

# A high density genetic map of tobacco (*Nicotiana tabacum* L.) obtained from large scale microsatellite marker development

Gregor Bindler · Jörg Plieske · Nicolas Bakaher ·  
Irfan Gunduz · Nikolai Ivanov · Rutger Van der Hoeven ·  
Martin Ganal · Paolo Donini

Received: 14 January 2011 / Accepted: 15 March 2011 / Published online: 2 April 2011  
© The Author(s) 2011. This article is published with open access at Springerlink.com

**Abstract** Tobacco (*Nicotiana tabacum* L.) is a species in the large family of the Solanaceae and is important as an agronomic crop and as a model system in plant biotechnology. Despite its importance, only limited molecular marker resources are available that can be used for genome analysis, genetic mapping and breeding. We report here on the development and characterization of 5,119 new and functional microsatellite markers and on the generation of a high-resolution genetic map for the tetraploid tobacco genome. The genetic map was generated using an F2 mapping population derived from the intervarietal cross of Hicks Broadleaf × Red Russian and merges the polymorphic markers from this new set with those from a smaller set previously used to produce a lower density map. The genetic map described here contains 2,317 microsatellite markers and 2,363 loci, resulting in an average distance between mapped microsatellite markers which is less than 2 million base pairs or 1.5 cM. With this new and expanded

marker resource, a sufficient number of markers are now available for multiple applications ranging from tobacco breeding to comparative genome analysis. The genetic map of tobacco is now comparable in marker density and resolution with the best characterized genomes of the Solanaceae: tomato and potato.

## Introduction

Tobacco (*Nicotiana tabacum* L.) is an important crop plant (Davis and Nielsen 1999) and a member of the nightshade (Solanaceae) family which is one of the largest and most diverse within the angiosperms. This family harbors 3,000–4,000 species (Olmstead et al. 2008; <http://www.nhm.ac.uk/research-curation/research/projects/solanaceae/source/solanaceae>), of which a considerable number are of major economic importance as crop, vegetable or ornamental species throughout the world such as potato (*Solanum tuberosum*), tomato (*Solanum lycopersicum*), eggplant (*Solanum melonena*), pepper (*Capsicum* species) and Petunia (*Petunia × hybrida*). Due to this and the fact that members of the Solanaceae are also important model species for systematics, plant biology, genetics and biotechnology, a current objective is to elucidate the genome structure of its most important species (Mueller et al. 2005; <http://solgenomics.net/solanaceae-project>) through various genome sequencing initiatives and comparative genome analyses.

Currently, the best characterized species within the Solanaceae are tomato and potato for which a large amount of information concerning their genome structure have been generated. This includes high density genetic and physical maps that contain several thousand anchored markers (Tanksley et al. 1992; van Os et al. 2006; <http://solgenomics.net>). At present, genome sequencing efforts for

Communicated by R. Visser.

**Electronic supplementary material** The online version of this article (doi:10.1007/s00122-011-1578-8) contains supplementary material, which is available to authorized users.

G. Bindler · N. Bakaher · I. Gunduz · N. Ivanov · R. Van der Hoeven · P. Donini (✉)  
Applied Research Department, Philip Morris International,  
Quai Jeanrenaud 5, 2000 Neuchatel, Switzerland  
e-mail: donini.paolo@gmail.com

J. Plieske · M. Ganal  
TraitGenetics GmbH, Am Schwabeplan 1b,  
06466 Gatersleben, Germany

*Present Address:*  
R. Van der Hoeven  
KWS SAAT AG, P.O. Box 1463, 37555 Einbeck, Germany

tomato and potato are approaching completion, with draft sequences being available (<http://solgenomics.net>; <http://www.potatogenome.net>). To obtain a more detailed picture of genome evolution in the Solanaceae, genetic maps of a number of other important species (e.g., pepper, eggplant) have been generated, and these were aligned through the mapping of conserved markers such as RFLP markers and conserved ortholog sequences (COS) markers (Fulton et al. 2002; Wu et al. 2006) to the genome of tomato as a reference. These data have indicated that at the chromosomal level, the Solanaceae genomes have undergone a number of inversions and translocations during speciation (Tanksley et al. 1992; Wu et al. 2009a, b), while at the gene sequence and gene order level, it could be shown through comparative BAC clone sequencing conducted in potato, eggplant, pepper, and petunia that most genes are conserved in their exon/intron structure and sequence order (Wang et al. 2008).

Beside the use of molecular markers for the characterization of genome structure and organization, they are also important tools for the development of new varieties through marker assisted breeding (Xu 2010). Since the level of polymorphism in elite material is frequently low, it is important to have access to a large number of markers (usually several thousand) so that in situations where a low level of polymorphism exists between closely related lines/varieties, sufficient polymorphic markers are still available for the mapping of traits that are either qualitatively or quantitatively inherited as well as for marker-assisted selection or marker-assisted backcrossing. While in many diploid species (especially with a high level of polymorphism) there is now a tendency to use single nucleotide polymorphism (SNP) for such applications (Ganal et al. 2009), microsatellite markers are still the object of active marker development efforts in other species (Ren et al. 2009). In polyploid species and in species with a low level of polymorphism such as tobacco, the current markers of choice are still microsatellites or simple sequence repeat (SSR) markers. SSR markers detect a higher level of polymorphism than other marker systems and they can be easily scored even in polyploid species (Röder et al. 1998). Scoring of SNP markers can be technically challenging in allopolyploid and autopolyploid species.

