using the 177 F_{7,8-9} recombinant

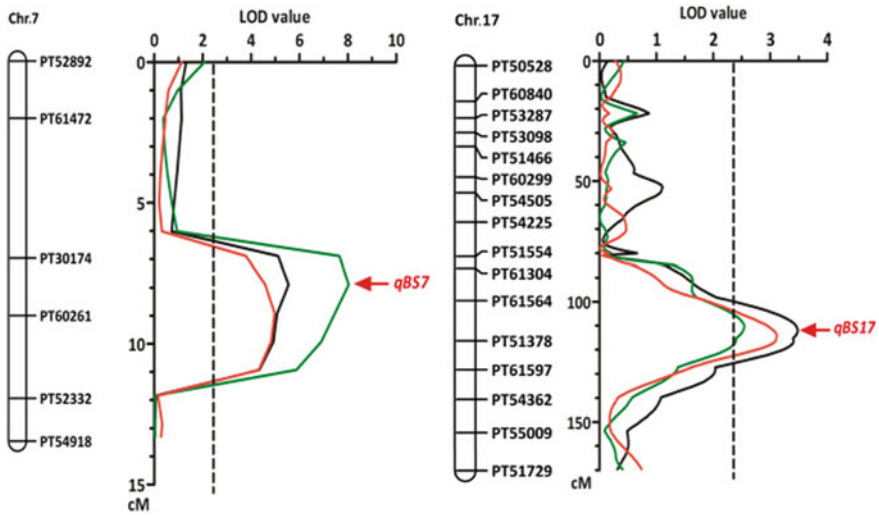


Fig. 10.4 Likelihood plots and position (cM) of the QTL associated with resistance to black shank in different conditions. Red line: LOD plots for the BC_1F_2 population; Green line: LOD plots for the $BC_1F_{2.3}$ population in the field; Black line: LOD plots for the $BC_1F_{2.3}$ population in the greenhouse; dashed lines represent the significant LOD threshold at the level of 2.35 (Zhang et al. 2018)

inbred lines generated from a cross between a resistant cultivar ‘Yunyan 85’ and a susceptible entry ‘Dabajin 599’. A total of 10 QTLs associated with resistance to *P. nicotianae* across multiple environments were detected and two major QTL *qBS7* and *qBS14* were repeatedly identified under all five environments explaining 56 and 6.78% of the mean phenotypic variance with high logarithm of the odds (LOD) scores, respectively.

10.7.9.3 Resistance to Brown Spot

The cigar tobacco line Beinhart 1000 and flue-cured tobacco variety Jingyehuang are considered important sources of partial resistance in the USA and China, respectively. Beinhart 1000 is derived from a selection of tobacco Quin Diaz and presents a high level of partial resistance to brown spot. Using SSR markers, Tong et al. (2012a), detected three QTL for resistance to brown spot in tobacco on 2, 3, and 5 linkage groups using a F_2 population derived from a cross between a brown spot susceptible variety Changbohuang and the resistant source Jingyehuang (Fig. 10.5). The major QTL mapped on the genetic linkage group three found to explain 14.3% of the phenotypic variation. In a study, Sun et al. (2018) evaluated F_2 , $F_{2.3}$ and $BC_3F_{2.3}$ populations developed from a cross between a source of brown spot resistance Jingyehuang and a brown spot susceptible flue-cured variety NC82 for field resistance under different environments and identified QTLs by linkage mapping.

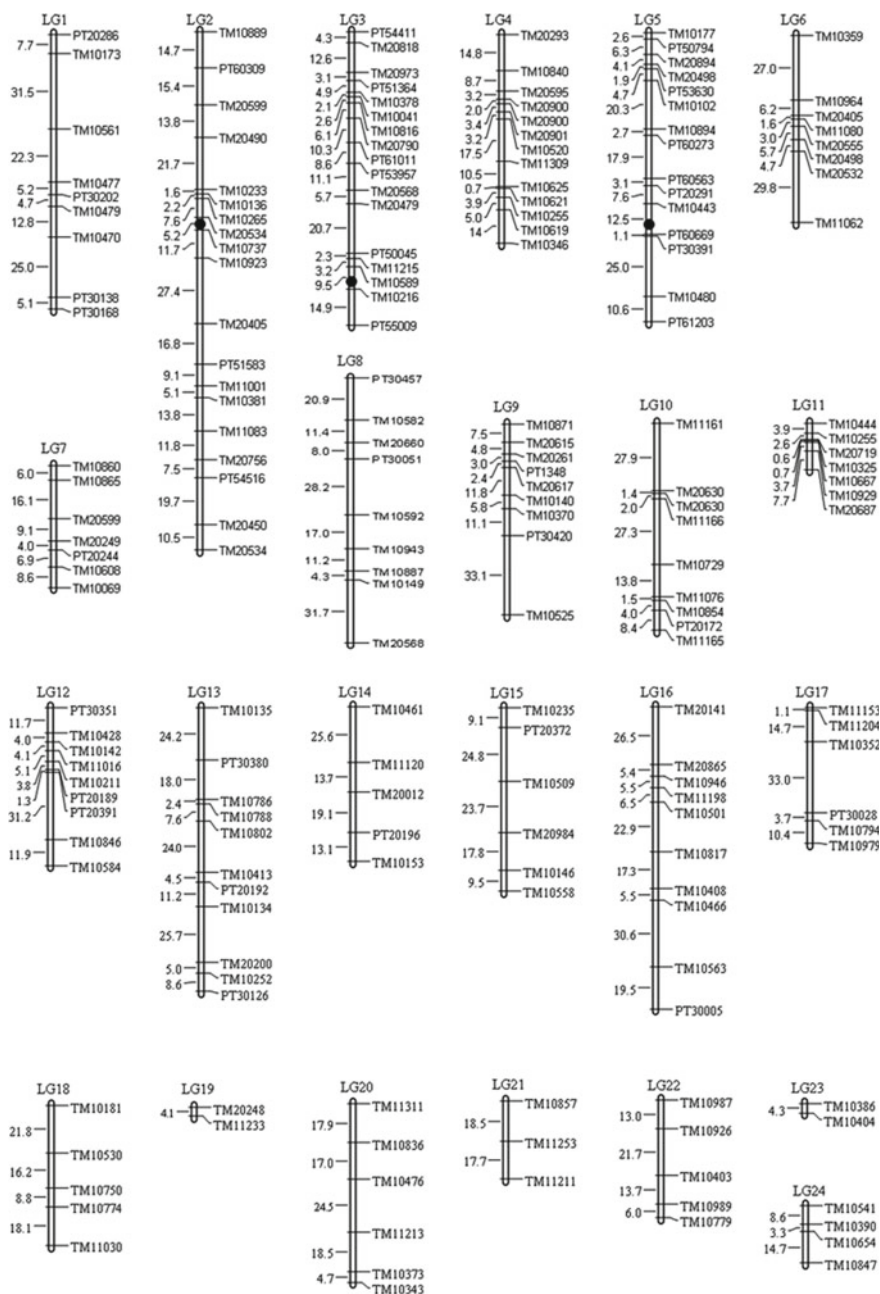


Fig. 10.5 Mapping of quantitative trait loci conferring resistance to brown spot in flue-cured tobacco (*N. tabacum*). Three QTLs for brown spot resistance were mapped on LG2a, LG3b and LG5, respectively (Tong et al. 2012a)

A major QTL mapped on chromosome 15 explained 8.6–18.0% of the phenotypic variation under diverse conditions. Furthermore, based on association mapping using 219 accessions for their responses to brown spot at two sites, six significant marker-trait associations were detected. Out of these markers, the marker Indel53 exhibited the most significant association with resistance to brown spot and explained around 21% of the phenotypic variation at the two sites. An approximately 2-Mb physical interval at the locus of marker Indel53 contained 31 predicted genes and two of these genes (*Nitab 4.5_0000264g0050.1* and *Nitab 4.5_0000264g0130.1*) were identified as probable candidate genes for resistance to brown spot.

10.7.9.4 Resistance to *Cucumber Mosaic Virus*

Cheng et al. (2019) identified seven QTLs, including two for incidence of disease and four for disease index for CMV in BC₁F₁ population at seedling stage. qID5 for incidence of disease was mapped to the interval mk6533–mk646 on LGs 5 explaining 7.70% of the total phenotypic variance. For disease index, qDI8 mapped on LG 8 explained 7.20% of total phenotypic variance indicating stable genetic effects in diverse environments.

10.7.10 Mendelization of QTL

The QTL, *Phn7.1* was found to have an additive effect on resistance to black shank disease and the corresponding QTL was localized to within a genetic interval of approximately 3 cM (Ma et al. 2019). A second QTL, *Phn15.1* on *N. tabacum* linkage group 15 in Beinhart 1000 was localized to a genetic interval of approximately 2.7 centimorgans using subNILs containing varying amounts of Beinhart 1000-derived *Phn15.1*-associated genetic material (Ma et al. 2020). Incorporation of this allelic variability into breeding programs could increase the level, range, and durability of genetic resistance to *P. nicotianae* in to be released newly tobacco cultivars. However, *Phn15.1* is very closely linked to the gene *NtCPS2* associated with *Z*-abienol biosynthesis (Vontimitta et al. 2010). *Z*-abienol is a trichome exudate that contributes to flavor and aroma characteristics of Oriental and some cigar tobacco but is considered undesirable for flue-cured and burley tobacco. For the effective utilization of *Phn15.1* in flue-cured and burley tobacco cultivar development, Ma et al. (2020) disassociated the favorable Beinhart 1000 *Phn15.1* alleles from the Beinhart 1000 *NtCPS2* allele.

10.8 Marker-Assisted Breeding for Resistance Traits

Marker-assisted breeding (MAB) refers to a breeding program in which detection of DNA markers and selection of desirable genotypes are integrated. The status and prospects of MAB are discussed here under.

10.8.1 *Germplasm Characterization and DUS*

Germplasm are to be characterized for desirable traits and tightly linked molecular markers for the traits need to be identified. Identification of reliable linked markers is critically important to initiate a marker-assisted breeding program. Polymorphism for DNA markers are available throughout the genome and DNA markers can be detected at any stage of plant growth. The presence or absence of these markers is not affected by environments and usually do not directly affect the phenotype. Identification and selection of markers located in close proximity to the target gene or within the gene will ensure the success in selection of the target gene. Therefore, DNA markers are the predominant types of genetic markers for MAB (Xu 2010). These tightly linked markers can be utilization in MAB while screening parents, F₁ and other segregating materials for selecting plants with desirable traits. Molecular-assisted breeding of pest/disease resistant tobacco plants can help to identify the resistant plants in early stage, cut down on workload by way of avoiding inoculation procedure of pests/disease, increase selection efficiency, accelerate rapid utilization of the resistant sources, and shorten the breeding cycle (Liu and Zhang 2008).

The information on marker trait associations in germplasm lines and mapping populations can be obtained through gene mapping, QTL analysis, association mapping, classical mutant analysis, linkage or recombination analysis, bulked segregant analysis, etc. It is also essential to know the linkage state i.e. cis/trans (coupling or repulsion) linkage with the desired allele of the trait. Most commonly used molecular markers in tobacco include RAPD, AFLP, SSR, SCAR, CAPS, dCAPS (derived CAPS), and KASP (Yang et al. 2019). Each type of DNA markers has advantages and disadvantages for specific purposes. Relatively speaking, SSRs have most of the desirable features and availability of large number of SSRs make them markers of choice in tobacco. SNPs require a detailed understanding of single nucleotide DNA changes responsible for genetic variation among individuals. Fairly large number of SNPs have become available in tobacco making them important choice of markers for MAB in tobacco.

Genetic mapping, QTL analysis and association mapping (AM) have accelerated the dissection of genetic control of biotic stress resistant traits in tobacco. Large number of studies for screening germplasm and sources of resistance are made in tobacco to identify closely linked markers to various biotic stresses for their utilization introgression into cultivated varieties. Tightly linked markers are important tools for DUS characterization of varieties also. Tightly linked molecular markers/genic

markers/QTLs were identified for some of the traits conferring resistance to various biotic stresses particularly diseases in tobacco are listed at Tables 10.6 and 10.7. This information has the potential to make marker-assisted selection (MAS) a successful option for tobacco improvement.

10.8.2 Marker-Assisted Gene Introgression

Marker-assisted backcrossing (MABC) is the simplest form of marker-assisted gene introgression that is most widely and successfully used in transferring biotic stress resistant genes into elite cultivars. MABC aims to transfer one or a few genes/QTLs for resistance from one genetic source (donor parent) into a superior cultivar or elite breeding line (recurrent parent) to improve the stress resistance. In contrast to traditional backcrossing, MABC depends on the molecular markers linked to gene(s)/QTL(s) of interest in the place of phenotypic performance of target trait. MABC program with two types of selections viz. foreground selection for the marker allele(s) of donor parent at the target locus (ex. Resistance) and background selection for the marker alleles of recurrent parent in all genomic regions of desirable traits (agronomic traits) except the target locus may be made effective transfer of resistance into elite genotypes (Hospital 2003). Foreground selection ensures the transfer of target trait from donor parent and background selection takes care of the genome recovery of recurrent parent.

MAS can be used when other characters are to be combined from two parents along with resistance trait. However, MAS will be more effective for simply inherited character controlled by a few genes than for a highly complex character governed by large number of genes. In tobacco, SCAR markers were used in marker-assisted breeding of black shank, black root rot, PVY, blue mold, and TMV disease resistance breeding (Whitham et al. 1994; Johnson et al. 2002; Milla et al. 2005; Julio et al. 2006; Li and Miller 2010).

10.8.3 Gene Pyramiding

Pyramiding of multiple genes/QTLs may be achieved through multiple-parent crossing or complex crossing, multiple backcrossing, and recurrent selection. A suitable breeding scheme for marker-assisted gene pyramiding depends on the number of genes/QTLs required to be transferred, the number of parents that contain the desired genes/QTLs, the heritability of traits of interest, and other factors (e.g. cost of genotyping). Pyramiding of three or four desired genes/QTLs existing separately in three or four lines can be realized by three-way, four-way or double crossing, convergent backcrossing or stepwise backcrossing. More than four genes/QTLs can be pyramided by way of complex or multiple crossing and/or recurrent selection.

Gene pyramiding through MABC may be achieved through three different strategies or breeding schemes viz. stepwise, simultaneous/synchronized and convergent backcrossing or transfer. In the stepwise backcrossing, the target genes/QTLs are transferred from donor parents into the recurrent parent (RP) in order one after the other. In the first step of backcrossing, one gene/QTL is targeted and transferred, followed by next step of backcrossing for another gene/QTL, until all target genes/QTLs are introgressed into the RP. The advantage is that gene pyramiding through stepwise backcrossing is more precise and easier to adopt as it involves only one gene/QTL at a time requiring small population size and lower genotyping cost. The disadvantage of this method is it takes a longer time to complete. In the simultaneous or synchronized backcrossing, the recurrent parent is first crossed to each of the donor parents and the resultant single-cross F_1 s are crossed with each other to produce two double-cross F_1 s and then the two double-cross F_1 s are crossed again to produce a hybrid integrating all target genes/QTLs in heterozygous state. The hybrid and/or progeny with heterozygous markers for all the target genes/QTLs are subsequently crossed back to the RP until the satisfactory recovery of the RP genome. Finally, homozygosity of the RP genome recovered line can be achieved through selfed. Simultaneous or synchronized backcrossing takes shorter time to transfer multiple genes, however, requires a large population and more number of genotyping as all target genes/QTLs are involved at the same time. Convergent backcrossing uses both stepwise and synchronized backcrossing strategies. First each of the target gene/QTLs from the donors are transferred separately into the recurrent parent through single crossing followed by backcrossing based on the linked markers to produce improved lines. The improved lines are crossed with each other and the resultant hybrids are then intercrossed to integrate all the genes/QTLs together to develop the final improved line with all the genes/QTLs pyramided. Convergent backcrossing not only reduce time (compared to stepwise transfer) but also easily fix and/or pyramid genes (compared to simultaneous transfer).

Marker-assisted complex or convergent crossing (MACC) can be undertaken to pyramid multiple genes/QTLs if all the parents are improved cultivars with complementary genes or favorable alleles for the traits of interest. In MACC, the hybrid of convergent crossing is self-pollinated for several consecutive generations along with MAS for target traits until genetically stable lines with desired marker alleles and traits have been developed. Detection and selection of most important genes/QTLs in early generations and less important markers in later generations can effectively reduce population size and avoid loss of important genes/QTLs.

Theoretically, application of MABC and MACC for pyramiding target genes/QTLs is possible in tobacco through various schemes discussed above. However, information is not currently available about the release of commercial cultivars resulted using these strategies.

10.8.4 *Limitations and Prospects of MAS and MABC*

Though MAS and MABC breeding has number of advantages, it may not be universally useful (Jiang 2013). Rapid DNA extraction technique and a high throughput system of marker detection are essential to handle a large number of samples and a large-scale screening of multiple markers in breeding programs. Development of suitable bioinformatics and statistical software packages are required for meeting the efficient and quick labeling, storing, retrieving, processing and analyzing large data set requirements, and even for integrating data sets available from other programs. Hence, the startup expenses and labor costs involved in MAS and MABC breeding are higher than conventional techniques making them not in the reach of all the researchers (Morris et al. 2003).

As the distance between the marker and the gene of interest increases, the chance of recombination between gene and marker increases thereby make the selection of resistant plants based on marker ineffective due to false positives. This may be avoided with the use of flanking markers on either side of the locus of interest in order to increase the probability that the desired gene is selected. Sometimes markers that were used to detect a locus may not be 'breeder-friendly'. Such markers viz. RFLP and RAPD may need to be converted into more reliable and easier to use markers. RFLP markers may be converted to STS (sequence tagged site) for detection via PCR protocols (Ribaut and Hoisington 1998) and RAPD markers into SCAR markers for reliable and repeatable amplifications (Milla et al. 2005; Lewis 2005). RAPD technique may be considered less reliable for MAS as RAPD results vary from lab to lab, largely due to low binding specificity of short (10-base) PCR primers. Hence, SCAR markers are developed by sequencing RAPD bands and designing more specific 18–25 base PCR primers to amplify the same DNA segment more reliably. Imprecise estimates of QTL locations and effects may result in slower progress than expected through MAS (Beavis 1998). Sometimes markers developed for MAS in one population may not be suitable for screening other populations due to absence of polymorphism for identified markers or lack of marker-trait association.

MAB is going to become a powerful and reliable tool in genetic manipulation of agronomically important traits in tobacco in view of the increasing utilization of molecular markers in various fields viz. germplasm evaluation, genetic mapping, map-based gene discovery, characterization of traits etc. Currently available high density linkage maps in tobacco provide a framework for identifying marker-trait associations and selecting markers for MAB. Markers linked to resistant traits discussed in previous sections can effectively be utilized in MAB in tobacco. However, only the markers that are closely associated with the target traits or tightly linked to the genes can offer adequate promise for the success in practical breeding. Availability of new high-throughput marker genotyping platforms for the detection of SSR and SNP markers along with the sequencing information of cultivated and wild relatives of *Nicotiana* going to have a great impact on discovering marker trait associations that can be used for MAS in the future. Array-based methods such as DArT (Lu et al. 2012) and single feature polymorphism (SFP) detection

(Rostoks et al. 2005) offer prospects for lower-cost marker technology that can be used for whole-genome scans in tobacco. Rapid growth in genomics research and huge data generated from functional genomics in tobacco in the recent years is leading to the identification of many candidate genes for numerous traits including biotic stress resistance. SNPs within candidate genes could be extremely useful for ‘association mapping’ and circumvents the requirement for construction of linkage maps and QTL analysis for the genotypes that have not been mapped previously. The availability of large numbers of publicly available markers and the parallel development of user-friendly databases (Sol genome network, NCBI etc.) for the storage of marker and QTL data, increasing number of studies on genes and marker trait associations will undoubtedly encourage the more widespread use of MAS in tobacco.

Selection for all kinds of traits at seedling stage in MAB helps to minimize the costs as undesirable genotypes are eliminated at early stages. Closely linked markers allow the selection of disease/pest resistance traits even without the incidence of pests and diseases. MAS based on reliable markers tightly linked to the multiple genes/QTLs for traits of interest can be more effective in pyramiding desirable genes than conventional breeding. Use of co-dominance markers (e.g. SSR and SNP) in MAB allow effective selection of recessive alleles in the heterozygous state without selfing or test crossing, thus saves time and accelerate breeding progress. As more and more newer techniques are available genotypic assays based on molecular markers may be faster, cheaper and more accurate than conventional phenotypic assays and thus MAB may result in higher effectiveness and higher efficiency in terms of time, resources and efforts saved in future.

Conventional breeding methodologies have extensively proven successful in development of tobacco cultivars and germplasm. Subjective assessment and empirical selection plays a significant role in conventional breeding. As a new addition to the whole family of plant breeding, MAB has brought great challenges, opportunities and prospects for breeding crops including tobacco. However, as transgenic breeding or genetic manipulation does, MAB cannot replace conventional breeding but is a supplementary addition to conventional breeding. High genotyping costs and technical/equipment requirements of MAB will be major limiting factors for its large-scale deployment in the near future, especially in the developing countries (Collard and Mackill 2008). Therefore, integration of MAB into conventional breeding programs will be an optimistic strategy for tobacco improvement in the future. It can be expected that the drawbacks of MAB will be gradually overcome leading to its wide spread adoption in practical breeding programs as its theory, technology and application are further developed and improved.

10.9 Map-Based Cloning of Resistance Genes

10.9.1 Traits and Genes

High levels of redundancy between genes in the large and complex genome of tobacco with the absence of molecular markers and genomic resources till recent years made the identification and subsequent mapping of interesting mutants a very difficult prospect. However, having anchored 64% of the genome assembly to chromosomal locations in recent years, a possibility now exists to apply map-based biotic stress resistant gene discovery approaches in the species (Edwards et al. 2017). First reported instance of successful map based cloning in tobacco was done by Edwards et al. (2017) and cloned *NtEGY1* and *NtEGY2* homeologous candidate genes for *YB1* and *YB2* loci conferring white stem phenotype in recessive condition in burley tobacco. However, map-based cloning of pest and disease resistant genes based on genetic maps are not yet reported in tobacco.

10.9.2 Strategies: Chromosome Landing and Walking

Chromosome landing and walking strategies are used in identification of clones carrying gene of interest for map based cloning. Recently available high density genetic maps, genome sequences and Bacterial artificial chromosome (BAC) clones are paving the way for map based cloning of resistance genes in tobacco.

In the only reported case of map based cloning in tobacco, Edward et al. (2017) used a specific technique to clone genotyped pairs of NILs carrying dominant or recessive alleles of the *YB1* and *YB2* loci (cultivars SC58, NC95, and Coker 1) with a custom 30 K Infinium iSelect HD BeadChip SNP chip (Illumina Inc., San Diego, CA) used in developing a high density genetic map (*N. tabacum* 30 k Infinium HD consensus map 2015; https://solgenomics.net/cview/map.pl?map_version_id=178). Genomic regions comprising SNP polymorphisms that distinguished the nearly isogenic lines were identified and SNP markers closely linked to the loci were aligned to the genome assembly and predicted potential candidate genes. Coding regions of candidate genes were then amplified from first-strand cDNA from tobacco cultivars K326 and TN90 using the primers specifically designed. Amplified fragments were then cloned into a vector.

10.9.3 Genomic Libraries

Availability high-capacity genomic libraries are essential resources for physical mapping, comparative genome analysis, molecular cytogenetics etc. Such libraries

are also powerful tools for large-scale gene discovery, elucidation of gene function and regulation, and map-based cloning of target trait loci or genes associated with important agronomic and resistant traits for their further study and use in crop improvement programs. BAC libraries are the large DNA insert libraries (inserts of DNA up to 200,000 base pairs) of choice for genomics research. Cloning of larger DNA segments (more than 1000 kb) are possible with Yeast Artificial Chromosome (YAC) libraries and greatly facilitates chromosome walking and physical mapping around the target locus. While, transformation-competent artificial chromosome (TAC) libraries make it possible to clone and transfer genes efficiently into plants. In recent years, BAC libraries are constructed and utilized in tobacco for genome sequencing, mapping and comparative genome analysis. However, construction of YAC and TAC libraries are not currently reported in tobacco.

Tobacco Genome Initiative (TGI) generated BACs library (9.7-fold genome coverage) for assembling the partial genome of Hicks Broadleaf variety (Opperman et al. 2003; Rushton et al. 2008). Sierro et al. (2013b) used 425,088 BAC clone library for construction of physical map and ancestral annotation of tobacco cultivar, Hicks Broadleaf. Edwards et al. (2017) constructed two libraries containing 150,528 BACs from K 326 variety using *Hind*III or *Eco*RI, with average insert sizes of 115 kb and 135 kb, respectively (representing $\sim 8 \times$ coverage of the genome) and used for generating a whole-genome profile (WGP) map from sequence reads at *Eco*RI and *Hind*III restriction sites. Jingjing (2018) reported a tobacco genome sequence of the HongDa cultivar, which has been produced by the combination of BAC-to-BAC and whole-genome shotgun technologies. Dong et al. (2020) constructed a BAC library of 414,720 clones using blank shank resistant flue-cured tobacco line, 14–60 with an average insert size of 123 kb ranging from 97.0–145.5 kb covering 11 times of genome equivalents. Further confirmed the utility of this library by screening the library with gene specific primers. A BAC library of wild tobacco, *N. tomentosiformis*, one of the parent of *N. tabacum* was constructed with inserted DNA size ranging from 50 to 200 kb and an estimated average size of 110 kb (Yuhe 2012). These libraries are important resources for map based cloning of resistant traits through various strategies.

10.9.4 Test for Expression (Mutant Complementation)

Transformation of a cloned gene into mutant plant and looking for wild phenotype rescue will help to validate the function of the target gene. However, currently mutant complementation studies with cloned genes are not reported in tobacco due to the absence of map based cloning of functional genes in general and biotic stress resistant genes particular.

10.10 Genomics-Aided Breeding for Resistance Traits

10.10.1 Structural and Functional Genomic Resources Developed

The aim of structural genomics is to characterize the structure of the genome. Understanding the genome structure of an individual can be advantageous in manipulating genes and DNA fragments in that species. In tobacco, huge data on genomic resources have been generated from in-depth genomic studies by various researchers. These resources are available for furthering the research in the area of genomics, gene tagging, identification, isolation and cloning for genome assisted breeding in tobacco as well as other crop species.

With advances in next-generation sequencing (NGS) technologies, decoding the genome sequences of 12 *Nicotiana* species viz. *N. tabacum*, *N. rustica*, *N. attenuata*, *N. benthamiana*, *N. knightiana*, *N. obtusifolia*, *N. otophora*, *N. paniculata*, *N. undulata*, *N. tomentosiformis*, *N. sylvestris* and *N. glauca* has been completed. Sequencing details of *Nicotiana* Spp. are presented at NCBI website (<https://www.ncbi.nlm.nih.gov/>) and Sol Genome Network (SGN) (Asaf et al. 2016).

Plastid genomes of 10 tobacco species: *N. tabacum*, *N. attenuata*, *N. tomentosiformis*, *N. sylvestris*, *N. otophora*, *N. knightiana*, *N. rustica*, *N. paniculata*, *N. obtusifolia* and *N. glauca* are sequenced and data made available (Asaf et al. 2016; Mehmood et al. 2020).

The developments in Transcriptomics have resulted in the development of large data sets and tools for the progression of functional genomics in tobacco. A database of 2513 tobacco (*N. tabacum*) TFs representing all of the 64 well-characterized plant TF families created using a dataset of 1,159,022 gene-space sequence reads (GSRs) (Rushton et al. 2008). Further, the transcriptional activity for thousands of tobacco genes in different tissues throughout the lifecycle of the tobacco from seed to senescence based on tobacco expression microarray from a set of over 40 k unigenes (a set of transcripts that appear to stem from the same transcription locus) and gene expression in 19 different tobacco samples has been generated (Edwards et al. 2010). 772 of 2513 transcription factors earlier recognized in tobacco were mapped to the array, with 87% of them being expressed in at least one tissue in the generated Tobacco Expression Atlas (TobEA). Based on the co-expression of these transcription factors, putative transcriptional networks could be identified. SGN contains the collection of transcriptome sequences of *N. sylvestris* (32,276), *N. tomentosiformis* (31,961) and *N. tabacum* (26,284) from transcriptome projects and unigenes data sets of *N. sylvestris* (6300), *N. tabacum* (84,602) and *N. benthamiana* (16,024).

The data hosted at NCBI and SGN databases includes majority of the available genomic resources information pertaining to cultivated tobacco and wild *Nicotiana* spp. Hence, the available information in these databases was discussed here as an indicator of genome resource availability in *Nicotiana*. Large collection of data on nucleotides, genes and protein sequences submitted by various researchers on *Nicotiana* are available at NCBI site (Table 10.8). More than 33 lakh nucleotide

Table 10.8 Nucleotide, gene, SRA, GEO and protein sequences of *Nicotiana* spp. available at NCBI website (as on 30.09.2021)

Species	Nucleotide sequences	Gene sequences	Sequence read archive (SRA) data	Gene expression omnibus (GEO) data	Proteins
<i>N. alata</i>	238	–	–	–	265
<i>N. amplexicaulis</i>	–	129	–	–	174
<i>N. attenuata</i>	185,245	40,175	455	496	78,686
<i>N. benthamiana</i>	97,525	–	1570	585	1009
<i>N. bonariensis</i>	–	–	13	14	–
<i>N. clevelandii</i>	–	–	12	14	–
<i>N. debneyi</i>	61	129	–	–	195
<i>N. glauca</i>	107	129	30	14	335
<i>N. glutinosa</i>	196	–	–	–	184
<i>N. knightiana</i>	72,026	–	50	–	–
<i>N. langsdorffii</i> × <i>N. sanderiae</i>	12,576	–	–	40	124
<i>N. langsdorffii</i>	59	–	–	–	–
<i>N. linearis</i>	–	–	6	–	–
<i>N. megalosiphon</i>	296	–	–	–	–
<i>N. miersii</i>	–	–	12	–	–
<i>N. obtusifolia</i>	–	–	17	–	116
<i>N. occidentalis</i>	–	–	7	–	–
<i>N. otophora</i>	–	155	8	–	240
<i>N. paniculata</i>	67,473	–	44	–	126
<i>N. pauciflora</i>	–	–	13	–	–
<i>N. plumbaginifolia</i>	710	–	22	23	404
<i>N. repanda</i>	133	129	–	–	260
<i>N. rosulata</i>	–	–	11	–	–
<i>N. rustica</i>	112,008	–	102	14	132
<i>N. stocktonii</i>	–	129	–	–	195
<i>N. suaveolens</i>	–	131	–	–	224
<i>N. suaveolens</i> × <i>N. tabacum</i>	140	–	–	–	–
<i>N. sylvestris</i>	385,277	40,544	41	22	49,040
<i>N. tabacum</i>	2,053,476	74,274	2589	1203	92,478
<i>N. tomentosiformis</i>	263,306	45,486	38	–	50,492
<i>N. undulata</i>	59,446	156	42	–	262
<i>Nicotiana</i> spp.	177	–	–	2438	–

(continued)

Table 10.8 (continued)

Species	Nucleotide sequences	Gene sequences	Sequence read archive (SRA) data	Gene expression omnibus (GEO) data	Proteins
<i>Total</i>	3,310,475	201,566	5082	4863	274,941

sequences of 20 *Nicotiana* spp. that includes genomic DNA/RNA, mRNA, cRNA, ncRNA, rRNA, tRNA and transcribed RNA is generated by various researchers and is available at NCBI website as on 30.09.2021 (Table 10.8). Among them, around 895,700 sequences are comprehensive, integrated, non-redundant, well-annotated set of reference sequences including genomic, transcript, and protein. These sequences also included 456,507 expressed sequence tags (ESTs) and 1,420,639 genomic survey sequences (GSS). Further, a total of over 201,560 records of gene sequences belonging to 12 *Nicotiana* spp. viz. *N. tabacum*, *N. tomentosiformis*, *N. sylvestris*, *N. attenuata*, *N. undulate*, *N. otophora*, *N. suaveolens*, *N. glauca*, *N. stocktonii*, *N. repanda*, *N. amplexicaulis* and *N. debneyi* are existing at NCBI (Table 10.8).

