Theor Appl Genet (2007) 114:341–349
DOI 10.1007/s00122-006-0437-5

ORIGINAL PAPER

A microsatellite marker based linkage map of tobacco

Gregor Bindler · Rutger van der Hoeven · Irfan Gunduz · Jörg Plieske · Martin Ganal · Luca Rossi · Ferruccio Gadani · Paolo Donini

Received: 16 May 2006 / Accepted: 17 October 2006 / Published online: 7 November 2006

Abstract We report the first linkage map of tobacco (*Nicotiana tabacum* L.) generated through microsatellite markers. The microsatellite markers were predominantly derived from genomic sequences of the Tobacco Genome Initiative (TGI) through bioinformatics screening for microsatellite motives. A total of 684 primer pairs were screened for functionality in a panel of 16 tobacco lines. Of those, 637 primer pairs were functional. Potential parents for mapping populations were evaluated for their polymorphism level through genetic similarity analysis. The similarity analysis revealed that the known groups of tobacco varieties (Burley, Flue-cured, Oriental and Dark) form distinct clusters. A mapping population, based on a cross between varieties Hicks Broad Leaf and Red Russian, and consisting of 186 F2 individuals, was selected for mapping. A total of 282 functional microsatellite markers were polymorphic in this population and 293 loci could be mapped together with the morphological trait flower color. Twenty-four tentative linkage groups spanning 1,920 cM could be identified. This map will provide the basis for the genetic mapping of traits in tobacco and for further analyses of the tobacco genome.

Communicated by A. Schulman.

Electronic supplementary material Supplementary material is available in the online version of this article at http://dx.doi.org/10.1007/s00122-006-0437-5 and is accessible for authorized users.

G. Bindler · R. van der Hoeven · L. Rossi · F. Gadani · P. Donini (✉)
Applied Research Department, Philip Morris International, R&D, Quai Jeanrenaud 56, 2000 Neuchâtel, Switzerland
e-mail: paolo.donini@pmintl.com

I. Gunduz
Philip Morris USA, Research Center, Richmond, VA 23261, USA

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

Introduction

The Solanaceae are one of the most complex and diverse plant families. Recently, considerable efforts have been made to catalog the biological diversity in this plant family. With important vegetable species such as tomato, potato, eggplants and pepper, as well as other important plants such as tobacco and petunia, it is economically the third most important plant taxon (http://www.apps.fao.org). Furthermore, it has recently been shown that coffee, a species valued as one of the world's most valuable agricultural commodities, is also very closely related to the Solanaceae with respect to its gene reservoir and genome structure (Lin et al. 2005).

Most Solanaceae are characterized by a basic chromosome number of $x = 12$ albeit having a varying genome size (Arumuganathan and Earle 1991) ranging from less than 950 Mb (tomato) to approximately 3,000 Mb (pepper). Extensive genetic analyses with molecular markers have been performed in the Solanaceae (Mueller et al. 2005; http://www.sgn.cornell.edu) and in comparison with other plants such as *Arabidopsis thaliana* (Fulton et al. 2002). The genome structure of tomato is one of the best characterized in the plant kingdom and efforts are underway to sequence the

euchromatic area of the tomato genome in a systematic manner. Comparative genetic mapping within the Solanaceae has demonstrated that, e.g., the genomes of tomato and potato differ by a small number of inversions (Tanksley et al. 1992). The genetic maps of pepper and eggplant differ from tomato by a larger number of rearrangements but nevertheless very large chromosomal sections have been conserved between these species (Livingstone et al. 1999; Doganlar et al. 2002a, b).

The genus *Nicotiana* (Goodspeed 1954) is a well-defined group of species of which tobacco (*Nicotiana tabacum* L.) is an important agricultural crop plant that plays a significant role in the economies of many countries (Davis and Nielsen 1999; http://www.fao.org/documents/show_cdr.asp?url_file=/DOCREP/006/Y4997E/ Y4997E00.HTM).

Furthermore, tobacco is considered to be one of the most important model systems in plant biotechnology. Being an easily transformable plant, tobacco serves the role of an experimental system that is frequently being used for pilot studies on the expression of novel transgenes that are later being used in important food crops. Through its high level of biomass accumulation, tobacco is a species that is considered to be highly promising for the production of commercially important substances (e.g., medical drugs and vaccines) in plants. Furthermore, *Nicotiana* species are investigated for aspects concerning the elucidation of principles of disease resistance, synthesis of secondary metabolites and basic questions of plant physiology.