Tobacco is an allopolyploid species ( $n = 24$ ) and shares its basic chromosome number of  $n = 12$  with many other solanaceous species such as tomato, potato, pepper and eggplant. The species is most likely the result of a tetraploidization event (Lim et al. 2004; Clarkson et al. 2005) involving *Nicotiana sylvestris* (S-genome) and a species closely related to modern day *Nicotiana tomentosiformis* (T-genome). With both genomes together, tobacco is at the high end of genome sizes (4.5 Gbp) in the Solanaceae (Arumuganathan and Earle 1991) and contains a large pro-

portion of repetitive sequences (Kenton et al. 1993; Zimmerman and Goldberg 1977). Until recently, little information was available in terms of genetic mapping and molecular marker development in the *Nicotiana* species (Lin et al. 2001; Rossi et al. 2001; Suen et al. 1997). This changed with the generation of 637 microsatellite markers and a first version of a genetic map covering 1,920 cM and 282 markers (Bindler et al. 2007). However, this map was not able to fully cover the tetraploid tobacco genome and the number of markers, while sufficient for diversity analyses (Moon et al. 2008), was still insufficient for many breeding purposes, although it was used for locating traits responsible for the formation of certain tobacco leaf surface exudates (Vontimitta et al. 2010). Recently a further study was published (Wu et al. 2009c), concerning the comparative mapping of the diploid ancestor of the T-genome (*N. tomentosiformis*) and a related species to the S-genome (*N. acuminata*) using 262 and 133 COS markers, respectively, together with a set of microsatellite markers that were mapped in these two species. The results from these mapping studies revealed that the tetraploid tobacco genome has undergone a number of chromosomal rearrangements compared to these diploid genomes. Furthermore, in the same study, with the COS markers that had also been mapped in the tomato genome, it could be shown that a number of reciprocal translocations and inversions (>10) differentiate the ancestral tobacco genomes from the tomato genome.

With respect to tobacco genome sequencing efforts, the tobacco genome initiative (TGI) (Gadani et al. 2003) aimed at sequencing of more than 90% of the *N. tabacum* genomic sequences containing open reading frames using methyl filtration technology (GeneThresher<sup>®</sup>), targeting unmethylated gene-rich regions. More than 1.3 million genome survey sequences (GSS) are now available in GenBank or through the public website <http://www.pngg.org/tgi/>. The goals of the research reported here, were to exploit the TGI sequences to generate a sufficient number of microsatellite markers for saturating the tobacco genome and to produce a high density, high resolution genetic map that could be used for tobacco breeding purposes and for further analysis of the tetraploid tobacco genome.

## Materials and methods

### DNA isolation and plant material

DNA was extracted from freeze-dried leaves of plants that were grown in the greenhouse as previously described by Bindler et al. (2007). Functionality tests of the developed markers and an initial analysis of polymorphism were performed using a panel of 16 tobacco varieties representing

**Table 1** List of accessions tested with 5,500 microsatellite primer pairs

Hicks Broad Leaf MP	FC
Red Russian MP	SP
TW142 <i>N. tomentosiformis</i>	A
TW138 <i>N. sylvestris</i>	A
Prilep12-2/1	OR
Basma Xanthi	OR
Izmir-Incekara TI1674	OR
TN 90	BU
Banket A1	BU
Kutsaga 35	FC
K326	FC
DacMataFina	DT
Amarillo Parado	DT
Criollo Misionero	DT
AA 37	DT
TN86/Big Cuban	BU/CW

All accessions were obtained from the tobacco germplasm collection of Philip Morris International

FC Flue-cured, SP speciality tobacco, A ancestral genome, OR Oriental tobacco, BU Burley tobacco, DT Dark tobacco, CW cigar wrapper, TN86 used for testing with the first 2,100 primer pairs, Big Cuban used for testing with the remaining 3,400 primer pairs

the main types of tobacco (Flue-cured, Burley, Oriental, Dark, and others) plus single accessions representing *N. sylvestris* and *N. tomentosiformis* (Table 1). Seeds for this material were obtained from the germplasm collection at Philip Morris International. The F2 population consisted of the same 186 individual plants as published by Bindler et al. (2007) and was originally provided by Dr. Ramsey Lewis (North Carolina State University).

#### Processing of TGI sequence data and primer generation

A total of 1,379,067 raw sequences and 55,411 expressed sequence tags (ESTs) obtained from the TGI (<http://www.pngg.org/tgi>) were surveyed for the presence of microsatellite sequences and further processed for primer development for 5,500 candidate SSR loci. The sequences and the corresponding quality files were initially prepared by screening for vector and *Escherichia coli* contamination and by removing redundant sequences using the cross\_match© program (<http://www.phrap.org/>). Sequences containing microsatellite motifs with at least eight repeats and a minimum match of 90% to the respective microsatellite motif were identified using Tandem Repeat Finder (Benson 1999) and BLAST (Altschul et al. 1990) searches. Oligonucleotide primer pairs flanking the respective microsatellite sequence were designed using the Primer 3.0 program (Rozen and Skaletsky 2000). Non-redundant primers

were selected to be approximately 20 bases long, to have a GC-content between 20 and 80%, and to have a melting temperature between 57 and 63°C (optimum 60°C). A second elimination step of potentially remaining duplicated sequences was performed by checking primer sequences, microsatellite motifs, and the cross\_match© scores.

#### Primer testing

Forward primers were labeled with FAM, HEX or ROX for fragment analysis on Applied Biosystems 3100 Genetic Analyzers. The respective fluorescent dyes were selected according to the expected size of the PCR fragment of the microsatellite marker. The FAM dye was used for fragments between approximately 90 and 170 bp, the ROX dye was used for fragments between approximately 170 and 207 bp and the HEX dye for fragments longer than 207 bp. Testing of primer pairs for functionality and linkage mapping were performed in PCR volumes of 10 µl with an annealing temperature of 55°C during 45 PCR cycles according to Bindler et al. (2007).

#### Mapping

Two different software packages were used for mapping of the SSRs: Map Manager QTXb20 (Manly et al. 2001, <http://www.mapmanager.org>) and JoinMap 3.0 (Plant Research International BV, Wageningen, Netherlands). MapManager was used for the main mapping procedure (settings: linkage evaluation F2 intercross, search linkage criterion  $P = 0.05$ , map function Kosambi, cross type line cross). The map position of some dominantly scored markers generated in the previous analysis and the phenotypic trait flower colour scored in the previous analysis (Bindler et al. 2007) were optimized manually taking into consideration the number of crossovers (as low as possible) and the length of the linkage group (as short as possible). Grouping of markers and traits as well as their segregation patterns were verified using the JoinMap® 3.0 program (Van Ooijen and Voorrips 2001) with the following settings: used linkages with REC smaller than 0.400 and LOD larger than 1.00, threshold for removal of loci with respect to jumps in goodness-of-fit 5.000, Kosambi mapping function. The final map was drawn using JoinMap.