Sequence Read Archive (SRA) data is the largest publicly available repository of high throughput sequencing data and is accessible through several cloud providers and NCBI servers. The archive admits data from different branches of life including metagenomic and environmental surveys. SRA stores raw sequencing data and alignment information to enhance reproducibility and facilitate new discoveries through data analysis (<https://www.ncbi.nlm.nih.gov/sra>). Nearly, 5080 records of SRA data of 20 *Nicotiana* spp. are available at NCBI website (as on 30.09.2021) (Table 10.8). Around 4860 curated gene expression datasets as well as original series and platform records of 11 *Nicotiana* spp. are available at Gene Expression Omnibus (GEO) repository of NCBI as on 30.09.2021 (Table 10.8; <https://www.ncbi.nlm.nih.gov/gds>). Around 275,000 collection of protein sequences from several sources, including translations from annotated coding regions in GenBank, RefSeq and Third Party Annotation (TPA) Sequence, as well as records from other data bases are available for 20 *Nicotiana* spp. at NCBI (Table 10.8).

Genomic resource collection of SGN include information on genome sequences, transcriptome sequences, mRNAs, and predicted proteins of one or the other five wild *Nicotiana* spp. namely *N. attenuata*, *N. benthamiana*, *N. tomentosiformis*, *N. sylvestris* and *N. otophora* and four versions of *N. tabacum*. SGN also hosts 39 transcript libraries of *N. tabacum* and two of *N. sylvestris* are also present. Though both NCBI and SGN utilize the information from same bio-projects, the number of sequences, predicted proteins, CDS and mRNA for different species are not similar in these databases due the use of different bioinformatics softwares and gene prediction model in analyzing the sequences.

The proteomic data generated globally is stored and accessed through the Universal Protein Resource (UniProt). UniProt provides several sets of proteins (proteomes) thought to be expressed by organisms whose genomes have been

completely sequenced (<https://www.uniprot.org/>). There are 73,606 protein entries associated with *Nicotiana tabacum* proteome (UP000084051) in the UniProt database as on 31.03.2021.

10.10.2 Details of Genome Sequencing

10.10.2.1 Nuclear Genome Sequencing

Though, tobacco have the same basic chromosome number of $n = 12$ as of other solanaceous crops such as tomato, potato, chili and eggplant, its genome size (4.5 Gbp) is largest in the Solanaceae (Arumuganathan and Earle 1991) with a large proportion of repetitive sequences (Zimmerman and Goldberg 1977; Kenton et al. 1993; Leitch et al. 2008; Sierro et al. 2014). *N. tabacum* genome is 50% larger than the human genome. Analysis of the tobacco genome has been started in the last decade. Tobacco Genome Initiative (TGI) have initiated the first tobacco genome sequencing project in 2003 by the with the aim of sequencing the open reading frames of *N. tabacum* utilizing the methyl-filtration technology for reducing genome complexity (Opperman et al. 2003; Rushton et al. 2008). The Hicks Broadleaf variety, ancestor for some of the currently cultivated flue-cured tobacco cultivars, was chosen as the genotype for generation of bacterial artificial chromosome (BAC) libraries used for sequencing because of its low introgression content (Sierro et al. 2013b). The project was completed in 2007 and the sequences are available from the National Center for Biotechnology Information (NCBI) GenBank (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA29349>). Around 689 Mb of genomic sequence generated through Sanger sequencing was assembled into 81,959 contigs with an average length of 1.2 kb and 871,255 singletons with an average length of 688 bp (Wang and Bennetzen 2015). However, the generated sequences comprised only a small portion of the tobacco genome, mainly because the sequencing technology (methylation filtration) was employed for only enriched genes, and for the portions of genes that are under methylated relative to TEs.

With the advances in next-generation sequencing (NGS) technologies, there is a rapid progress in sequencing of entire genomes of cultivated tobacco and important wild relatives in recent times (Table 10.9). Sequencing of three *N. tabacum* cultivars have been completed since 2014. Eleven complete genomes of wild *Nicotiana* species viz. *N. knightiana*, *N. paniculata*, *N. rustica*, *N. glauca*, *N. obtusifolia*, *N. otophora*, *N. attenuata*, *N. sylvestris*, *N. tomentosiformis*, *N. undulata* and *N. benthamiana* were sequenced and released since 2013. The scaffold level assemblies of *N. sylvestris* and *N. tomentosiformis* having $94.0\times$ and $146.0\times$ genome coverage and a length of 2222 and 1688 Mb, respectively being the first sequences released in 2013 (Sierro et al. 2013a). This was followed by the release of three genome assemblies of *N. tabacum*, one at Scaffold level (cv. TN90) and two at contig (cv. K326 and Basma Xanthi -BX) level by Philip Morris International in 2014 (Sierro et al. 2014). The length of these assemblies was around 3700 Mb with a GC content of 39% and genome coverages

Table 10.9 Genome sequencing details of *Nicotiana* species as (available at NCBI site)

Genome	Date published (dd-mm-yyyy)	Submitter	Assembly level	Assembly	Statistics	Genome coverage	Sequencing technology
<i>N. sylvestris</i>	08-05-2013	Philip Morris International R&D	Scaffold	Scaffolds: 253,918; Contigs: 392,998; N50: 26,744; L50: 21,230	Total length (Mb): 2221.99; GC%: 39.80	94.0x	Illumina HiSeq
<i>N. tomentosiformis</i>	10-05-2013	Philip Morris International R&D	Scaffold	Scaffolds: 159,548; Contigs: 215,609; N50: 34,051; L50: 13,932	Total length (Mb): 1688.47; GC%: 39.0999	146.0x	Illumina HiSeq
<i>N. benthamiana</i>	27-05-2014	National Academy of Agricultural Sciences	Contig	Scaffolds: 100,480; Contigs: 100,480; N50: 776; L50: 22,990	Total length (Mb): 61.9511; GC%: 39.6	79x	–
<i>N. tabacum</i> cv.TN90 (RefSeq Genome)	29-05-2014	Philip Morris International R&D	Scaffold	Scaffolds: 168,247; Contigs: 292,601; N50: 40,422; L50: 25,298	Total length (Mb): 3643.47; GC%: 39.2006	49.0x	Illumina HiSeq

(continued)

Table 10.9 (continued)

Genome	Date published (dd-mm-yyyy)	Submitter	Assembly level	Assembly	Statistics	Genome coverage	Sequencing technology
<i>N. tabacum</i> cv. K326 (sequencing & assembly)	29-05-2014	Philip Morris International R&D	Contig	Scaffolds: 582,565; Contigs: 711,416; N50: 24,673; L50: 40,137	Total length (Mb): 3732.64; GC%: 39	38.0x	Illumina HiSeq
<i>N. tabacum</i> cv.:BX (sequencing & assembly)	29-05-2014	Philip Morris International R&D	Contig	Scaffolds: 643,545; Contigs: 797,404; N50: 21,713; L50: 44,668	Total length (Mb): 3735.82; GC%: 39	29.0x	Illumina HiSeq
<i>N. otophora</i>	05-06-2014	Philip Morris International R&D	Contig	Scaffolds: 929,607; Contigs: 1,121,520; N50: 11,316; L50: 44,787	Total length (Mb): 2689.35; GC%: 39.1	66.0x	Illumina HiSeq
<i>N. attenuata</i> strain: UT (RefSeq, sequencing & assembly)	14-11-2016	Max Planck Institute for Chemical Ecology	Chromosome (12 haploid)	Scaffolds: 37,194; Contigs: 103,240; N50: 64,201; L50: 9648	Total length (Mb): 2,365.68; GC%: 41.333	30.0x	Illumina HiSeq

(continued)

Table 10.9 (continued)

Genome	Date published (dd-mm-yyyy)	Submitter	Assembly level	Assembly	Statistics	Genome coverage	Sequencing technology
<i>N. obtusifolia</i>	06-03-2017	Max Planck Institute for Chemical Ecology	Scaffold	Scaffolds: 53,128; Contigs: 93,877; N50: 39,219; L50: 9106	Total length (Mb): 1222.77; GC%: 38.4	50.0×	Illumina HiSeq
<i>N. attenuata</i> ecotype Arizona (sequencing & assembly)	06-03-2017	Max Planck Institute for Chemical Ecology	Scaffold	Scaffolds: 951,503; Contigs: 2,556,970; N50: 1153; L50: 283,638	Total length (Mb): 1827.78; GC%: 29.1	18.0×	Illumina HiSeq
<i>N. tabacum</i> cv. K326 (sequencing and assembly)	28-06-2017	British American Tobacco	Scaffold	Scaffolds: 937,112; Contigs: 1,257,801; N50: 9081 L50: 108,397	Total length (Mb): 4646.65; GC%: 33.5	86×	Illumina HiSeq
<i>N. glauca</i>	16-02-2018	Saint Petersburg University	Scaffold	Scaffolds: 514,289 Contigs: 1,214,276; N50: 5814; L50: 143,017	Total length (Mb): 3222.83; GC%: 36.6	35.0×	Illumina

(continued)

Table 10.9 (continued)

Genome	Date published (dd-mm-yyyy)	Submitter	Assembly level	Assembly	Statistics	Genome coverage	Sequencing technology
<i>N. knightiana</i>	02-05-2019	Philip Morris International R&D	Scaffold	Scaffolds: 160,415; Contigs: 368,149; N50: 28,120; L50: 23,129	Total length (Mb): 2298.94; GC%: 39.7	82.0x	Illumina HiSeq
<i>N. paniculata</i>	02-05-2019	Philip Morris International R&D	Scaffold	Scaffolds: 181,977 Contigs: 289,116; N50: 32,292; L50: 19,574	Total length (Mb): 2190.56; GC%: 39.4	100.0x	Illumina HiSeq
<i>N. rustica</i>	02-05-2019	Philip Morris International R&D	Scaffold	Scaffolds: 337,581 Contigs: 863,270; N50: 14,482; L50: 79,215	Total length (Mb): 4231.29; GC%: 40	90.0x	Illumina HiSeq
<i>N. undulata</i>	02-05-2019	Philip Morris International R&D	Scaffold	Scaffolds: 117,566 Contigs: 240,737; N50: 28,381; L50: 19,751	Total length (Mb): 1914.3; GC%: 40	112.0x	Illumina HiSeq

ranging from 29.0 to 49.0 \times , while the assembly of TN 90 is used as reference genome. Reference genome is a comprehensive, integrated, non-redundant and well-annotated set of sequences for a given genome. Later, an improved version of *N. tabacum* (cv. K 326) assembled at scaffold level covering the 86.0 \times of genome (4600 Mb) with a GC content of 33.5% was submitted by British American Tobacco in 2017 (Edward et al. 2017). This assembly covers over 4 Gb of non-N sequence (90% of predicted genome size), which is an increase from 3.6 Gb (81% of predicted genome size) in the previously published version (Sierro et al. 2014). This assembly achieved an N50 size of 2.17 Mb and anchored 64% of the genome to pseudomolecules; a significant increase from the previous value of 19% providing more complete coverage of the tobacco genome.

Currently, a total of 16 assemblies of 12 *Nicotiana* species viz. *N. sylvestris*, *N. tomentosiformis*, *N. benthamiana*, *N. tabacum*, *N. otophora*, *N. attenuata*, *N. obtusifolia*, *N. glauca*, *N. knightiana*, *N. paniculata*, *N. rustica* and *N. undulata* are obtainable at NCBI genbank with 18.0 \times to 146.0 \times genome coverages. Two assemblies are available for *N. attenuata* and four for *N. tabacum*. *N. attenuata* reference sequence (2366 Mb) is assembled at chromosomal level (12 haploid), while *N. tabacum* (K326, and Basma Xanthi) and *N. benthamiana* at contig level. All the other *Nicotiana* assemblies are available at scaffold level. Genome coverage of the sequences range from 18.0 \times in *N. attenuata* strain UT to 146.0 \times in *N. tomentosiformis*. GC content of the assemblies ranges from 29.1 to 41.3% and length ranges from 62 Mb in *N. benthamiana* (diploid genome) to 4647 Mb in *N. tabacum* cv. K326 (tetraploid). Most of the genomes were sequenced using high throughput Illumina Hiseq sequencing technology. The detail statistics including assembly level and their N50 and L50 values for each genome was also provided at Table 10.9.

Genomic resource collection of five wild *Nicotiana* spp. and four of *N. tabacum* are available at Sol Genome Network (SGN) (Fernandez-Pozo et al. 2015). They include contig level methylation filtered genome sequences generated under TGI project, and genomes, predicted proteins, and mRNAs *N. tabacum* cv. BX, *N. tabacum* cv. K326 and *N. tabacum* cv. TN90 and genome scaffolds, proteins and cDNA of improved K326 assembly (Sierro et al. 2014; Edwards et al. 2017). Further, scaffold level genome assemblies of five *Nicotiana* species namely *N. attenuata*, *N. tomentosiformis*, *N. benthamiana*, *N. sylvestris* and *N. otophora* also available at this database along with predicted proteins and mRNA for first four species. Chromosome level assembly is available for *N. attenuata*. Though the source of these assemblies is same as of NCBI, the sequence info is not similar due to differences in bioinformatics softwares employed in analysing the sequences.

10.10.2.2 Organelle Genome Sequencing

Plastid and mitochondrial genomes of tobacco are circular DNA molecules. Sequencing of chloroplast genome of tobacco, for the first time, was completed in 1986 (Shinozaki et al. 1986) and till then number of studies made to sequence chloroplast genomes of different entries of *N. tabacum* and its wild relatives. Currently,

Table 10.10 *Nicotiana* organelle genome resources available at NCBI (as on 31.03.2021)

Organelle reference genome	Sequence length (bp)	Archived complete genomes	PopSet	Genes	Proteins
Plastid					
<i>N. tabacum</i> (RefSeq)	155,943	124	10	144	98
<i>N. attenuata</i>	155,886	15	1	129	84
<i>N. tomentosiformis</i>	155,745	30	–	150	102
<i>N. sylvestris</i>	155,941	42	1	149	101
<i>N. otophora</i>	156,073	8	–	155	108
Mitochondrial					
<i>N. tabacum</i>	430,597	6	4	183	153
<i>N. attenuata</i>	394,341	1		68	40
<i>N. sylvestris</i>	430,597	1	1	64	37

sequencing of about 219 complete plastid genomes including 12 popset related to five *Nicotiana* spp. are completed (Table 10.10). The PopSet data is a collection of related DNA sequences derived from population, phylogenetic, mutation and ecosystem studies. The size of the plastid genome of *Nicotiana* species is around 0.156 Mb. In addition, recently, Mehmood et al. (2020) assembled the plastid genomes of five more tobacco species: *N. knightiana* (155,968 bp), *N. rustica* (155,849 bp), *N. paniculata* (155,689 bp), *N. obtusifolia* (156,022 bp) and *N. glauca* (155,917 bp) and made comparisons among themselves. Reference plastid genomes of five *Nicotiana* species namely *N. tabacum* (155,943 bp), *N. attenuata* (155,886 bp), *N. tomentosiformis* (155,745 bp), *N. sylvestris* (155,941 bp), and *N. otophora* (156,073 bp) are available at NCBI (Table 10.9).

Sequencing of mitochondrial complete genome of tobacco started in 2003 (Sugiyama et al. 2005). Eight mitochondrial complete genome sequences including 5 popsets from three *Nicotiana* species namely *N. tabacum* (430,597 bp), *N. attenuata* (394,341 bp) and *N. sylvestris* (430,597 bp) are completed till date and the details are available at NCBI site (Table 10.10). Further, reference mitochondrial genomes are made available for *N. tabacum*, *N. attenuata* and *N. sylvestris* at the NCBI.

10.10.3 Gene Annotation

Genome annotation is the process of identifying functional elements along the sequence of a genome. Assigning function to genome sequence is necessary because the sequencing of DNA produces sequences of unknown function. Once a genome is sequenced, it needs to be annotated to understand its functions for its further successful utilization in genetic manipulation. In tobacco, both nuclear and organelle genomes are successfully annotated and annotation reports are available at various

data bases. Gene annotation records available at NCBI and SGN databases are discussed here they host majority data of the *Nicotiana* sp.

NCBI Eukaryotic Genome Annotation Pipeline is an automated annotation pipeline that annotates genes, transcripts and proteins on draft and finished genome assemblies. Currently, annotation reports are available for four *Nicotiana* species viz. *N. sylvestris*, *N. tabacum* cv. TN 90, *N. attenuata* strain UT and *N. tomentosiformis* at NCBI website (Table 10.11). The first annotation was released in 2014 for *N. sylvestris* followed by *N. tabacum* cv. TN 90 (in 2016), *N. attenuata* strain UT (2016) and *N. tomentosiformis* (2020). Before annotation the repeat sequences in the reference sequences were masked with RepeatMasker making use of curated libraries of

Table 10.11 Gene annotation reports of *Nicotiana* species (as per NCBI)

Species	<i>N. sylvestris</i>	<i>N. tabacum</i>	<i>N. attenuata</i>	<i>N. tomentosiformis</i>
Genome assembly name	Nsyl	Ntab-TN90	NIATTr2	Ntom_v01
Annotation release number	100	100	100	102
Release date	22-10-2014	04-05-2016	06-12-2016	21-04-2020
RefSeq scaffolds	253,917	168,245	37,135	159,547
% Masked with RepeatMasker	1.33%	3.00%	1.33%	1.49%
% Masked with WindowMasker	48.35%	51.43%	47.93%	53.07%
Genes and pseudogenes	40,317	73,946	39,977	45,485
• Protein-coding	33,678	61,526	34,094	31,842
• Non-coding	4667	9019	3886	11,549
• Pseudogenes	1972	3401	1997	2094
• Genes with variants	9033	14,549	6,946	10,545
mRNAs	48,059	84,001	44,491	50,010
• Known RefSeq (NM_)	0	0	0	18
• Model RefSeq (XM_)	48,059	84,001	44,491	49,992
Other RNAs	10,984	17,907	7156	17,677
• Known RefSeq (NR_)	0	0	0	0
• Model RefSeq (XR_)	10,212	16,461	6069	16,867
Coding sequences (CDS)	48,059	84,001	44,491	50,112
• Known RefSeq (NP_)	0	0	0	18
Model RefSeq (XP_)	48,059	84,001	44,491	50,094
<i>Alignment of the annotated proteins to Ref Seq proteins</i>				
A protein with an alignment covering 50% or more of the query		51,484	27,475	25,881
An alignment covering 95% or more of the query		23,115	12,029	11,730

Table 10.12 Gene annotation reports of *Nicotiana* species (as per SGN)

<i>Nicotiana</i> species	Reference scaffolds	Predicted proteins	mRNA
<i>N. attenuata</i>	37,182	33,449	33,449
<i>N. benthamiana</i>	56,094	57,140	–
<i>N. tomentosiformis</i>	159,548	53,753	85,853
<i>N. sylvestris</i>	253,917	54,497	87,234
<i>N. tabacum</i> cv. BX	420,216	86,009	146,748
<i>N. tabacum</i> cv. K326	382,373	85,994	145,503
<i>N. tabacum</i> cv. TN90	249,104	122,388	189,413
<i>N. tabacum</i> v4.5 (Edwards et al. 2017)	1,084,432	69,500	–

repeats and WindowMasker software (Morgulis et al. 2006). RepeatMasker masked 1.33–3.0% of genomes while WindowMasker masked 47.93–53.07%. Based on gene prediction tools, genes and pseudogenes, protein-coding and non-coding sequences etc. were estimated for all four genomes. While, alignments with RefSeq and other known transcripts and proteins were made to predict mRNAs, other RNAs and proteins. The final set of annotated proteins from the *Nicotiana* species was aligned against the available known RefSeq proteins from *Arabidopsis thaliana*, utilizing the high-quality proteins as the target and the annotated proteins as the query. Protein with an alignment covering 50% or more of the query and 95% or more of the query are identified in each of the annotated *Nicotiana* species. The details of the annotations are given at Table 10.11.

Annotated gene predictions are made available for published genomes of *N. attenuata*, *N. benthamiana*, *N. tomentosiformis*, *N. sylvestris* and 4 versions of *N. tabacum* at SGN site also (Table 10.12). For *N. tabacum*, predicted proteins were ranging from 69,500 to 122,388 and mRNA from 145,503 to 189,413. However, less number of proteins (33,449–54,497) and mRNA (33,449–87,234) were predicted for *Nicotiana* species compared to *N. tabacum*.

Nicotiana tabacum proteome (UP000084051) in the UniProt database contains 73,606 protein entries as on 31.03.2021 (<https://www.uniprot.org/>). Biotic stress response related proteins are also included in this list. Edward et al. (2017) identified predicted proteins showing good cross-over with the related *Solanaceae* species tomato and potato in addition to other flowering plants based on gene ontology analysis.

Annotations are also done for published organelle genome sequences predicting genes and proteins (Table 10.10). For plastid genomes, predicted genes in five *Nicotiana* spp. vary from 129 to 155 and proteins from 84 to 108. In mitochondria, number of genes and proteins predicted are more for cultivated species, *N. tabacum* (183; 153, respectively) than for wild species, *N. attenuata* (68; 40) and *N. sylvestris* (64; 37).

10.10.4 Impact on Germplasm Characterization and Gene Discovery

Sequencing of *Nicotiana* spp. made it possible to compare the genomes of *Nicotiana* spp. among themselves and with other solanaceous crops. These studies assisted in identifying the relationships between cultivated and wild species, and their progenitor species in terms of sequence similarities and genome rearrangements (Wu et al. 2009; Sierro et al. 2014; Asaf et al. 2016; Gong et al. 2016; Edwards et al. 2017). The synteny between the genomes of *N. tabacum* cv. TN90, K326 and BX and those of tomato and potato could be evaluated at the protein level which leads to the detection of homologous genes (Sierro et al. 2014).

Annotation of published genome sequences assisted in the identification of functional sequences, and predicted mRNAs and proteins that can be expressed in tobacco. Though comparison of genome assemblies, genomic regions responsible for virus tolerance was recognized in draft genomes (Sierro et al. 2014). Nearly perfect match of *N. glutinosa* *N* gene (source of TMV resistance in many tobacco cultivar) sequence was found on the draft genome sequence of a TMV resistant cultivar, TN 90 and weak identity in susceptible genomes (K326 and BX genomes). Genes that are differentially expressed between PVY resistant and susceptible RIL plants were identified based on comparison with a reference transcriptome (Julio et al. 2014). Further, through annotation of differentiation expressed genes, Julio and co-workers confirmed that functional *eIF4E* gene mapped on chromosome 21 of the tobacco genetic map is responsible for PVY susceptibility in majority of the tobacco lines carrying dominant *Va* locus. One copy each of *eIF4E1*, *eIF4E2* and *eIF (iso) 4E* gene were identified in the *N. sylvestris* genome, whereas two copies of *eIF4E1* and one copy of the other genes in *N. tomentosiformis* (Sierro et al. 2014). Except the *N. sylvestris eIF4E1* gene in TN90, all identified *N. sylvestris* and *N. tomentosiformis* genes were observed in TN90, K326 and BX genomes. With this observation, Sierro et al. (2014) confirmed that the genomic deletion of the S-form *eIF4E1* locus is responsible for TVMV, TEV and PVY resistance in TN90 and its presence for susceptibility in K326 and BX.

Based on sequencing of *Nicotiana* genomes and EST data, large number of SSRs and SNPs were identified (Bindler et al. 2011; Tong et al. 2012b, 2020; Cai et al. 2015; Xiao et al. 2015; Thimmegowda et al. 2018; Wang et al. 2018). The identified markers have been used for characterization of germplasm, diversity studies, DUS testing and genetic relatedness of cultivated varieties etc. (Moon et al. 2008, 2009a, b; Davalieva et al. 2010; Fricano et al. 2012; Gholizadeh et al. 2012; Prabhakararao et al. 2012; Xiang et al. 2017; Binbin et al. 2020). Wang et al. (2021) used core markers developed based on genotyping-by-sequencing for varietal identification and fingerprinting of cigar tobacco accessions. The high-density maps developed based on SSR and SNP markers will be useful in characterization of germplasm, and identification of target traits including biotic stresses. Genome-wide SNP markers were used for simultaneous association analysis of leaf chemistry traits in natural populations of tobacco germplasms (Tong et al. 2020). SSRs and SNPs were used

to identify QTLs linked to biotic stress resistant traits viz. Bacterial wilt (Qian et al. 2013), Brown spot (Tong et al. 2012a; Sun et al. 2018), black shank (Ma et al. 2019, 2020; Gong et al. 2020) and CMV (Cheng et al. 2019).

The release of *N. tabacum* 30 k Infinium HD consensus map 2015 also provides the tobacco genetic research community with resources to detect genome-wide DNA polymorphisms, fine map and clone their trait of interest. Genome-wide DNA polymorphisms could be detected using the custom 30 K Infinium iSelect HD BeadChip SNP chip (Edwards et al. 2017). Map based cloning of target traits is becoming a reality in tobacco with cloning of two homeologous candidate genes conferring white stem phenotype in recessive condition in burley tobacco (Edwards et al. 2017).

Advances in tobacco genomics provide further means to advance the understandings of diversity at species and gene levels, and allows DNA markers to hasten the pace of genetic improvement. Discovery of novel genes/alleles for any given trait could be obtained through genotyping-by-sequencing, whole-genome re-sequencing, Sequence-based mapping etc. Genomics tools also enable rapid identification and selection of novel beneficial genes and their controlled incorporation into germplasm. Similarly, genome-wide association studies (GWAS) could help to identify the genomic regions controlling traits of interest in diverse collection of germplasms that are genotyped and phenotyped for traits of interest through statistical associations between DNA polymorphisms and trait variations. Genomics possesses the potential to increase the diversity of alleles available to breeders through mining the gene pools of crop wild relatives (CWRs).

10.10.5 Application of Structural and Functional Genomics in Genomics-Assisted Breeding

Recent advances in high-throughput sequencing and phenotyping platforms are transforming molecular breeding to genomics-assisted breeding (GAB). GAB is going to be the key in designing future tobacco cultivars through optimizing the tobacco genomes with accumulation of beneficial alleles and eradicating deleterious alleles (Varshney et al. 2021). Availability of draft *Nicotiana* genomes, transcriptome and metabolome profiles is paving the way for understanding the genomic areas, their ancestral origins, genes, gene products, expression patterns, control elements and associated allelic variations responsible for resistance to biotic stresses. Availability of molecular markers (Sects. 10.5.2 and 10.7.2), high density genetic maps (Sect. 10.7), structural and functional resources (Sect. 10.10), identification of DNA markers linked to traits of interest, identification and mapping of trait specific QTLs in recent years is going to be the starting points for the utilization of GAB an important tool for tobacco improvement.

Currently, application of structural and functional genomics in GAB in tobacco is in initial stage. Application of DNA markers to facilitate marker-aided selection (MAS), a preliminary form of GAB, for tobacco improvement is in commencement

stage. As progress is made in genomics, genotypic and phenotypic datasets on training populations can be used to develop models to predict the breeding value of lines for genomic selection (GS). All the available marker data for a population can be used as predictors of breeding value of a line. Breeding value serves as a predictor of how well a plant will perform as a parent for crossing and generation advance in a breeding program, based on the resemblance of its genomic profile to other plants in the training populations that are known to have performed better in the target environment(s). The genotypic data obtained from a seed or seedling on all the favorable alleles can be used to predict the phenotypic performance of mature individuals without the need for extensive phenotypic evaluation over years and environments during GAB (Varshney et al. 2014). The time required for breeding will be drastically come down due to selection of desirable lines in early generations with accuracy without the influence of environment.

10.11 Recent Concepts and Strategies Developed

Genome designing involves all biotechnological interventions that results in accumulation of desirable alleles and elimination of undesirable alleles and gene combinations for realizing maximum possible genetic potential with tolerance to various stresses. In addition to MAS and genetic engineering, gene editing and nanotechnology are emerging as key concepts in genome designing of crop plants. Gene modification with these techniques are briefly discussed here.

10.11.1 Gene Editing

Gene editing (GE) techniques have revolutionized biological sciences via precise modifications in the genome of both plants and animals to yield desirable changes in the phenotypes. These technologies make structural changes to DNA of target genes or epigenetic changes to alter gene expression. The techniques used to edit or change the genomes are evolved from the earlier attempts like nuclease technologies, homing endonucleases, and certain chemical methods (Khan 2019). Molecular techniques like meganuclease (MegaN), TALENs, and ZFNs initially emerged as genome-editing technologies. Currently, there are several open engineering platforms available for construction of ZFNs and TALENs (Townsend et al. 2009). These initial technologies suffer from lower specificity due to their off-targets side effects. The latest discovery of the CRISPR/Cas9 nuclease system seems more encouraging in view of its higher efficiency and feasibility advancing the genome-engineering techniques to the level of molecular engineering. GE is broadly categorized into three generations: MegaN and ZFNs are first-generation tools, TALENs are second-generation tools, and CRISPR associated system is considered the third-generation tool. Third-generation GEs, e.g., CRISPR/Cas9, CRISPR–CRISPR from

Prevotella and *Francisella* 1 (Cpf1) were found to be powerful tools for the successful modification of genome sequence in a precise and straightforward manner (Ahmar et al. 2021). In terms of delivery to the plant cell, techniques such as ZFN technology, TALENs and CRISPR/Cas9 typically use either *Agrobacterium*-mediated or protoplast transformation.

Meganuclease, also termed molecular DNA scissors originate from microorganisms such as bacteria and yeasts and recognize relatively large DNA sequences (18 to 30 building blocks) in comparison with standard nucleases (Daboussi et al. 2015). Only very specific fragments that seldom occur in plant DNA are cut by meganucleases. Their potential to excise large pieces of DNA sequences was recognized as a genetic tool to modify DNA (Khan 2019). Double-strand breaks could be obtained at the target region in the genome under consideration through a sequence-specific nuclease. However, Meganuclease technology is costly and time-consuming (Townsend et al. 2009). The advantage with meganucleases is less toxicity in view of their natural occurrence and site specific cleavage ability. Modifying the target genes with specifically engineered meganucleases was first demonstrated in tobacco (Puchta 1999; Honig et al. 2015).

ZFNs were demonstrated as site-specific nucleases for cutting DNA at strictly defined sites, for the first time, in 1996. ZFNs are proteins composed of a zinc finger part and a nuclease part. By coupling the nuclease to a zinc finger, a protein that binds with great accuracy on a specific DNA fragment, the nuclease will only be able cut DNA at that location. Repair of ZFN-induced double-strand breaks (DSBs) with error-prone non-homologous end-joining (NHEJ) will result in introduction of insertion or deletion mutations (indels) at the site of the DSB. Alternatively, homology-directed repair of a DSB with an exogenously introduced donor template can promote introduction of alterations or insertions at or near the break site (Sander et al. 2011). ZFNs have been used to introduce specific mutations and transgene insertions that confer herbicide resistance in tobacco (Townsend et al. 2009).

TALEN effectors for DNA targeting were discovered in 2009. TALENs almost resemble ZFNs in terms of manufacturing and mode of action (Khan 2019). They are made by a similar principle where a restriction nuclease (*FokI*) is bound to a DNA-binding protein domain called TAL effector that guides the nuclease to a specific DNA sequence. Fusing a nuclease and a TAL effector allow the nuclease to cut only at one specific place in the plant DNA. The construction of TALENs is quite easier and popular compared to ZFNs; however, repetitive sequences present in TALENs can enhance the rate of homologous recombination. TALENs can target 3 nt at a time thus making it slightly more site specific with fewer off-target effects compared to ZFNs that address only 1 nt. Methods were optimized for targeted modification of plant genomes using TALENs in tobacco (Zhang et al. 2013). Though, the mitochondrial genomes of higher plants, in general, are not transformable, TALENs could effectively be used to achieve targeted modification of the mitochondrial genomes of rice and *Arabidopsis* through mitochondrial localization signals (mitoTALENs; Arimura et al. 2020). Therefore, it is theoretically possible to engineer changes into any part of a plant's genome including tobacco using TALENs.

