Journal of the Korean Society of
Tobacco Science Vol. 30, No. 2, 85～93 (2008)
Printed in Republic of Korea

Identification of *Nicotiana tabacum* Cultivars using Molecular Markers

Yurry Um, Eunjeong Cho, Ha-Jeong Shin, Ho Bang Kim[1],
Yeong-Seon Seok, Kwan Suk Kim, and Yi Lee*

College of Agriculture, Life and Environment Sciences, Chungbuk National University, Chungbuk 361-763, Republic of Korea
[1]Department of Biological Sciences, Myongji University, Yongin 449-728, Republic of Korea
(Received November 17, 2008)

ABSTRACT: This report describes a set of seven informative single-nucleotide polymorphisms (SNPs) and one insertion-deletion (INDEL) distributed over 24 cultivars that can be used for tobacco (*Nicotiana tabacum* L.) cultivar identification. We analyzed 163,000 genomic DNA sequences downloaded from Tobacco Genome Initiative database and assembled 31,370 contigs and 60,000 singletons. Using relatively long contigs, we designed primer sets for PCR amplification. We amplified 61 loci from 24 cultivars and sequenced the PCR products. We found seven significant SNPs and one INDEL among the sequences and we classified the 24 cultivars into 10 groups. SNP frequency of tobacco, 1/8,380 bp, was very low in comparison with those of other plant species, between 1/46 bp and 1/336 bp. For exact identification of tobacco cultivars, many more SNP markers should be developed. This study is the first attempt to identify tobacco cultivars using SNP markers.

Key words: ***Nicotiana tabacum, Molecular marker, SNP***

Traditionally, plant classification has depended on the polymorphism of morphological, physiological, and biochemical traits of the organisms and it needs long experience and know-how. However, as the number of cultivars has increased, new cultivars are no longer distinguishable from older ones based on these traits. For these reasons, molecular markers based on DNA sequences were developed for making it easy to differentiate among plant species and cultivars. Common molecular markers include random amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), restriction fragment length polymorphism (RFLP), simple sequence repeat (SSR) and single nucleotide polymorphism (SNP), etc. These marker technologies diminished the need of highly trained human resources and cost-consuming field trials for cultivar identification.

Of the DNA-based markers, SSR markers have been used widely. Highly polymorphic SSR markers have been used for forensic and paternity

*연락저자 : 361-763 충북 청주시 흥덕구 성봉로 410번지 충북대학교 특용식물학과
*Corresponding author : *Department of Industrial Plant Science & Technology, Chungbuk National University, Cheongju, 361-763, Republic of Korea (phone: 82-43-261-3373; fax: 82-43-271-0413; e-mail: leeyi22@hanmail.net)*

analysis in humans (Grubwieser *et al.,* 2007; Jacewicz *et al.,* 2004; Rowold and Herrera, 2003). However, SNPs are more frequent than SSRs in animal and plant genomes and the mutation rate is lower, 10^{-8} (Kondrashov, 2003; Nachman and Crowell, 2000) than in SSRs, 10^{-3} (Brinkmann *et al.,* 1998); therefore, SNPs are more reliable as molecular markers for organism identification.

Nicotiana tabacum (2n=4x=48) is a eudiploid species derived from *N. sylvestris* (2n=24) and *N. tomentosiformis* (2n=24) (Kenton *et al.,* 1993; Lim *et al.,* 2004; Lim *et al.,* 2006). The genome size is about 4.5 x 10^9 bp (Arumuganathan and Earle, 1991). In spite of the economical importance of this plant, sequence data are very restricted and genetic analysis of tobacco is still in its infancy. Some molecular markers related to disease resistance genes and linkage maps of tobacco have recently been reported, but the markers are restricted to SSRs, RAPDs, and AFLPs (Bindler *et al.,* 2007; Julio *et al.,* 2006; Nishi *et al.,* 2003). Tobacco Genome Initiative completed a draft of the tobacco genome, but it is not yet available to the public.

These days, the cultivar determination of the tobacco leaf depends on the label of the stock and appearance of the leaves. However, cultivar determination of dried leaves is very difficult because the original shape is changed during the curing process and there are more than 10,000 cultivars in existence. Until now, no attempt had been made to find molecular markers to identify tobacco cultivars. In this study, we tried to find SNP or insertion-deletion (INDEL) markers to use in identifying of tobacco cultivars using selected 24 cultivars and primers designed from relatively long genomic DNA sequences.

MATERIALS AND METHODS

Plant Material and DNA Isolation

Twenty-four cultivars were used for the development of SNP markers. These 24 cultivars included traditional varieties collected from Europe, America, and Asia (Table 1). DNA was extracted from leaf disks of 2-cm diameter using a Plant Genomic DNA Mini Kit (RBC, Taiwan) according to the manufacturer's manual, then stored at -20℃.