#### Results

##### Microsatellite marker development from tobacco genomic sequences

A total of 1,379,067 raw sequences from the TGI and 55411 EST were analyzed for the presence of microsatellite

**Table 2** Functionality and number of detected loci for the investigated 5,500 tobacco microsatellite markers

	Functionality and loci detected per marker	Percent	Mapped loci	Percent
1 Locus	3,789	68.9	1,533	75.3
2 Loci	1,111	20.2	435	21.4
3 Loci	143	2.6	54	2.7
4 Loci	36	0.7	12	0.6
Multiple loci (<7)	40	0.7	2	0.1
Non functional	381	6.9	0	0
	5,500	100	2,036	100

motifs using the set of bioinformatics tools described in “Materials and methods”. After an initial screen of 288 primer pairs equally derived from ESTs and genomic sequences, EST derived microsatellites were excluded from further marker development due to the fact that in the ESTs, most of the identified microsatellite markers contained predominantly short repeats (<10 repeat motifs of mainly trinucleotide repeat-types) with a lower than expected level of polymorphism of the trinucleotide repeats between the parents of the mapping population (11 vs. 36% for the dinucleotide motifs from the genomic sequences). Thus, the developed markers were predominantly derived from single or low copy genomic sequences or introns. The 5,500 markers with the best primer design scores were selected for subsequent marker testing. Approximately 60% of the microsatellite markers contained predominantly TA repeats, 22% contained predominantly GA repeats, 14% contained predominantly GT repeats. The remaining 4% harbored mixed repeat types that could not be assigned clearly to any of the above types. 88% of the markers contained 8 (minimum) to 25 repeating units while the others were mostly larger and/or contained a mixture of repeat types.

#### Testing microsatellite markers for functionality

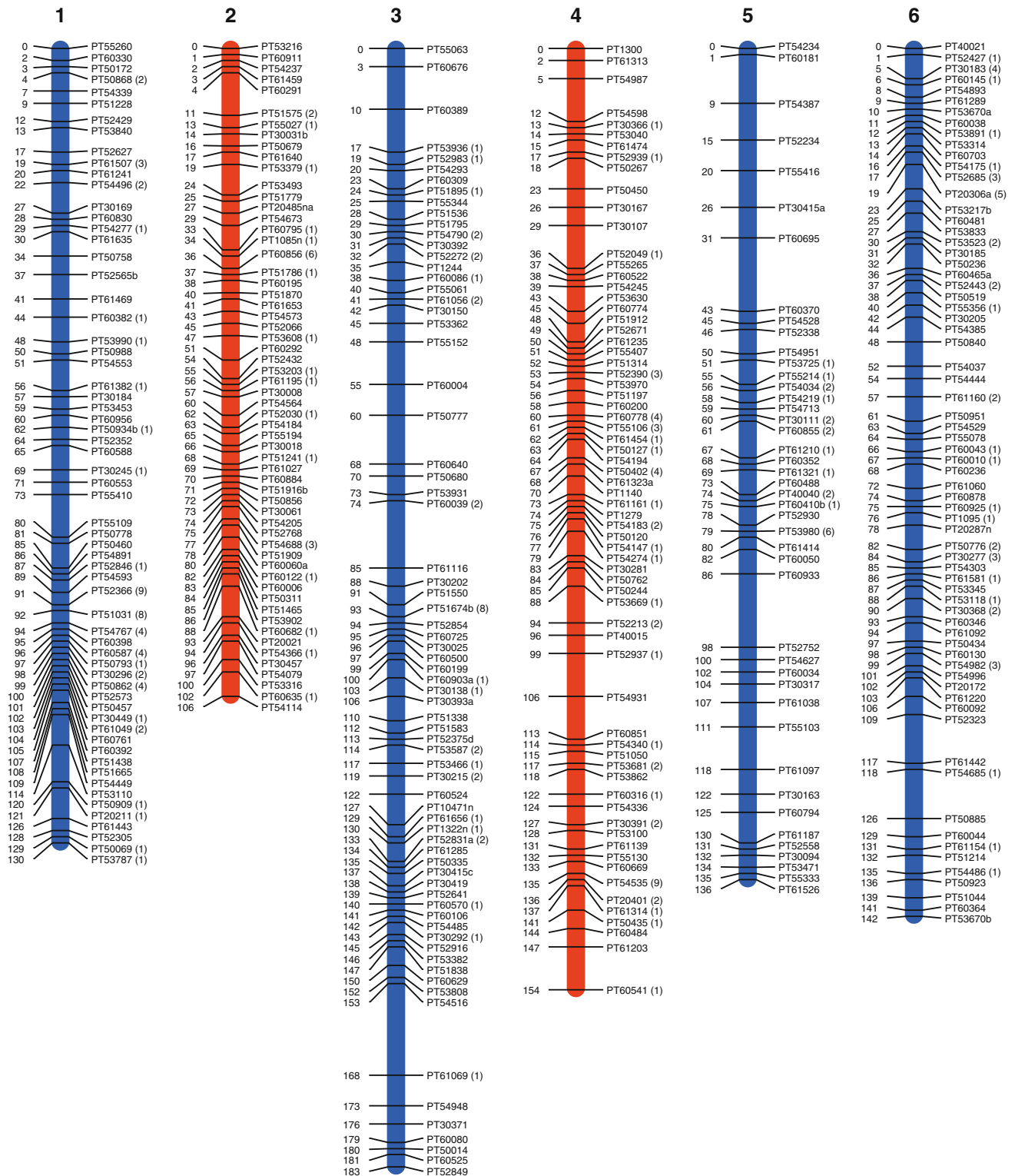
In an initial analysis, 5,500 primer pairs were evaluated for functionality on a test panel consisting of 16 different tobacco lines or varieties (Table 1). The panel included the mapping parents of the Hicks Broadleaf × Red Russian population and for the assignment of markers to the different genomes, representative accessions of *Nicotiana tomentosiformis* and *Nicotiana sylvestris*. From all 5,500 primer pairs tested, more than 93% (5,119) were functional under our analytical conditions. Functional markers were defined by no more than six amplified fragments with at least one fragment in the size range defined by the sequence used for primer design. Based on the results of the primer pair tests with the 16 sample panel, the markers were classified according to their number of loci (1, 2, 3, 4 and multiple loci), whereby the number of loci detected by the individual

primer pair was determined according to the number of fragments amplified in the samples of the test panel, provided that all samples displayed comparable fragment numbers (supplementary material S1, Table 2). From all functional primer pairs, 74% amplified only single fragments, 22% amplified two and 4% more than two fragments. Especially for those markers amplifying single fragments, they were assigned to the two different genomes (T/S) based on the presence of an amplification product in the *N. tomentosiformis* or *N. sylvestris* accessions respectively. All functional primer sequences and other relevant data are available from the electronic supplementary material (S2).