CRISPR Cas9/sgRNA system is a novel targeted genome-editing technique derived from bacterial immune system. It is an easy, less expensive, user friendly and quickly adopted genome editing tool and enables precise genomic modifications in many different organisms and tissues (Surender et al. 2016). Cas9 protein is an RNA guided endonuclease employed for creating targeted double-stranded breaks with a short RNA sequence to facilitate the recognition of the target site in animals and plants. Hence, CRISPR/Cas9 gene targeting requires a custom single guide RNA (sgRNA) that contains a targeting sequence (crRNA sequence) and a Cas9 nuclease-recruiting sequence (tracrRNA). The crRNA region is a 20-nucleotide sequence that is homologous to a region in target gene and will direct Cas9 nuclease activity. Development of genetically edited crops similar to those developed by conventional or mutation breeding using this technique makes it a promising and extremely versatile tool for genotype improvement. Inducible CRISPR/Cas9 system was developed to avoid constitutive expression of the Cas9 protein (Ren et al. 2020). Multiplexing could be performed using CRISPR/Cas9 simply through co-infiltration of multiple tobacco rattle viruses encoding different sgRNAs in tobacco (Ali et al. 2015). Through CRISPR/Cas9-mediated homologous recombination Hirohata et al. (2019) created three nucleotide substitutions (ATG to GCT) leading to herbicide chlorsulfuron (Cs) resistance. This technology has been used for improving seed oil in tobacco (Tian et al. 2020, 2021).

Engineered nuclease systems viz. ZFNs, TALENs, and CRISPR-Cas have emerged as innovative genome editing tools with their high genetic engineering efficiency and specificity. Successful demonstration of these techniques in tobacco promises its potential utilization in targeted editing of economically important traits in future.

10.11.2 Nanotechnology

The genetic engineering technique modifies plant cell genomes, involving the efficient delivery of modifier biomolecules as genetic cargo to targeted plant (Ahmar et al. 2021). Out of number of conventional tools for bimolecular deliveries, *Agrobacterium*-mediated transformation (AMT) and biolistic delivery of DNA are most widely used ones. AMT is being employed for the transformation of target DNA to the nuclear genome of a limited number of plant species. The AMT method results in random DNA integration, disruption of endogenous plant genes and alteration in gene expression arising from the inserted sites (Niazian et al. 2017). Biolistic delivery of DNA uses a high-pressure gene gun for directly targeting the plant tissues and randomly integrate the DNA into the chromosomal region across cell walls and membranes. This yet times results in destruction of tissues and multiple insertions in random portions of the plant genome (Toda et al. 2019). Because of the unavoidable high velocity of genetic cargo during biolistic delivery, the bombarded particles damage the cell wall through penetration and disrupt the homeostasis of target cell. Other less commonly used systems include electroporation, viral vectors

and chemical delivery. Polymer-based chemical transformation leads to cytotoxicity in plant cells because of the accumulation of high-density charged polymer-based genetic cargo. A decrease in charge results in the impairment of bioconjugated complex. The disadvantages with viral vectors are high host specificity and limited cargo size. In addition to biolistic methods, PEG-mediated transformation is one of the widely used methods for inserting genetic cargo into chloroplasts (Yu et al. 2020). This method allows to carry various genetic cargo types, such as DNA and RNAs (small interfering RNA [siRNA] and miRNA into target cells (Cunningham et al. 2018). However, it requires regeneration from protoplasts, which is highly challenging because of the limited number of plant species amenable to protoplast regeneration. Thus, plant transformation presents a major bottleneck for GE technology. Conventional biomolecule delivering technologies are time-consuming and involve complicated protocols. Thus, low efficiency of gene transmission, narrow species range for application, limited cargo types, and tissue damage are the critical drawbacks of conventional delivery methods.

Nanotechnology advancements have created opportunities to overcome the limitations of the above conventional methods. Nanoparticles (NPs) are promising for the species-independent passive delivery of various genetic cargo [DNA, RNA, proteins (site-specific recombinases or nucleases), and ribonucleoprotein (RNP)] across plant systems (Cunningham et al. 2018). Use of NPs for the transfer of biological molecules into plant cells overcome all the issues that are previously hindering the success of GE and is emerging as a promising technique for improving the effectiveness, robustness and flexibility of GE. Development of genetically modified (GM) organisms (GMOs) through nanotechnology involving the use of Nanoparticles (NPs) as nanocarriers by constructing a binding complex with biomodifier molecules (eg. CRISPR/Cas system) signifies a powerful technique for transgene delivery into plant cells (Demirer et al. 2021).

The unmatched potential of the NP-based delivery of biomolecules to plant cells has revolutionized the GE delivery process (Ahmer et al. 2021). The NP-bound GE nuclease can be efficiently transferred to plant cells without resultant damage to the target tissue. Thus, the use of NP-based methods for genetic cargo delivery has emerged as a cutting-edge technology with new insights and a robust GE (Cunningham et al. 2018). Smaller size of NPs facilitate them to transverse the cell wall and overcome the obstacles to delivering biomolecules to plant tissues without the species- and tissue-specific limitations. NPs can also be engineered to facilitate cargo delivery to any subcellular parts such as mitochondrial or chloroplast DNA that AMT cannot target.

NPs commonly used for biomolecule delivery in both animal and plant systems are classified according to the base material used: bio-inspired, carbon-based, silicon-based, polymeric, and metallic/magnetic (Cunningham et al. 2018). Each of the NP types delivers different genetic cargos. For example, carbon nanotubes (CNTs) can carry RNA and DNA, but metallic NPs can only deliver DNA as genetic cargo. Silicon-based NPs can deliver DNA and proteins, and polymeric NPs (PEG and polyethyleneimine) transfer RNA, DNA, and proteins into cells (Silva et al. 2010). Cationic NPs are preferred for gene delivery into plants as they can bind to the plant

cell wall (negatively charged) and perform gene transfer, whereas CNT NPs have been used to deliver plasmid DNA into various crops. Several NPs can penetrate the cell wall (eg. CNTs and mesoporous silica), whereas other NPs such as gold NPs and magnetic NPs (MNPs) require additional physical methods (e.g., magnetoinfection and electroporation) for genetic cargo delivery into the cells.

NP-mediated cargo delivery can be through physical and non-physical means. Creation of transient pores in the cell membrane with electric fields, soundwaves, or light, magnetofection, microinjection, and biolistic particle delivery are some of the physical methods. Utilisation of cationic carriers, incubation, and infiltration are nonphysical methods. Optimizing the use of NPs in different plant species in respect of their dose and spatiotemporal tuning is essential as NPs behave differently in specific plant cells (Ahmar et al. 2021). NPs considered to enable an efficient transformation of plants because of their ability to protect the genetic cargo from cellular enzymatic degradation (Ahmar et al. 2021).

Literature on NP–tobacco systems is mainly found to focus on genetically modified tobacco mosaic virus-based metallic nanomaterial synthesis, NP as pesticides, NP uptake, effects on plant growth, biomolecule delivery systems etc. (Burklew et al. 2012; Love et al. 2014, 2015; Wang et al. 2016; Tirani et al. 2019). Delivery of DNA into *Nicotiana tabacum* plants via biolistic delivery of 100–200-nm gold-capped mesoporous silica NPs MSNs was first demonstrated in 2007 by Torney and colleagues. Fu et al. (2012) used Zinc NPs to deliver DNA plasmid into tobacco. By contrast, NPs, such as silicon carbide whiskers (SCW) and MSN, have been effectively used to transfer genes into tobacco without using other physical methods (Golestanipour et al. 2018). However, the disadvantage with SCW method compared to other NP-mediated plant transformation is that an adequate protocol is required for plant regeneration from cell cultures. Silva et al. (2010) introduce siRNA into tobacco protoplasts using polymer NPs as an alternative gene knockout mechanism. Meanwhile, NP-mediated passive delivery of DNA plasmids has been reported with tobacco through CNTs (Burlaka et al. 2015; Kwak et al. 2019) and dsDNA through clay nano sheet NPs (Mitter et al. 2017). Demirer et al. (2019) have recently achieved passive delivery of DNA plasmids and protected siRNA using functionalized CNT NPs for transient silencing of constitutively expressed gene in transgenic *N. benthamiana* leaves with 95% efficiency.

In spite of significant advantages, few challenges are hampering the successful use of NPs in GE. The first relates to nanophytotoxicity effect of NPs on plant growth, causing damage to either the plant or the environment because of the subsequent release of NPs up to a toxic level (Ahmar et al. 2021). In general, significantly less, nontoxic level in terms of both the environment and the plant, amount of engineered NPs are required as genetic cargo. However, cell structural stability, metabolic pathway disturbance, deposition and dispersal of NPs to other plant cells after application necessitate further research to improve the use of NPs as genetic cargo. Another issue requiring attention for improving NP-mediated GE's efficacy relates to efficiency in binding of NPs to biomolecules and the breakdown of the binding complex in plant cells. Different biomolecules have a different binding affinities with various NPs based on their chemical composition, structure, surface area, and charge, making

them suitable for a bioconjugation complex. However, optimization of biomolecules specific binding requires further research to increase their versatility as genetic cargo.

10.12 Brief on Genetic Engineering for Resistance Traits

Tobacco is used widely as a model in transgenic research of plant for several reasons due to its well-studied molecular genetics, nearly complete genomic mapping, readily achievable genetic transformation, its survival ability under in vitro and greenhouse conditions, and large biomass yielding potential (Jube and Borthakur 2007). Transgenic tobacco plants are also ideal model organisms for the study of basic biological functions, such as plant-pathogen interactions, environmental responses, growth regulation, senescence, etc. In view of this, number of studies have been undertaken incorporating genes from bacteria, animals and other plant species into tobacco and their functional roles validated. Genetic engineering of tobacco plants for resistance related traits are reviewed here.

10.12.1 Target Traits and Alien Genes—Biotic Stress Resistance

Tobacco transgenics incorporated with transgenes from bacterial, other plants and virus were validated to confer resistance to various pests and diseases. These transgenic studies clearly proved that tobacco cultivars having resistant to biotic stresses can be successfully developed for commercial cultivation.

Number of bacterial genes found to confer resistance to diseases viz. *P. syringae* pv. *phaseolicola* (argK-ornithine carbamoyl transferase from *P. syringae*), *P. syringae* pv. *tabaci* (bO-Bacterio-opsin from *Halobacterium halobium*), *P. parasitica* var. *nicotianae* (popA-PopA protein from *R. solanacearum*), *E. carotovora* (expI-N-oxoacyl-homoserine lactone from *E. carotovora*; aiiA-Acyl-homoserine lactonase from *Bacillus* spp.), *Helicoverpa* spp. And *Spodoptera* spp. (cry genes-Crystal proteins from *B. thuringiensis*), boll weevil larvae (choM-ChoM protein from *Actinomyces*), tobacco hornworm (ipt-Cytokinin isopentenyl transferase from *A. tumefaciens*), etc. (Jube and Borthakur 2007). Further, tobacco transgenics having bacterial genes found to tolerance herbicides viz. Bialaphos (Bar gene—producing PPT acetyl transferase from *Streptomyces hygrosopicus*), Glyphosate (aroA-M1-EPSPS from *E. coli*), Phenmedipham (pcd-PMPH from *Arthrobacter oxydans*), 2, 4-D (tfdA-2, 4-D monooxygenase from *Ralstonia eutrophus*), Paraquat (pqrA-Paraquat resistant protein (PqrA) from *Ochrobactrum anthropic*), etc. (Jube and Borthakur 2007).

Exhaustive list of tobacco transgenics validating the effectiveness of plant pathogen related genes, lectins, proteinase inhibitors, trypsin inhibitor, transcription factor etc. from various plant sources conferring resistant to *Spodoptera*, *Heliothis*,

aphids, TMV, phytopathogens and nematodes are available in the literature (Sane et al. 1997; Luo et al. 2009b; Malone et al. 2009; Priya et al. 2011; Guo et al. 2013). Plant derived antifungal proteins found to give protection from fungal pathogens. Animal derived Antimicrobial magainin analogs, avidins, gamma-aminobutyrate (GABA), proteinase inhibitor etc. found to enhance resistance of tobacco to diseases, pests and nematodes (Jach et al. 1995; Li et al. 2001; Burgess et al. 2002; Christeller et al. 2002; McLean et al. 2003).

Transgenic tobacco plants showing resistant to CMV, TMV, TLCV were developed by the transfer of transgenes from virus, plant or other origins (Prins et al. 2008). The approaches such as viral coat protein mediated resistance, replicase protein, movement, proteases, and antisense sequences 'R'-genes from plants, plantibodies, double stranded RNA (dsRNA) etc. were used in conferring virus resistance in tobacco (Powell-Abel et al. 1986; Day et al. 1991; Audy et al. 1994; Xiao et al. 2000; Hofius et al. 2001; Spassova et al. 2001; Kalantidis et al. 2002). In a large number of transgenics developed using transgenes of viral origin, resistance is found to be conferred by post-transcriptional gene silencing (Kalantidis et al. 2002).

10.12.2 Review on Achievements of Transgenics

Tobacco has served as a model plant for producing large number of transgenics having pest and disease resistant and other economically important genes. However, no genetically transformed tobacco varieties (transgenic cultivars) are released for commercial cultivation in any of the countries, in view of the opposition faced by GM tobacco in the global market (Bowman and Sisson 2000). Though GM Approval Database of International Service for the Acquisition of Agri-biotech Applications (ISAAA) reports two GM tobacco events viz. (1) oxynil herbicide tolerance and (2) nicotine reduction, antibiotic resistance (GM approval database 2021), none of them are cultivated on commercial scale in any of the countries. In contrast, millions of hectares of genetically engineered soybean, corn, cotton and canola are being grown throughout the world (ISAAA 2019). Thus, tobacco breeding efforts lag behind those of other crops in genetic engineering. In addition, the strong opposition from the European countries to genetically modified organisms (GMOs) is also acting as hindrance in transgenic tobacco breeding. Thus, genetic engineering of tobacco cultivars is on hold until the trade related obstacles are alleviated. However, this methodology holds great promise for improving tobacco cultivars in terms of disease and pest resistance, and possibly health-related constituents in the cured leaf.

10.12.3 Organelle Transformation

Manipulation of nuclear genome through genetic engineering is performed widely in most economically important plant species. However, nuclear transformation has

several drawbacks including unpredictable expression of the gene of interest and gene silencing due to the random location of transfer DNA integration and/or position effects (Meyers et al. 2010). As organelles containing genetic materials in small DNA genomes, plastids (chloroplasts) and mitochondria provide an opportunity for transformation in plants (Butow and Fox 1990; Rascon-Cruz et al 2021).

Plastid genomes of tobacco are typically 150 kb, and codes for about 140 genes. Plastids are seat for some of the important biosynthetic pathways and processes that include photosynthesis, photorespiration, metabolism of amino acids, lipids, starch, carotenoids, other isoprenoids, phenol compounds, purines, pyrimidines, isoprenoids, starch, pigments, vitamins synthesis, and also are implicated in the metabolism of phytohormones such as cytokinins, abscisic acid, and gibberellins (Kuchuk et al. 2006; Rascon-Cruz et al. 2021). Compared with conventional nuclear genetic engineering, plastid genome transformation offers several advantages (Kuchuk et al. 2006; Li et al. 2021). High level of transgene expression is possible with chloroplasts as there are about 100 chloroplasts per cell, each containing about 100 copies of genome. Thus, there is possibility of 10,000 copies of transgenes per cell due to plastid transformation. Gene silencing or position effects were not defined for plastid genes. Thus, the level of expression is much more predictable. Unlike integration into the nuclear genome, integration of heterologous DNA through homologous recombination mechanism into a plastome allows very precise genetic manipulations. There exists the possibility of multigene engineering through stacking transgenes in synthetic operons in a single transformation event. Maternal inheritance of plastomes in majority of crop species reduce the risk of uncontrolled transgene release into the environment (Kuchuk et al. 2006; Li et al. 2021).

Stable transformation of the plastome was achieved first in unicellular alga, *Chlamydomonas reinhardtii* in 1988, and two years later for the dicotyledonous seed plant tobacco (*N. tabacum*) (Svab et al. 1990). Over the years, plastid transformation in tobacco has become more and more routine with a transformation efficiency equivalent to nuclear transformation (Svab and Maliga 1993; Daniell et al. 2016; Li et al. 2021). Plastids of *N. tabacum* var. Petit Havana (Svab et al. 1990), *N. benthamiana* (Davarpanah et al. 2009) and *N. sylvestris* (Maliga and Svab 2010) are transformed by different workers. The plastid transformation technology has lead transgene expression, genome editing, and RNA editing analysis in plastids.

The first biotechnological application of transplastomic technology for pest control was the expression of cry1A(c) gene from *B. thuringiensis* (Bt) in the tobacco plastid. The tobacco transplastomic plants accumulates higher amounts of the Bt insecticidal protein (3–5% of Total Soluble Proteins-TSP) and displayed high levels of resistance to herbivorous insects (McBride et al. 1995). When cry2Aa2 was transformed as an operon along with two small open reading frames, the Cry2Aa2 protein accumulated up to 45% of TSP and led to the formation of crystals (De Cosa et al. 2001). Higher expression levels of cry9Aa2 (10% of TSP) in tobacco plastid genome results in severe growth retardation of the transplastomic plants (Chakrabarti et al. 2006) indicating that the transgene expression level need to be cautiously optimized for providing sufficient protection without a yield penalty. The advances made in other crops clearly indicate that developing tobacco with high levels of resistance to

insects, bacterial, fungal and viral diseases, and different types of herbicides is quite possible with plastid transformation (Adem et al. 2017).

The comparison between plastids and mitochondria make it possible to transform mitochondrial genome with suitably designed constructs. However, reliable methods for the transformation of mitochondria using a biolistic device currently exist only for yeasts (Johnston et al. 1988) and green algae (Remacle et al. 2006) and no successful transformation of mitochondria in plant systems has been reported to date (Li et al. 2021). Most of the plant mitochondrial genomes composed of non-coding repeated sequences, gene spacing sequences and introns. A system for genetic transformation of plant mitochondria would facilitate functional analyses of the mitochondrial genome and its products, and also open the way for modification of mitochondrial metabolism, or to introduce cytoplasmic male sterility (CMS) into new crops and varieties (European Commission 1989; Wang et al. 2020).

10.12.4 Biosynthesis and Biotransformation

10.12.4.1 Biosynthesis

Alkaloids are important compounds synthesized in *Nicotiana* plants and essential in establishing the commercial quality of tobacco as well as its defense against herbivores (Zenkner et al. 2019). Both wild and domesticated forms of *Nicotiana* spp. accumulate nicotine, the content and composition of which vary among species. Nicotine, in general, is produced in tobacco roots and translocated to leaves. Some *Nicotiana* wild species produce *N*-acyl-nornicotine, an alkaloid with more potent insecticidal properties than nicotine. The regulation of nicotine biosynthesis has been considered a complex physiological response, and many TFs are directly or indirectly involved in its regulation (Kajikawa et al. 2017; Xu et al. 2017; Qin et al. 2020). Six TFs (from three TF families) found to affect nicotine metabolism, with two basic helix-loop-helix genes positively regulating the jasmonate activation of nicotine biosynthesis (Todd et al. 2010). Metabolic engineering for the biosynthesis of nicotine and its more potent *N*-acyl-nornicotine in tobacco without compromising the commercial quality of tobacco could be an option in developing biotic stress resistant cultivar. The developed high nicotine yielding cultivars can be used to extract nicotine for use as a pesticide.

10.12.4.2 Biotransformation

Biotransformation of applied xenobiotic (pesticides, herbicides etc.) chemicals for controlling biotic stresses is essential to maintain tobacco quality and reduction of pesticide residues, and ultimately for health of the tobacco users. Tobacco transgenics

having herbicide tolerance viz. bialaphos (having *Bar* gene—producing PPT acetyltransferase from *S. hygrosopicus*), glyphosate (*aroA*-M1-EPSPS from *E. coli*), phenmedipham (*pcd*-PMPH from *A. oxydans*), 2,4-D (*tfdA*-2,4-D monooxygenase from *R. eutrophus*), paraquat (*pqrA*-paraquat resistant protein (PqrA) from *O. anthropic*), etc. found to reduce the effect of herbicides on tobacco through their biotransformation in tobacco plant (Jube and Borthakur 2007). This ability enables to use these herbicides in controlling weeds in tobacco field without affecting the mail crop. Biotransformation of pesticides to their less toxic forms in a reasonable time frame in tobacco will assist in reducing pesticide residues. Research in this direction is essential in future for adding more number to the list of crop protection agents that can be used on tobacco.

10.12.5 Metabolic Engineering Pathways and Gene Discovery

“The improvement of cellular activities by manipulation of regulatory, enzymatic, and transport functions of the cell with the use of recombinant DNA technology” is defined as Metabolic engineering (Bailey 1991). Metabolic engineering is motivated by commercial applications by which one can improve the developing strains for production of useful metabolites. It is basically meant for altering the metabolic pathways for the production of chemicals, pharmaceuticals, fuels, and medicine. A metabolic pathway can be defined as any sequence of feasible and observable biochemical-reaction steps connecting a specified set of input and output metabolites. The rate at which various input metabolites are processed to form output metabolites is known as pathway flux. Metabolic engineering involves useful alteration of metabolic pathways to better understand and utilize the cellular pathways. This involves overexpression or down regulation of certain proteins in a metabolic pathway in such a way that the cell produces a new product.

First step for successful engineering requires the complete understanding of metabolic pathway and genes involved in the path way and host cell for genetic modifications (Fuentes et al. 2018). The engineering of metabolic pathways in plants frequently requires the concerted expression of more than one gene involved in that pathway. With traditional transgenic approaches, the expression of such multiple transgenes has been a challenge. Recent progress in transformation techniques has making it possible to integrate multiple transgenes into host genomes. New technological options include combinatorial transformation (large-scale co-transformation of the nuclear genome) and transformation of the chloroplast genome with synthetic operon constructs (Bock 2013). Metabolic pathway engineering of plastid (chloroplast) genome offers significant advantages, including straight forward multigene engineering by pathway expression through operons, higher levels of transgene expression, and transgene containment due to maternal inheritance. Further, it allows

direct access to the large number of diverse metabolite pools in chloroplasts and other non-green plastid types.

In contrast to most structural genes, TFs tend to control multiple pathway steps and hence, facilitates for the engineering of complex metabolic pathways for higher levels of metabolites (Broun 2004; Grotewold 2008). The TFs are often exists as gene families and regulate target genes in tissue- and species-specific patterns (Bovy et al. 2002). In most of the cases, detailed studies have not been made on the specificity of the regulatory genes. Studies on the variations in transcriptomes and metabolomes assists in understanding the regulation by transcription factors in heterologous systems. Flavonoids are a group of compounds involved in several aspects of plant growth and development, such as pigment production, pollen growth, seed coat development, pathogen resistance and UV light protection (Harborne 1986). Hence, manipulation of phenylpropanoid pathway responsible for flavonoid production can be a strategy for biotic and abiotic stress resistance.

Metabolic engineering using three monoterpene synthases from lemon altered fragrance of tobacco plants (Lucker et al. 2004). Engineering of synthetic operon constructs comprising three genes for the key enzymes of vitamin E (tocochromanol) biosynthesis resulted in an increase of up to tenfold in total tocochromanol accumulation in transplastomic tobacco (Lu et al. 2013). Astaxanthin content in the transplastomic tobacco plants was enhanced through plastid transformation of a synthetic operon consisting of three genes that redirect lycopene into the synthesis of β -carotene and ultimately astaxanthin, a high-value ketocarotenoid (Lu et al. 2017). Grafting of transplastomic tobacco onto the non-transformable *Nicotiana glauca* enabled the horizontal transfer of the transgenic chloroplast genomes through the graft junction (Lu et al. 2017). Thus, grafting may be helpful in the transplastomic engineering of plant species that are otherwise not amenable.

Metabolic engineering of artemisinic acid biosynthetic pathway provided a proof of concept for combining plastid and nuclear transformation to optimize product yields from complex biochemical pathways in chloroplasts (Fuentes et al. 2016). Transplastomic tobacco for two synthetic operons expressing the core artemisinic acid biosynthetic pathway accumulates only low levels of the metabolite. However, super transformation of the trasplastomics lines using the COSTREL (combinatorial super transformation of transplastomic recipient lines) approach, increased the artemisinic acid content up to 77-fold.

Photorespiration could be reduced in tobacco through the introduction of three distinct alternative glycolate metabolic pathways into tobacco chloroplasts (South et al. 2019). Coupling the reduced expression of a glycolate and glycerate transporter to limit glycolate flux out of the chloroplast with alternative photorespiratory pathway could raise the biomass productivity by >40% under field conditions (South et al. 2019). In this study, about 17 constructs were designed for nuclear transformation; these multienzyme pathways could effectively be introduced into the chloroplast in the form of operons.

Possibility of metabolic engineering for pest resistance was obtained with the modulation of transcriptome and metabolome of tobacco (*Nicotiana tabacum*) by *Arabidopsis* transcription factor, AtMYB12 (Misra et al. 2010). Expression of

AtMYB12 in tobacco resulted in enhanced expression of genes involved in the phenylpropanoid pathway, leading to severalfold higher accumulation of flavonols. The tobacco transgenic lines developed resistance against the insect pests *S. litura* and *H. armigera* due to enhanced accumulation of one of the flavonol, rutin. This study clearly indicates that metabolic engineering can be successfully employed in developing stress tolerant tobacco genotypes.

10.12.6 Gene Stacking

Gene stacking or Gene pyramiding or multigene transfer refers to incorporation of two or more genes of interest into a single plant. The combined traits obtained from this process are called stacked traits. A biotech crop variety that bears stacked traits is called a biotech stack or simply stack (ISAAA 2020). A biotech stack may be a plant transformed with two or more genes that code for proteins having different modes of action on a pest or a hybrid plant expressing both insect resistance and herbicide tolerance genes derived from two genetic sources. Biotech stacks are engineered to overcoming the myriad of problems that includes insect pests, diseases, weeds, and environmental stresses in order to increase farm level productivity. Insect resistance based on multiple genes confers stable resistance than single gene which may breakdown due to co-evaluation of pests. Gene stacks can be are generated through methods such as (i) the simultaneous introduction of transgenes through co-transformation, and (ii) the sequential introduction of genes using re-transformation processes or the sexual crossing of separate transgenic events.

Pyramiding of multiple genes found to impart resistance to insects and pathogens resistance and herbicide tolerance in tobacco. Stacking of genes conferring insect-resistance and herbicide (glyphosate) tolerance (Sun et al. 2012), dsRNAs silencing of chitin biosynthesis pathway genes for root-knot nematode resistance are reported in tobacco. Stacked insect resistant cry1Ac and cry2A genes (Bakhsh et al. 2018) and three codon optimized cry2Ah1 genes (Li et al. 2018) found to confer resistance to lepidopteran insects. Pyramiding of two pathogenesis-related genes imparted resistance to three filamentous fungus in tobacco (Boccardo et al. 2019). Stacking of protease inhibitors from sweet potato and taro found to impart resistance to both insects (*H. armigera*) and pathogens (damping-off disease caused by *P. aphanidermatum* and bacterial soft rot caused by *E. carotovora*) (Senthilkumar et al. 2010). Through, stacked products are promising and technically feasible in tobacco, till date, none of the stacks produced, through genetic engineering, are approved for commercial cultivation in tobacco mainly because of their transgenic tag. Gene pyramiding events in tobacco are mainly used as proof of concepts or for gene function and interaction studies.

Regulatory principles and procedures for approval and release of biotech stacks differ globally (ISAAA 2020). In USA and Canada, no separate or additional regulatory approval is required for commercialization of hybrid stacks that are products of

crossing a number of already approved biotech lines. This policy is based on the assertion that interactions between individual trait components in a stack that have not been shown environmental or health hazard would not give rise to new or altered hazards. The combinations of “plant incorporated protectants” or PIPs (eg. Bt insecticidal proteins) may give rise to higher or altered toxicity, hence, the US Environmental Protection Agency calls for a separate safety review of a stack in case of identification of any such hazard. Contrary to this, stacks are considered new events in Japan and European Union (EU) countries, even though individual events have commercial approval, and must pass through a separate regulatory approval process, including risk assessment of their safety, similar to mono-trait biotech events (ISAAA 2020). Risk assessment is focused on the identification of additional risks that could arise from the combined genes.

10.12.7 Gene Silencing

Gene silencing is the regulation of gene expression in a cell to prevent or reducing the expression of a certain gene. Gene silencing is used as a means for developing species-specific pest control methods that are alternatives to potentially harmful chemical methods in plants. RNA interference (RNAi) is a promising method for controlling insect pests and diseases by silencing the expression of vital pest and disease-causing organism genes to interfere with development and physiology. RNAi technology is based on the expression of dsRNA that shares nearly 100% sequence homology with a desired target gene for optimal silencing. The mechanism of RNAi mediated gene silencing is based mainly on the exogenous production of short interfering RNAs/microRNAs (siRNAs/miRNAs) in an organism to control gene expression. Expression or introduction of dsRNA in eukaryotic cells can trigger sequence-specific gene silencing of transgenes, endogenes, and viruses. Transgenic plants producing dsRNAs with homology to pest or viral sequences are likely to exhibit pathogen-derived resistance to the pests and diseases. *Tobacco Rattle Virus*-based virus-induced posttranscriptional gene silencing (termed virus-induced gene silencing or VIGS) found to be the widely used method to downregulate the expression of a target plant gene (Bachan and Dinesh-Kumar 2012; Senthil-Kumar and Mysore 2014). Temporal and spatial control of gene silencing could be achieved through inducible (ethanol) expression of double-stranded RNAs in tobacco (Chen et al. 2003). *In planta*-expressed dsRNA synthesized within the plastids are more effective than nuclear expressed circumventing the native RNAi paths in eukaryotes. *M. sexta* (tobacco hornworm) genes could be effectively silenced through plastid transformation of dsRNA genes targeting the hornworm genes (Burke et al. 2019).

As a model organism number of studies conducted in tobacco for silencing genes of various pest and disease-causing organisms there-by reducing their growth and development leading to resistance. Transgenic tobacco plants expressing a hairpin RNA (hpRNA) targeting a root-knot nematode (*M. javanica*) putative zinc finger transcription factor effectively suppressed the growth of nematodes feeding on the

roots of the transgenic plants (Fairbairn et al. 2007). Vietnamese scientists succeeded in breeding tobacco for virus resistance using gene silencing or RNAi technology using three expression vectors carrying single gene (TMV, TSWV etc.) or multiple genes (TMV, TSWV, CMV and TYLCV) (<http://agrobiotech.gov.vn/NewsDetail.aspx?ID=821&CatID=7>). Expressing of dsRNA homologous to pest genes found to enhance host plant resistance to whitefly (Thakur et al. 2014; Malik et al. 2017), *H. armigera* and *S. exigua* (Zhu et al. 2012), nematodes (Mani et al. 2020) and CMV (Kalantidis et al. 2002) in tobacco.

10.12.8 Prospects of Cisgenics

In cisgenesis, as in transgenesis, extra DNA is stably built into the plant DNA. The major difference between transgenesis and cisgenesis is the origin of the DNA (Schouten and Jacobsen 2008). With cisgenesis, the extra DNA originates from a plant with which the acceptor plant (the plant that will receive the extra DNA) can cross-breed. 'Cis' refers to within the same crossable group (Schouten et al. 2006a, b). Cisgenesis approach combines both traditional breeding techniques with modern biotechnology and dramatically speed up the breeding process. Cisgenic plants are presumably considered safer than those produced through conventionally bred plants because of the lack of linkage drag (Hou et al. 2014). Cisgenesis introduces just desired genes without the undesirable genes and hence, prevents hazards compared to induced translocation or mutation breeding (Telem et al. 2013). Through cisgenesis various biotic and abiotic stress resistance genes can be pyramided to provide wider and long-lasting forms of resistance. Several backcrossed generations are required to get rid of undesired genes in conventional hybridization programs, hence takes long time (Telem et al. 2013). Whereas the time taken for introducing a single gene or more so with multiple genes can be drastically reduced in cisgenesis due to incorporation of merely target traits.