Albeit an inbreeding, highly homozygous plant, as well as being of high economic and scientific importance, it is surprising that the genetic analysis of tobacco is still in its infancy. While detailed linkage maps are available for several solanaceous plants, the genetic mapping of tobacco has not made any significant progress in the last 15 years except for a genetic map in a population derived from a cross between the diploid *Nicotiana plumbaginifolia* and *Nicotiana longiflora* (Lin et al. 2001). Only punctual molecular marker analyses have been performed in allotetraploid tobacco (*Nicotiana tabacum*) such as the analysis of genetic relationships in wild and cultivated tobacco material (Bogani et al. 1997; Ren and Timko 2001; Rossi et al. 2001) or the tagging of individual disease resistance genes with molecular markers such as RAPDs and AFLPs (Nishi et al. 2003; Julio et al. 2005).

Furthermore, there is basically no information available regarding the relationship of the individual chromosomes (synteny) with those of other Solanaceae albeit many of these plants share the same basic chromosome number. Reasons for this are most likely that *Nicotiana tabacum* is a tetraploid species. Tobacco is considered to be an allopolyploid interspecific hybrid between *Nicotiana sylvestris* ($n = 12$) and *Nicotiana tomentosiformis* ($n = 12$) and has $n = 24$ chromosomes (Kenton et al. 1993; Lim et al. 2004). Another reason for the lack of a detailed genetic analysis of the tobacco genome is possibly that the tobacco genome is with approximately 4,500 Mbp at the high end of the genome size in the Solanaceae (Arumuganathan and Earle 1991). As a consequence of this large genome size, a large proportion of the tobacco genome is represented by highly repeated DNA sequences including rDNA (Borisjuk et al. 1997; Volkov et al. 1999), active transposons (Casacuberta et al. 1995) and retrotransposons (Gregor et al. 2004).

Only recently, through the efforts of the Tobacco Genome Initiative (TGI, http://www.tobaccogenome.org) efforts have been undertaken regarding a systematic characterization of the tobacco genome (Gadani et al. 2003). The aim of the TGI is the identification of more than 90% of all tobacco genes through a combination of EST sequencing, the sequencing of undermethylated low copy DNA and the generation of BAC libraries for physical mapping. Considerable progress has been made since the initiation of TGI but this project lacks until now a genetic component comprising the construction of a high-resolution genetic map that could provide a framework for anchoring the sequencing information to the tobacco chromosomes.

The aim of the research reported in this paper is the construction of a first draft of a tobacco genetic map which could serve as genetic backbone for the integration of the sequence data generated from the TGI while constituting a first step towards a comparative genome analysis with other solanaceous species.

Materials and methods

Plant material and DNA isolation

All plants were grown from seeds in the greenhouse. For the analysis of polymorphism a panel of 16 tobacco varieties (Table 1) representing the main types of tobacco (Flue-cured, Burley, Oriental and Dark/others) was used. For each variety, leaves from a pool of 6 plants were combined for DNA extraction. The mapping population (kindly provided by Prof. Ramsey Lewis, North Carolina State University) was derived from a cross between the varieties Red Russian and Hicks Broad Leaf (a Flue-cured tobacco type). The F2 population consisted of 186 individual plants that were

Table 1 List of tobacco varieties used for testing the variability of the tobacco microsatellite markers

No.	Variety
NT0001	OR Izmir Ege 64
NT0002	OR Prilep 12 2/1
NT0003	OR Basma Xanthi
NT0004	OR Samsum
NT0005	BU TN90
NT0006	BU TN86
NT0007	BU Kentucky 14
NT0008	BU Banket A1
NT0009	FC K326
NT0010	FC K346
NT0011	FC Kutsaga 35
NT0012	FC Hicks Broad Leaf
NT0013	Dark Dac Mata Fina
NT0014	Dark Amarillo Parado
NT0015	Dark Criollo Misionero
NT0016	Red Russian

grown to maturity in a greenhouse. Total genomic DNA was isolated according to the protocol of Rogers and Bendich (1988) from leaf tissue.

PCR conditions and allele detection

PCR was performed in 10 µl volumes containing approximately 25–50 ng of template DNA, 1 × PCR buffer (50 mM KCl, 10 mM Tris–HCl pH 8.3), 1.5 mM $MgCl_2$, 0.15 mM of each primer, 0.25 mM dNTPs and 0.3U of *Taq* polymerase (Applied Biosystems). Thermocycling started with a denaturation step for 3 min at 94°C followed by 45 cycles of 1 min at 94°C, 1 min at 55°C, 2 min at 72°C and stopped after a final extension step of 72°C for 7 min. The fragment analysis was performed as a multi-loading assay analyzing two or three markers simultaneously that were labeled by different ABI-dyes. Samples containing 0.5–1 µl PCR products of each marker, 1 µl internal size standard and 9 µl Hi-Di formamide were separated using 36 cm capillary arrays. Alleles were detected using the GeneScan/Genotyper® software package of Applied Biosystems.