#### Mapping of the microsatellite markers in the Hicks Broadleaf × Red Russian F2 population

A total of 2,415 (47%) of the functional markers were polymorphic between the parents of the mapping population. These candidate markers were then analyzed on the entire mapping population. 379 of these polymorphic markers could not be mapped due to insufficient data quality (low stability, difficult scoring due to strong stuttering, polymorphic between the parents but monomorphic within the mapping population) and thus had to be eliminated from the mapping experiment. This left 2,036 mapped markers, representing 37% of the newly identified primer pairs. Since the same plants were used as in Bindler et al. (2007), the 282 previously mapped markers were also included in the final map.

With the 282 previously mapped markers (including the morphological marker of flower colour), 2,318 markers detecting 2,363 loci were mapped on this high-resolution genetic map of tobacco. All mapped markers together generated 24 clearly defined linkage groups and the entire map of the tobacco genome covered 3,270 cM (Fig. 1). According to the amplification tests on the two samples from the ancestral genomes, the markers were assigned to the two different genomes of tobacco. 38% of all functional markers were assigned to the S-genome, 22% were assigned to the T-genome and those remaining could not be assigned due



**Fig. 1** High-density genetic map of the tobacco genome. *Red* chromosomal segments were assigned to the T-genome. *Blue* chromosomal segments were assigned to the S-genome. *Numbers on the left side* are

centiMorgan counted from the top of the chromosome. *Numbers in brackets* behind marker name on the *right side* display the numbers of additional cosegregating markers ( $\pm 0.5$  cM)

to lack of functionality in the ancestral genomes or because of simultaneous amplification from both genomes. In most of the linkage groups, the markers could be attributed pre-

dominantly to one of the two genomes. Eleven linkage groups could be clearly assigned to the S-genome and nine linkage groups to the T-genome. Four linkage groups

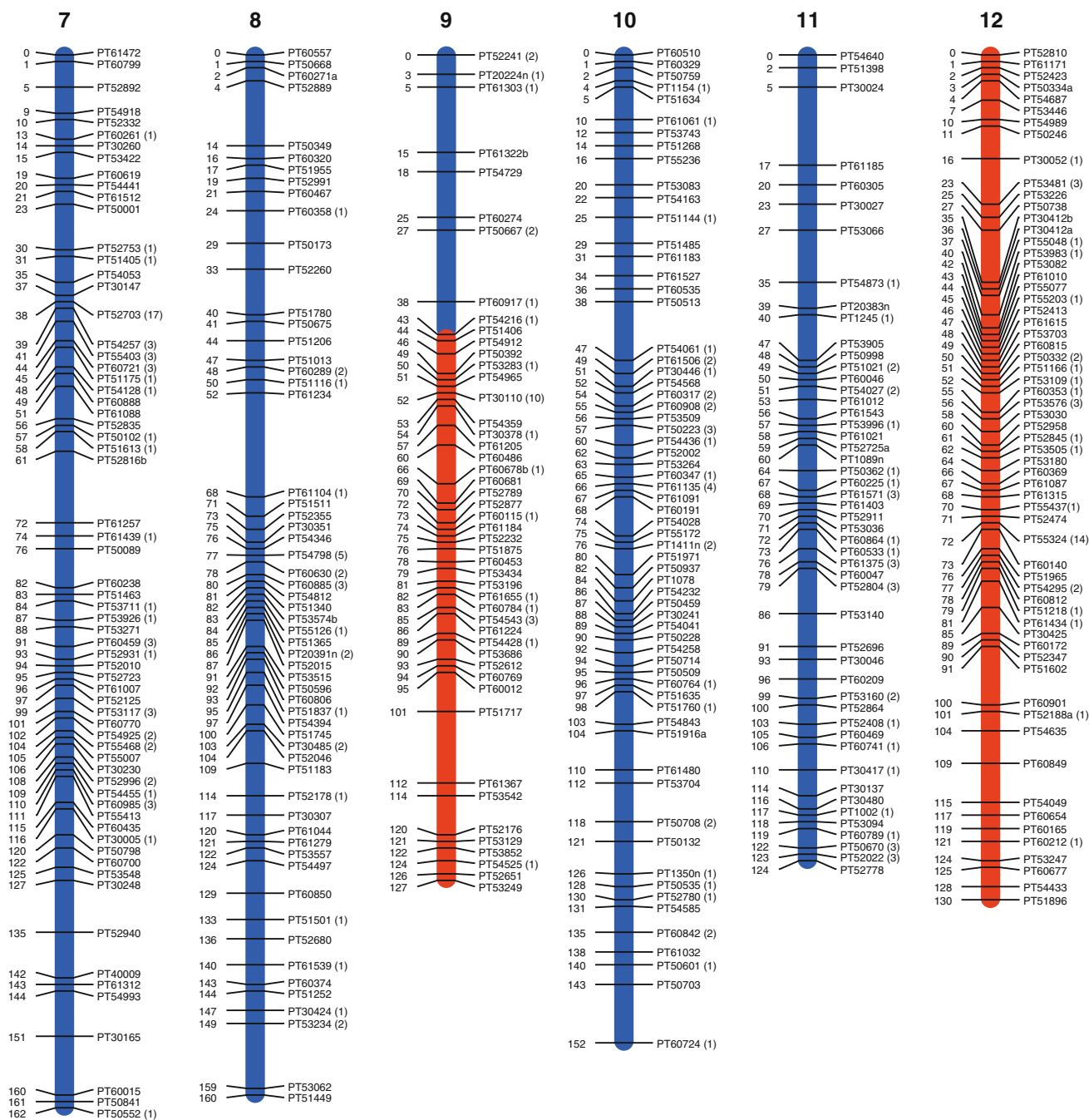


Fig. 1 continued

consisted of both S- and T-genome specific markers mainly grouped in different parts of the respective linkage group. These four linkage groups were stable up to very high LOD-scores. On the S-genome linkage groups, 1,163 markers were located covering 1,810 cM whereas on the T-genome linkage groups, 1,200 markers covered 1,460 cM. On average, each linkage group is represented by 98 SSR markers (Table 3). The marker density on the T-genome was higher than on the S-genome due to the shorter genetic distances.