Introduction of exogenous transfer process related genes through cisgenesis can be avoided through the use of new transformation protocols without bacterial selection markers (de Vetten et al. 2003; Schaart et al. 2004) and species-specific P-DNAs instead of bacterial T-DNAs for insertion of isolated genes (de Vetten et al. 2003; Rommens et al. 2004). Further, application of new methods, such as promoter trapping and RNA fingerprinting for the isolation of native regulatory elements can now be exploited for the precise expression of the desired traits (Meissner et al. 2000; Trindade et al. 2003). Majority of the methods for production of cisgenic crops without exogenous genes including removal of selectable marker genes, use of pDNAs and segregation of independently integrated T-DNAs have been patented, therefore scientists need either to use the existing patents or design new protocols to eradicate the undesired DNA sequences from host genomes (Holme et al. 2013).

Vast prospects exists for cisgenesis in tobacco crop. Availability of large number of wild species and germplasm resources, genome sequence information of cultivated

tobacco and few wild relatives, and comparative genomic techniques, the development of efficient gene isolation techniques like map-based cloning and allele mining are opening the avenues for identification and cloning of resistant traits from tobacco and their wild relatives. The cisgenes thus isolated can be used for transferring and imparting resistance for biotic stresses in tobacco. Cisgenesis was successfully taken up in tobacco through gene editing techniques, such as ZFNs (Townsend et al. 2009), TALENs (Zhang et al. 2013) and CRISPR-Cas (Upadhyay et al. 2013; Ali et al. 2015) induce targeted alleles in tobacco (Hou et al. 2014). Identification, cloning and transferring single or multiple biotic stress resistant cisgenes into tobacco can be achieved in future with the ever-improving gene technologies.

Though both transgenesis and cisgenesis employ the same genetic modification techniques for introducing gene(s) into a plant, cisgenesis use only genes of interest from the plant itself or from a crossable species which otherwise could also be transferred by traditional breeding techniques. The release of cisgenic plants into the environment doesn't evoke any environmental risk and they are as safe as that of traditionally bred plants. Hence, compared to transgenesis, cisgenesis is more similar to traditional plant breeding and may be considered as non-transgenic, in spite of use of molecular biology and plant genetic engineering methods in their development. Restrictions on any form on cisgenesis could delay further research on development of improved crop varieties, particularly at a time when a greater number of genes from crops and their crossable wild relatives are being isolated and are amenable to cisgenesis. Hence, it is necessity to discriminate cisgenesis from transgenesis.

Common people are found to be much satisfied with cis/intragenic crop than transgenic crops. Surveys indicated cisgenic plants are more likely to be acceptable to the public than transgenic plants (Viswanath and Strauss 2010; Gaskell et al. 2011; Mielby 2011). However, currently, GMO regulations to prevent any negative effects on the environment or human health are based on transgenic organisms and do not distinguish transgenic plants from cisgenic plants in majority of the countries. Canada has adopted a product-based regulation system instead of process-based one and making it legally possible to control cisgenic plants less strictly than transgenic plants. As per the Australian Gene Technology Regulations, "a mutant organism in which the mutational event did not involve the introduction of any foreign nucleic acid" is not specified as GMO and thus treats cisgenic plants differently (Russell and Sparrow 2008). European Food Safety Authority (EFSA 2012) states that cisgenic plants are similar to the traditionally bred plants and safe in terms of environment, food and feed.

Though, biotic stress resistant trasgenics are available, they could not be used for commercial cultivation due to worldwide GMO regulations. In case cisgenics are treated differently, that will be advantageous to tobacco in which GMOs are not acceptable. Lesser regulations on cisgenics will boost the cisgenesis research in tobacco for improving tobacco yields and resistant factors.

10.13 Brief Account on Role of Bioinformatics as a Tool

Recent technological advances have resulted in the accumulation of large volumes of biological data in terms of nucleic acid sequences and various details of biomolecules produced in tobacco and other crops under different situations and life stages. In order to store and analyze these data, number of general and crop specific databases were created. The databases may have the information related to one or more of omics types in an integrated way. The information pertaining to tobacco are being stored and accessed through quite a number of databases, globally. However, to name a few, some of the key databases that are covering most of the data related to tobacco are discussed here under.

10.13.1 Gene and Genome Databases

Ever-improving sequencing technologies, gene mapping and tagging projects, and phylogenetic studies have resulted in accumulation of large volumes of genomic data in tobacco. Computer databases are an increasingly necessary tool for organizing such vast amounts of biological data generated and for making it easier for researchers to access and analyze relevant information. The Genomic databases serve as hubs for storing, sharing and comparison of data across research studies, data types, individuals and organisms.

Among the various databases, the key genome databases harboring *Nicotiana* genome and gene information are NCBI Genome, Sol Genome Networks (SGN), Kyoto Encyclopedia of Genes and Genomes (KEGG genome), EnsemblPlants, *Nicotiana attenuata* data hub (NaDH), The International Nucleotide Sequence Database Collaboration (INSDC), Gramene etc. (Table 10.13). At present, genome sequences of 12 *Nicotiana* spp. viz. *N. tabacum*, *N. tomentosiformis*, *N. sylvestris*, *N. attenuata*, *N. undulate*, *N. otophora*, *N. suaveolens*, *N. glauca*, *N. stocktonii*, *N. repanda*, *N. amplexicaulis* and *N. debneyi* at scaffold or contigs level, chloroplast genomes of five species and mitochondrial genomes of three species are available with one or the other databases. Further, a total of over more than two lakh records of gene sequences belonging to 12 *Nicotiana* species are existing at various data bases. NCBI and SGN together are the principal databases that cover all the currently available information on genomes and genes of various *Nicotiana* spp. These databases are sharing the stored information with other databases and providing extensive tools for the analysis of sequences and annotation. INSDC is a long-standing foundational initiative that operates between DNA Databank of Japan (DDBJ), European Molecular Biology Laboratory-European Bioinformatics Institute (EMBL-EBI) and NCBI and covers the spectrum of data raw reads, from assemblies and alignments to functional annotation. Other databases mentioned above provide access to the *Nicotiana* resources, mostly, through collaboration with other databases along with additional analysis tools existing with their databases. *Nicotiana attenuata* data hub (NaDH) covers

Table 10.13 Some of the main genomic resource databases for *Nicotiana* species

S. No.	Name	Salient features	URL
1	National Center for Biotechnology Information (NCBI)	<ul style="list-style-type: none"> • An integrated database for gene loci, genomes, functional genomics data, gene expression, sequence sets from phylogenetic and population studies • Offers tools for genome, gene and protein analysis and VIGS Tool to design VIGS constructs 	http://www.ncbi.nlm.nih.gov/
2	Sol Genomics Networks (SGN)	<ul style="list-style-type: none"> • An integrated database that stores information on genome, genes, loci and phenotype data for the Solanaceae family and close relatives • Provides analysis tools for browsing genomes, expression and map data viewers, community annotation system, QTL analysis and to improve VIGS constructs etc 	http://solgenomics.net/
3	<i>Nicotiana attenuata</i> Data Hub (NaDH)	<ul style="list-style-type: none"> • NaDH maintains genomic, transcriptomic and metabolomic data and combines different data sources for their analysis relevant to <i>N. attenuata</i> 	http://nadh.ice.mpg.de
4	Kyoto Encyclopedia of Genes and Genomes (KEGG genome)	<ul style="list-style-type: none"> • An integrated database resource with 16 databases that includes genes, proteins, metabolites and other chemical substances, biochemical reactions, enzyme, disease-related network variations etc. • Provide tools for understanding high-level functions and utilities of the biological system covering genomic and molecular-level information related to an ecosystem, organism and cell 	https://www.genome.jp/kegg/

(continued)

Table 10.13 (continued)

S. No.	Name	Salient features	URL
5	EnsemblPlants	<ul style="list-style-type: none"> • An integrative resource presenting genome-scale information for sequenced plant species including <i>Nicotiana</i> • Data includes genome sequence, gene models, functional annotation, and polymorphic loci etc. and tools for their analysis 	https://plants.ensembl.org
6	Gramene	<ul style="list-style-type: none"> • A curated, integrated data resource for comparative functional genomics in crops and model plant species 	https://www.gramene.org/
7	VISTA	<ul style="list-style-type: none"> • A comprehensive suite of programs and databases for comparative analysis of genomic sequences 	https://genome.lbl.gov/
8	The Gene Ontology (GO) Resource	<ul style="list-style-type: none"> • Largest source of information on the functions and cellular locations of genes along with processes they may carry out 	http://geneontology.org/
9	Gene Expression Omnibus (GEO)	<ul style="list-style-type: none"> • A functional genomics data repository and accepts Array- and sequence-based data 	https://www.ncbi.nlm.nih.gov/geo/
10	ArrayExpress (AE)	<ul style="list-style-type: none"> • Repository that archive functional genomics data from microarray and sequencing platforms 	https://www.ebi.ac.uk/arrayexpress/
11	Genomic Expression Archive (GEA)	<ul style="list-style-type: none"> • A database of functional genomics data such as gene expression, epigenetics and genotyping SNP array 	https://www.ddbj.nig.ac.jp/gea/index-e.html
12	Universal Protein Resource (UniProt)	<ul style="list-style-type: none"> • Provide a comprehensive, high-quality and freely accessible resource of protein sequence, annotation data and functional information 	https://www.uniprot.org/

(continued)

Table 10.13 (continued)

S. No.	Name	Salient features	URL
13	Pfam database	<ul style="list-style-type: none"> • Large collection of protein families, each represented by multiple sequence alignments and hidden Markov models (HMMs) • Provides tools for protein alignments and annotation, domain organisation of a protein sequence etc. 	http://pfam.xfam.org/
14	SolCyc Biochemical Pathways	<ul style="list-style-type: none"> • A collection of Pathway Genome Databases (PGDBs) for Solanaceae species generated using pathway tools 	https://solcyc.solgenomics.net/
15	REACTOME	<ul style="list-style-type: none"> • A manually curated and peer-reviewed pathway database • Offers intuitive bioinformatics tools for the visualization, interpretation and analysis of pathway knowledge to genome analysis, modeling, systems biology, etc. 	https://reactome.org/
16	The Golm Metabolome Database (GMD)	<ul style="list-style-type: none"> • A metabolome database and exchange platform to develop and improve metabolomics by multidisciplinary cooperation • Provides access to custom mass spectral libraries, metabolite profiling experiments and tools with regard to methods, spectral information or compounds 	http://gmd.mpimp-golm.mpg.de/
17	MoNA (Massbank of North America)	<ul style="list-style-type: none"> • A centralized and collaborative database of metabolite mass spectra, metadata and associated compounds 	https://mona.fiehnlab.ucdavis.edu/
18	Database for Annotation, Visualization, and Integrated Discovery (DAVID)	<ul style="list-style-type: none"> • An integrated database for functional genomic annotations with intuitive graphical summaries 	https://david.ncifcrf.gov/

(continued)

Table 10.13 (continued)

S. No.	Name	Salient features	URL
19	The International Nucleotide Sequence Database Collaboration (INSDC)	<ul style="list-style-type: none"> • INSDC covers the spectrum of data raw reads, through alignments and assemblies to functional annotation, enriched with contextual information relating to samples and experimental configurations 	https://www.insdc.org/
20	The BioStudies Database	<ul style="list-style-type: none"> • A resource for encapsulating all the data associated with a biological study • Accept and archive “multi-omics”. data generated in experiments 	https://www.ebi.ac.uk/bio-studies/

the exclusive information on *N. attenuata* and its similarities with other *Nicotiana* species and 11 published dicot species. A website of Boyce Thompson Institute’s for *N. benthamiana* resources provide access to *N. benthamiana* genomic resources existing at SGN including gene and protein data, markers, genes to phenotypes database etc. (<https://btscience.org/our-research/research-facilities/research-resources/nicotiana-benthamiana>). It is also providing tools for alignment, annotation, designing siRNAs for VIGS, CRISPR designing etc.

In addition to above databases, the Gene Ontology resource database provides access to scientific information about the molecular functions of genes (or, more properly, the protein and non-coding RNA molecules produced by genes) from many different organisms, from humans to bacteria, their cellular locations and processes those gene products may carry out (Table 10.13). Currently, 25,761 genes and gene products are found to be associated with the term *Nicotiana* in The Gene Ontology resource database.

Most of the gene and genomic databases provide tools for searching, alignment and comparison of sequences with other *Nicotiana* spp. Solanaceous crops and other crops, annotation, marker designing, QTL analysis, VIGS Tool to design virus-induced gene silencing constructs, CRISPR designing etc. Genome analysis aims to describe the functions of genes and proteins, as well as the relationship that exists between a given genotype and phenotype. Gene and gene sequence alignments helps in the identification of inherited genetic variations like SNPs and alterations in gene sequences and their relationship to resultant phenotypes. Apart from analysis of genome sequence data, various genome databases are facilitating the analysis of gene variation and expression, prediction of gene/protein structure and function, prediction and detection of gene regulation networks, etc.

10.13.2 Comparative Genome Databases

The increasing availability of genomic sequence from multiple organisms has provided biomedical scientists with a large dataset for orthologous-sequence comparisons. The rationale for using cross-species sequence comparisons is to identify biologically active regions of a genome based on the observation that sequences that perform important functions are frequently conserved between evolutionarily distant species, distinguishing them from nonfunctional surrounding sequences. This is most readily apparent for protein-encoding sequences but also holds true for the sequences involved in the regulation of gene expression. However, examination of orthologous genomic sequences from several vertebrates has shown that the inverse is also true.

Genome sequences can be compared using pairwise and multiple whole-genome alignments and based on these alignments, synteny, sequence conservation scores and constrained elements can be determined. Comparison of whole-genome sequences at the level of nucleotide or protein provides a highly detailed narration of how organisms are related to each other at the genetic level. Comparative genome studies on genomic variations will identify the types of genes, gene families, and their location, as well as provide clues on the history of evolutionary gene rearrangements and duplications that might be responsible for the recognized genetic variations. By carefully comparing genome characteristics that define various organisms, researchers can pinpoint regions of similarity and difference. This information can be used to identify putative genes and regulatory elements for various traits that may lead to their cloning and further utilization.

A variety of tools to compare the complete genome sequences of within or among the different species are made available by different databases. All most all the gene and genome databases of tobacco namely NCBI, SGN, NaDH etc. are offering the tools for such comparative genome analysis. VISTA provides a comprehensive suite of programs and databases for comparative analysis of genomic sequences. There are two ways of using VISTA—one can submit their own sequences and alignments to VISTA servers for analysis or examine pre-computed whole-genome alignments of different species.

Gramene, a knowledgebase founded on comparative functional analyses of genomic and pathway data for model plants and major crops including tobacco. The current release, #64 (September 2021), hosts 114 reference genomes, and around 3.0 million genes from 90 plant genomes with 3,256,006 input proteins in 123,064 families with orthologous and paralogous classifications. Comparative genomics contains around 340 pairwise DNA alignments and 80 synteny maps. Plant Reactome portrays pathway networks using a combination of manual biocuration and orthology-based projections to 106 species. The Reactome platform enables comparison of reference and projected pathways, gene expression analyses and overlays of gene–gene interactions. Gramene integrates information extracted from plant-focused journals on genetic, epigenetic, expression, and phenotypic diversity; ontology-based protein structure–function annotation; and gene functional annotations.

Many online/web applications can be effectively used for comparative analyses at genomic and genic levels. Apart from tools provided by various genomic databases, applications such as BRIG (Alikhan et al. 2011), Mauve (Darling et al. 2004), Artemis Comparison Tool (ACT) (Carver et al. 2005), geneCo (Jung et al. 2019) etc. can be used for comparative genomics. At *Nicotiana attenuata* data hub, genes of 11 published dicot species were compared and found to cluster into 23,340 homologous groups (HG) based on their sequence similarity with at least two homolog sequences. The phylogenetic trees were also constructed for all these HG.

Comparative analyses of *Nicotiana* plastid genomes among themselves and with currently available Solanaceae genome sequences indicated the existence of similar GC and gene content, codon usage, simple sequence and oligonucleotide repeats, RNA editing sites, and substitutions (Asaf et al. 2016; Mehmood et al. 2020). Such analysis also revealed that *N. otophora* is a sister species to *N. tomentosiformis* within the *Nicotiana* genus, and *Atropa belladonna* and *Datura stramonium* are their closest relatives (Asaf et al. 2016).

Comparison of whole nuclear and plastid genomes facilitated the identification and confirmation wild progenitor species and their relative genome contributions in the evolution of *N. tabacum* and *N. rustica* genomes (Murad et al. 2002; Lim et al. 2005; Leitch et al. 2008; Sierra et al. 2014, 2018; Edwards et al. 2017). Whole-genomic sequence comparison indicated that the genomes of *N. sylvestris* and *N. tomentosiformis* contributes 53 and 47%, respectively to the genome of *N. tabacum* confirming a larger biased reduction of T genome (Sierra et al. 2014). Comparison of chloroplast genomes revealed that *N. tomentosiformis* is a sister species and *Atropa belladonna* and *Datura stramonium* are closest relatives to *N. otophora* (Asaf et al. 2016). Further, it also revealed that the maternal parent of the tetraploid *N. rustica* was a common ancestor even for *N. paniculata* and *N. knightiana*, and the later species is more closely related to *N. rustica*.

Comparative studies of *Nicotiana* genome sequences provided insights into how speciation impacts plant metabolism, and in particular alkaloid transport and accumulation. In a genome evolution study, Sierra et al. (2018) found that 41% of allotetraploid genome of *N. rustica* originated from the paternal donor (*N. undulata*) and 59% from the maternal donor (*N. paniculata/N. knightiana*). Gene clustering analysis revealed the commonality of 14,623 ortholog groups among the *Nicotiana* species and 207 specific to *N. rustica*. It was speculated from the results the higher nicotine content of *N. rustica* leaves is due to the combination of progenitor genomes and higher active transport of nicotine to the shoot. Comparative genomics of pests and diseases is helping in distinguishing isolates, races and strains (Hou et al. 2016).

10.13.3 Gene Expression Databases

With technological advancements large volumes of data is being generated on gene expression patterns in tobacco from seed to senescence and under different growing conditions including during the incidence of biotic stresses. The expression of genes

in plants are measured at transcriptome levels. In tobacco, the efforts of the Tobacco Genome Initiative (TGI) resulted in enrichment of the sequence information for the transcriptionally active regions of the tobacco genome. The information generated in the form of Expressed Sequence Tags (ESTs), short, single pass sequence reads derived from complementary DNA (cDNA) libraries and methyl filtered Genome Space Sequence Reads (GSRs) laid foundation for gene expression analysis in tobacco. Kamalay and Goldberg (1980) measured the extent to which structural gene expression is regulated in an entire tobacco plant. Later, Matsuoka et al. (2004) analyzed the change of gene expression during the growth of tobacco BY-2 cell lines and found 9200 EST fragments corresponding to about 7000 genes. Rushton et al. (2008) reported 2513 TFs representing all of the 64 well-characterized plant TF families and used to create a database of tobacco transcription factors (TOBFAC). Edwards et al. (2010) designed Affymetrix tobacco expression microarray from a set of over 40 k unigenes and measured the gene expression in 19 different tobacco samples to produce the Tobacco Expression Atlas (TobEA). TobEA provides a snapshot of the transcriptional activity for thousands of tobacco genes in different tissues throughout the lifecycle of the plant. Expression profiling of tobacco leaf trichomes resulted in the identification of putative genes involved in resistance to biotic and abiotic stresses (Harada et al. 2010; Cui et al. 2011). Differentially expressed long noncoding RNAs (lncRNAs) found to be involved regulating jasmonates (JA) mediated plant defense against *M. sexta* attack (Li et al. 2020).

Storing and integrating different types of expression data and making these data freely available in formats appropriate for comprehensive analysis is essential for their effective utilization. Analysis of gene expression data provide hints towards understanding various aspects of plant development and resistance to biotic and abiotic stresses along with defining the molecular and genetic pathways associated with these processes. The expression databases host the transcript/ RNA/ probe information of different genes under varied native or test conditions along with the relevant software tools for analysis and retrieval of the data. EST databases constructed in the years ago for this purpose has metamorphosed to host microarrays and RNA sequences with the advent of technology. ESTs provide an insight into transcriptionally active genes in a biological sample under a given set of conditions and is relatively expensive and time consuming. Microarrays, however, offer a faster and less expensive alternative for simultaneously measuring the expression of thousands of genes that can be easily and reproducibly applied to identify genes showing specific expression patterns or responses across a broad range of conditions or treatments.

At present the Gene expression data have been archived as microarray and RNA-seq datasets in the public databases such as Gene Expression Omnibus (GEO), ArrayExpress (AE) and Genomic Expression Archive (GEA) (Table 10.13). These databases act as useful resources for the functional interpretation of genes. GEO hosts around 4,860 curated gene expression data sets as well as original series and platform records of 11 *Nicotiana* spp. (Table 10.8; <https://www.ncbi.nlm.nih.gov/gds>). ArrayExpress database have gene expression profiles of around 75 experiments including *N. attenuate*, *N. benthamiana*, *N. langsdorffii* x *N. sanderae* and *N. tabacum*. Genomic Expression Archive has 205 gene expression records related to 11

Nicotiana species. SGN maintaining two transcript libraries of *N. sylvestris* and 39 of *N. tabacum*. Further, there are exclusive expression databases for *Nicotiana attenuata* (NaDH) and *N. benthamiana* (<https://btiscience.org/our-research/research-facilities/research-resources/nicotiana-benthamiana>) along with the Sol genome networks for expression analysis among solanaceous members.

10.13.4 Protein or Metabolome Databases

Proteins are the end products of some of the expressed genes in an organism. While, a proteome is the set of proteins thought to be expressed by an organism in its life cycle. The metabolome can be defined as the complete complement of all small molecule (<1500 Da) metabolites found in a specific cell, organ or organism. The metabolome of a species is the link between its genotype and phenotype. It indicates the stage specific and organ specific response of the plants through gene expression in response to its environment. Metabolomics can influence both the gene expression and the protein function of the plant, hence, make it a central component in elucidating cellular systems and deciphering gene functions.

Proteomics and metabolomics approaches play pivotal role in functional genomics and have been essentially required for understanding plant development and stress tolerance abilities. Proteome and Metabolome profiling is an attractive tool for phenotyping plants confronted by environmental changes and biotic stresses. Such studies contribute significantly to the study of stress biology by distinguishing different compounds such as auxiliary products of stress metabolism from biosynthetically unrelated pathways, stress induced signal molecules, molecules that are part of plant acclimation process etc. The resultant metabolic compounds could be further studied by direct measurement or correlating with the changes in transcriptome and proteome expression during stress condition and can be confirmed by mutant analysis. Thus, metabolome study aid in unravelling the different pathways related to plant development and response to stresses. With the advent of high-throughput based systems, proteome and metabolome profiling was extensively carry out in the model plant like tobacco to examine stress signaling pathways, cellular and developmental processes.

Principal databases hosting tobacco proteome information are UniProt, Pfam, KEGG, SGN, NCBI, etc. and metabolome databases are SolCyc, REACTOME, The Golm Metabolome Database (GMD), MoNA (Massbank of North America), etc. The salient features of various proteome data bases are briefed below.

The proteomic data generated globally in various organisms is being stored and accessed through the Universal Protein Resource (UniProt), a comprehensive resource for protein sequence and annotation data. UniProt was established as a collaboration between the European Bioinformatics Institute (EBI), the Swiss Institute of Bioinformatics (SIB), and the Protein Information Resource (PIR) and consists of three databases that are optimized for different uses. The UniProt Knowledgebase (UniProtKB) is the central access point for providing information on extensively curated proteins including their function, classification and cross-references. The

UniProt Reference Clusters (UniRef) pools closely related sequences into a single record to speed up sequence similarity searches. The UniProt Archive (UniParc) is a comprehensive repository of all protein sequences and consisting only of unique identifiers and sequences. UniProt provides several sets of proteins assumed to be expressed by organisms whose genomes have been completely sequenced, termed “proteomes”. There are 73,606 protein entries associated with proteome *N. tabacum* and a total of 154,728 entries for all *Nicotiana* species as on 30.09.2021.

The Pfam database contains a large collection of protein families, each represented by multiple sequence alignments. This database provides tools for protein alignments and annotation, domain organization of a protein sequence etc. There are about 4970 unique results for the search term *Nicotiana* in this data base as on 30.09.2021 indicating the protein entries in the database.

KEGG database in addition to protein info, projects biological processes from various organisms onto pathways consolidated in large network schemata. The KEGG database, at present, contains the information of annotated proteins of *N. tabacum* (61,780 No.), *N. tomentosiformis* (30,989) *N. sylvestris* (33,816) and *N. attenuata* (34,218). In the Sol genomics network database also providing the data of proteins annotated based on the draft genome sequences of *N. tabacum*, *N. benthamiana* and *N. attenuata*. Protein sequences collection of around 275,000, from several sources, are available for 20 *Nicotiana* spp. at NCBI along with annotated reports of four *Nicotiana* spp. (Tables 10.8 and 10.11).

SolCyc contains a collection of Pathway Genome Databases (PGDBs) related to Solanaceae species that are generated using Pathway Tools. It is a database hub at SGN for the manual curation of metabolic networks and includes annotated metabolic, regulatory and signaling processes in Solanaceous plants based on Omics data obtained from multiple resources. It has species-specific databases for Tomato (LycCyc), Potato (PotatoCyc), Pepper (CapCyc), Coffee (CoffeaCyc), Petunia (PetCyc), *N. tabacum* (K326Cyc), *N. attenuata* (NattCyc), *N. sylvestris* (NiSylCyc), *N. tomentosiformis* (NiTomCyc), *N. benthamiana* (BenthaCyc); and multi-species databases for Combined *Nicotiana* genus (NicotianaCyc) and Combined Solanaceae database (SolanaCyc). In addition to proteomic data, NaDH is also providing metabolome data of *N. attenuata* with analysis tools and facilities for the search of metabolites and fragments based on annotation and measured values.

REACTOME offers intuitive bioinformatics tools for the visualization, interpretation and analysis of pathway knowledge to support basic and clinical research, modeling, genome analysis, systems biology and education. GMD is an open access metabolome database and provides public access to custom mass spectral libraries, metabolite profiling experiments along with additional information and tools. MoNA (Massbank of North America) is a centralized and collaborative database of metabolite mass spectra, metadata and associated compounds. At present, MoNA contains over 200,000 mass spectral records from experimental and *in-silico* libraries besides from user contributions.

The proteomic studies in tobacco revealed the interesting facts about the different stress responses (Amme et al. 2005). Analysis of the proteome of glandular trichomes revealed the enrichment of proteins belongs to components of stress defense

responses. In another study comparative proteomics of tobacco mosaic virus-infected *N. tabacum* plants identified major host proteins involved in photosystems and plant defense (Das et al. 2019). Metabolome study under water stress in tobacco identified a useful marker for drought stress for members of Solanaceae (Rabara et al. 2017).

10.13.5 Integration of Different Data

Analysis of a single omics data (e.g. genome, proteome, transcriptome and metabolome) provides biological understanding at a specific molecular layer. However, many agronomic and quality traits embrace complex crosstalk between various molecular layers viz., genome, transcriptome, proteome, and metabolome. These four 'omes' collectively generate the building blocks of systems biology. An integrative analysis of multiple layers of molecular data or system biology helps to discover and elucidate underlying molecular mechanisms of complex traits and thus, provide clues for genome designing. In the holistic study of the complex biological processes, it is imperative to have an integrative approach that combines multi-omics data to highlight the interrelationships of the involved biomolecules and their functions. Advent of high-throughput techniques and accessibility of multi-omics data from a large set of samples, a number of promising methods and tools have been developed for data integration and interpretation. Most of the biological databases collect and integrate data from different sources.

Databases namely INSDC, NCBI, SGN, NaDH, KEGG genome, EnsemblPlants, DAVID (Database for Annotation, Visualization, and Integrated Discovery), The BioStudies Database etc. are some of the integrated databases and resources that are collecting and integrating the omics data from different plants and tobacco (Table 10.13). As a collaborative foundational initiative, INSDC covers the spectrum of data raw reads, through alignments and assemblies to functional annotation, enriched with contextual information relating to samples and experimental configurations. NCBI in addition to providing wide-ranging tobacco basic data in its various databases, offers tools for integration of structural and functional genomic data and their annotations. The curated proteome data of *Nicotiana* species in Uniport, and metabolome data from different resources are being integrated in new datahubs like SGN, KEGG, NaDH etc., to provide holistic information from gene to pathway for the researchers. KEGG is an integrated database resource consisting of sixteen databases that includes genes and proteins, metabolites and other chemical substances, biochemical reactions, enzyme, disease-related network variations etc.

EnsemblPlants is an integrative resource presenting genome-scale information for a growing number of sequenced plant species (currently 33). It provides data on genome sequence, gene models, polymorphic loci, and functional annotation. Also provide additional information on variation data comprising phenotype data, population structure, individual genotypes, and linkage. In each of its release, comparative analyses are performed on whole genome and protein sequences, and genome alignments and gene trees are made available showing the implied evolutionary

history of each gene family. DAVID is a web-accessible database that integrates functional genomic annotations with intuitive graphical summaries. According to shared categorical data for Gene Ontology, protein domain, and biochemical pathway membership, lists of gene or protein identifiers are rapidly annotated and summarized. Numerous public sources of protein and gene annotation have been parsed and integrated into DAVID database. This database currently contains information on over 1.5 million genes covering more than 65,000 species. European Molecular Biology Laboratory-European Bioinformatics Institute (EMBL-EBI) is building and maintaining the BioStudies Database as a resource for encapsulating all the data associated with a biological study. One of the goals of BioStudies is to accept and archive data generated in experiments that can be characterized as "multi-omics".

These efforts of building archives, databases, and analysis tools in an integrated approach have been successful at enabling a better understanding as well as comparison of the omic resources of tobacco.

10.14 Brief Account on Social, Political and Regulatory Issues

10.14.1 Concerns and Compliances

Tobacco is one of the important high value commercial crops and is valued for its potential to generate higher farm income and employment to farmers and farm labors, and revenue to national governments. Today tobacco sector is reeling under the whirlpool of diametrically conflicting concerns relating to the livelihood security of those associated with tobacco production, processing and marketing on one hand and the serious health risks to consumers on the other. Use of huge quantities of forest wood as source of energy for flue-curing of FCV tobacco, environmental pollution caused by tobacco smoking and associated spitting habits of chewing tobacco are issues of serious concern to tobacco production. Further, climate change impacts, emerging biotic and abiotic stresses, pesticide residues, consumer preferences and tobacco regulatory policies are becoming increasingly complex and represent future challenges for tobacco cultivation (ICAR 2015).

WHO-Framework Convention on Tobacco Control (FCTC), with overwhelming membership of 182 countries, envisages non-price, price and tax measures to reduce the supply and demand for tobacco in the world. The FCTC signatory countries are under obligation to support the measures for reduction of supply and demand for tobacco. On May 31st every year, the world observes World No Tobacco Day (WNTD) promoted by the World Health Organization (WHO) with a primary focus on encouraging users to refrain from tobacco consumption and its related products for a period of at least 24 h.

10.14.2 Patent and IPR Issues

Researchers are developing various management strategies for minimizing crop yield losses at field level due to biotic stresses. This includes development of biotic stress resistant varieties through conventional breeding and biotechnological interventions. The advances in biotechnology and bioinformatics generated various genome-based tools, techniques, genes and gene constructs in the field of stress resistance (Dangl and Jones 2001). Intellectual property rights (IPR) for plants help protect investments made in research and development of new tobacco varieties (CORESTA 2005). In turn, this encourages further investment and helps continue the development of new varieties that increases economic returns throughout the tobacco supply chain. Over the past 25 years, an increasing number of governments and international organizations have enacted laws, regulations, or policies that acknowledge the need for IPR. While these provide protection to inventors or breeders, some of them may also recognize a farmer's exemption for saving seed for their own use.

