Mol Genet Genomics (2015) 290:11–21
DOI 10.1007/s00438-015-0989-7

REVIEW

Current status and prospects for the study of *Nicotiana* genomics, genetics, and nicotine biosynthesis genes

Xuewen Wang · Jeffrey L. Bennetzen

Received: 5 April 2014 / Accepted: 5 January 2015 / Published online: 13 January 2015
© Springer-Verlag Berlin Heidelberg 2015

Abstract *Nicotiana*, a member of the *Solanaceae* family, is one of the most important research model plants, and of high agricultural and economic value worldwide. To better understand the substantial and rapid research progress with *Nicotiana* in recent years, its genomics, genetics, and nicotine gene studies are summarized, with useful web links. Several important genetic maps, including a high-density map of *N. tabacum* consisting of ~2,000 markers published in 2012, provide tools for genetics research. Four whole genome sequences are from allotetraploid species, including *N. benthamiana* in 2012, and three *N. tabacum* cultivars (TN90, K326, and BX) in 2014. Three whole genome sequences are from diploids, including progenitors *N. sylvestris* and *N. tomentosiformis* in 2013 and *N. otophora* in 2014. These and additional studies provide numerous insights into genome evolution after polyploidization, including changes in gene composition and transcriptome expression in *N. tabacum*. The major genes involved in the nicotine biosynthetic pathway have been identified and the genetic basis of the differences in nicotine levels among *Nicotiana* species has been revealed. In addition, other progress on chloroplast, mitochondrial, and NCBI-registered projects on *Nicotiana* are discussed. The challenges and prospects for genomic, genetic and application research are addressed. Hence, this review provides important resources and guidance for current and future research and application in *Nicotiana*.

Keywords Gene expression · Genome sequencing · Polyploidy · Nicotine biosynthesis · Tobacco · Genetic map

Communicated by J. Graw.

Electronic supplementary material The online version of this article (doi:10.1007/s00438-015-0989-7) contains supplementary material, which is available to authorized users.

X. Wang · J. L. Bennetzen
Germplasm Bank of Wild Species in Southwest China, Kunming Institute of Botany, Chinese Academy of Sciences, 132 Lanhei Road, Kunming 650201, People's Republic of China

X. Wang (✉)
China Tobacco Gene Research Center, Zhengzhou Tobacco Research Institute of CNTC, No. 2 Fengyang Street, High-Tech Zone, Zhengzhou 450001, People's Republic of China
e-mail: wangxuewen@mail.kib.ac.cn

The distribution and taxonomy of the genus *Nicotiana*

Species within the genus *Nicotiana* are annuals or perennials with tubular or trumpet-shaped flowers. Named after the sixteenth century French diplomat Jean Nicot, *Nicotiana* is the fifth largest genus of the Solanaceae family. The Solanaceae include important crops such as *Solanum tuberosum* (potato) and *Solanum lycopersicum* (tomato). An early taxonomy of the genus *Nicotiana* used morphological, cytological, bio-geographical and crossing data to investigate 60 species (Goodspeed 1954). This taxonomy was revised in 2004 (Knapp et al. 2004) to incorporate phylogenetic information from both chloroplast and nuclear DNAs, and to conform to the rules laid out in the Code of Botanical Nomenclature (Greuter 2000). Under the new classification, the genus *Nicotiana* was described as containing over 75 species for which the phylogenetic relationships are well documented. Some 75 % of the species are native to America and ~25 % to Australia (Chase et al. 2003; Clarkson et al. 2004). The genus *Nicotiana* is of scientific and economic importance (Ow et al. 1986). *N. benthamiana*, a species native to Australia, is used extensively as a model to study plant–pathogen interactions. *N. tabacum* L. (common

tobacco) and *N. rustica* L, two species native to America, are cultivated worldwide by the tobacco industry. Several other species, such as *N. alata* and *N. sylvestris,* are grown as ornamentals.

Genome evolution

Species in the genus *Nicotiana* have a basic haploid chromosome number of $n = 12$, although some species vary from this norm (e.g., *N. alata* with $n = 9$). Some 40 % of *Nicotiana* species are allopolyploids, all of which have been generated independently in six polyploidy events that occurred between ~10 million years ago (mya) and ~0.2 mya (Clarkson et al. 2004; Leitch et al. 2008). Some allopolyploids, such as species in *Nicotiana* section *Suaveolentes,* were formed >10 mya while others such as *N. tabacum* were formed in the last 200,000 years. The taxonomic relationship between the diploid genome donors that make up various allopolyploids varies widely, with some being closely related and others belonging to distantly related taxonomic sections (Clarkson et al. 2004; Leitch et al. 2008). These features make *Nicotiana* genomes ideal models to study the genome evolution of polyploids.

In most angiosperms, there is a trend for genomes to downsize following polyploidization (Leitch and Bennett 2004). Studies in maize, *Arabidopsis*, wheat, and *Tragopogon* have shown that gene content also changes after polyploidy formation, with frequent loss of some of the genes that were duplicated by the polyploidization (Ilic et al. 2003; Leitch and Bennett 2004; Freeling 2009; Buggs et al. 2012; Feldman et al. 2012). Usually, genes from one of the two progenitor species are more commonly lost than from the other (Edger and Pires 2009; Feldman et al. 2012), and this bias is also accompanied by a preferential retention of genes that encode proteins involved in multi-protein complexes that are dosage sensitive (Edger and Pires 2009). In *Nicotiana,* the allotetraploid *N. tabacum* genome size is estimated to be ~4.5 Gb, while its diploid ancestors *N. sylvestris* and *N. tomentosiformis* have genome sizes of ~2.7 and ~2.4 Gb, respectively (Leitch et al. 2008; Sierro et al. 2013). *N. tabacum* was estimated to have lost between 3.7 and 8 % of the repetitive DNA from its progenitors' since polyploidization (Leitch et al. 2008; Renny-Byfield et al. 2011; Sierro et al. 2014). Analysis of repeated DNA sequences in synthetic *N. tabacum*, generated by hybridization of the diploid species *N. sylvestris* (female) with *N. tomentosiformis* (male), revealed that the paternal, *tomentosiformis* (T) genome, was being rapidly eliminated (Skalická et al. 2005; Renny-Byfield et al. 2012; Sierro et al. 2014). This preferential elimination of the T genome also occurred in natural *N. tabacum* (Renny-Byfield et al. 2011). Whole genomic sequence comparison showed that the genome of *N. sylvestris* and *N. tomentosiformis* contributes 53 and 47 %, respectively, for that of *N. tabacum*, which indicates a larger biased genome reduction in T genome (Sierro et al. 2014). Surprisingly, none of the recent whole genome sequencing studies investigated gene loss or biased gene loss from the tetraploids, so this deserves a high priority once full genome assemblies are generated. By contrast, the maternal *sylvestris* (S) genome, underwent dynamic sequence amplification and/or homogenization in natural *N. tabacum* compared with that in the ancestral S genome (Leitch et al. 2008; Renny-Byfield et al. 2011). All natural *N. tabacum* showed translocation between the S and T genome which was considered to restore the fertility and reduce negative effects of the maternal cytoplasm on an alien paternal genome (Lim et al. 2004).