Only two linkage groups (9 and 11) showed major areas of disturbed segregation as already observed previously (Bindler et al. 2007).

With 186 F2 individuals, the generated map had an average resolution of approximately 0.27 cM per crossover and a 90% probability of separating two markers that are approximately 0.6 cM distant from each other, provided that the scoring of the markers was accurate and determined in most or all of the progeny plants. In order to determine

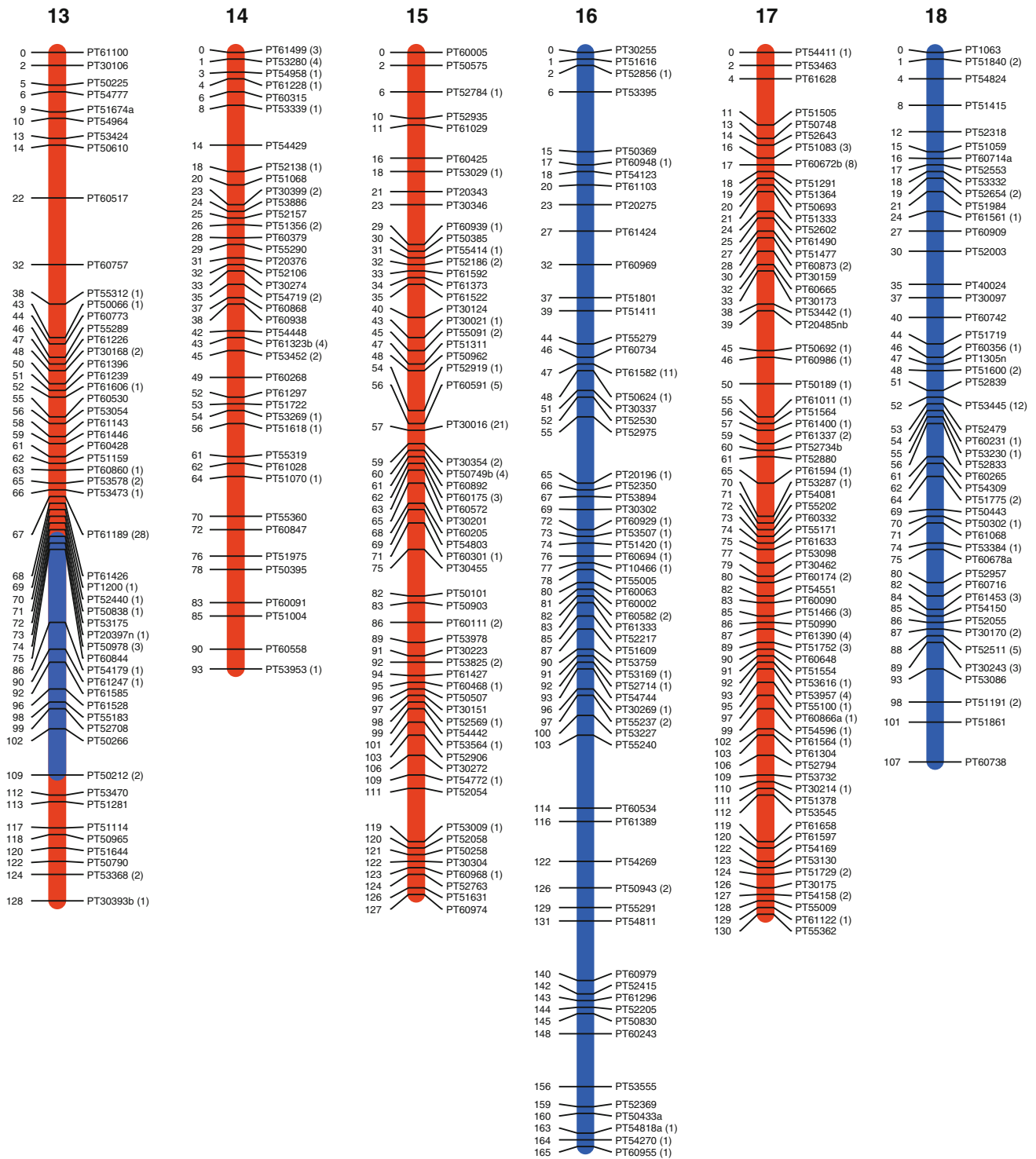


Fig. 1 continued

the quality of the map, a number of control analyses were performed. 5% of all data points generated throughout the mapping were randomly checked and 5% of markers classified as difficult to score were selectively checked. Less than 0.05% of the re-checked data points had technical or other problems in the data set. The final quality check included

the specific re-analysis of all double crossing over in the mapping data. After this analysis, 371 real double crossovers were still present considering all mapped markers (Table 3). On average, approximately two-thirds of the double crossovers within a chromosome were associated with regions of low marker density and large genetic