Patents offer the owner the exclusive right to make, use, sell, offer for sale, or import for those purposes a patented product. It also offers the rights to a patented process and to make, use, sell, offer for sale, or import for those purposes the direct product of the patented process. The ability to patent plant varieties is recognized in some countries, but in many countries like India it is disallowed. There have been few patents granted for tobacco varieties in the world (CORESTA 2005). Court decisions in the United States have permitted the use of utility patents to protect plant varieties, and to date the United States Patent and Trademark Office has issued over 200 utility patents with claims to seeds. European countries that are members of the European Patent Office (EPO) ban the patenting of plant varieties, but recent determinations support patent claims directed to plants of more than one variety. Thus, a utility patent and the ruling by the EPO may suggest a type of broad protection for a novel plant trait that is not recognized under plant variety protection (PVP) legislation. Rights granted under patents are also distinguishable from PVP in that exceptions are not allowed for using a patented variety for experiments, breeding purposes, or for private or non-commercial purposes. On the other hand, a patent provides explicit information on the processes and methods used to develop the variety, and it is often tied to a public notification of seed availability once the patent expires. The term of a patent varies across the world, but in the US a utility patent is valid for 20 years after the application date.

Recent breakthroughs in crop science, especially in the area of molecular biology, and plant genomics have opened unlimited number of opportunities in the way the tobacco plants can be utilized for commercial and medicinal uses. Genetic improvements in agronomic traits of crops such as yield attributes and, pest and disease resistance through molecular and plant genomic approaches has been accompanied by the great abundance of new patents issued in these fields. There is a scope for patenting of novel methods of making biotic stress resistant tobacco genotypes, methods of introgressing nucleic acid molecules associated with biotic stresses, genes conferring resistance to various tobacco pests, viruses, bacteria, fungi and nematodes (Hefferon

2010). Patenting activity in resistance genes in tobacco is initiated in 1992 and there is considerable progress in patenting from the year 2000 and it was more prominent from 2010 onwards (Prabhakararao et al. 2016). Majority of these patent documents (around 60%) are in the jurisdiction of USA and China.

The intellectual information generated in the frontier areas is available in non-patent and patent literature. Non-patent literature (NPL) consists of peer reviewed scientific paper, publications such as conference proceedings, databases (DNA structures, gene sequences, chemical compounds, etc.) and other literature (translation guides, statistical manuals, etc.). Nearly 80% of all the technical information available in the world is hidden in the patent documents and other IP assets (Prabhakararao et al. 2016). Patent mapping helps in retrieving and exploring the information protected in the intellectual property documents. Collections of patent documents are available in a number of patent information databases (<https://guides.library.queensu.ca/patents/databases>). Most patent offices provide free access to patent documents via public databases and some of the largest and most popular patent databases are given at Table 10.14. Patent information can be used to decide the patentability of an invention, avoid re-invention and infringement, provide the current state of the art in a given field of technology, find the latest trends in R and D being pursued by the peers and competitors etc.

Table 10.14 Patent databases

I	Patent office databases	
S. No	Databases	Producing patent office
1	Canadian Patents Database	Canadian Intellectual Property Office (CIPO)
2	DEPATISnet	German Patent and Trademark Office (DPMA)
3	Espacenet	European Patent Office (EPO)
4	JP-PlatPat	Japan Patent Office (JPO)
5	PatentScope	World Intellectual Property Organization (WIPO)
6	US Patent Databases	United States Patent and Trademark Office
II	Databases that index patent literature	
S. No	Database	Literature covered
1	FreePatentsOnline (FPO)	Patents from the European Patent Office, Germany and the US, Japanese patent abstracts and published Patent cooperation Treaty (PCT) applications
2	Google Patents	Full-text patent documents from Canada, China, EPO, Germany, PCT and the US
3	Patent Lens	Around 100 million patent documents from 90+ national and regional patent offices, and full-text EP, US and WO documents

10.14.3 Disclosure of Sources of GRs, Access and Benefit Sharing

Genetic Resources (GRs) are a key source of numerous biotechnology innovations (Steward 2018). Historical studies reveals that less than 1% of species have provided the necessary basic resources for the progress of all civilization so far. Therefore, unexplored GRs expected to have certain potential value in further advancement of civilization. The main characteristic of bioprospecting is uncertainty as it is seldom possible to forecast which genes, species or ecosystems will turn out to be valuable in the future. With an aim to improve the sustainable use of GRs to protect biodiversity, and support benefit sharing with originating countries Access-Benefit Sharing (ABS) systems/regulations have been developed in the recent decades. ‘Convention on Biological Diversity’ (CBD) of 1992 serves as starting point in many countries for biodiversity conservation and use. As a supplementary agreement to the CBD, The Nagoya protocol of 2010 is aimed at improving the fair and equitable sharing of benefits arising out of the utilization of genetic resources. ABS systems vary widely from country to country. GR-rich countries tend to organize their ABS systems more strictly and focus on securing an equitable share of the benefits from the products developed through the use of GRs. In order to enhance the ABS compliance, over the years, several governments have introduced disclosure requirements (DRs) in their patent systems. However, the study conducted by Steward (2018) in Brazil and India indicated that DRs might enhance the R&D costs with increased uncertainty and delay in using GRs. The uncertainty may relate to unclear definitions of GRs (Ex.: Brazil, India) and to the fact that the content of the Disclosure Requirements is not verified by the IP authorities in these countries. This may provides ample scope for challenging patents on ABS conditions even after approval. The extent of the DR-effects on R&D cycles depends largely on local market conditions and (efficiency of) ABS legislation (Steward 2018).

10.14.4 Farmers’ Rights

Since the dawn of agriculture, farmers around the world have been the custodians and innovators of agricultural biodiversity (FAO 2017; Craig 2004). Farmers are responsible for collecting the best seeds and cultivating different types and species of tobacco throughout the world. Through the careful selection of their best seeds and propagating material, and exchange with other farmers, it became possible to develop and diversify crop varieties. Over thousands of years of constant management and innovation by farmers, a small number of initial crops and varieties have progressed into an incredible treasure of plant diversity for food and farming. In certain countries like India, households traditionally raising different tobacco landraces in their kitchen gardens since generations from the seeds collected from their own crops, thus maintaining and protecting biodiversity.

Farmers' access to seed and propagating material and opportunities for exchanging the planting material are strongly influenced by seed regulations (variety release and seed marketing regulations), legislation linked to intellectual property rights (patents and plant breeders' rights), and regulations concerning the bio-prospecting of genetic resources. Farmers' Rights are the rights arising from the contributions of farmers in conserving, improving, and making available plant genetic resources, mainly those in the centers of origin/diversity, in the past, present and future.

The notion of Farmers' Rights was developed during the early 1980s to counter increased demands for Plant Breeders' Rights (PBR) being voiced in international negotiations. The aim was to draw attention to the unremunerated innovations of farmers, which were seen as the foundation of all modern plant breeding. The concept first emerged in international negotiations within FAO in 1986. Already in 1987, practical solutions were being proposed, serving as the foundation for all further negotiations on Farmers' Rights, and providing substantial input to the framing of current understanding of the issue. In 1989, Farmers' Rights gained formal recognition by the FAO Conference. In 1991, the Conference decided to set up a fund for the realization of these rights, but this has never materialized. In May 1992, the Convention on Biological Diversity (CBD) was adopted with a resolution on the interrelationship between the CBD and the promotion of sustainable agriculture. Through this resolution, FAO was advised to initiate negotiations for a legally binding international regime on the management of plant genetic resources, and in this context, to resolve the question of Farmers' Rights. Agenda 21, a dynamic program approved at the UN Conference on Environment and Development held in Rio de Janeiro in 1991, had voiced similar demands. In November 1996, Global Plan of Action for the Conservation and Sustainable Utilization of Plant Genetic Resources for Food and Agriculture (hereafter referred to as the Global Plan of Action) was endorsed by the FAO Council, by the Conference of the Parties to the CBD, and by the World Food Summit at FAO acknowledges the need to realize Farmers' Rights. The Second Global Plan of Action, prepared under the aegis of the Commission on Genetic Resources for Food and Agriculture, was adopted by the FAO Council on 29 November 2011 contains a set of recommendations and activities intended as a framework, guide and catalyst for action at community, national, regional and international levels. The International Treaty on Plant Genetic Resources for Food and Agriculture (hereafter referred as the International Treaty) adopted in 2001 addressed the issue of Farmers' Rights in its Article 9, and in its Preamble. This Treaty recommends the Contracting Parties to protect and promote Farmers' Rights in agreement with their national laws. In Article 9, the Contracting Parties of the International Treaty recognize the enormous contribution that farmers of all regions of the world have made, and will continue to make, for the conservation and development of plant genetic resources as the basis of food and agricultural production throughout the world. The responsibility for implementing Farmers' Rights lies with national governments, and that they can choose the measures to do so as per their needs and priorities. Measures covering the protection of traditional knowledge, benefit-sharing and participation in decision-making are suggested. Also, the farmers rights to save, use, exchange and sell farm-saved seeds and planting material are addressed, but without giving any specific direction

for execution. In addition, Farmers' Rights are addressed in the preamble, and other articles in the Treaty clearly support these rights, albeit not explicitly (for example, the provisions on conservation and sustainable use and on benefit sharing). There are no legally binding provisions in the International Treaty on how to implement Farmers' Rights at national level.

10.14.5 Traditional Knowledge

Indigenous peoples and local communities have to cope with extreme weather and adapt to environmental change for centuries in order to survive as often they live in harsh natural environments (Swiderska et al. 2011). They use their traditional knowledge (TK) generated using long standing traditions and practices in relation to adaptive ecosystem management and sustainable use of natural resources. Hence, the diversity of traditional varieties sustained by farmers around the world is increasingly valuable for adaptation to climate changes and to cope with emerging new pests and diseases, particularly as modern agriculture relies on a very limited number of crops and varieties. Landraces or traditional varieties are genetically more diverse than recent varieties and considered to be good sources of resistance to biotic stresses (<https://www.cbd.int/traditional/what.shtml>). Local communities use wild foods to supplement their diets and thus conserve wild species which are valuable sources of stress resistant genes. The tradition farmers are well placed to identify resilient crop species and resistant varieties for biotic stresses with the available accumulated TK, and in view of their long experience in cultivating crops under changing climates. Traditional Knowledge about resilient properties, such as drought and pest resistance traits and biotic stress resistant varieties and wild crop relatives (Jarvis et al. 2008) can be a valuable information in developing biotic stress tolerant varieties in tobacco also.

10.14.6 Treaties and Conventions

International agreements like Convention on Biological Diversity (CBD), International Treaty for Plant Genetic Resources for Food and Agriculture (ITPGRFA), The International Union for the Protection of New Varieties of Plants (UPOV) are of special significance in the context of agricultural sector in general and biotechnology particular. They are briefly discussed below.

10.14.6.1 Convention on Biological Diversity (CBD)

The Convention on Biological Diversity (CBD), known informally as the Biodiversity Convention, is a legally binding international treaty and was ratified in 1992 at the Rio

earth summit (<https://www.cbd.int/convention/>). The convention currently having 196 Parties (168 Signatures). The three main goals of the Convention are conservation of biological diversity (or biodiversity), sustainable use of its components and fair and equitable sharing of benefits arising from genetic resources. In other words, its objective is to develop national strategies for the conservation and sustainable use of biological diversity. It is considered as the key document with regards to sustainable development.

The convention, for the first time, recognized in the international law that the conservation of biological diversity is “a common concern of humankind” and is an integral part of the development process. All the ecosystems, species, and genetic resources are covered in the agreement. It associates traditional conservation efforts to the economic goal of sustainably using biological resources. It has put in place the principles for the fair and equitable sharing of the benefits arising from the use of genetic resources, notably those destined for commercial use.

The Contracting Parties shall promote international technical and scientific cooperation in the field of conservation and sustainable use of biological diversity, wherever necessary, through appropriate international and national institutions. In pursuance of the objectives of this Convention and in accordance with national legislation and policies, the Contracting Parties shall, encourage and develop methods of cooperation for the development and use of both indigenous and traditional technologies. For achieving this purpose, the Contracting Parties shall also promote cooperation in the training of personnel and exchange of experts.

In respect of handling of biotechnology and distribution of its benefits, Article 19 of the Convention states—

- Each Contracting Party shall take legislative, administrative or policy measures, as appropriate, to provide for the effective participation in biotechnological research activities by those Contracting Parties, especially developing countries, which provide the genetic resources for such research, and where feasible in such Contracting Parties.
- Each Contracting Party shall take all practicable measures to promote and advance priority access on a fair and equitable basis by Contracting Parties, especially developing countries, to the results and benefits arising from biotechnologies based upon genetic resources provided by those Contracting Parties.
- The Parties shall consider the need for and modalities of a protocol setting out appropriate procedures, including, in particular, advance informed agreement, in the field of the safe transfer, handling and use of any living modified organism resulting from biotechnology that may have adverse effect on the conservation and sustainable use of biological diversity.
- Each Contracting Party shall, directly or by requiring any natural or legal person under its jurisdiction providing the organisms referred to in paragraph 3 above, provide any available information about the use and safety regulations required by that Contracting Party in handling such organisms, as well as any available information on the potential adverse impact of the specific organisms concerned to the Contracting Party into which those organisms are to be introduced.

CBD has three Protocols viz. The Nagoya Protocol on Access and Benefit-sharing, The Cartagena Protocol on Biosafety and The Nagoya – Kuala Lumpur Supplementary Protocol on Liability and Redress to the Cartagena Protocol on Biosafety under CBD. The essence of these protocols is given below.

10.14.6.2 Cartagena Protocol on Biosafety

The Cartagena Protocol on Biosafety to the Convention on Biological Diversity is an international treaty governing the movements of living modified organisms (LMOs) resulting from modern biotechnology from one country to another (<http://bch.cbd.int/protocol>). It was adopted as a supplementary agreement to the Convention on Biological Diversity on 29 January 2000 and came in force on 11 September 2003. Currently, it has 173 Parties (103 Signatures). The objective of the Protocol is to contribute to ensuring an adequate level of protection in the field of the safe transfer, handling and use of ‘living modified organisms resulting from modern biotechnology’ that may have adverse effects on the conservation and sustainable use of biological diversity, taking also into account risks to human health, and specifically focusing on trans-boundary movements. The Protocol provides for Parties to enter into bilateral, regional and multilateral agreements and arrangements regarding intentional trans-boundary movements of living modified organisms.

The Protocol establishes a Bio-Safety Clearing -House to: (a) Facilitate the exchange of scientific, technical, environmental and legal information on, and experience with, living modified organisms; and, (b) Assist Parties to implement the Protocol, taking into account the special needs of developing country Parties, in particular the least developed and small island developing States among them, and countries with economies in transition as well as countries that are centres of origin and centres of genetic diversity.

10.14.6.3 Nagoya Protocol

The Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilization (ABS) to the Convention on Biological Diversity is a supplementary international agreement (<https://www.cbd.int/abs/>). The Protocol was adopted on 29 October 2010 in Nagoya, Japan and entered into force on 12 October 2014. Currently, it has 131 Parties (132 ratifications) (92 signatories). The protocol aims at fair and equitable sharing of benefits arising from the utilization of genetic resources, there-by contributing to the conservation and sustainable use of biodiversity. The protocol offers a transparent legal framework for the effective implementation of one of the three objectives of the CBD: the fair and equitable sharing of benefits arising out of the utilization of genetic resources. The Nagoya Protocol applies to genetic resources that are covered by the CBD, and to the benefits arising from their utilization. It also covers traditional knowledge (TK) associated with genetic resources that are covered by the CBD and the benefits arising from its utilization.

The Nagoya–Kuala Lumpur Supplementary Protocol on Liability and Redress to the Cartagena Protocol on Biosafety

This Supplementary Protocol entered into force on 5 March 2018 and currently, it has 49 Parties. Adopted as a supplementary agreement to the Cartagena Protocol on Biosafety, it aims to contribute to the conservation and sustainable use of biodiversity by providing international rules and procedures in the field of liability and redress relating to living modified organisms (<http://bch.cbd.int/protocol/supplementary/>). The Supplementary Protocol requires that response measures are taken in the event of damage resulting from living modified organisms which find their origin in a transboundary movement, or where there is sufficient likelihood that damage will result if timely response measures are not taken. This Protocol gives a definition of ‘damage’, referring to an adverse effect on the conservation and sustainable use of biological diversity that is measurable or otherwise observable and significant, taking also into account risks to human health. The Protocol requires that a causal link between the damage and the living modified organism be established. Along with imposing a requirement for response measures, the Protocol obliges Parties to continue to apply existing legislation on civil liability or to develop specific legislation concerning liability and redress for material or personal damage associated with damage to the conservation and sustainable use of biological diversity.

10.14.6.4 The International Treaty on Plant Genetic Resources for Food and Agriculture

This International Treaty was adopted in 2001 (FAO 2009). The objectives of the International Treaty on Plant Genetic Resources for Food and Agriculture are the conservation and sustainable use of all plant genetic resources for food and agriculture and the fair and equitable sharing of the benefits arising out of their use, in harmony with the Convention on Biological Diversity, for sustainable agriculture and food security. This legally-binding Treaty is in harmony with the Convention on Biological Diversity and covers all plant genetic resources relevant for food and agriculture. The Treaty is vital for ensuring the continuous availability of the plant genetic resources for feeding people from different countries. The Treaty recognizes the enormous contribution that the local and indigenous communities and farmers of all regions of the world and takes measures to protect and promote Farmers’ Rights. The Contracting Parties agree to establish a multilateral system, which is efficient, effective, and transparent, both to facilitate access to plant genetic resources for food and agriculture, and to share, in a fair and equitable way, the benefits arising from the utilization of these resources, on a complementary and mutually reinforcing basis. The treaty takes care of (a) protection of traditional knowledge relevant to plant genetic resources for food and agriculture; (b) the right to equitably participate in sharing benefits arising from the utilization of plant genetic resources for food and agriculture; and (c) the right to participate in making decisions, at the national level, on matters related to the conservation and sustainable use of plant genetic resources for food and agriculture.

10.14.6.5 The UPOV Convention

For the protection of plant variety, The UPOV Convention was adopted in Paris in 1961 and entered into force in 1968. It was first revised slightly in 1972 and more considerably in 1978 and 1991. The 1978 Act entered into force in 1981, and the 1991 Act in 1998 (www.upov.int). The Convention established an inter-governmental organization called the Union Internationale pour la Protection des Obtentions Végétales (UPOV). Internationally, the most prevalent means of protecting plant varieties is through Plant Variety Protection (PVP), which is enabled in the 77 countries that are members of the International Union for the Protection of New Varieties of Plants which is headquartered in Geneva, Switzerland. The stated mission of UPOV's is "To provide and promote an effective system for plant variety protection, with the aim of encouraging the development of new varieties of plants for the benefit of society". The current act of the convention was adopted in 1991 and recognizes breeder's rights to a variety if the variety is: (1) new; (2) distinct; (3) uniform; and (4) stable. The breeder's rights in the 1991 Act require authorization of the breeder to perform the following: (1) production or reproduction (multiplication) (2) conditioning for the purpose of propagation, (3) offering for sale, (4) selling or marketing, (5) exporting, (6) importing, (7) stocking for any purpose mentioned in (1)–(6) above. Breeder's rights to a variety remain in effect for a period of 20 years from the date on which the rights were granted.

For the first time, the Act of 1991 included protection against "essentially derived" varieties, which are derived from the protected variety, that is not clearly distinguishable from the protected variety, or which requires repeated use of the protected variety for production purposes. Examples of the manner in which an essentially derived variety could be developed from a protected variety include: (1) the selection of a natural or induced mutant, or of a somaclonal variant, (2) the selection of a variant individual from plants of the initial variety, (3) backcrossing, (4) transformation by genetic engineering. Exceptions to breeder's rights were granted for: (1) Acts done privately and for non-commercial purposes, (2) Acts done for experimental purposes (3) Acts done for the purpose of breeding other varieties, except for the generation of essentially derived varieties.

10.14.7 Participatory Breeding

Conservation of diversity and selection of naturally occurring high yielding and stress resistant variants by cultivators was the principal method of tobacco improvement, for thousands of years prior to 1800s, during and after the domestication of *Nicotiana* species. The systematic varietal improvement started by scientists in the later period in established research organizations in different countries led to release of number of biotic stress resistant tobacco varieties using conventional plant breeding techniques. This has resulted in erosion of genetic diversity in the farmers' fields due to large

scale cultivation of few improved varieties. The resultant genetic vulnerability to diseases and pests necessitates the development of stress resistant cultivars.

In view of the limitations to formal breeding and the threats to farmers' seed systems, participatory plant breeding (PPB) emerged as a means to bring farmers back into the breeding process as active participants (Greenberg 2018). The role played by farmers in agricultural biodiversity conservation and use is taken as an advantage while making them partners in breeding tobacco varieties. In the development of improved biotic stress resistant varieties, PPB ensures the improvement of adapted local genetic materials using the diversity available either with them or public gene banks to suit the farmer needs having resistance to races and biotypes prevailing under their conditions. This also empowers the farmer in terms of technical and organizational skills in preserving and evolving materials under their control, on-farm management, and local creativity/innovation. PPB involves the active participation of farmers in few or all of the set of sequenced breeding program activities viz. priority setting, genetic materials acquisition and selection, crossing, selection at early stages and advanced stages, in situ experimentation/testing, and production and sharing of genetic materials and knowledge. However, PPB may be a substitute for station-based research or scientist-managed on-farm trials; rather than a complementary breeding process (Hardon et al. 2005; Aguilar-Espinoza, 2007; Ceccarelli et al. 2009).

10.15 Future Perspectives

10.15.1 *Potential for Expansion of Productivity*

The genetic potential for increasing the crop yield through conventional approaches is still obtainable in tobacco (Sarala et al. 2016). Combination of traditional breeding techniques with genome designing strategies can further accelerate the genotype improvements in tobacco for increasing and stabilizing yields. Further advances in genomic research in terms of genomic, transcriptomic, proteomic and metabolomics and their integrated analysis would assist in designing of appropriate genome assisted breeding strategies for genome designing of tobacco for attaining maximum potential yields with good quality and pests and disease resistance in a short span of time.

10.15.2 *Potential for Expansion into Nontraditional Areas*

The genus *Nicotiana* with well-defined group of species of which tobacco (*N. tabacum* L.) is an important agricultural crop plant that plays a significant role in the economies of many countries (FAO 2019). *Nicotiana* species are also used in the elucidation of various principles related to disease resistance, synthesis of secondary metabolites and basic aspects of plant physiology. In view of its higher level of

biomass accumulation, tobacco is considered to be a promising crop for the production of commercially important substances (e.g., medical drugs and vaccines) through molecular farming and cultivation of tobacco for its valuable native phyto-chemicals viz. nicotine, solanesol, proteins and organic acids (Sarala 2019).

Tobacco is considered to be one of the most important model systems in plant biotechnology till date and going to continue further. In view of its easy transformation ability, tobacco plant serving as an experimental system various pilot studies on the expression of novel transgenes that are later being used in important food crops (Krishnamurthy et al. 2008), for the study of polyploidy, and for investigation of secondary product biosynthesis (Wang and Bennetzen 2015). Scientific research in *Nicotiana* going to be accelerated in a wide range of areas with the availability of entire genome sequences and transcriptomic profiles. As *Nicotiana* species are found to be one of the best plants for transient transgene expression in leaf via simple and *A. tumefaciens* infiltration, better knowledge of the *Nicotiana* genome will provide necessary raw material for studying the function of any transformed gene.

Both wild and domesticated forms of tobacco accumulate a wide variety of alkaloids and other phytochemicals, the content and composition of which vary among species. The work done so far has brought out the tremendous scope for exploiting the crop for extraction of many valuable native phytochemicals viz. nicotine, solanesol, organic acids, proteins (green leaf) and seed oil having pharmaceutical and industrial importance (Chakraborty et al. 1982; Chida et al. 2005; Narasimha Rao and Prabhu 2005; Patel et al. 1998). Solanesol is a major component of tobacco (from traces to 4.7%) and is used as an intermediate in the production of valuable pharmaceuticals viz. co enzyme Q9, co enzyme Q10, Vitamin K2, Vitamin E and N-solanesyl-N, N'-bis (3,4-dimethoxybenzyl) ethylenediamine (SDB) which has the potential for use in the treatment of migraines, osteoporosis, neurodegenerative diseases, hypertension, cardiovascular diseases, anti-aging, to improve brain health and as dietary supplement for type 2 diabetics (Sarala 2019). While, Nicotine is the principal alkaloid synthesized in roots and accumulated in the leaf (0.1–5%). Nicotine can be used as a pesticide for controlling many agriculturally important pests and in pharma industry as smoking cessation products. Fraction-1 protein is the most abundant protein in tobacco and constitutes about 50% of soluble protein and 25% of total protein. This can be used for manufacturing food supplements. Tobacco leaf contains malic acid (4.0–4.5%) and citric acid (0.5–2.0%). These acids can be extracted from leaf and can be used for solubilisation of Rock phosphate and as foods and beverages. Tobacco seed contains 32–42% oil which can be used as edible oil and for paint and soap industry. Thus, with the potential to generate considerable volumes of biomass per unit area, tobacco genotypes can be bred for higher amounts of native phytochemicals viz. nicotine, solanesol, organic acids, proteins (green leaf) and seed oil for pharmaceutical and industrial purposes.

Tobacco has an established history as a routine system for molecular farming and often is chosen as a production platform due to its easy genetic modification. Other advantages of tobacco are established technology for gene transfer and expression, higher biomass yield, potential for rapid scale-up in view of prolific seed production, and availability of required infrastructure for processing. Although many tobacco

cultivars produce high levels of toxic alkaloids, there are low-alkaloid ones that can be utilized for the production of pharmaceutical proteins (Ma et al. 2003).

Tobacco is the first and probably only plant, where plastid transformation is successfully established as a routine (Svab and Maliga 1993; Daniell et al. 2002). High level of transgene expression is possible with chloroplasts and hence, there is a remarkable potential for large scale production of active proteins through molecular farming. TMV genome can be rapidly manipulated and can be used as a vector as it has the ability to rapidly infect the tobacco plant. The transgenic tobacco plants express the target protein transiently. Molecular farming in tobacco hairy roots is another option where it triggers the secretion of a pharmaceutical antibody. Generally recombinant pharmaceutical proteins expressed in hairy root cultures can be secreted into the medium to improve product homogeneity and to facilitate purification.

A transgenic tobacco was the first plant used to be used for recombinant human antibody production (in 1989); this was quickly followed by production of human serum albumin (1990). Furthermore, it is not used as food or feed and therefore it is easier to manufacture active compounds in tobacco without fear of these compounds mixing with food or feed. Important pharmaceutical proteins viz. Human Growth hormone, Human serum albumin, Erythropoietin, Human-secreted alkaline phosphatase, Collagen, Protein C, Granulocyte–macrophage colony-stimulating factor, Epidermal growth factor, Hepatitis B virus envelope protein, Escherichia coli heat-labile enterotoxin, Diabetes autoantigen Disease, Cholera toxin B subunit, Immunoglobulin G1, Immunoglobulin M and Secretory immunoglobulin A etc. have been successfully found to produce in Tobacco (Thomas et al. 2002; Ma et al. 2003). In a review, Ashraful et al. (2014) reported that the bio-diesel from tobacco seed oil could be successfully used to run diesel engine that gave excellent performance and most effective regulated emissions. Tobacco genotypes with high biomass and seed yield can be identified and tailored to biosynthesize and accumulate of plant derived storage lipids such as triacylglycerol that can act as a sustainable, carbon–neutral alternative biofuels (Carlsson et al. 2011). Considering various potential uses of tobacco in farming of important biomolecules, tailoring tobacco for molecular farming going to be an important objective for tobacco improvement programs.