Studies on many plant species have revealed biased or differential homoeologous gene expression in polyploids (Chen and Ni 2006). In allotetraploid *Gossypium* (cotton), ~70 % of homoeologs were found to have biased expressed compared with those in the progenitor A and D genomes (Flagel et al. 2008; Chaudhary et al. 2009). In hexaploid wheat, *Triticum aestivum,* most genes investigated (79 out of 90) were expressed from all three ancestor genomes, while a few genes (11 out of 90) were silenced in one of these three genomes (Mochida et al. 2004). Recent full genome expression studies revealed that most homoeologous genes exhibit asymmetric expression in hexaploid wheat (Pfeifer et al. 2014). In tetraploid *N. tabacum*, comparison of transcriptomic expression in leaves revealed that only ~33 % of transcripts (968 homoeologous gene clusters) were parental genome distinguishable. However, expression was detected at only one of the homoeologous copies for ~90 % of those genes, while ~15 % of the genes that had both copies still expressed showed subfunctionalization (Bombarely et al. 2012a). It is interesting that such a rapid loss or change in homoeologous gene copy expression is observed in *N. tabacum*. It should be noted, however, that earlier studies of single genes in *Nicotiana* indicated that most show additive expression of parental homoeologous genes (Matassi et al. 1991; Kronenberger et al. 1993; Riechers and Timko 1999; Schenke et al. 2003), so the apparent disparity between these results and those of Bombarely and coworkers (Bombarely et al. 2012a) deserves further investigation.

In *N. repandae*, a polyploidy formed ~4.5 mya, the transposable element (TE) insertion sites are mostly different from those present in its diploid progenitor species, indicating that TEs played a major role in the differentiation of this genome that has occurred during post-polyploid evolution (Parisod et al. 2012). Other repetitive sequences, such as satellite repeats, were also found to be quite dynamic (Koukalova et al. 2010). Some polyploids exhibit genome size increases relative to their diploid ancestors, likely an

outcome of TE amplification. Although polyploidy has not been seen to be a cause of TE activation or amplification in plants, the general processes of genome size increase in angiosperms are so frequent that any lineage can exhibit these changes in very short time frame regardless of its ploidy history (Ewing and Green 1998; Estep et al. 2013). The increases or decreases in genome size do not correlate with the age of the polyploidization event. Among nine measured genome sizes of polyploid species of *Nicotiana*, four showed downsized genomes while five had increased genome sizes compared to the summed sizes of their proposed genome progenitors (Leitch et al. 2008). For example, polyploid *N. repanda* showed an increased genome size of ~28.6 % while polyploid *N. nudicaulis* exhibited a genome size decreased by ~14.3 %, even though these two species have the same two progenitors (*N. sylvestris* and *N. obtusifolia)* (Leitch et al. 2008). Recent next-generation sequence data analysis revealed that three out of four investigated allopolyploids showed increased genome sizes relative to the sum of the two diploid progenitors that they all shared (Renny-Byfield et al. 2013).

In the genus *Nicotiana*, *N. tabacum, N. rustica* and *N. benthamiana* are the best-studied species. The first two are cultivated for the production of tobacco and hence are of agricultural importance. They are easily transformable; therefore, they represent important research models for the characterization of gene function (Senthil-Kumar and Mysore 2011). Both *N. tabacum* and *N. rustica* are allotetraploid species ($2n = 48$) with approximately the same genome sizes and dates of origin but with different parental genomes (Goodspeed 1954; Leitch et al. 2008). *N. tabacum* originated from a hybridization event between diploid ancestor species *N. sylvestris* ($2n = 24$) and *N. tomentosiformis* ($2n = 24$) as the maternal and paternal parents, respectively. Recent whole genome sequence analysis revealed that the genome composition of these two ancestors evolved quite differently after their divergence ~15 mya. *N. tomentosiformis* has many more TEs of the Copia-type. Furthermore, *N. tomentosiformis* has less gene-rich DNA (425 Mb) than *N. sylvestris* (625 Mb) (Sierro et al. 2013), indicating fewer insertions of TEs into gene-containing regions in *N. sylvestris*. The genome of *N. rustica* is derived from the paternal U genome of *N. undulata* ($2n = 24$) and the maternal P genome of a species belonging to *Nicotiana* section *Paniculatae* ($2n = 24$) (Goodspeed 1954; Chase et al. 2003; Clarkson et al. 2004; Lim et al. 2004; Yukawa et al. 2006; Leitch et al. 2008). *N. benthamiana* is an important model for research on plant–pathogen interactions. This species is also allotetraploid ($2n = 38$; $1C = 3.5$ Gb) and arose >10 mya from a hybridization event between species belonging to *N. Sect. Sylvestres* ($2n = 24$) and *N. Sect. Noctiflorae* ($2n = 24$) (Goodspeed 1954; Goodin et al. 2008; Leitch et al. 2008). The genome characteristics and available web links for the sequences of different *Nicotiana* species are provided in Table 1.

Nicotiana genome sequence resources

Nicotiana species represent economically important crops as well as model systems for plant research; therefore, both the tobacco industry and academic researchers have had a long-standing interest in understanding the structure and function of the tobacco genome. The first tobacco genome sequencing project was initiated in 2003 by the Tobacco Genome Initiative (TGI) with the goal of sequencing the open reading frames of *N. tabacum* L. cultivar Hicks Broadleaf. The project was completed in 2007 and the sequences are available from the National Center for Biotechnology Information (NCBI) GenBank (links given in Table 1). Around 689 Mb of genomic sequence was produced by Sanger sequencing and assembled into 81,959 contigs with an average length of 1.2 kb and 871,255 singletons with an average length of 688 bp. However, the obtained sequences covered only a small portion of the tobacco genome, partly because the sequencing technology (methylation filtration) that was employed enriched for genes, and only for the portions of genes that are under methylated relative to TEs.

With advances in next-generation sequencing (NGS) technologies, great progress has been made recently in decoding the entire tobacco genome. Seven complete genome sequences from *Nicotiana* were released between 2012 and 2014. These sequences are derived from three diploids (*N. otophora*, *N. sylvestris* and *N. tomentosiformis*) and four allotetraploids (*N. benthamiana* plus three cultivars of *N. tabacum:* TN90, K326, and *Basma Xanthi* (BX)) (Bombarely et al. 2012b, 2013, 2014).

The *N. benthamiana* genome was sequenced using an Illumina HiSeq-2000 platform. A draft assembly was released in 2012 for online analysis and downloading (Sol Genomics Network: http://solgenomics.net, Table 1). The draft genome assembly, generated from a 63-fold depth of sequence data, consisted of 141,000 scaffolds, totaling ~2.9 Gb of an estimated total genome size of 3.5 Gb, with a scaffold N50 length of 89 kb. Approximately 90 % of the available ~16,000 unigenes of *Nicotiana* in the Sol Genomics Network can be mapped to the assembly (Bombarely et al. 2012b). However, there was no gene prediction in the draft genome when it was released. A later RNA-sequencing study helped fill in this gap and revealed that there were ~119,000 unigenes present in *N. benthamiana* (accessible at http://www.benthgenome.com) (Nakasugi et al. 2013).