Table 1. Tobacco varieties used for SNP finding

Number	Cultivars	Type
1	PV-08	Flue Cured
2	MSKF0102-37	Flue Cured
3	Hicks	Flue Cured
4	BY4	Flue Cured
5	Orinoco Tobacco	Flue Cured
6	KB103	Burley
7	VA528	Burley
8	White burley	Burley
9	Goose Creek Red	Burley
10	Xanthi	Orient
11	Basma	Orient
12	Ismir	Orient
13	Shirey	Dark
14	Lizard Tail Turtlefoot	Dark
15	Greenwood Dark	Dark
16	Little Crittenden	Chewing
17	Havana 608	Cigar Binder
18	Walker's Broadleaf	-
19	Monte Calme tobacco	-
20	Mokgicho	Korean Local
21	Hyangcho	Korean Local
22	Chilean Tobacco	-
23	Perique Tobacco	-
24	Iranian Tobacco	-

Tobacco Genomic DNA Sequences

Tobacco genomic DNA sequences (AGN_OF3_05222007.all.fasta) were obtained from Tobacco Genome Initiative (www.tgi.ncsu.edu/data/new_raw/2007-11-NOV/AGN_OF/) and assembled using PCAP program (Huang *et al.,* 2003). From the original 163,000 sequences, we assembled 31,370 contigs and about 60,000 singlets.

Primer Design and PCR Reaction

Gene-specific primer pairs were designed according to the following criteria: length, 20~22 bp; optimum melting temperature, 60℃; length of PCR product, 500~2,000 bp (Table 2). The PCR was performed in 40-μL reaction mixtures that contained 10~30 ng of genomic DNA. After the initial determination of the amplification *via* agarose gel electrophoresis, the PCR primers appeared to produce a single amplicon from the genomic DNA used for amplification of the amplicon from the 24 varieties.

Table 2. Accession numbers, primer sequences, and PCR product sizes of the DNA fragments analyzed for *Nicotiana tabacum* SNP identification

	Contig #	Accession #	Forward primer	Reverse primer	Product size (bp)
1	contig303	AJ718105	GACCAAAACGCTTGGGTAAA	TGCGAAGAAGGTTCTGAGGT	1270
2	contig309	X76982	GTGAACGGAGGGAGTACCAA	GGGGGTTTGATTGCCTATTT	871
3	contig360	AJ718286	GCCATGTAACAGCCAAAGGT	AGATGACAGGCTCTGCAGGT	1382
4	contig371	BA000042	TCTCCGGTATACCCTGCAAC	TTGGGTTGGGACAGTAGGAG	1155
5	contig387	AJ717841	ATAAGGGGAAGGGGTTTCCT	CACCAACAACAGCACCAATC	1792
6	contig442	Y13861	GCGAGCAACCTTTTTCAGTC	GGGTAAGGCCTGCGTACATA	716
7	contig468	BA000042	GCCCCTATGTTAGCAGGACA	CTCAACCCCTTCACCAAAAA	1120
8	contig474	AB289449	TTTGGTACGTCCTGCCTTTC	CTCACAACTCTCGGATGCAA	954
9	contig487	AJ718809	TTGTGCCCTCCATATCACAC	CGACATGAAAGGACAGGACA	898
10	contig492	AY351386	GATGGGCTACAACCTCTCCA	TCGTCTGAGCTCCGGTTACT	1228
11	contig495	AJ007903	CTTGCCTGAGCTGAAGGTCT	AGCCCTCGGAGCTCTAAATC	1114
12	contig500	AB193039	GATAGCCTGCAGCTCGAAGT	TGTATTTGAGGGGTTCTTCCA	1361
13	contig502	AY998018	CCGAGTGGAACCGTTTTAGA	GACATGCATGGGTTGAGATG	1389
14	contig507	AB029327	TCAAGCAATCCCATCACAAA	TGTGACCTTGCAACCGATTA	1005
15	contig511	AJ538622	TTTCCCTCCAAAACACTTGG	GGGTCGGAAACTGACCAATA	1774
16	contig514	U35619	TCTGGTAGGGGGTGACCATA	CAAATGCTTTTGCCACCTTT	991
17	contig522	AJ538507	GCCCAACGTTTTTGAGAAGA	TGTGTAGGTAAGCAGGCAGGA	1404
18	contig525	X52742	GGCTCCCATCAACACAACTT	CACTCACACGTGGCAACTCT	810
19	contig535	AY383599	TGAGCAGCCCAAAAGAAGAT	GCCTAAGCGAGGATGTGAAG	1055
20	contig570	AJ718542	TATCGCCCAAAACTCCATTC	CTTTCCACAAGCAGTCCACA	864
21	contig572	AB445397	GTCCTTGTGGGAGGACAAAA	AGCAGCAATTCCAGCTTGTT	1109
22	contig371	BA000042	TGGCTTCCAATTCCTTTTCA	ATAGTACTCCCGCCGTCTCA	1549
23	contig474	AB289449	GCACATCTTGGAGCCATTCT	CGTATGCCCAAATTCATTCC	1339
24	cintig522	AJ538507	AATTGTTGGTTCCCTGGACA	CGTTTACATAATTCCCCGAAG	1765
25	cintig502	AY998018	AGATCATCTCAGCCGTCCAT	CCACATGAAGCAACGTTCAG	1361
26	contig572	AB445397	TGAGGCTGACGAGTTCATTC	CCCCCACTTTTCAATTTCCT	1444
27	contig511	AJ538622	TTTCCCTCCAAAACACTTGG	TGATCACGACTCGGAGATTG	1309
28	contig511	AJ538622	CTCCGAGTCGTGATCATTCA	AGCAGCTGAAGCTCCAATGT	1656
29	contig371	BA000042	CCCTTTTTCTTCGCTCTTCC	AAAGAATCGGCAAGTCATGG	2047
30	contig687	AB024600	TGGACGTGTGAAAAACTTGC	TCGAGCCCTCGAAATTAAAA	1541
31	contig688	AJ718359	GCAGCTCTGTGCATGGTTTA	TGCTGGAAGAAGCCAGAGAT	1373
32	contig689	X70903	ACATAGCAGGGGATGACCAG	GTTGTTGGGCTGCGTATTTT	1449
33	contig693	AJ627183	GGTTCGGCCGGTTATTTTAT	GGATTGGCTGACAAGACCAT	1567