Molecular markers

A total of 187 SSR markers had previously been developed from a microsatellite-enriched genomic library of the Burley tobacco variety TN90 (Philip Morris International, unpublished data). Further microsatellite identification and marker development employed data from the TGI (http://www.tobaccogenome.org). A pre-screen on approximately 50,000 non-redundant methyl filtered tobacco sequences was performed to identify tandem repeats using the program Tandem Repeat Finder (Benson 1999). The program settings permitted the identification of di-, tri- (ten repeat units minimum), tetra-, hexa- and penta-nucleotide repeats (five repeat units minimum) with at least 80 flanking bases. Then, the resulting sequences were used to identify microsatellite containing DNA sequences as candidate target sites for development of SSR markers. The sequences were received as multiple fasta files with the corresponding quality files. Subsequently, vector-, *E. coli*- and duplicated sequences were masked in the microsatellite containing sequences by using the cross_match© program (University of Washington, distributed by CodonCode Inc.). Following a transformation into experimental sequencing files, di- to penta-nucleotide microsatellite containing sequences were selected with a minimum of 20 bp length and a minimum match of 90% of the virtual microsatellite motif. Oligonucleotide primer pairs flanking the microsatellite sequence were designed using the Primer 3.0 program (Steve Rozen and Helen J. Skaletsky, Whitehead Institute for Biomedical Research). Primers were selected to be approximately 20 bp long, to have a GC content between 20 and 80%, and with a melting temperature between 57 and 63°C (optimum 60°C). Forward primers were labeled with either FAM, HEX or ROX for fragment analysis on Applied Biosystems 3100 Genetic Analyzers. Primer sequences are available from the electronic supplementary material (S1).

Genetic mapping

The genetic mapping of polymorphic markers was performed 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, mapping function Kosambi. Flower color was scored visually in all F2 plants and integrated into the genetic map as a dominant trait.

Cluster analysis

The genetic relationship of the 16 tobacco varieties was investigated using the NTSYSpc 2.11j program (Exeter Software, Setauket, NY 11733–2870) with the following settings: Qualitative data, Dice coefficient (Dice 1945), Sequential Agglomerative Hierarchical and Nested clustering method (SAHN), Unweighted Pair-Group Method, Arithmetic average (UPGMA).

Results

Marker development

The microsatellite markers employed in this study were derived from two different sources. One set of 187 markers (PT1... and PT2...) has been developed previously (Philip Morris International, unpublished data). The majority and second set of markers (PT3... and PT4...) was generated from screening the TGI sequences for di- to hexanucleotide motives using bioinformatic means. Five hundred such microsatellite motifs containing sequences were used for subsequent primer development. A total of 684 primer pairs were tested on the panel of 16 lines. After the analysis on the 16 lines, markers were defined as either functional or non-functional. Functional markers were defined by a maximum of six amplification fragments, with at least one of the fragments being precisely in the expected size range as deduced from the genomic sequence. Non-functional were those markers for which amplification failed in all the lines, or resulted in amplification of more than six fragments, or the marker showed an excess of stuttering on the chromatogram. The majority of the functional markers (77%) amplified either one or two fragments in the tobacco genome. Figure 2 shows examples of two high-quality markers that amplify either one or two loci from the tobacco genome. Amplification of one fragment can be interpreted as a primer pair that binds specifically to only one of the two genomes of tobacco and not to the other while two fragments are amplified when a primer pair binds to the (probably homologous) region in the two different genomes that constitute the allotetroploid genome of *Nicotiana tabacum*. Table 3 shows a summary of the results for the 684 investigated microsatellite markers regarding their functionality, number of amplified fragments and polymorphism in the actual mapping population whereby no significant differences were observed between the two different sources.

Parental survey and selection of mapping population

One of the prerequisites for efficient genetic mapping in a given plant species is the availability of a highly polymorphic mapping population. For this, a set of 16 selected varieties/lines was screened to investigate the level of polymorphisms between potential parents for the mapping population. This set of lines comprised of different tobacco types such as Flue-cured, Burley, Oriental and Dark/other tobaccos. A set of 90 high-quality microsatellite markers out of all tested functional microsatellite markers was used for the evaluation of potential parents. The results of the study are shown in Table 2 and Fig. 1. The genetic similarity coefficient within these samples ranged from a very low value of 0.212 for the combination of Oriental Prilep and Burley Banket to a very high value of 0.918 for the combination of Flue-cured tobacco line K326 and Flue-cured tobacco line K346 indicating that a wide range of genetic diversity was present in these selected 16 samples. Furthermore, a cluster analysis for these 16 lines shows that the various tobacco types clustered together