Full genome sequencing of three diploids, *N. sylvestris*, *N. tomentosiformis,* and *N. otophora,* was accomplished recently and the data are available to the public at GenBank (links given in Table 1). The whole genome sequences of

Table 1 Genomic information for *Nicotiana*

Species	Chromosomes	Genome type	C value (pg)	Estimated genome size	Sequence size	Sequencing technology	Organization/web link
N. tomentosiformis	$2n = 24$	T genome	2.7	~2.4 Gb (Leitch et al. 2008)	2.36 Gb complete genome (Sierro et al. 2013)	Illumina HiSeq	GenBank accession: ASAF00000000.1 www.ncbi.nlm.nih.gov/nuccore/ASAG00000000
N. sylvestris	$2n = 24$	S genome	2.7	~2.4 Gb (Leitch et al. 2008)	2.68 Gb Complete genome (Sierro et al. 2013)	Illumina HiSeq	GenBank accession: ASAG00000000.1 http://www.ncbi.nlm.nih.gov/nuccore/ASAF00000000
N. tabacum	$2n = 48$	S and T genomes	5.2	~4.5 Gb (Leitch et al. 2008)	partial genome	Sanger sequencing	http://www.ncbi.nlm.nih.gov/bioproject/PRJNA29349
	$2n = 48$ (var. TN90, Burley)	S and T genomes	5.2	~4.4 Gb (Sierro et al. 2014)	3.7 Gb, genome coverage 84.3 %	Illumina HiSeq	http://www.ncbi.nlm.nih.gov/nuccore/AYMY00000000.1/
	$2n = 48$ (var. K326, Flue-cured)	S and T genome	5.2	~4.6 Gb (Sierro et al. 2014)	3.7 Gb, genome coverage 81.1 %	Illumina HiSeq	http://www.ncbi.nlm.nih.gov/nuccore/AWOJ00000000.1/
	$2n = 48$ (var. Basma Xanthi, Oriental)	S and T genomes	5.2	~4.5 Gb (Sierro et al. 2014)	3.7 Gb, genome coverage 81.8 %	Illumina HiSeq	http://www.ncbi.nlm.nih.gov/nuccore/AWOK00000000.1/
N. benthamiana	$2n = 38$	B and S genomes	3.2	~3.1 Gb (Narayan 1987; Bennett and Leitch 1995)	2.899 Gb (V0.3) Complete genome	Illumina HiSeq	http://solgenomics.net http://www.ncbi.nlm.nih.gov/nuccore/CBMM000000000.1/
N. otophora	$2n = 24$			2.7 Gb (Sierro et al. 2014)	2.5 Gb, complete genome	Illumina HiSeq	http://www.ncbi.nlm.nih.gov/nuccore/AWOL000000000.1/
N. rustica	$2n = 48$	U and P genomes	5.3	~4.5 Gb (Leitch et al. 2008)	None	–	–

Table 2 Other available *Nicotiana* sequence resources

Specie name	Platform	NCBI accession no. or data set	Sequence size
Synthetic *N. tabacum*	Illumina Genome Analyzer IIx	SRX096296	2.7 Gb
N. sylvestris	Illumina Genome Analyzer IIx	SRX096295	1.5 Gb
N. tomentosiformis	Illumina Genome Analyzer IIx	SRX096294	1.6 Gb
N. tabacum mitochondrial	Sanger sequencing	BA000042	430,597 bp (complete genome)
N. tabacum chloroplast	Sanger sequencing	Z00044	155,943 bp (complete genome)
N. sylvestris chloroplast	Sanger sequencing	AB237912	155,941 bp (complete genome)
N. tomentosiformis chloroplast	Sanger sequencing	AB240139	155,745 bp (complete genome)
N. tabacum EST	–	430,842 records at NCBI[a]	–
N. benthamiana EST	–	56,180 records at NCBI[b]	–
N. sylvestris EST	–	8,583 records at NCBI[b]	–

[a] The summary was calculated at NCBI on April 2nd 2014

N. sylvestris and *N. tomentosiformis* were released in 2013, about a half year after the release of the *N. benthamiana* genome. Genome sequences at 94× and 146× coverage were used for the genome assembly, and the N50 lengths of scaffolds were 79 and 82 kb, covering 82.9 and 71.6 % of the genome, in *N. sylvestris* and *N. tomentosiformis,* respectively. Most DNA sequences (72–75 %) in both species are repeat elements. An RNA-sequencing study revealed that there were around 44,000–53,000 transcripts in the diploids (Sierro et al. 2013). About 2.5 Gb of assembled sequences of *N. otophora* from 66x sequence data, yielding an N50 contig length of 11 kb, were released in 2014 (Sierro et al. 2014).

Most recently, the full genome sequences of three agriculturally important *N. tabacum* cultivars were released (Sierro et al. 2014). Currently, these genomes are only assembled at the contig level. The genome sizes were estimated to be 4.4 Gb for *N. tabacum* TN90, 4.6 Gb for *N. tabacum* K326, and 4.6 Gb for *N. tabacum* BX. About 3.7 Gb of assembled sequences are available for each variety at GenBank (links available in Table 1). Genome sequences at 38x, 49x and 29x coverage were used for the contig assembly, and the N50 lengths of the contigs were 39.4, 24.7, and 21.7 kb, covering ~81, ~84, and ~73 % of the genomes of *N. tabacum* TN90, K326, and BX, respectively. Transcriptome analysis of nine tissues of *N. tabacum* identified 134,694–188,510 transcripts and the gene number was estimated to be >93,000 in each cultivar. Gene ontology analysis predicted roles for >50,000 potential proteins (Sierro et al. 2014).

The complete mitochondrial and chloroplast genome sequences have been available for quite a while for several *Nicotiana* species (Table 2). The mitochondrial genome (mtDNA) of *N. tabacum* is 430,597 bp. It encodes 36 proteins, 3 ribosomal RNAs and 21 tRNAs (Sugiyama et al. 2005). The chloroplast genome of *N. tabacum* is 155,943 bp. A comparison of the chloroplast sequences of *N. sylvestris*, *N. tomentosiformis* and *N. tabacum* revealed that *N. sylvestris* was the maternal genome donor to *N. tabacum* (Yukawa et al. 2006).

In addition to genome assemblies, more than 8.8×10^9 genomic and transcriptomic sequence reads from *Nicotiana* species, including *N. benthamiana*, *N. tabacum*, *N. sylvestris, N. tomentosiformis* and *N. attenuata,* were generated on NGS platforms and were deposited in the Short Read Archive (SRA) at NCBI (Supplemental File 1). NCBI also contains 430,842 EST records as of April 2nd 2014, most generated from *N. tabacum* (77.8 %) and the remainder from *N. sylvestris* and *N. tomentosiformis* (Table 2). We know that many more transcriptome reads will be added to current data sets because many ongoing projects were registered with NCBI (http://www.ncbi.nlm.nih.gov/bioproject/?term=nicotiana).