References

- Abdul Wajid SM, Sheno MM, Moses JSL (1987) Screening *Nicotiana* germplasm for resistance to Black shank disease. *Tob Res* 13:111–115
- Abdurakhmonov IY, Abdukarimov A (2008) Application of association mapping to understanding the genetic diversity of plant germplasm resources. *Int J Plant Genomics*. <https://doi.org/10.1155/2008/574927>
- Adem M, Beyene D, Feyissa T (2017) Recent achievements obtained by chloroplast transformation. *Plant Methods* 13(30). <https://doi.org/10.1186/s13007-017-0179-1>
- Aguilar-Espinoza E (2007) Evaluating a participatory plant breeding approach in a local seed system of Northern Nicaragua: challenges for upscaling technologies and institutions. Master's thesis, Norwegian University of Life Sciences

- Ahmar S, Mahmood T, Fiaz S et al (2021) Advantage of nanotechnology-based genome editing system and its application in crop improvement. *Front Plant Sci* 12:663849. <https://doi.org/10.3389/fpls.2021.663849>
- Ali Z, Abul-Faraj A, Piatek M et al (2015) Activity and specificity of TRV-mediated gene editing in plants. *Plant Signal Behav* 10(10):e1044191. <https://doi.org/10.1080/15592324.2015.1044191>
- Alikhan NF, Petty NK, Zakour NLB et al (2011) BLAST ring image generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* 12:402
- Aly R (2007) Conventional and biotechnological approaches for control of parasitic weeds. *In Vitro Cell Develop Biol Plant* 43(4):304–317. <https://doi.org/10.1007/s11627-007-9054-5>
- Amme S, Rutten T, Melzer M et al (2005) A proteome approach defines protective functions of tobacco leaf trichomes. *Proteomics* 5(10):2508–2518. <https://doi.org/10.1002/pmic.200401274>
- Anderson E, De Winton D (1931) The genetic analysis of an unusual relationship between self-sterility and self-fertility in *Nicotiana*. *Ann Mo Bot Gard* 18(1):97–116. <https://doi.org/10.2307/2394046>
- Apple JL, Chaplin JF, Mann TJ (1963) Comparison of the *Nicotiana glutinosa* resistance factor and milk in controlling mosaic of flue-cured tobacco. *Tob Sci* 7:46–53
- Arimura S, Ayabe H, Sugaya H et al (2020) Targeted gene disruption of ATP synthases 6–1 and 6–2 in the mitochondrial genome of *Arabidopsis thaliana* by mitoTALENs. *Plant J* 104(6):1459–1471. <https://doi.org/10.1111/tjp.15041>
- Arslan B, Okumus A (2006) Genetic and geographic polymorphism of cultivated tobaccos (*Nicotiana tabacum*) in Turkey. *Genetika* 42(6):818–823. <https://doi.org/10.1134/S1022795406060123>
- Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. *Plant Mol Biol Rep* 9:208–218
- Asaf S, Khan AL, Khan AR et al (2016) Complete chloroplast genome of *Nicotiana otophora* and its comparison with related species. *Front Plant Sci* 7:843. <https://doi.org/10.3389/fpls.2016.00843>
- Ashraf AM, Masjuki HH, Kalam IMA et al (2014) Production and comparison of fuel properties, engine performance, and emission characteristics of biodiesel from various non-edible vegetable oils: a review. *Energy Conversion Manage* 80:202–228
- Audy P, Palukaitis P, Slack SA et al (1994) Replicase mediated resistance to *Potato Virus Y* in transgenic tobacco plants. *Mol Plant Microbe Interact* 7:15–22
- Bachan S, Dinesh-Kumar SP (2012) *Tobacco Rattle Virus (TRV)*-based virus-induced gene silencing. *Methods Mol Biol* 894:83–92. https://doi.org/10.1007/978-1-61779-882-5_6
- Baghyalakshmi K, Sarala K, Prabhakara Rao K et al (2018) Genetic diversity in tobacco genotypes evaluated under All India Network Project. *Multilogic Sci* 8(issue special):30–34
- Bai D, Reeleder R, Brandie JE (1995) Identification of two RAPD markers tightly linked with the *Nicotiana debneyi* gene for resistance to black root rot of tobacco. *Theor Appl Genet* 91(8):1184–1189. <https://doi.org/10.1007/bf00220927>
- Bailey JE (1991) Toward a science of metabolic engineering. *Science* 21:252(5013):1668–75. <https://doi.org/10.1126/science.2047876>. PMID: 2047876
- Bairwa A, Patel HR (2016) Estimation of loss and analysis of nicotine, reducing sugar and chloride in bidi tobacco due to reniform nematode under pot conditions. *J Parasit Dis* 40(2):240–247. <https://doi.org/10.1007/s12639-014-0483-x>
- Bakhsh A, Dinç T, Hussain T (2018) Development of transgenic tobacco lines with pyramided insect resistant genes. *Turk J Biol* 42(2):174–186. <https://doi.org/10.3906/biy-1710-71>
- Barchi L, Bonnet J, Boudet C et al (2007) A high-resolution, intraspecific linkage map of pepper (*Capsicum annuum* L.) and selection of reduced recombinant inbred line subsets for fast mapping. *Genome* 50(1):51–60
- Barrett BA, Kidwell KK (1998) AFLP-based genetic diversity assessment among wheat cultivars from the Pacific Northwest. *Crop Sci* 38:1261–1271
- Basimia A, Maleki HH, Darvishzadeh R et al (2014) Mixed linear model association mapping for low chloride accumulation rate in oriental-type tobacco (*Nicotiana tabacum* L.) germplasm. *J Plant Interact* 9:666–672

- Beavis WD (1998) QTL analysis: Power, precision, and accuracy. In: Paterson A (ed) Molecular dissection of complex traits. CRC Press, Boca Raton, Florida
- Bebber DP, Ramotowski MAT, Gurr SJ (2013) Crop pests and pathogens move polewards in a warming world. *Nat Clim Chang* 3:985–988
- Berbec A, Doroszewska T (2020) The use of *Nicotiana* species in tobacco improvement. In: Ivanov NV et al (eds) The tobacco plant genome, compendium of plant genomes. Springer Nature Switzerland AG. https://doi.org/10.1007/978-3-030-29493-9_8
- Binbin H, Geng R, Cheng L et al (2020) Genetic diversity and fingerprinting of 33 standard flue-cured tobacco varieties for use in distinctness, uniformity and stability testing. *BMC Plant Biol* 20:378
- 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
- Blancard D, Delon R, Blair BW et al (1999) Virus diseases. In: Davis DL, Nielsen MT (eds) Tobacco production, chemistry and technology. Blackwell Science Ltd., London, pp 198–215
- Blume NB, Bergervoet JHW, Maiss E (2017) Development of molecular assay for the detection of *Cucumber Mosaic Virus* and the discrimination of its subgroups I and II. *J Virol Methods* 243:35–43
- Boccardo NA, Segretin ME, Hernandez I et al (2019) Expression of pathogenesis-related proteins in transplastomic tobacco plants confers resistance to filamentous pathogens under field trials. *Sci Rep* 9:2791. <https://doi.org/10.1038/s41598-019-39568-6>
- Bock R (2013) Strategies for metabolic pathway engineering with multiple transgenes. *Plant Mol Biol* 83(1–2):21–31. <https://doi.org/10.1007/s11103-013-0045-0>
- Bovy A, de Vos R, Kemper M et al (2002) High-flavonol tomatoes resulting from the heterologous expression of the maize transcription factor genes LC and C1. *Plant Cell* 14:2509–2526
- Bowman DT, Sisson V (2000) A historical overview of flue-cured tobacco breeding in the U.S.A. *Tob Sci* 44:59–64. <https://doi.org/10.3381/0082-4623-44.1.59>
- Bradbury PJ, Zhang Z, Kroon DE et al (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23(19):2633–2635
- Braun AC (1955) A study on the mode of action of the wildfire toxin. *Phytopathology* 45:659–664
- Bretting PK, Widrechner MP (1995) Genetic markers and plant genetic resource management. *Plant Breed Rev* 13:11–86
- Brieger FG (1935) Genetic analysis of the cross between self-fertile *N. Langsdorffii* and the self-sterile *N. Sanderae*. *J Genet* 30(1):79–100
- Broun P (2004) Transcription factors as tools for metabolic engineering in plants. *Curr Opin Plant Biol* 2:202–209
- Brown JKM (2002) Yield penalties of disease resistance in crops. *Curr Opin in Plant Biol* 5:339–344
- Burbridge TN (1960) The Australian species of *Nicotiana* L. (Solanaceae). *Aust J Bot* 8:342–380
- Burgess EPJ, Malone LA, Christeller JT (2002) Avidin expressed in transgenic tobacco leaves confers resistance to two noctuid pests, *Helicoverpa armigera* and *Spodoptera litura*. *Transgenic Res* 11:185–198. <https://doi.org/10.1023/A:1015297302990>
- Burke WG, Kaplanoglu E, Kolotilin I et al (2019) RNA interference in the tobacco hornworm, *Manduca sexta*, using plastid-encoded long double-stranded RNA. *Front Plant Sci* 10:313. <https://doi.org/10.3389/fpls.2019.00313>
- Burklew CE, Ashlock J, Winfrey WB et al (2012) Effects of aluminum oxide nanoparticles on the growth, development and microRNA expression of Tobacco (*Nicotiana tabacum*). *PLoS ONE* 7(5):e34783. <https://doi.org/10.1371/journal.pone.0034783>
- Burlaka OM, Pirko YV, Yemets AI et al (2015) Plant genetic transformation using carbon nanotubes for DNA delivery. *Cytol Genet* 49:3–12. <https://doi.org/10.3103/S009545271506002X>
- Buschmann C, Lichtenthaler HK (1998) Principles and characteristics of multi-colour fluorescence imaging of plants. *J Plant Physiol* 152:297–314. [https://doi.org/10.1016/s0176-1617\(98\)80144-2](https://doi.org/10.1016/s0176-1617(98)80144-2)

- Butow RA, Fox TD (1990) Organelle transformation: Shoot first, ask questions later. *Trends Biochem Sci* 15(12):465–468. [https://doi.org/10.1016/0968-0004\(90\)90299-Q](https://doi.org/10.1016/0968-0004(90)90299-Q)
- CABI (2021a) *Orobanche cernua* (nodding broomrape). In: *Invasive species compendium*. CAB International, Wallingford. www.cabi.org/isc. Accessed 28 Aug 2021
- CABI (2021b) *Globodera tabacum* (Tobacco cyst nematode) In: *Invasive species compendium*. CAB International, Wallingford. www.cabi.org/isc. Accessed 28 Aug 2021
- Cai CC, Chai LG, Wang Y et al (2009) Construction of genetic linkage map of burley tobacco (*Nicotiana tabacum* L.) and genetic dissection of partial traits. *Acta Agron Sin* 35:1646–1654
- Cai C, Yang Y, Cheng L et al (2015) Development and assessment of EST-SSR marker for the genetic diversity among tobaccos (*Nicotiana tabacum* L.). *Russian J Genet* 51:591–600
- Carlsson AS, Yilmaz JL, Green AG et al (2011) Replacing fossil oil with fresh oil-with what and for what? *Eur J Lipid Sci Technol* 113:812–831
- Carver TJ, Rutherford KM, Berriman M et al (2005) ACT: the artemis comparison tool. *Bioinformatics* 21(16):3422–3423. <https://doi.org/10.1093/bioinformatics/bti553>
- Ceccarelli S, Guimaraes E, Weltzien E (eds) (2009) *Plant breeding and farmer participation*. FAO, Rome
- Chakrabarti SK, Lutz KA, Lertwiriawong B et al (2006) Expression of the cry9Aa2 B.t. gene in tobacco chloroplasts confers resistance to potato tuber moth. *Transgenic Res* 15:481–488
- Chakraborty MK, Patel BK, Tewari MN et al (1982) Alternative use of tobacco raw material for agro-based industry. *Ind Tob J* 14:9–16
- Chaplin JF (1962) Transfer of black shank resistance from *Nicotiana pulmaginifolia* to flue cured *N. tabacum*. *Tob Sci* 6:184–189
- Chaplin JF (1966) Comparison of tobacco black shank resistance from four sources. *Tob Sci* 10:55–58
- 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, Matzinger DF, Mann TJ (1966) Influence of homozygous and heterozygous mosaic-resistance factor on quantitative characters of flue-cured tobacco. *Tob Sci* 10:81–84
- Chari MS, Patel NM, Patel AR (1983) Losses due to insect pests on tobacco. In: *All India Workshop on Crop losses due to insect pests*, Hyderabad
- Chari MS, Rao RSN, Sreedhar U et al (1996) Bio-intensive IPM of *Spodoptera litura* F. in field crop of tobacco in Andhra Pradesh. In: Goel SC (ed) *IPM-an Entomological approach to sustainable agriculture*. UPZS, Muzaffarnagar, India, pp 51–58
- Chari MS, Nagarajan K, Ramaprasad G (1999) Pest and disease management in tobacco. In: Upadhyay RK, Mukerji KG, Dubey OP (eds) *IPM systems in agriculture*, vol 6. Cash Crops, Aditya Books Private Ltd, New Delhi
- 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 S, Hofius D, Sonnewald U (2003) Temporal and spatial control of gene silencing in transgenic plants by inducible expression of double-stranded RNA. *Plant J* 36:731–740
- Cheng L, Chen X, Jiang C et al (2019) High-density SNP genetic linkage map construction and quantitative trait locus mapping for resistance to *Cucumber Mosaic Virus* in tobacco (*Nicotiana tabacum* L.). *Crop J* 7(4):539–547. <https://doi.org/10.1016/j.cj.2018.11.010>
- Chida AS, Vani PVSN, Singh AK (2005) Tobacco phytochemicals. In: Deo Singh, Jain KC (eds) *Alternative uses of tobacco*. Central Tobacco Research Institute (ICAR), Rajahmundry
- Chitwood BG (1949) Root-Knot nematodes-part I. A revision of the genus *Meloidogyne* Goeldi, 1887. *Proc Helminthol Soc Wash* 16:90–104
- Christeller JT, Burgess EPJ, Mett V et al (2002) The expression of a mammalian proteinase inhibitor, bovine spleen trypsin inhibitor in tobacco and its effects on *Helicoverpa armigera* larvae. *Transgenic Res* 11:161–173
- Chuanyin D, Zhizhong H, Xuliang Y (2009) AFLP analysis of genetic diversity of tobacco germplasms. In: Paper presented at 63rd tobacco science research conference, Florida, 27–30 Sept 2009

- 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
- Clarkson JJ, Lim KY, Kovarik A et al (2005) Long-term genome diploidization in allopolyploid *Nicotiana* section *Repandae* (*Solanaceae*). *New Phytol* 168:241–252
- Clausen RE, Goodspeed TH (1925) Interspecific Hybridization in *Nicotiana*. II. A Tetraploid GLUTINOSA-TABACUM hybrid, an experimental verification of winge's hypothesis. *Genetics* 10(3):278–284
- Clausen RE, Cameron DR (1944) Inheritance in *Nicotiana tabacum*. Xviii. Monosomic analysis. *Genetics* 29(5):447–477
- Clausen RE, Goodspeed TH (1926) Inheritance in *Nicotiana tabacum*. VII. the monosomic character, fluted. *Univ Calif Pub Bot* 11:61–82
- Clayton EE (1947) A wildfire resistant tobacco. *J Hered* 38:35–40
- Clayton EE (1958) The genetics and breeding progress in tobacco during the last 50 years. *Agron J* 50:352–356
- Clayton EE (1967) The transfer of blue mold resistance to tobacco from *Nicotiana debneyi*. Part III. Development of a blue mold resistant cigar wrapper variety. *Tob Sci* 11:107–110
- Clayton EE (1968) The transfer of blue mold resistance tobacco from *Nicotiana debneyi*. Part 4. 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, Graham TW, Todd FA et al (1958) Resistance to the root knot nematode disease of tobacco. *Tob Sci* 2:53–63
- Clayton EE, Heggestad HE, Grosso JJ et al (1967) The transfer of blue mold resistance to tobacco from *Nicotiana debneyi*. *Tob Sci* 11:91–97
- Collard BC, Mackill DJ (2008) Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Phil Trans Roy Soc Lond B Biol Sci* 363:557–572
- Collard BCY, Jahufer MZZ, Brouwer JB et al (2005) An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. *Euphytica* 142:169–196
- CORESTA (2005) Tobacco seed integrity task force final report, May 2005, pp 1–14. https://www.coresta.org/sites/default/files/technical_documents/main/Seed-Integrity-TF-May05.pdf. Accessed 13 Jul 2021
- CORESTA (2020) Guide no 1: the concept and implementation of CPA guidance residue levels. CORESTA, France
- Craig B (2004) Farmers' rights: intellectual property regimes and the struggle over seeds. *Polit Soc* 32(4):511–543. <https://doi.org/10.1177/0032329204269979>
- Crews JW, Wills WH, La Prade JL (1964) Black shank disease reactions of six flue cured tobacco varieties and the F₁ hybrids among them. *Tob Sci* 8:128–132
- Cui H, Zhang ST, Yang HJ et al (2011) Gene expression profile analysis of tobacco leaf trichomes. *BMC Plant Biol* 11:76. <https://doi.org/10.1186/1471-2229-11-76>
- Cunningham FJ, Goh NS, Demirer GS et al (2018) Nanoparticle-mediated delivery towards advancing plant genetic engineering. *Trends Biotechnol* 36:882–897. <https://doi.org/10.1016/j.tibtech.2018.03.009>
- D'hoop BB, Paulo MJ, Kowitwanich K, et al (2010) Population structure and linkage disequilibrium unravelled in tetraploid potato. *Theor Appl Genet* 121:1151–1170
- Daboussi FTJ, Stoddard, Zhang F (2015) Engineering meganuclease for precise plant genome modification. In: Zhang et al (ed) *Advances in new technology for targeted modification of plant genomes*. https://doi.org/10.1007/978-1-4939-2556-8_5
- Dam SK, Roy S, Dutta S (2010) Biological control of brown spot caused by *Alternaria alternata* in Motihari tobacco (*Nicotiana rustica*) in West Bengal, India. *J Mycol and Plant Pathol* 40:356–359
- Danesh D, Aarons S, McGill GE et al (1994) Genetic dissection of oligogenetic resistance to bacterial wilt in tomato. *Mol Plant-Microbe Interact* 7:464–471
- Dangl JL, Jones JDG (2001) Plant pathogens and integrated defense responses to infection. *Nature* 411:826–833

- Daniell H, Khan MS, Allison L (2002) Milestones in chloroplast genetic engineering: an environmentally friendly era in biotechnology. *Trends Plant Sci* 7:84–91. [https://doi.org/10.1016/S1360-1385\(01\)02193-8](https://doi.org/10.1016/S1360-1385(01)02193-8)
- Daniell H, Lin CS, Yu M et al (2016) Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biol* 17:134. <https://doi.org/10.1186/s13059-016-1004-2>
- Darling AC, Mau B, Blattner FR et al (2004) Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* 14(7):1394–1403. <https://doi.org/10.1101/gr.2289704>
- Darlington CD, Janaki Ammal EK (1945) *Chromosome atlas of cultivated plants*. George Allen and Unwin Ltd., London
- Darvishzadeh R (2016) Genetic variability, structure analysis, and association mapping of resistance to resistance to broomrape (*Orobanche aegyptiaca* Pers.) in Tobacco. *J Agric Sci Technol* 18:1407–1418
- Darvishzadeh R, Alavi R, Sarrafi A (2010) Resistance to powdery mildew (*Erysiphe cichoracearum* DC.) in oriental and semi-oriental tobacco germplasm under field condition. *J Crop Improv* 24(2):122–130
- Darvishzadeh R, Alavi SR, Sarrafi A (2011) Genetic variability for chlorine concentration in oriental tobacco genotypes. *Arch Agron Soil Sci* 57(2):167–177
- Das PP, Lin Q, Wong SM (2019) Comparative proteomics of *Tobacco Mosaic Virus*-infected *Nicotiana tabacum* plants identified major host proteins involved in photosystems and plant defence. *J Proteomics* 194:191–199. <https://doi.org/10.1016/j.jprot.2018.11.018>
- Davalieva K, Maleva I, Filiposki K et al (2010) Genetic variability of Macedonian tobacco varieties determined by microsatellite marker analysis. *Diversity* 2(4):439–449. <https://doi.org/10.3390/d2040439>
- Davarpanah SJ, Jung SH, Kim YJ et al (2009) Stable plastid transformation in *Nicotiana benthamiana* J. *Plant Biol* 52(3):244–250. <https://doi.org/10.1007/s12374-009-9023-0>
- Day AG, Bejarano ER, Buck KW et al (1991) Expression of an antisense viral gene in transgenic tobacco confers resistance to the DNA virus tomato golden mosaic virus. *Proc Natl Acad Sci USA* 1:88(15):6721–6725. <https://doi.org/10.1073/pnas.88.15.6721>
- De Cosa B, Moar W, Lee SB et al (2001) Hyper-expression of Bt Cry2Aa2 operon in chloroplasts leads to formation of insecticidal crystals. *Nat Biotechnol* 19:71–74
- de Vetten N, Wolters AM, Raemakers K et al (2003) A transformation method for obtaining marker free plants of a cross-pollinating and vegetatively propagated crop. *Nat Biotechnol* 21:439–442
- del Campo ML, Renwick JAA (1999) Dependence on host constituents controlling food acceptance by *Manduca sexta* larvae. *Entomol Exp Appl* 93:209–215
- Del Piano L, Abet M, Sorrentino C et al (2000) Genetic variability in *Nicotiana tabacum* and *Nicotiana* species as revealed by RAPD markers: 1. Development of the RAPD procedure. *Beitr Tab Int/Contrib Tob Res* 19:1–15
- Demirer GS, Zhang H, Matos JL et al (2019) High aspect ratio nanomaterials enable delivery of functional genetic material without DNA integration in mature plants. *Nat Nanotechnol* 14:456–464. <https://doi.org/10.1038/s41565-019-0382-5>
- Demirer GS, Silva TN, Jackson CT et al (2021) Nanotechnology to advance CRISPR-Cas genetic engineering of plants. *Nat Nanotechnol* 12:1–8
- Denduangboripant J, Piteekan T, Nantharat M (2010) Genetic polymorphism between tobacco cultivar-groups revealed by amplified fragment length polymorphism analysis. *J Agric Sci* 2(2):41–48
- Drüge KL, Song Z, Wang B (2018) Characterization of *Nicotiana tabacum* genotypes possessing deletion mutations that affect *Potyvirus* resistance and the production of trichome exudates. *BMC Genomics* 19(1):484. <https://doi.org/10.1186/s12864-018-4839-y>
- Doganlar S, Frary A, Daunay MC et al (2002) A comparative genetic linkage map of eggplant (*Solanum melongena*) and its implications for genome evolution in the Solanaceae. *Genetics* 161:1697–1711

- Dong QY, Tong ZJ, Yu HQ et al (2020) Construction and characterization of a BAC library for flue-cured tobacco line with high resistance to blank shank. *Acta Agron Sin* 46(6):869–877. <https://doi.org/10.3724/SP.J.1006.2020.94110>
- Doody A (2020) Pests and diseases and climate change: is there a connection? <https://www.cimmyt.org/news/pests-and-diseases-and-climate-change-is-there-a-connection/>. Accessed 30 Sept 2021
- Doroszewska T, Berbec A, Czarnecka D et al (2013) Diseases and pests of tobacco. Institute of Soil Science and Plant Cultivation, Pulawy, Poland
- Eagles HA, Bariana HS, Ogonnaya FC et al (2001) Implementation of markers in Australian wheat breeding. *Crop Pasture Sci* 52(12):1349–1356
- East EM (1932) Studies of self-sterility. IX. The behavior of crosses between self-sterile and self-fertile plants. *Genetics* 17:175–202
- Edwards KD, Bombarely A, Story GW et al (2010) TobEA: an atlas of tobacco gene expression from seed to senescence. *BMC Genomics* 11:142
- 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 Genomics* 18:448. <https://doi.org/10.1186/s12864-017-3791-6>
- EFSA (2012) Scientific opinion addressing the safety assessment of plants developed through cisgenesis and intragenesis. *EFSA J* 10(2):2561
- El-Morsy SHI, Dorra MDM, Elham AAAE (2009) Comparative studies on diploid and tetraploid levels of *Nicotiana glauca*. *Acad J Plant Sci* 2(3):182–188
- European Commission (1989) Genetic transformation of plant mitochondria : development of a general strategy. https://cordis.europa.eu/project/id/BAP*0102. Accessed 21 Sept 2021
- Evanno G, Reganant E, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* 14:2611–2620
- Fairbairn D, Cavallaro A, Bernard M et al (2007) Host-delivered RNAi: an effective strategy to silence genes in plant parasitic nematodes. *Planta* 226:1525–1533
- Fan W, Sun X, Yang A (2015) Exploring high-potassium favorable allele mutation of tobacco based on genome-wide association analysis. *Acta Tab Sin* 25:11–19
- FAO (2009) International treaty on plant genetic resources for food and agriculture. Food and Agriculture Organization of the United Nations, Rome
- FAO (2017) The international treaty on plant genetic resources for food and agriculture: farmers' rights. Food and Agriculture Organization of the United Nations, Rome. <http://www.fao.org/3/I7820EN/i7820en.pdf>
- FAO (2019) FAOSTAT statistical database. Food and Agriculture Organization of the United Nations, Rome. <http://www.fao.org/faostat/en/#data/QCL>
- Fernandez-Pozo N, Menda N, Edwards JD et al (2015) The Sol Genomics Network (SGN)—from genotype to phenotype to breeding. *Nucleic Acids Res* 43:1036–1041
- Flint-Garcia SA, Thornsberry JM, Buckler ES (2003) Structure of linkage disequilibrium in plants. *Annu Rev Plant Biol* 54:357–374. <https://doi.org/10.1146/annurev.arplant.54.031902.134907>
- Fortnum BA, Martin SB (1998) Disease management strategies for control of bacterial wilt of tobacco in the South eastern USA. In: Prior P, Allen C, Elphinstone J (eds) *Bacterial wilt disease*. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-03592-4_60
- 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. doi:<https://doi.org/10.1186/1471-2156-13-18>
- Friebe B, Jiang J, Raupp WJ et al (1996) Characterization of wheat-alien translocations conferring resistance to diseases and pests: current status. *Euphytica* 91:59–87
- Fu YQ, Li LH, Wang PW et al (2012) Delivering DNA into plant cell by gene carriers of ZnS nanoparticles. *Chem Res Chin Univ* 28:672–676
- Fuentes P, Armarego-Marriott T, Bock R (2018) Plastid transformation and its application in metabolic engineering. *Curr Opin Biotechnol* 49:10–15. <https://doi.org/10.1016/j.copbio.2017.07.004>

- Fuentes P, Zhou F, Erban A 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:e13664. <https://doi.org/10.7554/eLife.13664>
- Fujimura T, Sato S, Tajima T et al (2015) Powdery mildew resistance in the Japanese domestic tobacco cultivar Kokubu is associated with aberrant splicing of MLO orthologs. *Plant Pathol* 65(8):1358–1365. <https://doi.org/10.1111/ppa.12498>
- Gajos Z (1988) Polalta—odmiana tytoniu odporna na wirus brązowej plamistości pomidora (TSWV) i czarną zgniliznę korzeni (*Thielaviopsis basicola* Ferr.). [Polalta—a tobacco variety resistant to *Tomato Spotted Wilt Virus* (TSWV) and black root rot (*Thielaviopsis basicola* Ferr.)]. *Biul CLPT* 1–4:7–25
- Ganal MW, Altmann T, Roder MS (2009) SNP identification in crop plants. *Curr Opin Plant Biol* 12:211–217
- Ganesh CT, Saiprasad GVS, Mohan Raju B et al (2014) Genetic structure in FCV tobacco population as assessed by multi-locus genotyping using SSR markers. *Adv Crop Sci Tech* 2:3. <https://doi.org/10.4172/2329-8863.1000127>
- Gaskell G, Allansdottir A, Allum N, Revuelta G, Stares S, Torgersen H, Wager W et al (2011) The 2010 Eurobarometer on the life sciences. *Nat Biotechnol* 29:113–114
- Gevezova M, Dekalska T, Stoyanov K et al (2012) Recent advance in broomrapes research. *J Biosci Biotechnol* 1(2):91–105
- Gholizadeh S, Darvishzadeh R, Mandoulakani BA et al (2012) Molecular characterization and similarity relationships among flue-cured tobacco (*Nicotiana tabacum* L.) genotypes using simple sequence repeat markers. *Not Bot Horti Agrobot* 40(1): 247–253
- Ghosh S, Malhotra P, Lalitha PV et al (2002) Novel genetic mapping tools in plants: SNPs and LD-based approaches. *Plant Sci* 162(3):329–333
- 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 Agri Animal Husb* 17(87):652–658
- GM approval database (2021) International service for the acquisition of agri-biotech applications (ISAAA), New York. GM Crops List- GM Approval DatabaselISAAA.org. Accessed 21Sept 2021
- Gnanamanickam SS, Samuel S, Kandaswamy TK (1977) Tobacco wildfire disease in India. *Curr Sci* 46(13):450–451
- Goggin FL, Lorence A, Topp CN (2015) Applying high-throughput phenotyping to plant–insect interactions: picturing more resistant crops. *Curr Opin Insect Sci* 9:69–76. <https://doi.org/10.1016/j.cois.2015.03.002>
- Golestanipour A, Nikkha M, Aalami A et al (2018) Gene delivery to tobacco root cells with single-walled carbon nanotubes and cell-penetrating fusogenic peptides. *Mol Biotechnol* 60:863–878. <https://doi.org/10.1007/s12033-018-0120-5>
- Gong D, Huang L, Xu X et al (2016) Construction of a high-density SNP genetic map in flue-cured tobacco based on SLAF-seq. *Mol Breed* 36:100
- Gong D, Chen M, Sun Y et al (2020) Fine mapping of QTLs for resistance to *Phytophthora nicotianae* in flue-cured tobacco using a high-density genetic map. *Mol Breed* 40:45. <https://doi.org/10.1007/s11032-020-01126-8>
- Gooding JGV (1969). Epidemiology of *Tobacco Mosaic Virus* on tobacco in North Carolina. *North Carolina Agric Exp Stn Tech Bull* 195:24
- Goodspeed TH (1954) The genus *Nicotiana*: origins, relationships and evolution of its species in the light of their distribution, morphology and cytogenetics. *Chronica Botanica*, Waltham, Mass, USA
- Goodspeed TH, Thompson MC (1959) Cytotaxonomy of *Nicotiana* II. *Bot Rev* 25:385–415
- Gopalachari NC (1984) Tobacco. Publications and information division, Indian Council of Agricultural Research, New Delhi

- Greenberg S (2018) A review of participatory plant breeding and lessons for African seed and food sovereignty movements. Published by The African Centre for Biodiversity, Johannesburg, South Africa. <https://www.researchgate.net/publication/323771637>
- Greuter W, McNeill J, Barrie FR et al (2000) International code of botanical nomenclature (Saint Louis Code). Koeltz, Königstein, Regnum Veg 138:1–474
- Grotewold E (2008) Transcription factors for predictive plant metabolic engineering: are we there yet? *Curr Opin Biotechnol* 19(2):138–144. <https://doi.org/10.1016/j.copbio.2008.02.002>
- Guo P, Wang Y, Zhou X et al (2013) Expression of soybean lectin in transgenic tobacco results in enhanced resistance to pathogens and pests. *Plant Sci* 211:17–22. <https://doi.org/10.1016/j.plantsci.2013.06.001>
- Harada E, Kim JA, Meyer AJ et al (2010) Expression profiling of tobacco leaf trichomes identifies genes for biotic and abiotic stresses. *Plant Cell Physiol* 51(10):1627–1637. <https://doi.org/10.1093/pcp/pcq118>
- Harborne JB (1986) *The flavonoids*. Advances in Research Since 1986. Chapman Hall, London
- Hardon J, Norbu U, Nguyen S et al (2005) Evaluation report of the biodiversity use and conservation in Asia program (BUCAP). NORAD, Oslo
- Hayward AC (1991) Biology and epidemiology of bacterial wilt caused by *Pseudomonas solanacearum*. *Annu Rev Phytopathol* 29:65–87
- Hefferon KL (2010) Plant biotechnology patents: applications in agriculture and medicine. *Recent Pat Biotechnol* 4. <https://doi.org/10.2174/187220810791110723>
- Heist EP, Schardl N, Nesmith WC, Schardl CL (2002) Interactions of *Peronospora tabacina* with roots of *Nicotiana* spp. in Gnotobiotic associations. *Phytopathology* 92:400–405
- Hirohata A, Sato I, Kaino K et al (2019) CRISPR/Cas9-mediated homologous recombination in tobacco. *Plant Cell Rep* 38:463–473
- Hofius D, Herbers K, Melzer M et al (2001) Evidence for expression level-dependent modulation of carbohydrate status and viral resistance by the *Potato Leaf Roll Virus* movement protein in transgenic tobacco plants. *Plant J* 28(5):529–543. <https://doi.org/10.1046/j.1365-3113x.2001.01179.x>
- Holme IB, Wendt T, Holm PB (2013) Intragenesis and cisgenesis as alternatives to transgenic crop development. *Plant Biotechnol J* 11(4):395–407. <https://doi.org/10.1111/pbi.12055>
- Holmes FO (1938) Inheritance of resistance to tobacco-mosaic disease in tobacco. *Phytopathol* 28:553–561
- Holmes FO (1946) A comparison of the experimental host ranges of *Tobacco Etch* and *Tobacco Mosaic Viruses*. *Phytopathology* 36:643–657
- Holmes FO (1960) Inheritance in tobacco of an improved resistance to infection by TMV. *Virology* 12:59–67
- Holmes FO (1961) Concomitant inheritance of resistance to several viral diseases in tobacco. *Virology* 13:409–413
- Honig A, Marton I, Rosenthal M et al (2015) Transient expression of virally delivered meganuclease in planta generates inherited genomic deletions. *Mol Plant* 8(8):1292–1294. <https://doi.org/10.1016/j.molp.2015.04.001>
- Hospital F (2003) Marker-assisted breeding. In: Newbury HJ (ed) *Plant molecular breeding*. Blackwell Publishing and CRC Press, Oxford and Boca Raton, pp 30–59. https://doi.org/10.1007/978-3-642-21201-7_10
- Hossain DM (2020) Determination of economic threshold level (ETL) of *Tobacco Mosaic* and *Leaf Curl Viruses*. *Int J Sci: Basic Appl Res* 43:147–165
- Hou H, Atlihan N, Lu Z-X (2014) New biotechnology enhances the application of cisgenesis in plant breeding. *Front Plant Sci* 5:389. <https://doi.org/10.3389/fpls.2014.00389> <https://www.cimmyt.org/news/pests-and-diseases-and-climate-change-is-there-a-connection/>. Accessed 24 July 2021
- Hou Y, Ma X, Wan W et al (2016) Comparative genomics of pathogens causing brown spot disease of tobacco: *Alternaria longipes* and *Alternaria alternata*. *PLoS ONE* 11(5):e0155258. <https://doi.org/10.1371/journal.pone.0155258>