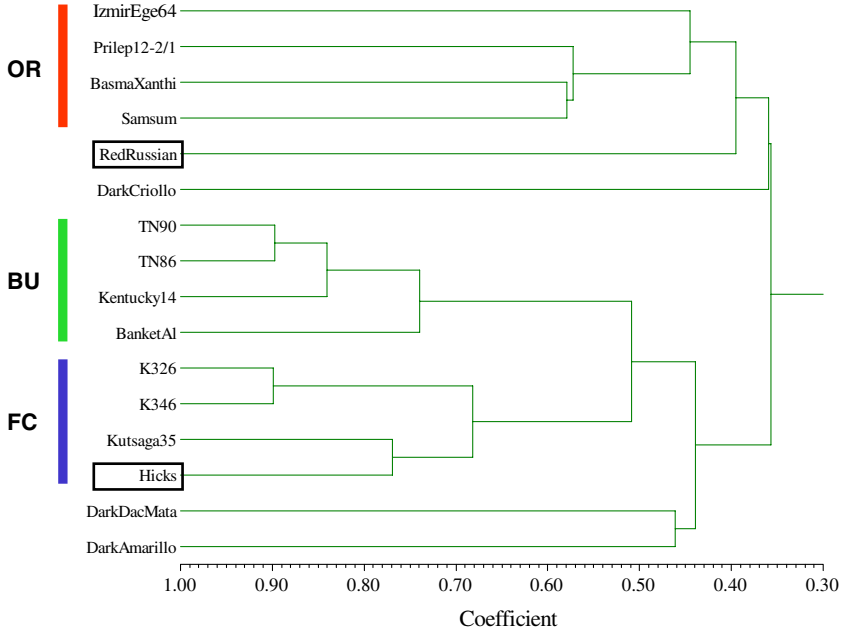

Fig. 1 Cluster analysis of the 16 tobacco varieties based on the data of 90 polymorphic markers using the similarity coefficient of Dice (1945). OR = Oriental; BU = Burley; FC = Flue-cured