Nicotiana genetics

Genetic marker discovery is considered a first step for many genomic applications. Genetic markers have been used in the construction of genetic maps, for varietal genotyping, genome comparisons, gene mapping, quantitative trait locus (QTL) analysis and marker-assisted breeding (Davey et al. 2011; Bennetzen et al. 2012; Mauro-Herrera et al. 2013). In the era of cost-efficient sequencing, genetic maps and markers remain important tools to validate sequence assemblies (Bennetzen et al. 2012). Currently, all of the de novo-assembled tobacco genomes mentioned above lack comprehensive validation by detailed genetic mapping, so this step in genome assembly assessment and enrichment deserves a high priority.

Significant progress has been made in the development of genetic markers and genetic map construction in tobacco over the past 2–3 years. Two sets of simple sequence repeat (SSR) markers, 5,119 named PT markers and 4,886 named

TM or TME markers, were developed from the tobacco genomic DNA or EST sequences by Bindler et al. and by Tong and coworkers, respectively (Bindler et al. 2011; Tong et al. 2012b). For the genomic markers, Tong et al. mainly focused on tri-nucleotide SSR repeats, while Bindler and colleagues used dinucleotide SSR repeats for marker development. Hence, there was little overlap between the two sets of SSRs (Tong et al. 2012b). Although the PT markers were developed from genomic sequences and verified experimentally, there was no predicted PCR product found when use of these primers was computer simulated on tobacco EST templates (Tong et al. 2012b). This result is probably caused by a lack of sufficient depth in EST sequences that would provide matches to the genomic sequences during a computational search.

There are currently two genetic maps available for cultivated tetraploid tobacco (*N. tabacum L.*) and another two maps for wild diploid *Nicotiana* species. Both diploid maps yielded 12 linkage groups and spanned ~1,000 cM. The *N. tomentosiformis* map was constructed with a mixture of 489 SSR and cleaved amplified polymorphic sequence (CAPS) markers, while the *N. acuminata* (closely related to *N. sylvestris*) map contained a mixture of 308 SSR and CAPS markers (Wu et al. 2010). The average distance between adjacent markers was in the range of 2.2–3.4 cM, so the maps were relatively low density for this ~2.4 Gb genome. The first genetic map for tetraploid tobacco, published in 2011, was generated with an F2 mapping population derived from a cross between *N. tabacum* cv. Hicks Broadleaf and cv. Red Russian. This map consisted of 2,318 SSR markers, 282 of which had previously been included in a preliminary map (Bindler et al. 2007). The more detailed map yielded 24 linkage groups and covered 3,270 cM (Bindler et al. 2011). The length of individual linkage groups in this high-density map varied from 86 to 199 cM, and the average genetic distance between adjacent markers was 1.4 cM. However, despite the high marker density, some gaps, one of which spanned ~16 cM, remained in the map. A second genetic map of tetraploid tobacco was constructed from a cross between two tobacco varieties, Honghua Dajinyuan and Hicks Broadleaf. This map comprised 611 SSR marker loci (590 SSR markers) distributed over 24 linkage groups and covering 1,882 cM. The number of markers per linkage group varied from 13 to 58 and the size of individual linkage groups ranged from 28 to 120 cM. The average marker distance genome-wide was 3.1 cM (Tong et al. 2012b). This map contained some regions of up to 20 cM that were devoid of markers. The two tetraploid genetic maps shared 218 markers and a common parent; therefore, it may be possible to integrate both maps to generate a consensus map. The genetic markers and maps mentioned above provide important tools for basic and applied research in *Nicotiana*, but maps with even greater marker density and more easily assayed markers, like single nucleotide polymorphisms (SNPs) are needed.

Genes associated with the nicotine pathway

Species within the genus *Nicotiana* contain a high level of nicotine, with up to 90–95 % of the total alkaloid content playing a role in the defense against insect herbivores (Steppuhn et al. 2004; Siminszky et al. 2005). The biosynthesis of nicotine involves pyrrolidine ring formation, pyridine ring formation, and the coupling of both rings. The enzymes and corresponding genes involved in the major steps of the nicotine synthetic pathway in tobacco have been well characterized. Recent whole genome sequencing provided full sequences of genes involved in the synthetic pathway (Sierro et al. 2014). Arginine decarboxylase (*NtADC*), ornithine decarboxylase (*NtODC*), putrescine N-methyltransferase (*NtPMT*) and N-methylputrescine oxidase (*NtMPO*) are involved in the formation of the pyrrolidine ring (Fig. 1) (Hibi et al. 1994; Imanishi et al. 1998; Bortolotti et al. 2004; Xu et al. 2004; Heim et al. 2007; Katoh et al. 2007). Aspartate oxidase (*AO*), quinolinate synthase (*QS*) and quinolinic acid phosphoribosyltransferase (*NtQPT*) are responsible for the biosynthesis of the pyridine ring (Sinclair et al. 2000; Katoh et al. 2006). These three enzymes are also involved in the early steps of nicotinic acid dinucleotide (NAD) biosynthesis (Fig. 1). A PIP (pinoresinol-lariciresinol reductase, isoflavone reductase, and phenylcoumaran benzylic ether reductase) family isoflavone reductase-like protein A622 and berberine bridge enzyme-like protein (*NtBBL*) are required for nicotine ring coupling (Fig. 1) (DeBoer et al. 2009; Kajikawa et al. 2009, 2011). However, it remains unclear how the two rings are joined to form nicotine. The nicotine synthesis pathway has been described in detail in recent reviews (Shoji and Hashimoto 2011; Dewey and Xie 2013).

Nicotine is synthesized in the roots of *Nicotiana* plants and then transferred into leaves by a multidrug and toxic compound extrusion (MATE) transporter (Baldwin 1999; Morita et al. 2009). Recent work has shown that nicotine uptake permease (NUP1) affects nicotine metabolism, localization and root growth but not the transport of nicotine. *NUP1* transcripts accumulate in root tips, and reduced mRNA levels for this gene correlate with reduced nicotine levels in the leaves of *NUP1-RNAi* plants (Hildreth et al. 2011). Data from a range of studies revealed that the complex regulation of nicotine biosynthesis involves hormones, transcription factors (TF), protein kinases and other proteins. Jasmonic acid (JA) is a positive regulator of nicotine biosynthesis while ethylene and auxin are negative regulators. Abscisic acid may regulate nicotine biosynthesis both negatively and positively though how