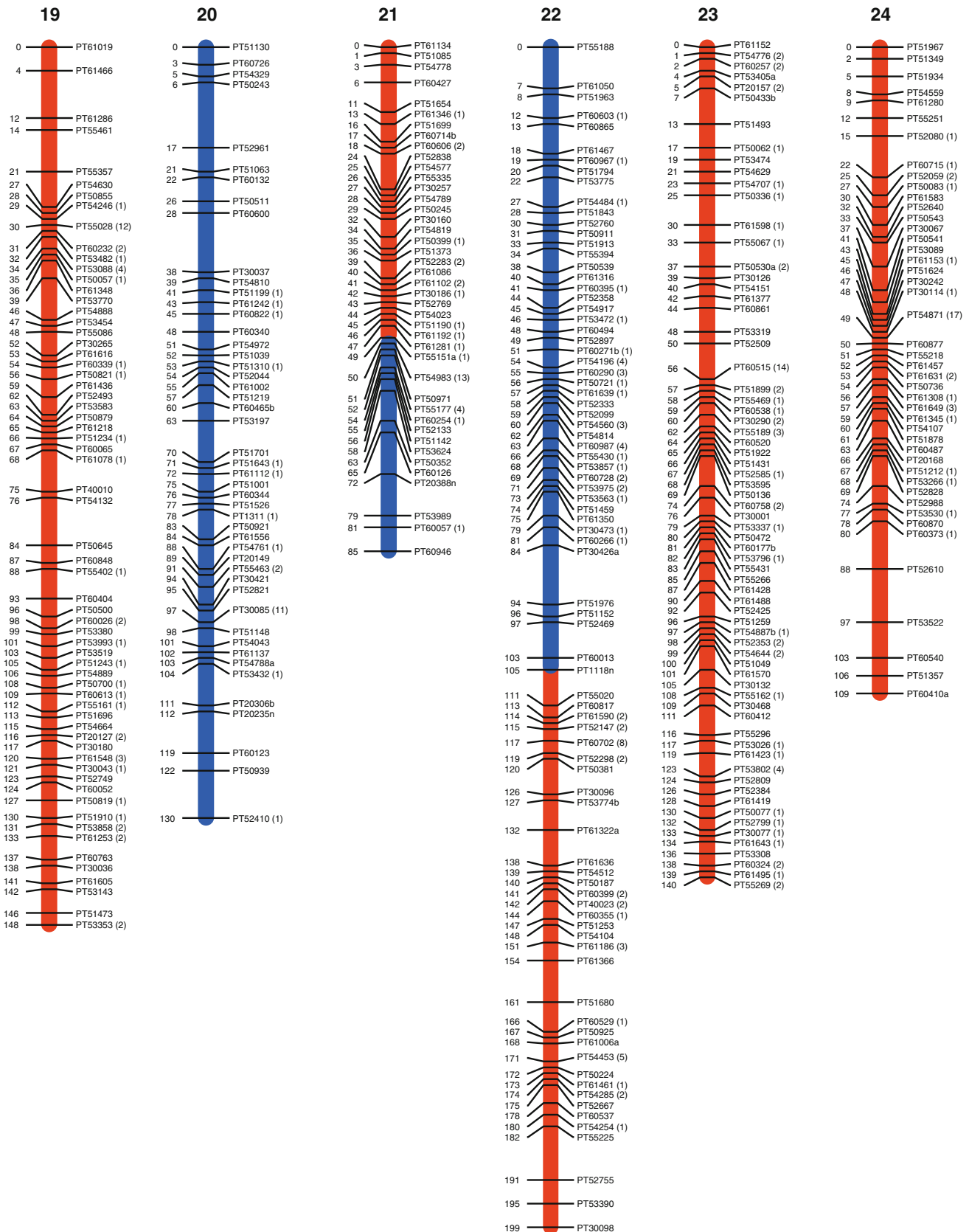


Fig. 1 continued



**Table 3** Tobacco genetic map data displayed per linkage group

	Linkage group (LG)	Genome	Loci/LG	LG length (cM)	Average marker distance	Maximal marker distance	Number of double crossing over
	1	S	116	130	1.12	6.7	18
	2	T	85	106	1.25	6.5	16
	3	S	106	183	1.73	14.7	24
	4	T	116	154	1.33	7	16
	5	S	64	136	2.13	12	19
	6	S	111	142	1.28	7.9	20
	7	S	121	162	1.34	10.9	18
	8	S	86	160	1.86	16	13
	9	S/T	77	127	1.65	11.3	14
	10	S	97	143	1.47	8.8	13
	11	S	83	124	1.49	12	17
	12	T	100	130	1.30	8.3	13
	13	S/T	104	128	1.23	10.6	14
	14	T	67	93	1.39	6.3	7
	15	T	115	127	1.10	8	10
	16	S	94	165	1.76	11.3	21
	17	T	121	130	1.07	6.7	19
	18	S	88	107	1.22	5.6	6
	19	T	110	148	1.35	8.1	12
	20	S	71	130	1.83	11.1	15
	21	S/T	74	85	1.15	7.6	14
Presumed linkage group assignment	22	S/T	146	199	1.36	10.7	23
S, <i>N. sylvestris</i> ; T, <i>N. tomentosiformis</i>	23	T	131	140	1.07	6.1	14
	24	T	80	109	1.36	8.5	10

distances between these markers. In addition, each linkage group had one major region of markers separated by small genetic distances whereby frequently a significant number of markers (up to 28) showed essentially no recombination between each other (electronic supplementary material S3 and Fig. 1).

Although the largest distance between two markers is still 16 cM and a number of markers are cosegregating, the average marker distance of mapped markers is 1.41 cM based on a genome size of 4,500 Mbp and a genetic distance of 3,270 cM. This is roughly equivalent to an average distance of 2 million base pairs between individual mapped SSR markers.

435 markers amplified two clearly defined fragments in the tetraploid *N. tabacum* genome (Table 2). In these cases identification of the homeologous linkage groups of the T- and S-genome was not possible since both fragments need to be polymorphic and mapped. For 389 markers (89%) only one fragment was polymorphic and could be mapped while the other fragment was monomorphic. Only for 46 markers could both fragments be mapped. From these, six duplicated loci were located on the same linkage group and 40 mapped to different linkage groups producing inconsistent homeologous pair assignments.

## Discussion

With the development of 5,119 new and functional microsatellite markers for the *N. tabacum* tobacco genome, the number of available functional microsatellite markers has increased nearly tenfold (from 637 to 5,756). This demonstrates the value of the TGI sequences for development of large numbers of SSR markers. Given a genome size of approximately 4,500 Mbp for the tobacco genome, it is estimated that the 5,736 markers increase the marker density to, on average, more than one available microsatellite marker (4,500 Mbp/5,756 marker) per million base pairs and now makes tobacco the species with the largest number of tested SSR or sequence tagged markers in the Solanaaceae. The tobacco genome was previously less well mapped than species such as pepper and eggplant. It is now comparable to tomato and potato in terms of number of available sequence tagged markers. It also makes tobacco a species for which more tested SSR markers are available compared to other allopolyploid or autopolyploid plant species (Cheng et al. 2009; Somers et al. 2004; Guo et al. 2007).

The established map in the Hicks Broadleaf × Red Russian F2 population is of high quality and high resolution. Most of the observed and confirmed double crossovers are

present only in regions with a low marker density and relatively large distances between markers. Furthermore, skewed segregation patterns are found only in limited chromosomal regions (chromosome 9 and 11). At an average genetic resolution of 0.27 cM and with most markers being scored in a codominant manner, the map is more accurate than the current saturated tomato map consisting of approximately half as many individuals (Tanksley et al. 1992; Fulton et al. 2002; <http://solgenomics.net>).