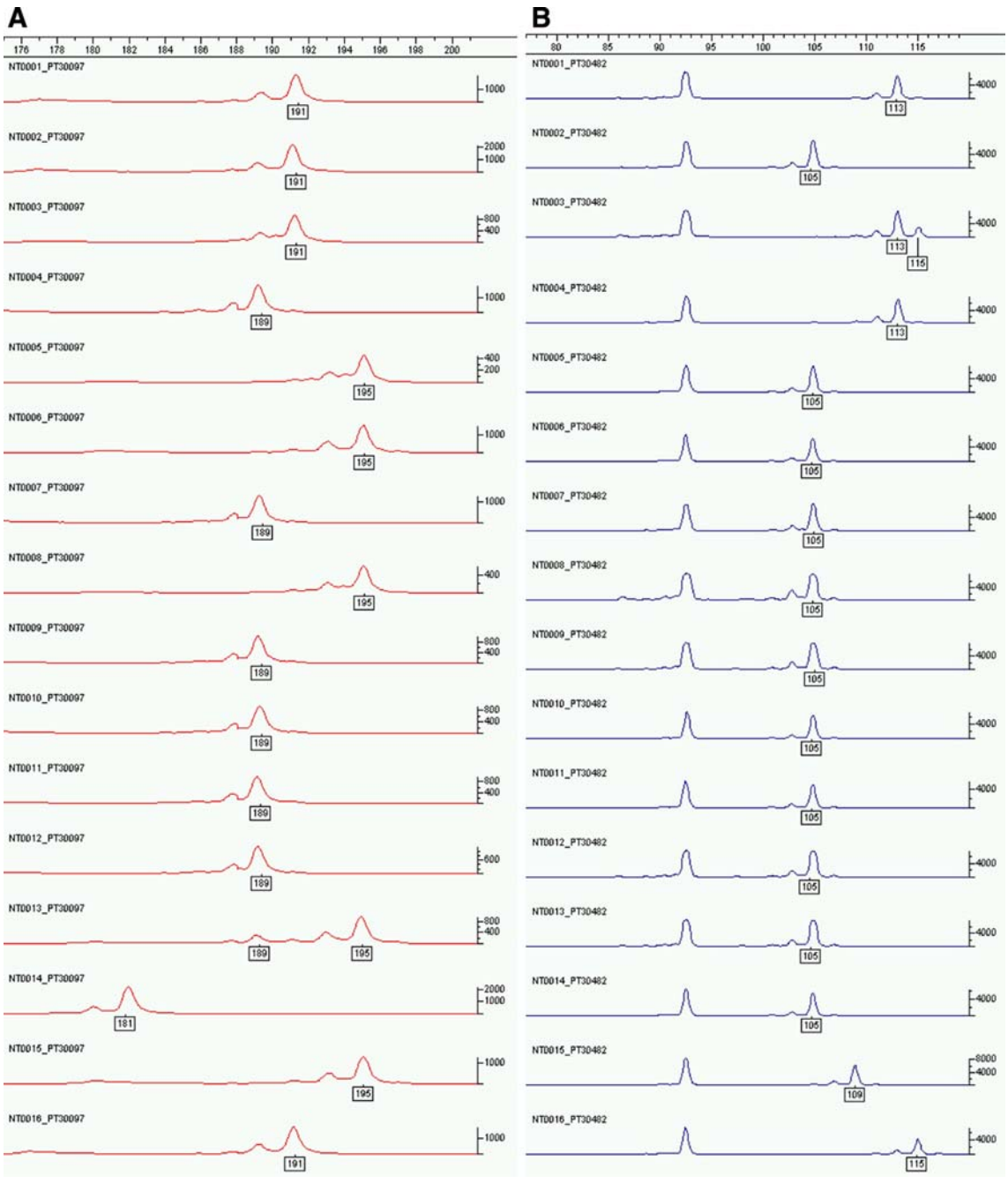

Fig. 2 Examples for microsatellite patterns in the investigated 16 tobacco varieties. **a** Dinucleotide repeat containing marker with slight stutter that detects one locus on one of the tobacco genomes and is polymorphic between the parents of the mapping population. **b** Dinucleotide repeat containing marker with slight stutter that detects two loci. One of the two loci is polymorphic between the parents of the mapping population

in almost all cases. The three main tobacco types Burley, Flue-cured and Oriental formed three distinct groups with defined clusters. Only the Dark tobacco types did not cluster in a distinct group. In no case, however, Dark tobaccos were found to group within a cluster of the other tobacco types, indicating that the three groups of Burley, Flue-cured and Oriental tobaccos were clearly genetically distinct.

Generation of a tobacco genetic map with microsatellite markers

Based on the availability of mapping populations; the cross between the Flue-cured variety Hicks Broad Leaf and Red Russian was chosen as a mapping population for the generation of a genetic map. The genetic similarity coefficient between these two lines was 0.385

Table 2 Similarity table of the 16 investigated tobacco varieties based on data for 90 microsatellite markers, indicated as similarity coefficient of Dice (1945)

	NT01	NT02	NT03	NT04	NT05	NT06	NT07	NT08	NT09	NT10	NT11	NT12	NT13	NT14	NT15	NT16
NT01	1.000															
NT02	0.459	1.000														
NT03	0.415	0.606	1.000													
NT04	0.379	0.606	0.595	1.000												
NT05	0.313	0.246	0.325	0.260	1.000											
NT06	0.337	0.251	0.346	0.300	0.918	1.000										
NT07	0.298	0.304	0.358	0.300	0.727	0.726	1.000									
NT08	0.328	0.212	0.316	0.245	0.876	0.853	0.695	1.000								
NT09	0.344	0.276	0.302	0.312	0.513	0.505	0.500	0.423	1.000							
NT10	0.352	0.274	0.290	0.277	0.528	0.497	0.516	0.449	0.902	1.000						
NT11	0.341	0.311	0.353	0.378	0.556	0.534	0.516	0.545	0.636	0.658	1.000					
NT12	0.342	0.235	0.294	0.333	0.545	0.507	0.525	0.495	0.719	0.694	0.774	1.000				
NT13	0.393	0.436	0.432	0.370	0.405	0.424	0.435	0.382	0.437	0.433	0.402	0.373	1.000			
NT14	0.332	0.367	0.380	0.309	0.419	0.431	0.428	0.381	0.448	0.433	0.479	0.458	0.437	1.000		
NT15	0.289	0.306	0.333	0.284	0.262	0.302	0.273	0.299	0.276	0.249	0.324	0.267	0.333	0.392	1.000	
NT16	0.372	0.376	0.505	0.352	0.355	0.351	0.363	0.330	0.373	0.382	0.396	0.385	0.419	0.434	0.346	1.000

which is close to the low end of all pair-wise combinations. For mapping, a total of 186 F2 plants were used. Mapping data could be generated for a total of 282 polymorphic microsatellite markers. More than 50% of the mapped markers amplified only a single fragment (Table 3) which is indicative for amplification from only one of the two tobacco genomes. Approximately, one quarter of the mapped markers amplified two fragments in the tobacco genome suggesting that one fragment is being amplified from each of the two tobacco genomes. However, in nearly all cases, for markers amplifying two fragments, only one fragment could be mapped due to the fact that the other fragment was monomorphic in the mapping parents and frequently also in the other panel plants, so that the total number of mapped loci added up to only 293. The resulting genetic map of tobacco is presented in Fig. 3. A total of 24 linkage groups could be identified, although three putative linkage groups (3a/3b, 8a/8b and 14a/14b) were each still divided into two sub-linkage groups of markers each, whereby JoinMap, however, suggested that they should group together. At present, five markers cannot be linked to any of the identified linkage groups, indicating that the genetic map is not yet covering the entire length of the tobacco genome. The entire map covers approximately 1,930 cM. Since tobacco is an allotetraploid species, the genetic length of each genome is in the present map close to 1,000 cM. The morphological trait of flower color could be integrated into linkage group 5.