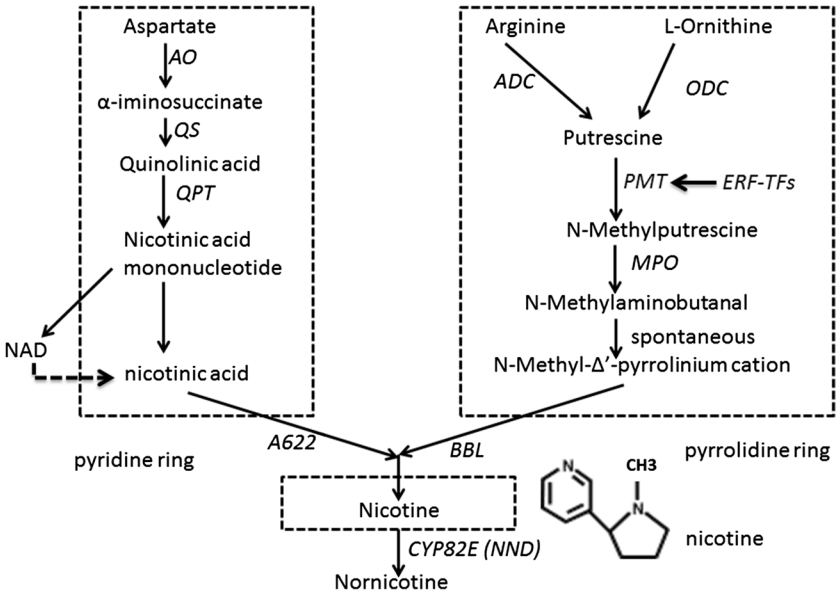

Fig. 1 Schematic showing the enzymes known to be involved in the synthesis of nicotine and nornicotine *ADC* arginine decarboxylase, *ODC* ornithine decarboxylase, *PMT* putrescine N-methyltransferase, *MPO* N-methylputrescine oxidase, *AO* aspartate oxidase, *QS* quinolinate synthase, *QPT* quinolinic acid phosphoribosyl transferase, *ERF-TFs* ERF transcription factors, *BBL* berberine bridge enzyme-like, *NND* nicotine N-demethylase

this occurs is not well understood (Shoji et al. 2000, 2010; De Sutter et al. 2005; Shi et al. 2006; Todd et al. 2010; Lackman et al. 2011; Shoji and Hashimoto 2011, 2014). In response to JA, the JA receptor coronatine insensitive 1 (NtCOI1) mediates degradation of the jasmonate ZIM domain (NtJAZs) protein, a transcriptional repressor, resulting in inducible expression of *PMT* (Shoji et al. 2008). A series of ethylene response factor (ERF) TFs that regulate nicotine levels were identified in tobacco (Shoji et al. 2010; Todd et al. 2010). The deletion of a cluster of seven TF genes was found to be the cause of the low nicotine levels in the *nic2* mutant (Shoji et al. 2010). Abscisic acid increases expression of a subset of some ERF genes (Shoji and Hashimoto 2014). In *N. benthamiana,* the basic helix-loop-helix TFs bHLH1 and bHLH2, which are homologs of MYC2 (a Z-box binding factor), bind to the *PMT* promoter to provide positive regulation of nicotine synthesis (Todd et al. 2010). An auxin response factor, ARF1, also plays a role in regulating nicotine content (Todd et al. 2010). As an added layer of complexity, regulation of nicotine synthesis is affected by abiotic and biotic factors. Agricultural topping, suckering and herbivory all increase nicotine accumulation in tobacco leaves (Baldwin 1988; Wang et al. 2008). Nicotine biosynthesis can be diverted to the alkaloids anabasine and anatabine, which share part of the nicotine biosynthetic pathway. Another alkaloid, nornicotine, is directly converted from nicotine by nicotine N-demethylase (NND) or CYP82E (Chintapakorn and Hamill 2003; Siminszky et al. 2005; Lewis et al. 2010; Shoji and Hashimoto 2011). A mutation in CYP82E was found to cause a lower conversion rate of nicotine to nornicotine in *N. tabacum* compared to wild type (Chakrabarti et al. 2007).

With the availability of entire *Nicotiana* genome sequences in 2013, the mysterious origin of dramatic difference in nicotine contents was uncovered in comparative analyses between *N. sylvestris* and *N. tomentosiformis*. Nicotine and nornicotine are the predominant alkaloids in *N. sylvestris* and *N. tomentosiformis,* respectively. Gene expression and genomic studies found that the high level of nicotine in *N. sylvestris* is caused by up-regulated expression of key genes, including *AO*, *QS*, *QPT* and *PMT*, in the nicotine synthesis pathway when compared with *N. tomentosiformis.* In addition, the multiple copies of the *PMT* genes may contribute to the higher levels of nicotine. There are three *NtPMT* genes (*NtPMT-1*, *NtPMT-3* and *NtPMT-4*) in *N. sylvestris* while there is only *NtPMT-2* in *N. tomentosiformis* (Sierro et al. 2013). Nornicotine is converted from nicotine by NND during leaf senescence and curing. Nornicotine is a potent carcinogen, so understanding and minimizing its content is a high priority in the tobacco industry. Ortholog searches revealed that three highly expressed NND genes called *CYP82E3*, *CYP82E4*, and *CYP82E5,* in *N. tomentosiformis*, are responsible for its high level of nornicotine (Sierro et al. 2013).

Prospects

Future genomic research

With the advanced NGS sequencing and assembly technologies, it is feasible to sequence and fully assemble the entire *Nicotiana* genome. It is expected that fully assembled genome sequences of numerous *Nicotiana* species will become available in the near future. The genomes of four

other species belonging to the family *Solanaceae*, potato (*Solanum tuberosum*), tomato (*Solanum lycopersicum*), pepper (*Capsicum annuum*) and eggplant (*Solanum melongena*), have recently been sequenced (The Potato Genome Sequencing Consortium 2011; The Tomato Genome Consortium 2012; Hirakawa et al. 2014; Kim et al. 2014). Although a syntenic comparison of those genomes was reported (Sierro et al. 2014), a deep comparison between these genomes will provide very meaningful information on the evolutionary events that have shaped the genomes and the very different properties of their leaves, roots and fruit.

Currently, the genomic sequences of diploid *Nicotiana* ancestors have been assembled to scaffold level, while tetraploid genomic sequences have been assembled at the contig level only. Knowledge of the structure of the diploid parental genomes will facilitate the full assembly and sequence analysis of the commercial allotetraploid, *N. tabacum*. Sequence comparison of the diploids with that of the tetraploids will also enhance our understanding of *Nicotiana* genome evolution. More than 90 % of SSR genetic markers mapped in *N. tabacum* were amplified in only one of the two ancient S and T genomes, indicating substantial sequence divergence between the S and T genomes after polyploidization (Bindler et al. 2011). What are the roles of homoeologous genes in polyploid *N. tabacum*? Further research into the genome sequences and transcriptomes for both the tetraploid and the diploid progenitor genomes will provide answers to these questions. The genome sequences of *Nicotiana* will also help to address other scientific questions related to tobacco performance, quality, durability and physiology, and this will assist with practical breeding applications in tobacco and related crops.