With a total size of 3,270 cM, the genetic map has increased by 1,350 cM compared to the previously published map of 1,920 cM which was incomplete since not all markers could be linked to each other (Bindler et al. 2007). In the current map, all 2,318 markers could be attributed unequivocally to the 24 linkage groups or chromosomes, indicating that the map should cover the entire tobacco genome, although the telomeres are not included in the map due to the lack of suitable markers. Considering that tobacco is an allotetraploid species with two genomes, each segregating in a diploid inheritance mode, the genome length of each genome in centiMorgan is comparable to the full map of tomato of approximately 1,500 cM (Shirasawa et al. 2010). The position of the centromeres of the 24 tobacco chromosomes is not known, although most chromosomes show a single region with a considerable clustering of markers. In other Solanaceous species such as tomato, such a marker cluster is usually associated with the centromere and its constitutive heterochromatin (Tanksley et al. 1992).

It is interesting to note that the vast majority (>90%) of the developed markers are either specific to one of the ancestral genomes or detect two loci (albeit usually only one locus was polymorphic and could be mapped) indicating that the two genomes are significantly different from each other and that in case two loci are identified, they might be located at the same position on the homeologous chromosomes. It was expected that the SSR markers for which two loci could be mapped should lead to the identification of the homeologous groups in the two genomes. This was however not the case since only a very limited number (46) of SSR markers were polymorphic on both chromosomes. Furthermore, in a number of cases linked marker pairs that were polymorphic in both genomes were definitely not located on the same chromosome, indicating that the two genomes have been rearranged in the current forms of the *N. tomentosiformis* and *N. sylvestris* genomes and/or after the speciation event which led to the generation of tetraploid tobacco. For both cases there is evidence based on this study and the results of Wu et al. (2009c). There, it could be shown that the *N. tomentosiformis* and *N. acuminata* (a closely related species to *N. sylvestris*) have undergone a number of chromosomal rearrangements since their separation. Based on the amplification of genome-specific

markers in the two ancestral genomes of *N. tomentosiformis* and *N. sylvestris*, followed by the assignment of these markers to linkage groups in the current map, it is likely that translocations have also occurred after the polyploidization event since four linkage groups contain genome-specific markers from both linkage groups. This is independently supported by cytogenetic analyses (Lim et al. 2007) where it could be shown by GISH (genomic in situ hybridization) that numerous intergenomic translocations exist in natural *N. tabacum*. It can also not be excluded that the S and T genome may have undergone deletion of specific chromosomal segments since the polyploidization event (Lim et al. 2004; Doyle et al. 2008) and recent data (G.B. unpublished data) suggest that on linkage group 9, some genes are absent in the S-genome.

In terms of genetic mapping and molecular marker assisted breeding, this map represents a significant improvement over existing marker resources. While it has not been previously possible to identify polymorphic markers spread at roughly similar genetic distances over the entire tobacco genome even in distantly related material, it is now possible to identify sufficient (hundreds) of polymorphic markers in crosses within the four main tobacco types: Burley, Flue-cured, Oriental and Dark, so that genes and quantitative trait loci can be mapped with a marker density/spacing which is suitable for marker-assisted breeding (Xu 2010).

The SSR markers and the sequences from which they have been derived will in the future be important tools to further advance tobacco genome analysis, since they can be used as anchor points in the physical mapping of bacterial artificial chromosome (BAC) clones and as anchor points to align the genetic map with a future genome sequence of tobacco. Furthermore, since the SSR markers are located in predominantly single copy sequences, the flanking DNA sequence from which the respective marker has been generated can be used to further align the tomato and potato genome sequence to the tobacco genome via sequence homology analysis, once their genome sequences are fully available, and if sequence homology is sufficient.