Table 3 Functionality and number of amplified loci of the investigated tobacco microsatellite markers

	All	Percentage	Mapped	Percentage
1 locus	357	52.2	165	58.5
2 loci	171	25.0	82	29.1
3 loci	21	3.1	12	4.3
4 loci	5	0.7	4	1.4
Multiple loci (<7)	83	12.1	19	6.7
Non-functional	47	6.9	0	0
	684	100	282	100

Discussion

Microsatellite markers as markers for mapping allopolyploid plants

Microsatellite markers have been used for the generation of a molecular linkage map of tobacco. The general advantages of microsatellite markers for the generation of molecular linkage maps and marker analyses in plants have been reviewed extensively. In addition to these general advantages, microsatellite markers are currently the marker system of choice for the genetic analysis of allopolyploid plant species due to several characteristics.

First, microsatellite markers are usually derived from genomic sequences residing outside genes. Because of the high variability of such non-coding sequences, a large number of markers amplify only one fragment. For example, in the allohexaploid wheat genome approximately 40% of the developed microsatellite markers are genome specific (Röder et al. 1998). In the allotetraploid, *Brassica napus* 30% of all markers are genome-specific (Plieske and Ganal, unpublished results) and in tobacco nearly 60% of all markers amplify only a single fragment from one of the

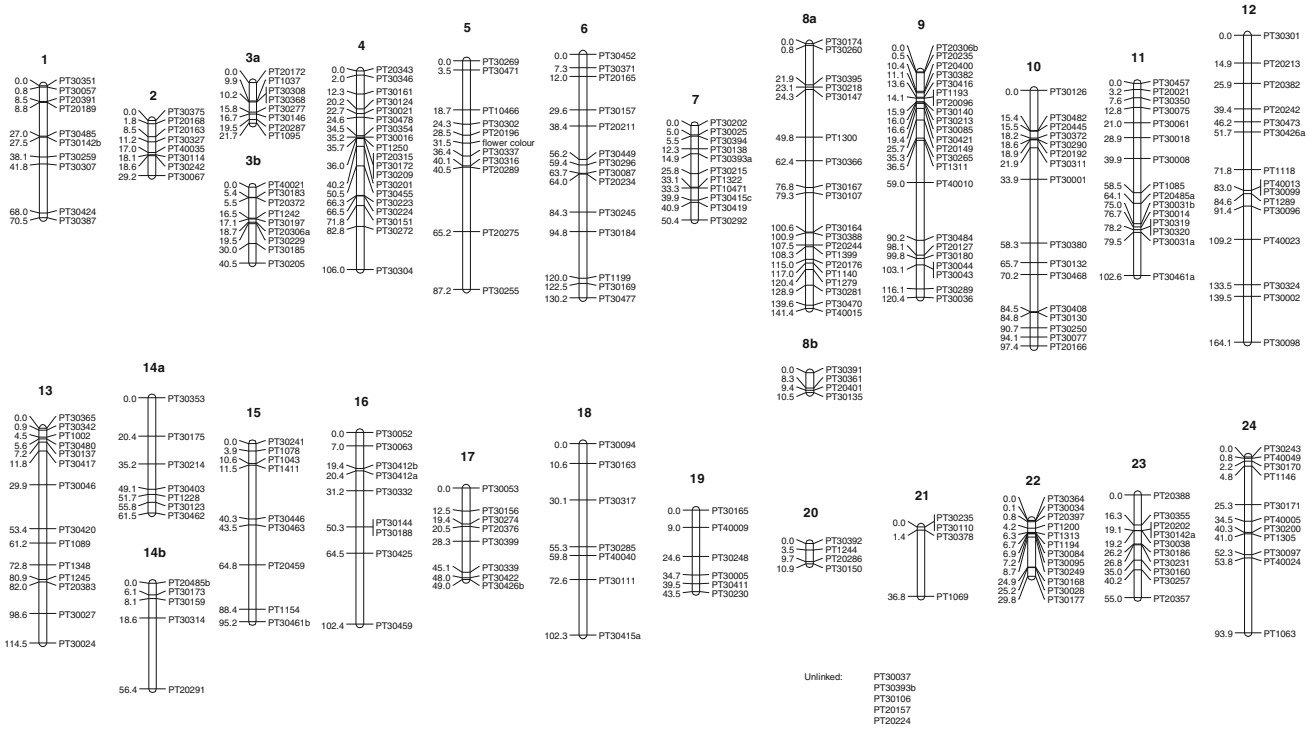

Fig. 3 Genetic map of *Nicotiana tabacum* consisting of 282 microsatellite markers and 293 mapped SSR loci and one phenotypic trait (flower color) on 24 linkage groups. Five markers are still unlinked. The markers of three groups (3, 8 and 14) are each distributed on two subgroups, but probably belong to one linkage group (LOD score > 4). If two loci were mapped with one marker, the two loci were labeled *a* and *b*

two genomes. This feature permits the assignment of markers, linkages to traits and linkage maps to specific chromosomes in a more reliable way than with other marker systems. Microsatellite markers that amplify a single fragment are the most suitable markers for the analysis of genetic relationships in allopolyploid species since they do not have problems with perturbations in the data analysis due to loci amplified from other genome(s). On the other hand, the main disadvantage of microsatellite markers is that they cannot be used for synteny studies due to their genome-specificity and origin from non-coding sequences since it is very likely that they will not amplify fragments from even closely related genomes unless they are derived from coding sequences.