Future genetic and breeding research

Most marker studies published to date have been carried out in cultivated varieties with closely related genetic backgrounds (Bindler et al. 2011; Tong et al. 2012b). For future marker development and breeding studies, some different wild species, especially those in different taxonomic sections within the *Nicotiana* genus, should be included in these studies to provide more polymorphism and to broaden the germplasm base. To date, polymorphism levels often have been too low to conduct detailed genotyping that can differentiate between varieties within the same taxonomic section. For example, only 4 % (227 out of 5,718) of markers were polymorphic between the cultivated varieties Changbohuang and Jinyehuang (Tong et al. 2012a).

SNPs are highly abundant and distributed throughout a genome, and are therefore highly valued as molecular markers. Many SNP detection systems are high throughput, thus providing high genotyping efficiency, a high genome coverage and high accuracy. SNPs have therefore become the markers of choice for many genomic and genetic applications. With the advances of NGS-associated technologies, it should be easy to obtain SNPs through full genome sequencing ("re-sequencing"), genotyping-by-sequencing (GBS), or reduced complexity sequencing from the cultivated species at reduced cost, although the current cost is still relatively high (Davey et al. 2011). For future marker-assisted breeding applications, SNP-based genome-scale application will be the trend, especially in tetraploid tobacco research. The available genome sequences of *N. sylvestris* and *N. tomentosiformis*, *N. otophora*, *N. tabacum, and N. benthamiana* will provide informative sequences for genetic marker development. It should be possible to breed a super-tobacco plant with a cluster of economically important traits and minimized negative characteristics.

Dissection of the genetic basis of some complex traits will be the important research in *Nicotiana* for the future. Identification of optimal alleles of QTL or major genes underlying key traits like alkaloid production will be very useful for breeding. Linkage-based QTL mapping, genome-wide association analysis (GWAS), and nested association mapping (NAM) are all important tools that need to be applied to *Nicotiana* to dissect important physiological, genetic and agronomic characteristics. GWAS is a powerful approach to examine whether a genetic variant is associated to a particular trait in a broad germplasm sampling, assessed at a whole genome scale. GWAS typically focuses on association between particular SNPs and particular traits. GWAS has been very successful in finding association between SNPs and disease traits in human since its first report for human age-related macular degeneration in 2005 (Klein et al. 2005). An online database called NHGRI GWAS Catalog (http://www.genome.gov/gwastudies/) has been curated with >14,000 SNP–trait associations in humans, published in 2002 articles, as of Oct 22nd 2014 (Welter et al. 2014). GWAS has also proven to be a very powerful tool in plant studies on such species as maize, rice and *Arabidopsis* and has been summarized in a recent review (Huang and Han 2014). For GWAS in *Nicotiana*, the first step should be to gather comprehensive SNP data from a broad germplasm panel, and an appropriate diversity panel can then be selected for comprehensive phenotyping. Especially for some rare alleles association, NAM is a powerful joint linkage-association approach. Like the NAM populations in maize, a NAM population needs to be created for *Nicotiana* by crossing the appropriate germplasm, as identified in the GWAS research (Kump et al. 2011; Poland et al. 2011). Better genetic maps, GWAS and NAM will all help to capture and mobilize the use of the *Nicotiana* genetic diversity for mapping and employing superior agronomic traits.

Research on tobacco as a model plant

The *Nicotiana* genus has long been used as a research model plant. The genus is particularly useful for transgenic studies (because of its ease of transformation), for the study of polyploidy, and for investigation of secondary product biosynthesis. Availability of entire genome sequences and transcriptomic profiles will benefit and accelerate a wide range of scientific research in *Nicotiana*. One particularly exciting area of study is in plant–microbe interactions, including research on how orthologs of resistance genes and other immunity-associated genes from different species function in *N. benthamiana* (Bombarely et al. 2012b). Since *Nicotiana* species are some of the best plants for transient transgene expression in leaf via simple *Agrobacterium tumefaciens* infiltration, better knowledge of the *Nicotiana* genome will provide more raw material for studying the function of any transformed gene. Another example is heavy metal transport, an important research topic in plants, *Nicotiana* species display a high capacity to accumulate heavy metals in their leaves. Some genes responsible for cadmium/zinc accumulation in *Nicotiana* were identified in the genome sequences of *N. sylvestris* and *N. tomentosiformis* (Sierro et al. 2013). Further research can be done based on these newly identified genes to understand heavy metal transport and accumulation in *Nicotiana*, and the results of this research will provide useful clues with regard to metal accumulation in other crops.

Conclusion

Nicotiana is an important commercial crop and model plant. Excellent progress has been made in recent years in the generation of whole genome sequence data, understanding genome evolution, investigating gene expression, and unraveling the genetics and genes related to nicotine biosynthesis. The genetic and genomic advances reviewed and future prospect discussed here will guide both basic and applied research in tobacco for years to come. Progress in tobacco will also facilitate genetic, physiological, developmental and evolutionary studies in other plants, especially crops that belong to the same taxonomic family, such as potato and tomato.

Acknowledgments The authors thank Prof. Katrien M. Devos of the University of Georgia, USA for her efforts, suggestions, and assistance in editing during manuscript preparation.

References

Baldwin IT (1988) The alkaloidal responses of wild tobacco to real and simulated herbivory. Oecologia 77:378–381

Baldwin IT (1999) Inducible nicotine production in native *Nicotiana* as an example of phenotypic plasticity. J Chem Ecol 25:3–30

Bennett MD, Leitch IJ (1995) Nuclear DNA amounts in Angiosperms. Ann Botany 76:113–176

Bennetzen JL, Schmutz J, Wang H, Percifield R, Hawkins J, Pontaroli AC, Estep M, Feng L, Vaughn JN, Grimwood J, Jenkins J, Barry K, Lindquist E, Hellsten U, Deshpande S, Wang X, Wu X, Mitros T, Triplett J, Yang X, Ye C-Y, Mauro-Herrera M, Wang L, Li P, Sharma M, Sharma R, Ronald PC, Panaud O, Kellogg EA, Brutnell TP, Doust AN, Tuskan GA, Rokhsar D, Devos KM (2012) Reference genome sequence of the model plant *Setaria*. Nat Biotech 30:555–561

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

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

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

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

Bortolotti C, Cordeiro A, Alcazar R, Borrell A, Culianez-Macia FA, Tiburcio AF, Altabella T (2004) Localization of arginine decarboxylase in tobacco plants. Physiol Plant 120:84–92

Buggs RJA, Renny-Byfield S, Chester M, Jordon-Thaden IE, Viccini LF, Chamala S, Leitch AR, Schnable PS, Barbazuk WB, Soltis PS, Soltis DE (2012) Next-generation sequencing and genome evolution in allopolyploids. Am J Bot 99:372–382

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

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

Chaudhary B, Flagel L, Stupar RM, Udall JA, Verma N, Springer NM, Wendel JF (2009) Reciprocal silencing, transcriptional bias and functional divergence of homeologs in polyploid cotton (*Gossypium*). Genetics 182:503–517

Chen Z, Ni Z (2006) Mechanisms of genomic rearrangements and gene expression changes in plant polyploids. BioEssays 28:240–252