**Acknowledgments** TraitGenetics is thankful to Doris Kriseleit, Anika Küttner and Katja Wendehake for technical assistance and Andreas Polley for providing the tools for the identification of the SSRs in the TGI sequences. In Philip Morris International, we acknowledge Luca Rossi and Ferruccio Gadani for their support and input during the project and Philip Morris International, for providing the funding. In addition, we thank Florian Martin and Nicolas Sierro from the Bioinformatics team as well as Audrey Cordier and Louis Renaud from the Biotechnology team for their technical assistance.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410
- Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. *Plant Mol Biol Rep* 9:208–218
- Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* 27–2:573–580
- Bindler G, van der Hoeven R, Gunduz I, Plieske J, Ganai M, Rossi L, Gadani F, Donini P (2007) A microsatellite marker based linkage map of tobacco. *Theor Appl Genet* 114:341–349
- Cheng X, Xu J, Xia S, Gu J, Yang Y, Fu J, Qian X, Zhang S, Wu J, Liu K (2009) Development and genetic mapping of microsatellite markers from genome survey sequences in *Brassica napus*. *Theor Appl Genet* 118:1121–1131
- Clarkson JJ, Lim KY, Kovarik A, Chase MW, Knapp S, Leitch AR (2005) Long-term genome diploidization in allopolyploid *Nicotiana* section *Repandae* (Solanaceae). *New Phytol* 168:241–252
- Davis DL, Nielsen MT (1999) Tobacco—production. Chemistry and Technology, Blackwell
- Doyle JJ, Flagel LE, Paterson AH, Rapp RA, Soltis DE, Soltis PS, Wendel JF (2008) Evolutionary genetics of genome merger and doubling in plants. *Ann Rev Genet* 42:443–461
- Fulton TM, van der Hoeven R, Eannetta NT, Tanksley SD (2002) Identification, analysis, and utilization of conserved ortholog set markers for comparative genomics in higher plants. *Plant Cell* 14:1457–1467
- Gadani F, Hayes A, Opperman CH, Lommel SA, Sosinski BR, Burke M, Hi L, Brierly R, Salstead A, Heer J, Fuelner G, Lakey N (2003) Large scale genome sequencing and analysis of *Nicotiana tabacum*: the tobacco genome initiative. In: Proceedings, 5èmes Journées Scientifiques du Tabac de Bergerac—5th Bergerac Tobacco Scientific Meeting, Bergerac, pp 117–130. [http://www.altadis-bergerac.com/pdf/5\\_JS\\_Bergerac.pdf](http://www.altadis-bergerac.com/pdf/5_JS_Bergerac.pdf)
- Ganal MW, Altmann T, Röder MS (2009) SNP identification in crop plants. *Curr Opin Plant Biol* 12:211–217
- Guo W, Cai C, Wang C, Han Z, Song X, Wang K, Niu X, Wang C, Lu K, Shi B, Zhang T (2007) A microsatellite-based, gene rich linkage map reveals genome structure, function and evolution in *Gossypium*. *Genome* 176:527–541
- Kenton A, Parokony AS, Gleba YY, Bennett MD (1993) Characterization of the *Nicotiana tabacum* L. genome by molecular cytogenetics. *Mol Gen Genet* 240:159–169
- Lim KY, Matyasek R, Kovarik A, Leitch AR (2004) Genome evolution in allotetraploid *Nicotiana*. *Biol J of Linn Soc* 82:599–606
- Lim KY, Kovarik A, Matyasek R, Chase MW, Clarkson JJ, Grandbastien MA, Leitch AR (2007) Sequence of events leading to near-complete genome turnover in allopolyploid *Nicotiana* within five million years. *New Phytol* 175:756–763
- Lin TY, Kao YY, Lin S, Lin RF, Chen CM, Huang CH, Wang CK, Lin YZ, Chen CC (2001) A genetic linkage map of *Nicotiana plumbaginifolia*/*Nicotiana longiflora* based on RFLP and RAPD markers. *Theor Appl Genet* 103:905–911
- Manly KF, Cudmore RH Jr, Meer JM (2001) Map Manager QTX, cross platform software for genetic mapping. *Mamm Genome* 12:930–932
- 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
- Mueller LA, Solow TH, Skwarecki B, Buels R, Binns J, Lin C, Wright MH, Ahrens R, Wang Y, Herbst EV, Keyder ER, Menda N, Zamir D, Tanksley SD (2005) The SOL Genomics Network: a comparative resource for Solanaceae biology and beyond. *Plant Physiol* 138:1310–1317
- Olmstead RG, Bohs L, Migid HA, Santiago-Valentin E, Garcia V, Collier SM (2008) A molecular phylogeny of the Solanaceae. *Taxon* 57:1159–1181
- Ren Y, Zhang Z, Liu J, Staub JE, Han Y, Cheng Z, Li X, Lu J, Miao H, Kang H, Xie B, Gu X, Wang X, Du Y, Jin W, Huang S (2009) An integrated genetic and cytogenetic map of the cucumber genome. *PLoS ONE* 4:e5795. doi:10.1371/journal.pone.0005795
- Röder MS, Korzun V, Wendehake K, Plaschke J, Tixier M-H, Leroy P, Ganal MW (1998) A microsatellite map of wheat. *Genetics* 149:2007–2023
- Rossi L, Bindler G, Pijnenburg H, Isaac PG, Giraud-Henry I, Mahe M, Orvain C, Gadani F (2001) Potential of molecular marker analysis for variety identification in processed tobacco. *Plant Var Seeds* 14:89–101
- Rozen S, Skaletsky H (2000) Primer3 on the www for general users and for biologist programmers. *Methods Mol Biol* 132:365–386
- Shirasawa K, Asamizu E, Fukuoka H, Ohyama A, Sato S, Nakamura Y, Tabata S, Sasamoto S, Wada T, Kishida Y, Tsuruoka H, Fujishiro T, Yamada M, Isobe S (2010) An interspecific linkage map of SSR and intronic polymorphism markers in tomato. *Theor Appl Genet*. doi:10.1000-1344-37/s00122-01
- Somers DJ, Isaac P, Edwards K (2004) A high-density microsatellite consensus map for bread wheat (*Triticum aestivum* L.). *Theor Appl Genet* 109:1105–1114
- Suen DF, Wang CK, Lin RF, Kao YY, Lee FM, Chen CC (1997) Assignment of DNA markers to *Nicotiana glauca* chromosomes using monosomic alien addition lines. *Theor Appl Genet* 94:331–337
- Tanksley SD, Ganal MW, Prince JP, de Vicente MC, Bonierbale MW, Broun P, Fulton TM, Giovannoni JJ, Grandillo S, Martin GB, Messaguer R, Miller JC, Miller L, Paterson AH, Pineda O, Röder MS, Wing RA, Wu W, Young ND (1992) High density molecular linkage maps of the tomato and potato genomes. *Genetics* 132:1141–1160
- Van Ooijen JW, Voorrips RE (2001) JoinMap<sup>®</sup> 3.0. Software for calculation of genetic linkage maps. Plant Research International, Wageningen
- van Os H, Andrzejewski s, Bakker E, Barrena I, Bryan GJ, Caromel B, Ghareeb B, Isidore E, de jong W, van Koert P, Lefebvre V, Milbourne D, Ritter E, van der Voort JN, Rousselle-Bourgeois F, van Vliet J, Waugh R, Visser RG, Bakker J, van Eck HJ (2006) Construction of a 10, 000-marker ultradense genetic recombination map of potato: providing a framework for accelerated gene isolation and a genomewide physical map. *Genetics* 173:1075–1087
- Vontimitta V, Daneshmandi DA, Steede T, Moon HS, Lewis RS (2010) Analysis of a *Nicotiana tabacum* L. genomic region controlling two leaf surface chemistry traits. *J Agric Food Chem* 58:294–300
- Wang Y, Diehl A, Wu FN, Vrebalov J, Giovannoni J, Siepel A, Tanksley SD (2008) Sequencing and comparative analysis of a conserved syntenic segment in the Solanaceae. *Genetics* 180:391–408
- Wu FN, Mueller LA, Crouzillat D, Petiard V, Tanksley SD (2006) Combining bioinformatics and phylogenetics to identify large sets of single-copy orthologous genes (COSII) for comparative, evolutionary and systematic studies: a test case in the euasterid plant clade. *Genetics* 174:1407–1420
- Wu FN, Eannetta NT, Xu YM, Durrett R, Mazourek M, Jahn MM, Tanksley SD (2009a) 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 YM, Tanksley SD (2009b) A detailed synteny map of the eggplant genome based on conserved ortholog set II (COSII) markers. *Theor Appl Genet* 118:927–935

- Wu F, Eanetta NT, Xu Y, Plieske J, Ganai M, Pozzi C, Bakaher N, Tanksley SD (2009c) COSII genetics 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 Y (2010) *Molecular plant breeding*. CAB International, Wallingford
- Zimmerman JL, Goldberg RB (1977) DNA sequence organization in tobacco. *Chromosoma* 59:227–252