Second, due to their multiallelic structure, microsatellite markers are the only reliable marker system that can be analyzed in a predominantly codominant fashion in a segregating population even in allopolyploid species. Most other PCR-based marker systems such as RAPDs or AFLPs can in polyploids only be analyzed as dominant markers. The same is true for the biallelic SNP (single nucleotide polymorphism) markers since they are mostly derived from highly conserved genes and, thus, are usually not genome-specific and only scorable as presence/absence markers.

Genetic relationships of tobacco varieties

The data on genetic relatedness of tobacco varieties clearly suggest that the established groups of tobacco varieties are based on different germplasm pools. The three groups of Flue-cured, Burley and Oriental tobacco varieties form clear clusters. The fourth group represented by the Dark tobacco varieties does not form a clear monophyletic group due to the fact that the genetic relationships between the investigated lines are very limited and thus, this cluster might represent an ancestral group containing a large amount of genetic variability. Previous AFLP analyses have suggested a similar population structure but have not provided sufficient resolution to substantiate such a clear distinction of the individual groups since only eight primer combinations were used (Ren and Timko 2001; Rossi et al. 2001). In the future, the analysis of additional tobacco lines using a set of high quality microsatellite markers should provide a very detailed picture of the genetic relatedness of tobacco lines within and between individual groups of germplasm. Furthermore, these data show that a limited set of microsatellite markers should be sufficient for variety identification and grouping of tobacco lines and varieties through the use of microsatellite marker databases.

The tobacco microsatellite map

This genetic map of *Nicotiana tabacum* represents the first version of a linkage map that covers most of the tobacco genome. With 293 loci covering a genetic distance of 1,930 cM, each of the two ancestral genomes should cover approximately 1,000 cM. It is clear that the current version of the tobacco linkage map is not yet complete since some unlinked markers exist as well as some potentially large gaps in at least three linkage groups. For a more complete map, it will be necessary to map additional microsatellite markers onto the tobacco linkage map. Based on the mapping experience in tomato and other diploid solanaceous plants, a genetic map that covers the entire genome with molecular markers will need at least 300–350 markers per genome resulting in a need of an excess of 600–700 microsatellite markers to sufficiently cover the two genomes of tobacco without major gaps. It is interesting to note that the total genetic distance of approximately 1,000 cM per genome is very close to the total genetic distance that has been observed for the diploid tomato with a similar set of markers (approx. 150), suggesting that the total length of the map for each of the two tobacco genomes might be in the range of 1,400–1,500 cM, thus being similar to the tomato genome (Bernatzky and Tanksley 1986; Tanksley et al. 1992).

An assignment of the individual linkage groups to the two genomes derived from *Nicotiana sylvestris* and *Nicotiana tomentosiformis* is not possible at present since the ancestral genomes have not yet been included into the microsatellite analysis. Due to the genome specificity of more than 50% of the mapped microsatellite markers, it will probably be possible to assign each linkage group to one of the two genomes in the future. Furthermore, through the availability of monosomic alien addition lines, it will be possible to merge the genetic and cytogenetic maps of tobacco with its two ancestral genomes (Suen et al. 1997). Another important point would be the identification of the homeologous chromosome pairs for the *sylvestris* and *tomentosiformis* genome of *Nicotiana tabacum* through the mapping of markers on both genomes. However, with only ten markers that amplified two polymorphic fragments and could be mapped on two different chromosomes and considering the incompleteness of the map, such an assignment would be speculative until more such markers have been identified and mapped.

The mapping of a phenotypic trait (flower color) to linkage group 5 demonstrates that with this microsatellite-based linkage map it is possible to localize monogenic and possibly also polygenic traits in the tobacco genome thus offering scope for marker-assisted selection of traits in the frame of marker-assisted breeding programs. Previously, the use of anonymous RAPD or AFLP markers has not permitted such a precise mapping and chromosomal assignment. Furthermore, in the long-term it will be possible to integrate and compare the chromosomal position and structure of flower color genes with other Solanaceae (De Jong et al. 2004)

Future developments

The microsatellite marker development and analysis reported in this paper would not have been possible without the availability of genomic and especially single-copy genomic sequences generated through the sequencing of methyl-filtrated sequences in the frame of the TGI, since these sequences were the source of most of the developed markers. Together with the construction of BAC-libraries and BAC-end sequencing, it will in the future be possible to generate additional microsatellite markers for the completion and further saturation of the tobacco genetic map to a density that is comparable to that of other solanaceous species such as tomato, potato and pepper, with more than 1,000 markers per genome (Tanksley et al. 1992). Only such marker density will result in sufficient markers for the genetic analysis within defined germplasm pools of tobacco. Furthermore, the integration of a large number of genetic markers into future physical maps based on BAC-fingerprinting will be necessary to generate a reliable physical map of the tobacco genome.