- Huang WK, Guo JY, Wan FH et al (2008) AFLP analyses on genetic diversity and structure of *Eupatorium adenophorum* populations in China. *Chin J Agric Biotechnol* 5:33–41
- ICAR (2015) Vision 2050 Central Tobacco Research Institute. Indian Council of Agricultural Research, New Delhi
- ICAR-CTRI (2005) Annual report 2004–05. ICAR-Central Tobacco Research Institute, Rajahmundry, India
- ICAR-CTRI (2006) Annual report 2005–06. ICAR-Central Tobacco Research Institute, Rajahmundry, India
- ICAR-CTRI (2010) Annual report 2009–2010. ICAR-Central Tobacco Research Institute, Rajahmundry, India
- ICAR-CTRI (2021a) Management of insect pests. https://ctri.icar.gov.in/for_controlPests.php. Accessed 31 Sept 2021a
- ICAR-CTRI (2021b) List of released/identified varieties of various tobacco types for cultivation in India. <https://ctri.icar.gov.in/files/varieties.pdf>. Accessed 31 Sept 2021
- ISAAA (2019) Global status of commercialized biotech/GM crops. ISAAA Brief 55–2019. <https://www.isaaa.org/resources/publications/briefs/55/default.asp>. Accessed 30 Sept 2021
- ISAAA (2020) Pocket K No. 42 Stacked traits in biotech crops. <https://www.isaaa.org/resources/publications/pocketk/42/default.asp>. Accessed 30 Sept 2021
- Jach G, Görnhardt B, Mundy J et al (1995) Enhanced quantitative resistance against fungal disease by combinatorial expression of different barley antifungal proteins in transgenic tobacco. *Plant J* 8(1):97–109. <https://doi.org/10.1046/j.1365-313X.1995.08010097.x>
- Jafar M, Aggarwal R, Ramesh B et al (2012) SNP markers and their impact on plant breeding. *Int J Plant Genomics* 2012. <https://doi.org/10.1155/2012/728398>
- Jarvis A, Lane A, Hijmans RJ (2008) The effect of climate change on crop wild relatives. *Agric Ecosyst Environ* 126:13–23
- Jiang G-L (2013) Molecular markers and marker-assisted breeding in plants: plant breeding from laboratories to fields. In: Andersen SB (ed) *IntechOpen*. <https://doi.org/10.5772/52583>
- Jingjing JIN (2018) Tobacco genomic resource. <https://www.coresta.org/abstracts/tobacco-genomic-resource-31854.html>. Accessed 30 Sept 2021
- Johnson J (1923) A bacterial leaf spot of tobacco. *J Agril Res* 23(6):481–493
- Johnson CS, Main CE, Gooding JGV (1983) Crop loss assessment for flue-cured tobacco cultivars infected with *Tobacco Mosaic Virus*. *Plant Dis* 67:881–885
- Johnson ES, Wolff MF, Wernsman EA (2002) Origin of the black shank resistance gene, Ph, in tobacco cultivar coker 371-Gold. *Plant Dis* 86:10. <https://doi.org/10.1094/PDIS.2002.86.10.1080>
- Johnston S, Anziano P, Shark K et al (1988) Mitochondrial transformation in yeast by bombardment with microprojectiles. *Science* 240:1538–1541
- Joshi BG, Sitaramaiah S (1975) Note on relative toxicity of some *Nicotiana* species to tobacco caterpillar, *Spodoptera litura*. *Tob Res* 1:86–87
- Joshi BG, Ramaprasad G, Sitaramaiah S (1978) Relative toxicity of some *Nicotiana* species to green peach aphid, *Myzus persicae*. *Tob Res* 4(1):65–66
- Jube S, Borthakur D (2007) Expression of bacterial genes in transgenic tobacco: methods, applications and future prospects. *Electron J Biotechnol* 10(3):452–467. <https://doi.org/10.2225/vol10-issue3-fulltext-4>
- Julio E, Verrier JL, de Dorlhac BF (2006) Development of SCAR markers linked to three disease resistances based on AFLP within *Nicotiana tabacum* L. *Theor Appl Genet* 112(2):335–346. <https://doi.org/10.1007/s00122-005-0132-y>
- Julio E, Cotucheau J, Decors C et al (2014) A eukaryotic translation initiation factor 4E (eIF4E) is responsible for the “va” tobacco recessive resistance to *Potyvirus*. *Plant Mol Biol Rep* 33:609–623. <https://doi.org/10.1007/s11105-014-0775-4>
- Jung J, Kim JJ, Yi G (2019) GeneCo: a visualized comparative genomic method to analyze multiple genome structures. *Bioinformatics* 35(24):5303–5305. <https://doi.org/10.1093/bioinformatics/btz596>

- Kajikawa M, Siervo N, Kawaguchi H et al (2017) Genomic insights into the evolution of the nicotine biosynthesis pathway in tobacco. *Plant Physiol* 174(2):999–1011
- Kalantidis K, Psaradakis S, Tabler M et al (2002) The occurrence of CMV-specific short RNAs in transgenic tobacco expressing virus-derived double stranded RNA is indicative of resistance to the virus. *Mol Plant Microbe Interact* 15:826–833
- Kamalay JC, Goldberg RB (1980) Regulation of structural gene expression in tobacco. *Cell* 19:935–946. [https://doi.org/10.1016/0092-8674\(80\)90085-9](https://doi.org/10.1016/0092-8674(80)90085-9)
- Kelman A (1953) The bacterial wilt caused by *Pseudomonas solanacearum*. *NC Agric Exp Stn Tech Bull* 99
- Kenton A, Parokonny AS, Gleba YY et al (1993) Characterization of the *Nicotiana tabacum* L. genome by molecular cytogenetics. *Mol Gen Genet* 240:159–169
- Khan SH (2019) Genome-editing technologies: concept, pros and cons of various genome-editing techniques and bioethical concerns for clinical application. *Mol Ther Nucleic Acids* 16:326–334. <https://doi.org/10.1016/j.omtn.2019.02.027>
- Knapp S, Chase MW, Clarkson JJ (2004) Nomenclatural changes and a new sectional classification in *Nicotiana* (Solanaceae). *Taxon* 52:73–82
- Komarnitsky SI, Komarnitsky IK, Cox A et al (1998) Molecular phylogeny of the nuclear 5.8S ribosomal RNA genes in 37 species of *Nicotiana* genus. *Genetika* 34:883–889
- Komatsu T, Horita H, Kitayama M (2002) Bacterial wilt of China Aster caused by *Erwinia chrysanthemi*. *J Gen Plant Path* 68(1):105–107
- Kostoff D (1943) An androgenic *Nicotiana* haploid. *Zeits. F Zelloysch U Mikr Anat* 9:640–642
- Krishnamurthy GVG (1994) *Orobanche* problem and management in India. In: Pieterse, AH, JAC, Verkleij et al (eds) *Biology and management of Orobanche*. Proceedings of the third international workshop on *Orobanche* and related *Striga* research. Royal Tropical Institute, Amsterdam, The Netherlands
- Krishnamurthy V, Sarala K, Murthy TGK (2008) Tobacco: an ideal plant for biological research. *Tob Res* 34:1–4
- Kuchuk N, Sytnyk K, Vasylenko M et al (2006) Genetic transformation of plastids of different Solanaceae species using tobacco cells as organelle hosts. *Theor Appl Genet* 113:519–527. <https://doi.org/10.1007/s00122-006-0318-y>
- Kumar A, Singh V (2016) Scenario of insect-pests under climate change situation and future challenges in India. *J Exp Zool India* 19(Supplement 1):1209–1213
- Kwak SY, Lew TTS, Sweeney CJ et al (2019) Chloroplast-selective gene delivery and expression in planta using chitosan-complexed single-walled carbon nanotube carriers. *Nat Nanotechnol* 14:447–455. <https://doi.org/10.1038/s41565-019-0375-4>
- LaMondia JA (2015) Fusarium wilt of tobacco. *Crop Prot* 73:73–77. <https://doi.org/10.1094/PDIS-09-12-0835-PDN>
- Lan T, Zheng S, Yang L et al (2014) Mapping of quantitative trait loci conferring resistance to bacterial wilt in tobacco (*Nicotiana tabacum* L.). *Plant Breed* 133:672–677
- Lander ES, Green P, Abrahamson J et al (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* 1(2):174–181. [https://doi.org/10.1016/0888-7543\(87\)90010-3](https://doi.org/10.1016/0888-7543(87)90010-3)
- Lapham DF (1976) Banket A1, a new burley tobacco cultivar resistant to brown spot, wildfire, and *Tobacco Mosaic Virus*. *Tob Sci* XX:111–114
- Lea HW (1963) The transfer of resistance against blue mold (*Peronospora tabacina* Adam) from *Nicotiana debneyi* to cultivated tobacco. *CORESTA Inf Bull* 1963(3):13–15
- Legg PD, Smeeton BW (1999) Breeding and genetics. In: David DL, Nielsen MT (eds) *Tobacco production. Chemistry and technology*. Blackwell Science, Malden, MA, USA
- Legg PD, Litton CC, Collins GB (1981) Effects of the *Nicotiana debneyi* black root rot resistance factor on agronomic and chemical traits in burley tobacco. *Theor Appl Genet* 60:365–368
- Leitch IJ, Hanson L, Lim KY et al (2008) The ups and downs of genome size evolution in polyploid species of *Nicotiana* (Solanaceae). *Ann Bot* 101:805–814

- Lester RN and Hawkes JG (2001) Solanaceae. In: Hanelt P (ed) Mansfeld's encyclopedia of agricultural and horticultural crops (except ornamentals), vol 4. Springer, Berlin
- 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 (2011) *Nicotiana*. In: Kole C (ed) Wild crop relatives: genomic and breeding resources, plantation and ornamental crops. Springer, Berlin, Heidelberg
- Lewis RS, Milla SR, Kемodle SP (2007) Analysis of an introgressed *Nicotiana tomentosa* genomic region affecting leaf number and correlated traits in *Nicotiana tabacum*. *Theor Appl Genet* 114:841–854
- Li D, Miller RD (2010) Marker assisted selection (MAS) in tobacco disease resistance breeding. In: Paper presented at 44th tobaccoworkers conference, CORESTA, 2010
- Li Q, Lawrence C, Xing HY et al (2001) Enhanced disease resistance conferred by expression of an antimicrobial magainin analog in transgenic tobacco. *Planta* 212:635–639. <https://doi.org/10.1007/s004250000480>
- Li S, Wang Z, Zhou Y et al (2018) Expression of *cry2Ahl* and two domain II mutants in transgenic tobacco confers high resistance to susceptible and Cry1Ac-resistant cotton bollworm. *Sci Rep* 8:508. <https://doi.org/10.1038/s41598-017-19064-5>
- Li R, Jin, Jingjing X, Jie W et al (2020). Long non-coding RNAs associate with jasmonate-mediated plant defense against herbivores. *Plant Cell Environ* 44. <https://doi.org/10.1111/pce.13952>
- Li S, Chang L, Zhang J (2021) Advancing organelle genome transformation and editing for crop improvement. *Plant Commun* 2(2):100141. <https://doi.org/10.1016/j.xplc.2021.100141>
- Liharska TB, Koornneef M, van Wordragen M et al (1996) Tomato chromosome 6: effect of alien chromosomal segments on recombinant frequencies. *Genome* 39:485–491
- Lim KY, Matyasek R, Lichtenstein CPAR et al (2000) Molecular cytogenetic analyses and phylogenetic studies in the *Nicotiana* section *Tomentosae*. *Chromosoma* 109:245–258
- Lim KY, Matyasek R, Kovarik A et al (2004) Genome evolution in allotetraploid *Nicotiana*. *Biol J Linn Soc* 82:599–606
- Lim KY, Matyasek R, Kovarik A et al (2005) Molecular cytogenetics and tandem repeat sequence evolution in the allopolyploid *Nicotiana rustica* compared with diploid progenitors *N. paniculata* and *N. undulate*. *Cytogenet Genome Res* 109:298–309. <https://doi.org/10.1159/000082413>
- Lin TY, Kao YY, Lin S et al (2001) A genetic linkage map of *Nicotiana plumbaginifolia*/*Nicotiana longiflora* based on RFLP and RAPD markers. *Theor Appl Genet* 103:905–911
- Liu XZ, Zhang HY (2008) Advance of molecular marker application in the tobacco research. *African J Biotechnol* 7(25):4827–4831
- Liu XZ, He CS, Yang YM et al (2009a) Genetic diversity among flue-cured tobacco cultivars on the basis of AFLP markers. *Czech J Genet Plant Breed* 45(4):155–159
- Liu XZ, Yang YM, He CS (2009b) A RAMP marker linked to the tobacco black shank resistant gene. *African J Biotechnol* 8(10):2060–2063
- Livingstone KD, Lackney VK, Blauth JR et al (1999) Genome mapping in *Capsicum* and the evolution of genome structure in the Solanaceae. *Genetics* 152:1183–1202
- Love AJ, Makarov VV, Yaminsky, et al (2014) The use of *Tobacco Mosaic Virus* And *Cowpea Mosaic Virus* for the production of novel metal nanomaterials. *Virology* 449:133–139
- Love AJ, Makarov VV, Sinitsyna OV et al (2015) A genetically modified *Tobacco Mosaic Virus* that can produce gold nanoparticles from a metal salt precursor. *Front Plant Sci* 6:984. <https://doi.org/10.3389/fpls.2015.00984>
- Lu JP (1997) The application of PAGE in the cultivars identification of flue-cured tobacco. *Seed* 5:30–32
- Lu XP, Gui YJ, Xiao BG et al (2012) Development of DArT markers for a linkage map of flue-cured tobacco. *Chin Sci Bull* 58:641–648
- Lu YH, Rijzaani H, Karcher D et al (2013) Efficient metabolic pathway engineering in transgenic tobacco and tomato plastids with synthetic multigene operons. *Proc Natl Acad Sci USA* E623–E632

- Lu YH, Stegemann S, Agrawal S et al (2017) Horizontal transfer of a synthetic metabolic pathway between plant species. *Curr Biol* 19:3034–3041
- Lucas GB (1975) Diseases of tobacco. Biological Consulting Association, Raleigh, NC
- Lucker J, Schwab W, Hautum BV et al (2004) Increased and altered fragrance of tobacco plants after metabolic engineering using three monoterpene synthases from lemon. *Plant Physiol* 34(1):510–519. <https://doi.org/10.1104/pp.103.030189>
- Luo M, Wang Z, Li H (2009a) Overexpression of a weed (*Solanum americanum*) proteinase inhibitor in transgenic tobacco results in increased glandular trichome density and enhanced resistance to *Helicoverpa armigera* and *Spodoptera litura*. *Int J Mol Sci* 10(4):1896–1910. <https://doi.org/10.3390/ijms10041896>
- Luo Y, Zhu M, Yang J et al (2009b) A spreading colony formed method for rapid evaluation of dicarboximide fungicides resistance level of field tobacco brown spot disease. *Ann Microbiol* 59:173–177
- Ma JKC, Drake PMW, Christou P (2003) The production of recombinant pharmaceutical proteins in plants. *Genet* 4:794–805
- Ma HB, Qi JM, Li YK et al (2008) Construction of a molecular genetic linkage map of tobacco based on SRAP and ISSR markers. *Acta Agron Sin* 34:1958–1963
- Ma JM, Heim C, Humphry M et al (2019) Genetic analysis of *Phn7.1*, a major QTL conferring partial resistance to *Phytophthora nicotianae* in *Nicotiana tabacum*. *Mol Breed* 39:11. <https://doi.org/10.1007/s11032-018-0923-x>
- Ma JM, Heim CB, Humphry M et al (2020) Characterization of *Phn15.1*, a newly identified *Phytophthora nicotianae* resistance QTL in *Nicotiana tabacum*. *Plant Dis* 104 (6):1633–1646. <https://doi.org/10.1094/PDIS-10-19-2257-RE>
- Madhav MS, Raju KS, Gaikwad K et al (2015) Development of new set of microsatellite markers in cultivated tobacco and their transferability in other *Nicotiana* spp. *Mol Plant Breed* 6(16):1–13
- Maliga P, Svab Z (2010) Engineering the plastid genome of *Nicotiana sylvestris*, a diploid model species for plastid genetics. In: Birchler J (ed) Plant chromosome engineering: methods and protocols. Humana Press, Totowa. https://doi.org/10.1007/978-1-61737-957-4_2
- Malik H, Raza A, Amin I et al (2017) RNAi-mediated mortality of the whitefly through transgenic expression of double-stranded RNA homologous to acetylcholinesterase and ecdysone receptor in tobacco plants. *Sci Rep* 6:38469. <https://doi.org/10.1038/srep38469>
- Malone LA, Barraclough EI, Kui LK et al (2009) Effects of red-leaved transgenic tobacco expressing a MYB transcription factor on two herbivorous insects, *Spodoptera litura* and *Helicoverpa armigera*. *Entomol Exp Appl* 133(2):117–127. <https://doi.org/10.1111/j.1570-7458.2009.00910.x>
- Mani V, Reddy CS, Lee SK (2020) Chitin biosynthesis Inhibition of *Meloidogyne incognita* by RNAi-mediated gene silencing increases resistance to transgenic tobacco Plants. *Int J Mol Sci* 21:6626. <https://doi.org/10.3390/ijms21186626>
- Manly KF, Cudmore RH Jr, Meer JM (2001) Map Manager QTX, cross-platform software for genetic mapping. *Mamm Genome* 12(12):930–932. <https://doi.org/10.1007/s00335-001-1016-3>
- Margin B, Thouquet P, Olivier J et al (1999) Temporal and multiple quantitative trait loci analysis of resistance to bacterial wilt in tomato permit resolution of linked loci. *Genetics* 151:1165–1172
- Markert CL, Moller F (1959) Multiple forms of enzymes: tissue, ontogenetic and species specific pattern. *Proc Nat Acad Sci USA* 45:753–763
- Markose VT, Patel GJ (1977) Effect of soil fumigation and nitrogen manuring on bidi tobacco. *Tob Res* 3:7–10
- Martinez N, Miller RD, Li D et al (2011) Selection of resistance to multiple pathogens in tobacco assisted by markers and greenhouse screenings. https://www.coresta.org/sites/default/files/abstracts/2011_TSRC46_Martinez.pdf
- Mateu-Andres I, de Paco L (2005) Allozymic differentiation of the *Antirrhinum majus* and *A. siculum* species groups. *Ann Bot* 95:465–473
- Matsuda T (1977) Fundamental studies on the breeding of bacterial wilt resistant varieties in tobacco (in Japanese). *Bull Utsunomiya Tobacco Exp Stn* 15:1–95

- Matsuoka K, Demura T, Galis I et al (2004) A comprehensive gene expression analysis toward the understanding of growth and differentiation of tobacco BY-2 cells. *Plant Cell Physiol* 45(9):1280–1289. <https://doi.org/10.1093/pcp/pch155>
- Matyasek R, Lim KY, Kovarik A et al (2003) Ribosomal DNA evolution and gene conversion in *Nicotiana rustica*. *Heredity* (edinb) 91(3):268–275. <https://doi.org/10.1038/sj.hdy.6800333>
- McBride KE, Svab Z, Schaaf DJ et al (1995) Amplification of a chimeric *Bacillus* gene in chloroplasts leads to an extraordinary level of an insecticidal protein in Tobacco. *Biotechnol* 13:362–365
- McLean MD, Yevtushenko DP, Deschene A et al (2003) Overexpression of glutamate decarboxylase in transgenic tobacco plants confers resistance to the northern root-knot nematode. *Mol Breed* 11:277–285
- Mehmood F, Abdullah, Ubaid Z et al (2020) Plastid genomics of *Nicotiana* (Solanaceae): insights into molecular evolution, positive selection and the origin of the maternal genome of Aztec tobacco (*Nicotiana rustica*). *Peer J* 8:e9552. <https://doi.org/10.7717/peerj.9552>
- Meissner R, Chague V, Zhu Q et al (2000) Technical advance: a high throughput system for transposon tagging and promoter trapping in tomato. *Plant J* 22:265–274. <https://doi.org/10.1046/j.1365-3113x.2000.00735.x>
- Melton TA, Shew HD (2000) Brown spot. <http://www.ces.ncsu.edu/depts/pp/notes/oldnotes/tb3.htm>. Accessed 15 Sept 2021
- Meyers B, Zaltsman A, Lacroix B et al (2010) Nuclear and plastid genetic engineering of plants: comparison of opportunities and challenges. *Biotechnol Adv* 28:747–756
- Mielby H (2011) Public attitudes to Cisgenic crops. PhD thesis, University of Copenhagen
- Milla SR, Levin JS, Lewis RS et al (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
- Miller RD (1987) TN 86: A burley tobacco resistant to TVMV, TEV, and PVY. In: *Bulletin* 10–1987, University of Tennessee, Knoxville. https://trace.tennessee.edu/utk_agbulletin/454
- Misra P, Pandey A, Tiwari M et al (2010) Modulation of transcriptome and metabolome of tobacco by *Arabidopsis* transcription factor, AtMYB12, leads to insect resistance. *Plant Physiol* 152(4):2258–2268. <https://doi.org/10.1104/pp.109.150979>
- Mitter N, Worrall EA, Robinson KE et al (2017) Clay nanosheets for topical delivery of RNAi for sustained protection against plant viruses. *Nat Plants* 3:1–10
- Mo J, Chen Q, Yu F (2012) Analysis of meteorological condition for infection of the pathogen of tobacco brown spot (*Alternaria alternata*). *Plant Dis Pests* 3:24–27
- Monga D, Dobhal VK (1987) Note on reaction of cigar wrapper tobacco (*Nicotiana tabacum* L.) germplasm to brown spot disease. *Tob Res* 13:69–70
- Moon H (2006) Identification of AFLP markers linked to tomato spotted wilt resistance in tobacco. Ph D thesis, North Carolina State University
- Moon HS, Nicholson JS, Lewis RS (2008) Use of transferable *Nicotiana tabacum* L. microsatellite markers for investigating genetic diversity in the genus *Nicotiana*. *Genome* 51:547–559
- Moon HS, Nicholson JS, Heineman A et al (2009a) Changes in genetic diversity of U.S. Flue-cured tobacco germplasm over seven decades of cultivar development. *Crop Sci* 49(2):498–508. <https://doi.org/10.2135/cropsci2008.05.0253>
- Moon HS, Nifong JM, Nicholson JS et al (2009b) Microsatellite-based analysis of tobacco (*Nicotiana tabacum* L.) genetic resources. *Crop Sci* 49:1–11. <https://doi.org/10.2135/cropsci2009.01.0024>
- Moore EL, Powell NT (1959) Dominant and modifying genes for resistance to black shank of tobacco. *Phytopathology* 49:318–319
- Mora F, Quiral YA, Matus I et al (2016) SNP-based QTL mapping of 15 complex traits in Barley under rain-fed and well-watered conditions by a mixed modeling approach. *Front Plant Sci* 7:909
- Morgulis A, Gertz EM, Schäffer AA et al (2006) Window-based masker for sequenced genomes. *Bioinformatics* 22(2):134–141. <https://doi.org/10.1093/bioinformatics/bti774>

- Morris M, Dreher K, Ribaut JM (2003) Money matters (II): costs of maize inbred line conversion schemes at CIMMYT using conventional and marker-assisted selection. *Mol Breed* 11:235–247. <https://doi.org/10.1023/A:1022872604743>
- Murad L, Lim KY, Christopodulou V et al (2002) The origin of tobacco's T genome is traced to a particular lineage within *Nicotiana tomentosiformis* (Solanaceae). *Am J Bot* 89:921–928
- Murthy TGK, Sreedhar U, Siva Raju K et al (2014) TBST-2: A high yielding multiple resistant FCV line for traditional black soils of Andhra Pradesh. *Tob Res* 40(2):73–78
- Myles S, Peiffer J, Brown PJ et al (2009) Association mapping: critical considerations shift from genotyping to experimental design. *Plant Cell* 21:2194–2202. <https://doi.org/10.1105/tpc.109.068437>
- Nadeem MA, Nawaz MA, Shahid MQ et al (2018) DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing. *Biotechnol Equip* 32(2):261–285. <https://doi.org/10.1080/13102818.2017.1400401>
- Narasimha Rao CV, Prabhu SR (2005) Solanesol. In: Singh KD and Jain KC (eds) *Alternative uses of tobacco*. Central Tobacco Research Institute, Rajahmundry, India
- Narayan RK (1987) Nuclear DNA changes, genome differentiation and evolution in *Nicotiana* (Solanaceae). *Plant Syst Evol* 157(04):161–180
- Narayanan AI, Murthy TGK, Sarala K et al (2003) Division of crop improvement. Information bulletin. Central Tobacco Research Institute, Rajahmundry
- Nesmith WC (1984) The North American blue mold warning system. *Plant Dis* 11:933–936
- Niazian M, Sadatnoori SA, Galuszka P et al (2017) Tissue culture-based *Agrobacterium*-mediated and in Planta transformation methods. *Czech J Genet Plant Breed* 53:133–143. <https://doi.org/10.17221/177/2016-CJGPB>
- Nishi T, Tajima T, Noguchi S et al (2003) Identification of DNA markers of tobacco linked to bacterial wilt resistance. *Theor Appl Genet* 106(4):765–770. <https://doi.org/10.1007/s00122-002-1096-9>
- Noguchi S, Tajima T, Yamamoto Y et al (1999) Deletion of a large genomic segment in tobacco varieties that are resistant to *Potato Virus Y* (PVY). *Mol Gen Genet* 262:822–829
- Nolla JSB, Roque A (1933) A variety of tobacco resistant to ordinary tobacco mosaic. *J Puerto Rico Dept Agric* 17:301–303
- Nombela G, Muniz M (2010) Host plant resistance for the management of *Bemisia tabaci*: a multi-crop survey with emphasis on tomato. In: Stansly PA, Naranjo SE (eds) *Bemisia: bionomics and management of a global pest*. Springer Science Business Media, Heidelberg, Germany
- Okumus A, Gulumser A (2001) Chiasma frequency chromosome pairing at tobacco genotypes in the Middle Black Sea region. *Online J Biol Sci* 1(7):612–613
- Opperman CH, Lommel SA, Sosinski BR et al (2003) The tobacco genome initiative. In: Paper presented at CORESTA meeting, agronomy/phytopathology, Bucharest, Romania, 12–17 Oct 2003
- Pace J, Gardner C, Romay C (2015) Genome-wide association analysis of seedling root development in maize (*Zea mays* L.). *BMC Genomics* 16:47
- Paran I, Gidoni D, Jacobsohn R (1997) Variation between and within broomrape (Orobanch) species revealed by RAPD markers. *Heredity* 78:68–74
- Patel SK, Patel DJ, Patel HV (1986) Estimation of losses due to root-knot nematodes in bidi tobacco field. In: Paper presented at national conference on plant parasitic nematodes. Problems and Progress, Indian Agricultural Research Institute, New Delhi, India, 17–20 Dec 1986
- Patel JA, Patel BK, Chakraborty MK (1998) Production potential and quality aspects of tobacco seed oil. *Tob Res* 24:44–49
- Patel BN, Patel HR, Patel PM et al (2001) Assessment of avoidable yield loss due to frog eye disease in bidi tobacco. *Tob Res* 27:187–189
- Paterson AH, DeVerna JW, Lanini B et al (1990) Fine mapping of quantitative trait loci using selected overlapping recombinant chromosomes, in an interspecies cross of tomato. *Genetics* 124(3):735–742. <https://doi.org/10.1093/genetics/124.3.735>

- Pérez L, Rodríguez ME, Rodríguez F et al (2003) Efficacy of acibenzolar-S-methyl, an inducer of systemic acquired resistance against tobacco blue mould caused by *Peronospora hyoscyami* f. sp. *tabacina*. *Crop Prot* 22:405–413
- Powell-Abel P, Nelson RS, De B et al (1986) Delay of disease development in transgenic plants that express the *Tobacco Mosaic Virus* coat protein gene. *Sci* 232:738–743
- Prabhakararao K, Sarala K, Murthy TGK et al (2012) Application of simple sequence repeat (SSR) markers in diversity analysis of tobacco types. *Tob Res* 38(1):29–33
- Prabhakararao K, Sarala K, Murthy TGK (2016) Patent mapping of disease resistance genes in tobacco. *J Agric Crop Sci* 3:01–07
- Prins M, Laimer M, Noris et al (2008) Strategies for antiviral resistance in transgenic plants. *Mol Plant Path* 9(1):73–83. <https://doi.org/10.1111/J.1364-3703.2007.00447.X>
- Pritchard JK, Stephens M, Rosenberg NA et al (2000) Association mapping in structured populations. *Am J Hum Genet* 67(1):170–181
- Priya DB, Somasekhar N, Prasad J et al (2011) Transgenic tobacco plants constitutively expressing *Arabidopsis* NPR1 show enhanced resistance to root-knot nematode, *Meloidogyne incognita*. *BMC Res Notes* 4:231. <https://doi.org/10.1186/1756-0500-4-231>
- Puchta H (1999) Use of I-Sce I to induce DNA double-strand breaks in *Nicotiana*. *Methods Mol Biol* 113:447–451
- Purrington CB (2000) Costs of resistance. *Curr Opin Plant Biol* 3:305–308
- Qi JM, Wang T, Chen SH et al (2006) Genetic diversity and genetic relatives analysis of tobacco germplasm based on inter-simple sequence repeat (ISSR). *Acta Agron Sin* 32:373–378
- Qian YL, Wanga ZS, Wang DZ et al (2013) The detection of QTLs controlling bacterial wilt resistance in Tobacco (*N. tabacum* L.). *Euphytica* 192:259–266
- Qin Q, Li Y, Ding N et al (2018) Development of user-friendly markers for disease resistance to black root rot of tobacco through genotyping by sequencing. *Mol Breed* 38:76. <https://doi.org/10.1007/s11032-018-0834-x>
- Qin Y, Bai S, Li W et al (2020) Transcriptome analysis reveals key genes involved in the regulation of nicotine biosynthesis at early time points after topping in tobacco (*Nicotiana tabacum* L.). *BMC Plant Biol* 20:30. <https://doi.org/10.1186/s12870-020-2241-9>
- Rabara RC, Tripathi P, Rushton PJ (2017) Comparative metabolome profile between tobacco and soybean grown under water-stressed conditions. *BioMed Res Int*. <https://doi.org/10.1155/2017/3065251>
- Rafalski JA (2010) Association genetics in crop improvement. *Curr Opin Plant Biol* 13:174–180. <https://doi.org/10.1016/j.pbi.2009.12.004>
- Ramaprasad G, Sitaramaiah S, Sreedhar U (2000) Efficacy of *Bacillus thuringiensis* Berliner formulations against tobacco caterpillar, *Spodoptera litura* F. in tobacco nurseries. *Tob Res* 26(2):1–6
- Ramavarma KT, Apparao K, Sitharamaiah S (1980) Interspecific transfer of resistance to tobacco caterpillar (*Spodoptera litura* F.) from *Nicotiana benthamiana* to *N. tabacum*. *Cytologia* 45(1/2):103–111
- Rao RSN, Sreedhar U, Rao SG et al (1994) Bio environmental control of *Heliothis armigera* in FCV tobacco. *Tob Res* 20(1):36–39
- Rascon-Cruz Q, Gonzalez-Barriga CD, Blanca FIBF et al (2021) Plastid transformation: advances and challenges for its implementation in agricultural crops. *Electron J Biotechnol* 51:95–109. <https://doi.org/10.1016/j.ejbt.2021.03.005>
- Rastas P (2017) Lep-MAP3: robust linkage mapping even for low-coverage whole genome sequencing data. *Bioinformatics* 33(23):3726–3732. <https://doi.org/10.1093/bioinformatics/btx494>
- Reddy TSN, Nagarajan K (1981) Studies on *Tobacco Mosaic Virus* on tobacco (*Nicotiana tabacum* L.). I. Evaluation of germplasm and loss in yield and Quality of FCV tobacco. *Tob Res* 7:75–80
- Reddy TSN, Nagarajan K, Chandwani GH (1976) Sources of resistance in *Nicotiana* species to important tobacco diseases in India. *Tob Res* 2:63–68