Chintapakorn Y, Hamill JD (2003) Antisense-mediated down-regulation of putrescine N-methyltransferase activity in transgenic *Nicotiana tabacum* L. can lead to elevated levels of anatabine at the expense of nicotine. Plant Mol Biol 53:87–105

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

Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat Rev Genet 12:499–510

De Sutter V, Vanderhaeghen R, Tilleman S, Lammertyn F, Vanhoutte I, Karimi M, Inze D, Goossens A, Hilson P (2005) Exploration of jasmonate signalling via automated and standardized transient expression assays in tobacco cells. Plant J 44:1065–1076

DeBoer K, Lye J, Aitken C, Su AK, Hamill J (2009) The A622 gene in *Nicotiana glauca* (tree tobacco): evidence for a functional role in pyridine alkaloid synthesis. Plant Mol Biol 69:299–312

Dewey RE, Xie J (2013) Molecular genetics of alkaloid biosynthesis in *Nicotiana tabacum*. Phytochemistry 94:10–27

Edger P, Pires JC (2009) Gene and genome duplications: the impact of dosage-sensitivity on the fate of nuclear genes. Chromosome Res 17:699–717

Estep MC, DeBarry JD, Bennetzen JL (2013) The dynamics of LTR retrotransposon accumulation across 25 million years of panicoid grass evolution. Heredity 110:194–204

Ewing B, Green P (1998) Base-calling of automated sequencer traces using Phred. II. error probabilities. Genome Res 8:186–194

Feldman M, Levi A, Fahima T, Korol A (2012) Genomic asymmetry in allopolyploid plants: wheat as a model. J Exp Bot 63:5045–5059

Flagel L, Udall J, Nettleton D, Wendel J (2008) Duplicate gene expression in allopolyploid *Gossypium* reveals two temporally distinct phases of expression evolution. BMC Biol 6:16

Freeling M (2009) Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. Annu Rev Plant Biol 60:433–453

Goodin MM, Zaitlin D, Naidu RA, Lommel SA (2008) *Nicotiana benthamiana*: its history and future as a model for plant–pathogen interactions. MPMI 21:1015–1026

Goodspeed TH (1954) The genus *nicotiana*: Origins, relationships and evolution of its species in the light of their distribution, morphology and cytogenetics. Chronica Botanica Company, Waltham

Greuter W (2000) International Code of Botanical Nomenclature: Saint Louis Code : adopted by the Sixteenth International Botanical Congress. Koeltz Scientific Books, St Louis

Heim WG, Sykes KA, Hildreth SB, Sun J, Lu R-H, Jelesko JG (2007) Cloning and characterization of a *Nicotiana tabacum* methylputrescine oxidase transcript. Phytochemistry 68:454–463

Hibi N, Higashiguchi S, Hashimoto T, Yamada Y (1994) Gene expression in tobacco low-nicotine mutants. Plant cell 6:723–735

Hildreth SB, Gehman EA, Yang H, Lu R-H, Ritesh KC, Harich KC, Yu S, Lin J, Sandoe JL, Okumoto S, Murphy AS, Jelesko JG (2011) Tobacco nicotine uptake permease (NUP1) affects alkaloid metabolism. PNAS 108:18179–18184

Hirakawa H, Shirasawa K, Miyatake K, Nunome T, Negoro S, Ohyama A, Yamaguchi H, Sato S, Isobe S, Tabata S, Fukuoka H (2014) Draft genome sequence of eggplant (*Solanum melongena* L.): the representative solanum species indigenous to the old world. DNA Res

Huang X, Han B (2014) Natural variations and Genome-Wide Association Studies in crop plants. Annu Rev Plant Biol 65:531–551

Ilic K, SanMiguel PJ, Bennetzen JL (2003) A complex history of rearrangement in an orthologous region of the maize, sorghum, and rice genomes. PNAS 100:12265–12270

Imanishi S, Hashizume K, Nakakita M, Kojima H, Matsubayashi Y, Hashimoto T, Sakagami Y, Yamada Y, Nakamura K (1998) Differential induction by methyl jasmonate of genes encoding ornithine decarboxylase and other enzymes involved in nicotine biosynthesis in tobacco cell cultures. Plant Mol Biol 38:1101–1111

Kajikawa M, Hirai N, Hashimoto T (2009) A PIP-family protein is required for biosynthesis of tobacco alkaloids. Plant Mol Biol 69:287–298

Kajikawa M, Shoji T, Kato A, Hashimoto T (2011) Vacuole-localized berberine bridge enzyme-like proteins are required for a late step of nicotine biosynthesis in tobacco. Plant Physiol 155:2010–2022

Katoh A, Uenohara K, Akita M, Hashimoto T (2006) Early steps in the biosynthesis of NAD in *Arabidopsis* start with Aspartate and occur in the plastid. Plant Physiol 141:851–857

Katoh A, Shoji T, Hashimoto T (2007) Molecular cloning of N-methylputrescine oxidase from tobacco. Plant Cell Physiol 48:550–554

Kim S, Park M, Yeom S-I, Kim Y-M, Lee JM, Lee H-A, Seo E, Choi J, Cheong K, Kim K-T, Jung K, Lee G-W, Oh S-K, Bae C, Kim S-B, Lee H-Y, Kim S-Y, Kim M-S, Kang B-C, Jo YD, Yang H-B, Jeong H-J, Kang W-H, Kwon J-K, Shin C, Lim JY, Park JH, Huh JH, Kim J-S, Kim B-D, Cohen O, Paran I, Suh MC, Lee SB, Kim Y-K, Shin Y, Noh S-J, Park J, Seo YS, Kwon S-Y, Kim HA, Park JM, Kim H-J, Choi S-B, Bosland PW, Reeves G, Jo S-H, Lee B-W, Cho H-T, Choi H-S, Lee M-S, Yu Y, Do Choi Y, Park B-S, van Deynze A, Ashrafi H, Hill T, Kim WT, Pai H-S, Ahn HK, Yeam I, Giovannoni JJ, Rose JKC, Sorensen I, Lee S-J, Kim RW, Choi I-Y, Choi B-S, Lim J-S, Lee Y-H, Choi D (2014) Genome sequence of the hot pepper provides insights into the evolution of pungency in *Capsicum* species. Nat Genet 46:270–278

Klein RJ, Zeiss C, Chew EY, Tsai J-Y, Sackler RS, Haynes C, Henning AK, SanGiovanni JP, Mane SM, Mayne ST, Bracken MB, Ferris FL, Ott J, Barnstable C, Hoh J (2005) Complement factor H polymorphism in age-related macular degeneration. Science 308:385–389

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

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

Kronenberger J, Lepingle A, Caboche M, Vaucheret H (1993) Cloning and expression of distinct nitrite reductases in tobacco leaves and roots. Mol Gen Genet 236:203–208

Kump KL, Bradbury PJ, Wisser RJ, Buckler ES, Belcher AR, Oropeza-Rosas MA, Zwonitzer JC, Kresovich S, McMullen MD, Ware D, Balint-Kurti PJ, Holland JB (2011) Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population. Nature Genet 43:163–168