The TGI has also resulted in a large set of ESTs from the tobacco genome which will permit a comparison of the gene repertoire of tobacco with that of other, well characterized, solanaceous species such as tomato, potato and pepper, as well as closely related species outside of the Solanaceae such as coffee. Mapping of conserved ESTs into the tobacco genetic map through microsatellite markers that are generated from ESTs or through the mapping of SNPs in conserved genes will in the long term provide a picture of the differences in gene order between tobacco and other Solanaceae and extend our knowledge concerning genome evolution and structure in this important plant group.

Acknowledgments Prof. Ramsey Lewis is acknowledged for providing the seeds for the mapping population. TraitGenetics GmbH is thankful to Anika Küttner and Doris Kriseleit for their qualified technical assistance. Furthermore, the efforts of Alec Hayes (Philip Morris USA), C. Opperman and S. Lommel (both North Carolina State University) within the TGI are acknowledged.

References

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

Bernatzky R, Tanksley SD (1986) Towards a saturated linkage map in tomato based on isozymes and random cDNA sequences. Genetics 112:887–898

Bogani P, Liò P, Intriri C, Buratti M (1997) A physiological and molecular analysis of the genus *Nicotiana*. Mol Phylogenet Evol 7:62–70

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

Casacuberta JM, Vernhettes S, Grandbastien MA (1995) Sequence variability within the tabacco retrotransposon *Tnt1* population. EMBO J 14:2670–2678

Davis DL, Nielsen MT (1999) Tobacco—production, chemistry and technology. Blackwell, Oxford

De Jong WS, Eannetta NT, De Jong DM, Bodis M (2004) Candidate gene analysis of anthocyanin pigmentation loci in the Solanaceae. Theor Appl Genet 108:423–432

Dice LR (1945) Measures of the amount of ecologic association between species. Ecology 26:297–302

Doganlar S, Frary A, Daunay M-C, Lester R, Tanksley SD (2002a) A comparative genetic linkage map of eggplant (*Solanum melongena*) and its implications for genome evolution in the Solanaceae. Genetics 161:1697–1711

Doganlar S, Frary A, Daunay M-C, Lester R, Tanksley SD (2002b) Conservation of gene function in the Solanaceae as revealed by comparative mapping of domestication traits in eggplant. Genetics 161:1713–1726

Fulton T, van der Hoeven R, Eannetta NT, Tanksley SD (2002) Identification, analysis and utilization of conserved ortholog set (COS) 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

Goodspeed TH (1954) The genus *Nicotiana*. Chronica Botanica, Waltham, MA

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

Julio E, Verrier J-L, de Borne FD (2005) Development of SCAR markers linked to three disease resistances based on AFLP within *Nicotiana tabacum* L. Theor Appl Genet 112:335–346

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

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

Lin TY, Kao YY, Lin S, Lin RF, Chen CM, Fuang 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

Lin C, Mueller LA, McCarthy J, Crouzillat D, Petiard V, Tanksley SD (2005) Coffee and tomato share common gene repertoires as revealed by deep sequencing of seed and cherry transcripts. Theor Appl Genet 112:114–130

Livingstone KD, Lackney VK, Blauth JR, van Wijk R, Jahn MK (1999) Genetic mapping in *Capsicum* and the evolution of genome structure in the Solanaceae. Genetics 152:1183–1202

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

Nishi T, Tajima T, Noguchi S, Ajisaka H, Nedishi H (2003) Identification of DNA markers of tobacco linked to bacterial wilt resistance. Theor Appl Genet 106:765–770

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

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

Rogers SO, Bendich AJ (1988) Extraction of DNA from plant tissues. Plant Mol Biol Manual A6:1–10

Rossi L, Bindler G, Pijnenburg H, Isaac P, Giraud-Henry I, Mahe M, Orvain C, Gadani F (2001) Potential of molecular marker analysis for variety identification in processed tobacco. Plant Varieties and Seeds 14:89–101

Suen DF, Wang CK, Lin RF, Kao YY, Lee FM, Chen CC (1997) Assignment of DNA markers to *Nicotiana sylvestris* chromosomes using monosomic alien addition lines. Theor Appl Genet 94:331–337

Tanksley SD, Ganal MW, Prince JP, de Vicente MC, Bonierbale MW et al (1992) High density molecular linkage maps of the tomato and potato genomes. Genetics 132:1141–1160

Van Ooijen JW, Voorrips RE (2001) JoinMap® 3.0, software for calculation of genetic linkage maps. Plant Research International, Wageningen

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