Table 2. continued

	Contig #	Accession #	Forward primer	Reverse primer	Product size (bp)
34	contig696	AJ719101	TTGGCTAAGCATGCTGTTTCT	AATAGAATCGCAGCCGTCAT	1519
35	contig713	X70903	CTCGTGTGGGAGGTTGTTTT	TGCACCTGCCTGTCTATCAG	1422
36	contig715	L04680	TCGAGACCCGATTCTTCATC	CACTGCACAACAGCTCCCTA	1589
37	contig719	D89635	TTGCCTGTTTTACCGGATTC	GCAAAAGAAGAAGGCTGGTG	1440
38	contig722	AJ718105	TGCGAAGAAGGTTCTGAGGT	TCGCTTCTGCTTTGCACTAA	1610
39	contig740	AB032539	TGGTCACCAACATGGCTAAA	TCTGCGTCAAGTCTTCATGG	1446
40	contig742	DQ350333	ATGTGCTGGAAAGGGTGAAG	GAGCTGCAAAGTTGCATTGA	1491
41	contig765	X66784	AAATGGATGCGATGACACAA	GGAATCACGGCATGAAGTTT	1564
42	contig772	AJ538640	GCAAAACTCAGACGCACTTG	GTGCATGCAGACTTGTGTCC	1463
43	contig785	U64925	TCATCCCCTCCAGTGAAGAC	CACGAACATTCCACGCTAGA	1426
44	contig801	AB032529	CAAACCTTTTCCCAGTTCCA	AGCAGCTGAAGCTCCAATGT	1579
45	contig824	AY219234	AATTCCACAAGTGGGGTCTG	TTGGATTTAGGGGTGGTTCA	782
46	contig824	AY219234	GCGTTGATTTCGGAAGTAGC	TAGCAGGTTCTGCAACATGG	900
47	contig825	AF401735	GCAGGTGTTGAGAAACAGCA	TGGGAATAAACCCCCTACAA	999
48	contig842	AB025714	CGCAGAAAGCATGAAATGAG	GCAGTGGTACATCCCATAGTGA	1408
49	contig871	U22260	TGGCTGAAAATGGAATCACA	GCCCAATCTAAAATGCGAAA	1577
50	contig880	AY528645	AGCAGCTGAAGCTCCAATGT	AATGGTGGTTGCAGGAAGAG	1406
51	contig886	D84238	TATCGGTTTTGTTCGGGTGT	TCCTTTTCCCAGAACCTCCT	1576
52	contig899	AJ012363	GTTTATAGCCTTGGCCGACA	CAGCTCGTCACTGTCCACAT	1415
53	TOPO2	AY169239	ACCCTAAAATGGACGCAGTG	CAGGAATGCGTTGTTCATTG	1181
54	BBF1	AJ009594	AATCCAGTTCGTTGGTCCAG	TTGAGGGTTCCACAGCTTTC	880
55	contig563	AY528645	AAGCTCGGTGAAGTTCGAGA	CCCCACTATTTGCCCTACAA	685
56	contig570	AJ718542	AAGCCATCCCGAATACTGTG	TTGCTTTGTGCTGGAGAATG	1007
57	contig583	U66401	CTTGATCCTCGCTTCTCGAC	ACTGATTTTTGGCCGTCTGT	1387
58	contig591	AB102806	TAGGGATTGAGGCAGTGGAG	GAACTGGTCTGAACCGAAGC	1215
59	contig592	AC215407	TGAAGACGACGATGATGGAA	TTGAAGACGGAGGAGTTGCT	1157
60	contig628	AF061106	GCCCACTGTTACGTTCCCTA	TTGGTTATGCCAATCAAGCA	1119
61	contig648	AF368380	TCTGCCCAATTACTCGATCC	TGTTGGACGAGCTGTTCTTG	1063