- Remacle C, Cardol P, Coosemans N et al (2006) High-efficiency biolistic transformation of *Chlamydomonas* mitochondria can be used to insert mutations in complex I genes. *Proc Natl Acad Sci USA* 103:4771–4776
- Ren N, Timko MP (2001) AFLP analysis of genetic polymorphism and evolutionary relationships among cultivated and wild *Nicotiana* species. *Genome* 44:559–571
- Ren M, Zhang C, Jiang C (2014) Association analysis of tobacco aroma constituents based on high density SSR linkage group. *Acta Tab Sin* 20:88–93
- Ren C, Liu Y, Wang X et al (2020) Targeted genome editing in *Nicotiana tabacum* using inducible CRISPR/Cas9 system. *bioRxiv* 2020.03.13.990085. <https://doi.org/10.1101/2020.03.13.990085>
- Ribaut J-M, Hoisington D (1998) Marker-assisted selection: new tools and strategies. *Trends Plant Sci* 3:236–239. [https://doi.org/10.1016/S1360-1385\(98\)01240-0](https://doi.org/10.1016/S1360-1385(98)01240-0)
- Rommens CM, Humara JM, Ye J et al (2004) Crop improvement through modification of the plant's own genome. *Plant Physiol* 135:421–431. <https://doi.org/10.1104/pp.104.040949>
- Rossi L, Bindler G, Pijnenburg H et al (2001) Potential of molecular marker analysis for variety identification in processed tobacco. *Plant Var Seeds* 14:89–101
- Rostoks N, Borevitz J, Hedley P et al (2005) Single-feature polymorphism discovery in the barley transcriptome. *Genome Biol* 6:R54. <https://doi.org/10.1186/gb-2005-6-6-r54>
- Roy S, Amarnath S, Arya RL (2008) Symptomatology and integrated management of bacterial diseases of Jati (*N. tabacum*) and Motihari (*N. rustica*) tobacco in terai agro-ecological zone of West Bengal. *Environ Ecol* 26:528–532
- Roy S, Amarnath S, Bhattacharyya PM (2012) screening of *Nicotiana rustica* tobacco germplasm for resistance to hollow stalk disease in North Bengal. *Tob Res* 38(1):9–14
- Ruffy RC, Main CE (1989) Components of partial resistance to blue mold in six tobacco genotypes under controlled environmental conditions. *Phytopathol* 79:606–609
- Rushton PJ, Bokowiec MT, Laudeman TW et al (2008) TOBFAC: the database of tobacco transcription factors. *BMC Bioinform* 9:53
- Russell AW, Sparrow R (2008) The case for regulating intragenic GMOs. *J Agr Environ Ethic* 21:153–181. <https://doi.org/10.1007/s10806-007-9074-5>
- Sander JD, Dahlborg EJ, Goodwin MJ et al (2011) Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). *Nat Methods* 8(1):67–69. <https://doi.org/10.1038/nmeth.1542>
- Sane VA, Nath P, Sane A et al (1997) Development of insect-resistant transgenic plants using plant genes: expression of cowpea trypsin inhibitor in transgenic tobacco plants. *Curr Sci* 72(10):741–747
- Sarala K (2019) Next Gen Breeding Strategies for Enhancing FCV Tobacco Productivity and Quality. In: Paper presented at national symposium on approaches and strategies for augmenting tobacco farmers' income, Indian Society of Tobacco Sciences, Rajahmundry, 19–20 July 2019
- Sarala K, Rao RVS (2008) Genetic diversity in Indian FCV and burley tobacco cultivars. *J Genet* 87:159–163
- Sarala K, Nagarajan K, Murthy MS et al (2005) Yield improvement in black soil Virginia tobacco (*Nicotiana tabacum* L.). *Indian J Genet Plant Breed* 65(1):33–36
- Sarala K, Murthy TGK, Krishnamurthy V (2012) Tobacco (*Nicotiana tabacum* L.). In: Bharadwaj DN (ed) Breeding of field crops. Agrobios (India), Jodhpur
- Sarala K, Prabhakara Rao K, Murthy TGK et al (2016) Yield gains in Indian flue-cured virginia tobacco: a case study. *Int J Agril Sci* 8(52):2538–2541
- Sarala K, Baghyalakshmi K, Prabhakara Rao K et al (2018) Genetic diversity among mutant germplasm accessions of *Nicotiana tabacum* as determined by morphological parameters. *Tob Res* 44(2):47–53
- Sasser JN, Freckman DW (1987) A world perspective on nematology: the role of society. In: Veech JA, Freckman DW (eds) *Vistas on nematology: a commemoration of the twenty fifth anniversary of the society of nematologist*, pp 1–4
- Sastri AB, Krishnamurthy KV, Nagarajan K (1981) Performance of *Tobacco Mosaic Virus* resistant flue-cured tobacco varieties under field conditions. *Tob Res* 7:145–149

- Schaart JG, Krens FA, Pelgrom KTB et al (2004) Effective production of marker-free transgenic strawberry plants using inducible site-specific recombination and a bifunctional selectable marker gene. *Plant Biotechnol J* 2:233–240. <https://doi.org/10.1111/j.1467-7652.2004.00067.x>
- Schiltz P (1981) Downy mildew of tobacco. In: Spencer DM (ed) *The downy mildews*. Academic Press, London, UK, pp 577–599
- Schouten HJ, Jacobsen E (2008) Cisgenesis and intragenesis, sisters in innovative plant breeding. *Trends Plant Sci* 13(6):260–261
- Schouten HJ, Krens FA, Jacobsen E (2006a) Cisgenic plants are similar to traditionally bred plants: international regulations for genetically modified organisms should be altered to exempt cisgenesis. *EMBO Rep* 7:750–753
- Schouten HJ, Krens FA, Jacobsen E (2006b) Do cisgenic plants warrant less stringent oversight? *Nat Biotechnol* 24:753
- Seebold K, Townsend L, Green JD (2007) *Pest Management*. In: Seebold K, Pearce B (eds) 2013–14 Kentucky and Tennessee tobacco production guide. University of Kentucky, Lexington, Kentucky and The University of Tennessee Knoxville, Tennessee
- Senthilkumar R, Cheng CP, Yeh KW (2010) Genetically pyramiding protease-inhibitor genes for dual broad-spectrum resistance against insect and phytopathogens in transgenic tobacco. *Plant Biotechnol J* 8:65–75. <https://doi.org/10.1111/j.1467-7652.2009.00466.x>
- Senthil-Kumar M, Mysore K (2014) *Tobacco Rattle Virus*-based virus-induced gene silencing in *Nicotiana benthamiana*. *Nat Protoc* 9:549–1562. <https://doi.org/10.1038/nprot.2014.092>
- Shah HM, Patel DJ, Valand GB (1983) Assessment of losses due to root-knot disease in bidi tobacco nursery. *Tob Res* 9:108–109
- Shenoi MM, Srinivas SS, Karunakaramurthy K et al (2002) A new rating scale for assessment of brown spot disease in tobacco. *Tob Res* 28:142–145
- Shew HD, Lucas GB (1991) *Compendium of tobacco diseases*. APS Press, USA
- Shinozaki K, Ohme M, Tanaka M et al (1986) The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. *EMBO J* 5(9):2043–2049
- Sierro N, Battey JN, Ouadi S et al (2013a) Reference genomes and transcriptomes of *Nicotiana sylvestris* and *Nicotiana tomentosiformis*. *Genome Biol* 4(6):R60. <https://doi.org/10.1186/gb-2013-14-6-r60>
- Sierro N, Oeveren JV, Eijk MJTV et al (2013b) Whole genome profiling physical map and ancestral annotation of tobacco Hicks Broadleaf. *Plant J* 75(5):880–889. <https://doi.org/10.1111/tpj.12247>
- 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. <https://doi.org/10.1038/ncomms4833>
- Sierro N, Battey JND, Bovet L et al (2018) The impact of genome evolution on the allotetraploid *Nicotiana rustica*—an intriguing story of enhanced alkaloid production. *BMC Genomics* 19(1):855. <https://doi.org/10.1186/s12864-018-5241-5>
- Silva AT, Nguyen A, Ye C et al (2010) Conjugated polymer nanoparticles for effective siRNA delivery to tobacco BY-2 protoplasts. *BMC Plant Biol* 10:291. <https://doi.org/10.1186/1471-2229-10-291>
- Sitaramaiah S, Ramaprasad G, Murthy NS et al (1994) Estimation of avoidable yield losses in FCV tobacco due to insect pests. *Tob Res* 20:47–52
- Sitaramaiah S, Ramaprasad G, Sreedhar U (1999) Management of tobacco ground beetle, *Mesomorpha villiger* Blanch with insecticide baits on FCV tobacco. *Ind J Agric Sci* 69(9):660–663
- Sitaramaiah S, Sreedhar U, Gunneswararao S et al (2002) Bio-intensive management of leaf eating caterpillar, *Spodoptera litura* F. in tobacco nurseries. In: Koul O, Dhaliwal GS, Marwaha S et al (eds) *Biopesticides and Pest management*. Campus Books International, New Delhi
- Sivaraju K, Madhav MS, Sharma RK et al (2008) Molecular diversity in Indian tobacco types as revealed by randomly amplified DNA polymorphism. *J Plant Biochem Biotechnol* 17:51–56
- Smeeton BW, Ternouth RAF (1992) Sources of resistance to powdery mildew, wildfire, angular leaf spot, *Alternaria*. *CORESTA Info Bull* 3(4):127–135
- Smith KM (1937) *A textbook of plant virus diseases*. P Blakiston's Son & Co Inc., Philadelphia

- Smith TE, Clayton EE (1948) Resistance to bacterial wilt and black shank in flue-cured tobacco. *Phytopathology* 38:227–229
- South PF, Cavanagh AP, Liu HW et al (2019) Synthetic glycolate metabolism pathways stimulate crop growth and productivity in the field. *Sci* 363(6422):eaat9077. <https://doi.org/10.1126/science.aat9077>
- Spassova MI, Prins TW, Folkertsma RT et al (2001) The tomato gene Sw-5 is a member of the coiled coil, nucleotide binding, leucine-rich repeat class of plant resistance genes and confers resistance to TSWV in tobacco. *Mol Breed* 7:151–161
- Sreedhar U (2011) Evaluation of tobacco genotypes for resistance to insect pests. *Ind J Plant Prot* 39(3):183–185
- Sreedhar U (2016) Climate change effects on dynamics of insect pests: management options in tobacco. In: Chattopadhyay C, Prasad D (eds) *Dynamics of crop protection and climate change*. Studera Press, New Delhi
- Sreedhar U (2020) Comparative effectiveness of new insecticides against *Spodoptera litura* in tobacco nurseries. *Innov Farm* 5(2):98–102
- Sreedhar U, Sitaramaiah S, Venkateswarlu P (2005) Validation of IPM module in FCV tobacco. Paper presented at national conference on tobacco-scientific strategies for sustainable growth of farm economy and export, Andhra Pradesh Agricultural University, Hyderabad, 3–7 October 2005
- Sreedhar U, Sitaramaiah S, Padmaja K (2007) Evaluation of trap crops for management of budworm, *Helicoverpa armigera* in flue cured Virginia tobacco. *Ind J PI Protection* 35(2):213–217
- Sreedhar U, Gunneswararao S, Krishnamurthy V (2014) Biopesticides in tobacco pest management: problems and prospects. In: Koul O, Dhaliwal G S, Khokhar S et al (eds) *Biopesticides in sustainable agriculture: progress and potential*. Scientific Publishers (India), Jodhpur
- Sreedhar U, Subbarao R (2014) Development and evaluation of pest management modules in burley tobacco. *Tob Res* 40 (2):97–102
- Sreeramulu KR, Onkarappa T, Narayana Swamy H (1998) Biocontrol of damping off and black shank disease in tobacco nursery. *Tob Res* 24:1–4
- Stam P, Zeven C (1981) The theoretical proportion of the donor genome in near-isogenic lines of self-fertilizers bred by backcrossing. *Euphytica* 30:227–238
- Stavely JR, Skoog HA (1976) transfer of resistance to a virulent strain of *Pseudomonas tabaci* from *Nicotiana rustica* to *Nicotiana tabacum* breeding. *Proc Am Phytopathol Soc* 3:231
- Steward R (2018) Economic impact of disclosure requirements in patent applications for ‘genetic resources’-based innovation. International Federation of Pharmaceutical Manufacturers and Associations (IFPMA), Geneva Switzerland and Crop Life International, Brussels, Belgium
- Suen DF, Wang CK, Lin RF et al (1997) Assignment of DNA markers to *Nicotiana sylvestris* chromosomes using monosomic alien addition lines. *Theor Appl Genet* 94:331–337
- Sugiyama Y, Watase Y, Nagase M et al (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. <https://doi.org/10.1007/s00438-004-1075-8>
- Sun H, Lang Z, Zhu L et al (2012) Acquiring transgenic tobacco plants with insect resistance and glyphosate tolerance by fusion gene transformation. *Plant Cell Rep* 31:1877–1887. <https://doi.org/10.1007/s00299-012-1301-5>
- Sun M, Cheng L, Jiang C et al (2018) Identification of a major QTL affecting resistance to brown spot in tobacco (*Nicotiana tabacum* L.) via linkage and association mapping methods. *Euphytica* 214:195
- Surender K, Kirti B, Nishat P et al (2016) The CRISPR/Cas genome-editing tool: application in improvement of crops. *Front Plant Sci* 7:506. <https://doi.org/10.3389/fpls.2016.00506>
- Svab Z, Maliga P (1993) High-frequency plastid transformation in tobacco by selection for a chimeric aadA gene. *Proc Natl Acad Sci USA* 90(3):913–917. <https://doi.org/10.1073/pnas.90.3.913>

- Svab Z, Hajdukiewicz P, Maliga P (1990) Stable transformation of plastids in higher plants. *Proc Natl Acad Sci USA* 87(21):8526–8530. <https://doi.org/10.1073/pnas.87.21.8526>
- Swiderska K, Reid H, Song Y et al (2011) The role of traditional knowledge and crop varieties in adaptation to climate change and food security in SW China, Bolivian Andes and coastal Kenya. <https://pubs.iied.org/sites/default/files/pdfs/migrate/G03338.pdf>. Accessed 30 Sept 2021
- Tahmasbali M, Darvishzadeh R, Fayaz MA (2021) Association analysis of morpho-phenological traits in oriental tobacco (*Nicotiana tabacum* L.) with SSR markers under presence and absence of *Orobanche* conditions. *J Mol Cell Res (Iranian J Biol)*. https://cell.ijbio.ir/article_2000.html
- Takakura Y, Udagawa H, Shinjo A et al (2018) Mutation of a *Nicotiana tabacum* L. eukaryotic translation-initiation factor gene reduces susceptibility to a resistance-breaking strain of *Potato Virus Y*. *Mol Plant Pathol* 19:2124–2133
- Tanksley SD, Ganai MW, Prince JP et al (1992) High density molecular linkage maps of the tomato and potato genomes. *Genetics* 132(4):1141–1160. <https://doi.org/10.1093/genetics/132.4.1141-1160>
- Tbbnovsky MF (1941) Methods of breeding tobacco varieties resistant to tobacco mosaic and powdery mildew. The AI Mikoyan pan-soviet Sci Res Inst Tob Indian Tob Ind Krasnodar, Krasnodar
- Telem RS, Wani SH, Singh NB et al (2013) Cisgenics—a sustainable approach for crop improvement. *Curr Genomics* 14(7):468–476. <https://doi.org/10.2174/13892029113146660013>
- Ternouth RAF, MacKenzie J, Shepherd JA (1986) Introduction of *Meloidogyne javanica* resistance into flue cured tobacco in Zimbabwe. Paper presented at CORESTA Symposium, Taormina, Sicily, Italy, 26–30 Oct 1986
- 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(2). <https://doi.org/10.1371/journal.pone.0036204>
- Tezuka T, Tsutomu K, Toshiaki M et al (2010) Seven of eight species in *Nicotiana* section *Suaveolentes* have common factors leading to hybrid lethality in crosses with *Nicotiana tabacum*. *Ann Bot* 106(2):267–276. <https://doi.org/10.1093/aob/mcq114>
- Thakur N, Upadhyay SK, Verma PC et al (2014) Enhanced whitefly resistance in transgenic tobacco plants expressing double stranded RNA of *v-ATPase A* Gene. *PLoS One* 9(3):e87235. <https://doi.org/10.1371/journal.pone.0087235>
- Thimmgowda GC, Ramadoss SK, Kaikala V et al (2018) Whole genome resequencing of tobacco (*Nicotiana tabacum* L.) genotypes and high-throughput SNP discovery. *Mol Breed* 38:121
- Thomas BR, Van Deynze A, Bradford KJ (2002) Production of therapeutic proteins in plants. Agricultural biotechnology in california series, Publication 8078, University of California, Davis
- Thornsberry JM, Goodman MM, Doebley J et al (2001) Dwarf8 polymorphisms associate with variation in flowering time. *Nat Genet* 28(3):286–289
- Thurston R (1961) Resistance in *Nicotiana* to the green peach aphid and some other tobacco insect pests. *J Econ Entomol* 54(5):946–949. <https://doi.org/10.1093/jee/54.5.946>
- Tian Y, Chen K, Li X et al (2020) Design of high-oleic tobacco (*Nicotiana tabacum* L.) seed oil by CRISPR-Cas9-mediated knockout of NtFAD2–2. *BMC Plant Biol* 20: 233
- Tian Y, Liu X, Fan C et al (2021) Enhancement of Tobacco (*Nicotiana tabacum* L.) seed lipid content for biodiesel production by CRISPR-Cas9-Mediated Knockout of NtAn1. *Front Plant Sci* 11:599474. <https://doi.org/10.3389/fpls.2020.599474>
- Tirani MM, Haghjou MM, Ismaili A (2019) Hydroponic grown tobacco plants respond to zinc oxide nanoparticles and bulk exposures by morphological, physiological and anatomical adjustments. *Funct Plant Biol* 46:360–375
- Toda E, Koiso N, Takebayashi A et al (2019) An efficient DNA- and selectable-marker-free genome-editing system using zygotes in rice. *Nat Plants* 5:363–368. <https://doi.org/10.1038/s41477-019-0386-z>
- Todd AT, Liu EW, Polvi SL et al (2010) A functional genomics screen identifies diverse transcription factors that regulate alkaloid biosynthesis in *Nicotiana benthamiana*. *Plant J* 62(4):589–600

- Tong Z, Jiao, T, Wang F et al (2012a) Mapping of quantitative trait loci conferring resistance to brown spot in flue-cured tobacco (*Nicotiana tabacum* L.) Plant Breed <https://doi.org/10.1111/j.1439-0523.2011.01940.x>
- Tong ZJ, Yang ZM, Chen XJ et al (2012b) 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. <https://doi.org/10.1270/jsbbs.15129>
- Tong Z, Zhou J, Xiu Z (2020) Construction of a high-density genetic map with whole genome sequencing in *Nicotiana tabacum* L. Genomics 112:2028–2033. <https://doi.org/10.1016/j.ygeno.2019.11.015>
- Tong Z, Jiang S, He W et al (2021) Construction of high-density genetic map and QTL mapping in *Nicotiana tabacum* backcrossing BC4F3 population using whole-genome sequencing. Czech J Genet Plant Breed 57:102–112. <https://doi.org/10.17221/8/2021-CJGPB>
- Torney F, Trewyn BG, Lin VSY et al (2007) Mesoporous silica nanoparticles deliver DNA and chemicals into plants. Nat Nanotechnol 2:295–300
- Townsend JA, Wright DA, Winfrey RJ et al (2009) High-frequency modification of plant genes using engineered zinc-finger nucleases. Nature 459:442–445
- Trindade LM, Horvath B, Bachem C et al (2003) Isolation and functional characterization of a stolon specific promoter from potato (*Solanum tuberosum* L.). Gene 303:77–87. [https://doi.org/10.1016/S0378-1119\(02\)01147-2](https://doi.org/10.1016/S0378-1119(02)01147-2)
- Tso TC, Chaplin JF, Adams JD et al (1983) Simple correlation and multiple regression among leaf and smoke characteristics of burley tobaccos. Beitr Tabakforsch International 11(03):141–150
- United Nations (2019) Goal 13: Take urgent action to combat climate change and its impacts. <https://www.un.org/sustainabledevelopment/climate-change/>. Accessed 31 Sept 2021
- Upadhyay SK, Kumar J, Alok A et al (2013) RNA guided genome editing for target gene mutations in wheat. G3 (Bethesda) 3:2233–2238
- Valand GB, Muniyappa V (1992) Epidemiology of *Tobacco Leaf Curl Virus* in India. Ann of App Bio 120:257–267
- Valleau WD (1952) Breeding tobacco for disease resistance. Econ Bot 6:69–102
- Valleau WD, Johnson EM (1927) The effect of a strain of tobacco mosaic on the yield and quality of burley tobacco. Phytopathology 17:523–527
- Valleau WD, Stokes GW, Johnson EM (1960) Nine years experience with *Nicotiana longiflora* factor for resistance to *Phytophthora parasitica* var. *nicotianae* in control of black shank. Tob Sci 4:92–94
- Van Ooijen JW (2006) Joinmap® 4, Software for the calculation of genetic linkage maps in experimental population. Plant Research International, Wageningen
- Van Ooijen JW, Voorrips RE (2001) JoinMap® version 3.0: software for the calculation of genetic linkage maps. Plant Research International, Wageningen
- Varshney RK, Terauchi R, McCouch SR (2014) Harvesting the promising fruits of genomics: applying genome sequencing technologies to crop breeding. PLoS Biol 12(6):e1001883. <https://doi.org/10.1371/journal.pbio.1001883>
- Varshney RK, Bohra A, Yu J et al (2021) Designing future crops: genomics-assisted breeding comes of age. Trends Plant Sci 26(6):631–649. <https://doi.org/10.1016/j.tplants.2021.03.010>
- Verrier JL, Doroszewska T (2018) Tobacco virus collaborative study (1996–2011). VIR-030-CTR Final Report. CORESTA, Paris, France
- Viswanath V, Strauss SH (2010) Modifying plant growth the cisgenic way. Inform Syst Biotechnol News Rep 2010:1–4
- Vontimitta V, Lewis RS (2012a) Growth chamber evaluation of a *Nicotiana tabacum* ‘Beinhart 1000’ × ‘Hicks’ mapping population for QTL affecting resistance to multiple races of *Phytophthora nicotianae*. Crop Sci 52:91–98. <https://doi.org/10.2135/cropsci2011.06.0303>

- Vontimitta V, Lewis RS (2012b) Mapping of quantitative trait loci affecting resistance to *Phytophthora nicotianae* in tobacco (*Nicotiana tabacum*) line Beinhart-1000. *Mol Breed* 29:89–98
- Vontimitta V, Danehower DA, Steede T et al (2010) Analysis of a *Nicotiana tabacum* genomic region controlling two leaf surface chemistry traits. *J Agric Food Chem* 58:294–300
- Wan H, Chen CH, Wu JK (1983) Breeding tobacco resistant to *Cucumber Mosaic Virus* in Taiwan. *Euphytica* 33:17–25
- Wang X, Bennetzen JL (2015) Current status and prospects for the study of *Nicotiana* genomics, genetics, and nicotine biosynthesis genes. *Mol Genet Genomics* 290(1):11–21. <https://doi.org/10.1007/s00438-015-0989-7>
- Wang Y, Sun C, Xu C (2016) Preliminary experiments on nano-silver against *Tobacco Mosaic Virus* and its mechanism. *Tob Sci Technol* 49:22–30
- Wang X, Yang S, Chen Y et al (2018) Comparative genome-wide characterization leading to simple sequence repeat marker development for *Nicotiana*. *BMC Genomics* 19(1):500. <https://doi.org/10.1186/s12864-018-4878-4>
- Wang RW, Cai X, Hu S et al (2020) Comparative analysis of the mitochondrial genomes of *Nicotiana tabacum*: hints toward the key factors closely related to the cytoplasmic male sterility mechanism. *Front Genet* 11:257. <https://doi.org/10.3389/fgene.2020.00257>
- Wang Y, Lv H, Xiang X et al (2021) Construction of a SNP Fingerprinting Database and Population Genetic Analysis of Cigar Tobacco Germplasm Resources in China. *Front Plant Sci* 12:618133. <https://doi.org/10.3389/fpls.2021.618133>
- Wark DC (1970) Development of flue-cured tobacco cultivars resistant to a common strain of blue mold. *Tob Sci* 14:147–150
- Waugh R, Jannik JL, Muehlbauer GJ et al (2009) The emergence of whole genome association scans in barley. *Curr Opin Plant Biol* 12:218–222. <https://doi.org/10.1016/j.pbi.2008.12.007>
- Weil C, Schaart J, Niks R et al (2010) Traditional plant breeding methods. Wageningen UR Plant Breeding report 338, Wageningen
- Wenping LI, Sabaghnia N, Ranjbar R et al (2009) Correlation and path coefficient analysis and euclidean distance clustering for several characters in tobacco germplasm resource. *Chin Tob Sci* 30:59–63
- Whitham S, Dinesh-Kumar SP, Choi D et al (1994) The product of the *Tobacco Mosaic Virus* resistance gene N: similarity to toll and the interleukin-1 receptor. *Cell* 78(6):1101–1115. [https://doi.org/10.1016/0092-8674\(94\)90283-6](https://doi.org/10.1016/0092-8674(94)90283-6)
- Winter P, Kahl G (1995) Molecular marker technologies for plant improvement. *World J Microbiol Biotechnol* 11(4):438–448. <https://doi.org/10.1007/BF00364619>
- Wolf FA, Foster AC (1917) Bacterial leaf spot of tobacco. *Science* 46:361–362
- Wolf FA, Moss EG (1933) Effect of mosaic of flue-cured tobacco on yield and quality. *Phytopathology* 23:834–836
- Woodend JJ, Mudzengerere E (1992) Inheritance of resistance to wildfire and angular leaf spot derived from *Nicotiana rustica* var. *Brasilea*. *Euphytica* 64:149–156
- Worland AJ, Gale MD, Law CN (1987) Wheat genetics. In: Lupton FJH (ed) *Wheat breeding: its scientific principals*. Chapman and Hill, London
- Wu FN, Eannetta NT, Xu YM et al (2009) A COSII genetic map of the pepper genome provides a detailed picture of synteny with tomato and new insights into recent chromosome evolution in the genus *Capsicum*. *Theor Appl Genet* 118:1279–1293
- Wu FN, 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(4):809–27. <https://doi.org/10.1007/s00122-009-1206-z>
- Wu X, Li D, Bao Y (2015) Genetic dissection of disease resistance to the blue mold pathogen, *Peronospora tabacina*, in Tobacco. *Agron* 5:555–568
- Xiang D, Yao Z, Yanhua LY et al (2017) Analysis on solanesol content and genetic diversity of chinese flue-cured tobacco (*Nicotiana tabacum* L.) *Crop Sci* 57(2):847–855. <https://doi.org/10.2135/cropsci2016.07.0568>

- Xiao XW, Chu PWG, Frenkel MJ et al (2000) Antibody-mediated improved resistance to CIYVV and PVY infections in transgenic tobacco plants expressing a single-chain variable region antibody. *Mol Breed* 6:421–431
- Xiao BG, Xu ZL, Chen XJ et al (2006) Genetic linkage map constructed by using a DH population for the flue-cured tobacco. *Acta Tab Sin* 12:35–40
- Xiao B, Drake K, Vontimitta V et al (2013) Location of genomic regions contributing to *Phytophthora nicotianae* resistance in tobacco cultivar Florida 301. *Crop Sci* 53:473–481
- Xiao B, Yuntao TY, Ni LN et al (2015) SNP-based genetic linkage map of tobacco (*Nicotiana tabacum* L.) using next-generation RAD sequencing. *J Biol Res (Thessalon)* 22:11. <https://doi.org/10.1186/s40709-015-0034-3>
- Xu MH, Zheng MH, Liu GT (1998) Genetic diversity in cultivars of *Nicotiana tabacum* revealed by RAPD markers. *J Agri Biotechnol* 6:281–285
- Xu S, Brockmoller T, Navarro-Quezada A et al (2017) Wild tobacco genomes reveal the evolution of nicotine biosynthesis. *Proc Natl Acad Sci USA* 114(23):6133–6138
- Xue M, Pang YH, Wang HT et al (2010) Effects of four host plants on biology and food utilization of the cutworm, *Spodoptera litura*. *J. Insect Sci* 10:14–22. <https://doi.org/10.1673/031.010.2201>
- Yang BC, Xiao BG, Chen XJ et al (2005) Genetic diversity of flue-cured tobacco varieties based on ISSR markers. *Hereditas (Beijing)* 27(5):753–758
- Yang BC, Xiao B, Chen XJ et al (2007) Assessing the genetic diversity of tobacco germplasm using inter simple sequence repeat and inter-retrotransposon amplification polymorphism markers. *Ann Appl Biol* 150:393–401
- Yang S, Li D, Fisher AJ (2019) CORESTA guide no. 16 molecular markers in tobacco. CORESTA, France
- Yao FL, Zheng Y, Huang XY et al (2017) Dynamics of *Bemisia tabaci* biotypes and insecticide resistance in Fujian province in China during 2005–2014. *Sci Rep* 7:40803
- Yi YH, Rufty RC, Wernsman EA (1998a) Identification of RAPD markers linked the wildfire resistance gene of tobacco using bulked segregant analysis. *Tob Sci* 42:52–57
- Yi YH, Rufty RC, Wernsman EA (1998b) Mapping the root-knot nematode resistance gene (Rk) in tobacco with RAPD markers. *Plant Dis* 82(12):1319–1322
- Young ND, Tanksley SD (1989) Restriction fragment length polymorphism maps and the concept of graphical genotypes. *Theor Appl Genet* 77(1):95–101
- Yu YW, Xia YS, Li RH (2014) Association analysis of tobacco specific nitrosamines content in tobacco germplasm. *Acta Tab Sin* 20:48–55
- Yu Y, Yu PC, Chang WJ et al (2020) Plastid transformation: how does it work? Can it be applied to crops? What can it offer? *Int J Mol Sci* 21:4854
- Yuhe S (2012) BAC library construction of Villi tobacco (*Nicotiana tomentosiformis*). *Chin Tob Sci*. Corpus ID: 87250212
- Zeba N, Isbat M (2011) Multivariate analysis for yield and yield contributing traits in F0 and F1 generations in tobacco (*Nicotiana tabacum*). *J Exp Biosci* 2(1):101–106
- Zenkner FF, Margis-Pinheiro M, Cagliari A (2019) Nicotine biosynthesis in *Nicotiana*: a metabolic overview. *Tob Sci* 56(1):1–9. <https://doi.org/10.3381/18-063>
- Zeven AC, Knott DR, Johnson R (1983) Investigation of linkage drag in near isogenic lines of wheat by testing for seedling reaction to races of stem rust, leaf rust and yellow rust. *Euphytica* 32:319–327
- Zhang DM (1994) Study on surface structure scanning of some tobacco seeds with electrical-glass observing. *Chin Tob Sin* 2:2–15
- Zhang HY, Liu XZ, He CS et al (2005) Random amplified DNA polymorphism of *Nicotiana tabacum* L. cultivars. *Biol Plant* 49:605–607
- Zhang HY, Liu XZ, He CS et al (2008a) Genetic diversity among flue-cured tobacco cultivars based on RAPD and AFLP markers. *Braz Arch Biol Technol* 51(6):1097–1101. <https://doi.org/10.1590/S1516-89132008000600003>
- Zhang HY, Yang YM, Li FS et al (2008b) Screening and characterization a RAPD marker of Tobacco brown-spot resistant gene. *African J Biotechnol* 7(15):2559–2561

- Zhang JS, Wang RG, Yang CY et al (2012a) Genetic diversity of agronomic traits and association analysis with SRAP markers in flue-cured tobacco (*Nicotiana tabacum*) varieties from China and abroad. *Acta Agron Sin* 38:1029–1041
- Zhang S, Gao M, Zaitlin D (2012b) Molecular linkage mapping and marker-trait associations with NIRPT, a downy mildew resistance gene in *Nicotiana langsdorffii*. *Front Plant-Microbe Interact* 3:185. <https://doi.org/10.3389/fpls.2012.00185>
- Zhang Y, Zhang F, Li X et al (2013) Transcription activator-like effector nucleases enable efficient plant genome engineering. *Plant Physiol* 161(1):20–27. <https://doi.org/10.1104/pp.112.205179>
- Zhang Y, Guo X, Yan X et al (2018) Identification of stably expressed QTL for resistance to black shank disease in tobacco (*Nicotiana tabacum* L.) line Beinhart 1000–1. *The Crop J* 6(3):282–290
- Zhu C, Gore M, Buckler ES, Yu J (2008) Status and prospects of association mapping in plants. *Plant Genome* 1:1–20. <https://doi.org/10.3835/plantgenome2008.06.0007ed>
- Zhu JQ, Liu S, Ma Y (2012) Improvement of pest resistance in transgenic tobacco plants expressing dsRNA of an insect-associated Gene *EcR*. *PLoS One* 7(6):e38572. <https://doi.org/10.1371/journal.pone.0038572>
- Zimmerman JL, Goldberg RB (1977) DNA sequence organization in tobacco. *Chromosoma* 59:227–252
- Xu Y (2010) *Molecular plant breeding*. CABI International, Wallingford, Oxfordshire