Lackman P, Gonzalez-Guzman M, Tilleman S, Carqueijeiro I, Perez AC, Moses T, Seo M, Kanno Y, Hakkinen ST, Van Montagu MC, Thevelein JM, Maaheimo H, Oksman-Caldentey KM, Rodriguez PL, Rischer H, Goossens A (2011) Jasmonate signaling involves the abscisic acid receptor PYL4 to regulate metabolic reprogramming in *Arabidopsis* and tobacco. PNAS 108:5891–5896

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

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

Lewis RS, Bowen SW, Keogh MR, Dewey RE (2010) Three nicotine demethylase genes mediate nornicotine biosynthesis in *Nicotiana tabacum* L.: functional characterization of the CYP82E10 gene. Phytochemistry 71:1988–1998

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

Matassi G, Melis R, Macaya G, Bernardi G (1991) Compositional bimodality of the nuclear genome of tobacco. Nucleic Acids Res 19:5561–5567

Mauro-Herrera M, Wang X, Barbier H, Brutnell TP, Devos KM, Doust AN (2013) Genetic control and comparative genomic analysis of flowering time in Setaria (Poaceae). G3(3):283–295

Mochida K, Yamazaki Y, Ogihara Y (2004) Discrimination of homoeologous gene expression in hexaploid wheat by SNP analysis of contigs grouped from a large number of expressed sequence tags. Mol Genet Genomics 270:371–377

Morita M, Shitan N, Sawada K, Van Montagu MCE, Inze D, Rischer H, Goossens A, Oksman-Caldentey KM, Moriyama Y, Yazaki K (2009) Vacuolar transport of nicotine is mediated by a multidrug and toxic compound extrusion (MATE) transporter in *Nicotiana tabacum*. PNAS 106:2447–2452

Nakasugi K, Crowhurst RN, Bally J, Wood CC, Hellens RP, Waterhouse PM (2013) *De Novo* transcriptome sequence assembly and analysis of RNA silencing genes of *Nicotiana benthamiana*. PLoS ONE 8:e59534

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

Ow DW, De Wet JR, Helinski DR, Howell SH, Wood KV, Deluca M (1986) Transient and stable expression of the firefly Luciferase gene in plant cells and transgenic plants. Science 234:856–859

Parisod C, Mhiri C, Lim KY, Clarkson JJ, Chase MW, Leitch AR, Grandbastien MA (2012) Differential dynamics of transposable elements during long-term diploidization of *Nicotiana* Section *Repandae* (Solanaceae) allopolyploid genomes. PLoS ONE 7:e50352

Pfeifer M, Kugler KG, Sandve SR, Zhan B, Rudi H, Hvidsten TR, Consortium IWGS, Mayer KFX, Olsen O-A (2014) Genome interplay in the grain transcriptome of hexaploid bread wheat. Science 345

Poland JA, Bradbury PJ, Buckler ES, Nelson RJ, Goodman MM (2011) Genome-wide nested association mapping of quantitative resistance to northern leaf blight in maize. In. National Academy of Sciences, p 6893

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

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

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

Riechers DE, Timko MP (1999) Structure and expression of the gene family encoding putrescine N-methyltransferase in *Nicotiana tabacum*: new clues to the evolutionary origin of cultivated tobacco. Plant Mol Biol 41:387–401

Schenke D, Sasabe M, Toyoda K, Inagaki YS, Shiraishi T, Ichinose Y (2003) Genomic structure of the *NtPDR1* gene, harboring the two miniature inverted-repeat transposable elements, *NtToya1* and *NtStowaway101*. Genes Genet Syst 78:409–418

Senthil-Kumar M, Mysore KS (2011) New dimensions for VIGS in plant functional genomics. Trends Plant Sci 16:656–665

Shi Q, Li C, Zhang F (2006) Nicotine synthesis in *Nicotiana tabacum* L. induced by mechanical wounding is regulated by auxin. J Exp Bot 57:2899–2907

Shoji T, Hashimoto T (2011) Nicotine biosynthesis In: Plant metabolism and biotechnology. John Wiley & Sons, Ltd, pp 191–216

Shoji T, Hashimoto T (2014) Stress-induced expression of NICOTINE2-locus genes and their homologs encoding Ethylene Response Factor transcription factors in tobacco. Phytochemistry. doi:10.1016/j.phytochem.2014.05.017

Shoji T, Nakajima K, Hashimoto T (2000) Ethylene suppresses Jasmonate-induced gene expression in nicotine biosynthesis. Plant Cell Physiol 41:1072–1076

Shoji T, Ogawa T, Hashimoto T (2008) Jasmonate-induced nicotine formation in tobacco is mediated by tobacco COI1 and JAZ genes. Plant Cell Physiol 49:1003–1012

Shoji T, Kajikawa M, Hashimoto T (2010) Clustered transcription factor genes regulate nicotine biosynthesis in tobacco. Plant cell 22:3390–3409

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

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

Siminszky B, Gavilano L, Bowen SW, Dewey RE (2005) Conversion of nicotine to nornicotine in *Nicotiana tabacum* is mediated by CYP82E4, a cytochrome P450 monooxygenase. PNAS 102:14919–14924

Sinclair SJ, Murphy KJ, Birch CD, Hamill JD (2000) Molecular characterization of quinolinate phosphoribosyltransferase (QPRTase) in *Nicotiana*. Plant Mol Biol 44:603–617

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

Steppuhn A, Gase K, Krock B, Halitschke R, Baldwin IT (2004) Nicotine's defensive function in nature. PLoS Biol 2:1074–1080

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

The Potato Genome Sequencing Consortium (2011) Genome sequence and analysis of the tuber crop potato. Nature 475:189–195

The Tomato Genome Consortium (2012) The tomato genome sequence provides insights into fleshy fruit evolution. Nature 485:635–641

Todd AT, Liu E, Polvi SL, Pammett RT, Page JE (2010) A functional genomics screen identifies diverse transcription factors that regulate alkaloid biosynthesis in *Nicotiana benthamiana*. Plant J 62:589–600

Tong Z, Jiao T, Wang F, Li M, Leng X, Gao Y, Li Y, Xiao B, Wu W (2012a) Mapping of quantitative trait loci conferring resistance to brown spot in flue-cured tobacco (*Nicotiana tabacum* L.). Plant Breeding 131:335–339

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

Wang SS, Shi QM, Li WQ, Niu JF, Li CJ, Zhang FS (2008) Nicotine concentration in leaves of flue-cured tobacco plants as affected by removal of the shoot apex and lateral buds. J Integr Plant Biol 50:958–964

Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, Klemm A, Flicek P, Manolio T, Hindorff L, Parkinson H (2014) The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res 42:D1001–D1006

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

Xu B, Sheehan MJ, Timko MP (2004) Differential induction of ornithine decarboxylase (ODC) gene family members in transgenic tobacco (*Nicotiana tabacum* L. cv. bright yellow 2) cell suspensions by methyl-jasmonate treatment. Plant Growth Regul 44:101–116

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