Sequence Analysis and SNP Discovery

The PCR products were cleaned up using Wizard SV Gel and PCR Clean-Up kit (Promega, USA) and sequenced using both of the primers used for amplification in the labeling reaction. Sequences were analyzed on an ABI3730xl DNA Analyzer (Applied Biosystems, Foster City, CA). The sequence data from each amplicon were analyzed with sequence analyzing software: CLC Workbench (CLC Bio, Denmark), Clustal X (Larkin *et al.*, 2007), and Sequencher™ 4.8 (Gene Code Corporation). The minimum and maximum PCR product length was 716 bp and 2,047 bp, respectively, and the average length was 1,295 bp (Table 3). We removed the first 25 bp from the sequence and compared only the first 26 bp to 525 bp region.

Table 3. Sequence information of SNP or INDEL identification

Parameters	Values
Number of cultivars	24
Number of alleles	61
Minimum PCR product length	716 bp
Maximum PCR product length	2,047 bp
Average PCR product length	1,295 bp
Total sequenced DNA length/cultivar	58,662 bp
Number of SNPs	7
Frequency of SNPs	1 per 8,380 bp
Number of INDELs	1
Frequency of INDELs	1 per 58,662 bp

RESULTS AND DISCUSSION

Overall, we compared 58,662 bp of genomic DNA sequence of each variety. We found 7 SNPs and 1 INDEL from 24 varieties (Tables 3, 4). Therefore, the SNP frequency was one per 8,380 bp.

Table 5 summarizes the haplotypes of tobacco varieties. The 24 varieties were classified into 10 groups. Group I has Hicks, BY4, Walker's Broadleaf, Lizard Tail Turtlefoot, Shirey, Greenwood Dark, and Little Crittenden, group II has PV-08, MSKF0102-37, Xanthi, and Goose Creek Red. Group III, containing Korean local varieties Mokgicho and Hyangcho, is most closely related to group II. Burley varieties, KB103, VA528, and White

Table 4. SNPs and INDELs identified from tobacco varieties

SNP ID	Contig # (Accession #)	Primer sequence (5′→3′)	Annealing Temperature(℃)	Product size (bp)	SNP Motif	SNP position from F primer
NTSL1	contig514 (U35619)	F: TCTGGTAGGGGGTGACCATA R: CAAATGCTTTTGCCACCTTT	60	991	A/C	375
NTSL2	contig522 (AJ538507)	F: AATTGTTGGTTCCCTGGACA R: CGTTTACATAATTCCCCGAAG	60	1765	C/T	313
NTSL3	contig687 (AB024600)	F: TGGACGTGTGAAAAACTTGC R: TCGAGCCCTCGAAATTAAAA	60	1541	A/C	65
NTSL4	contig785 (U64925)	F: TCATCCCCTCCAGTGAAGAC R: CACGAACATTCCACGCTAGA	60	1426	G/C	223
NTSL5	TOPO2 (AY169239)	F: ACCCTAAAATGGACGCAGTG R: CAGGAATGCGTTGTTCATTG	60	501	T/C	307
NTSL6	contig570 (AJ718542)	F: AAGCCATCCCGAATACTGTG R: TTGCTTTGTGCTGGAGAATG	60	1007	AT/-	292
NTSL7	contig563 (AY528645)	F: AAGCTCGGTGAAGTTCGAGA R: CCCCACTATTTGCCCTACAA	60	685	C/T	149
NTSL8	contig570 (AJ718542)	F: TCTCCGAGTGATGAAGCTGA R: ATGAGGAGCAGCGTTCAGAT	60	1179	T/C	243

Table 5. Haplotypes of tobacco varieties

Group	Varieties	SNP ID							
		NTSL 1	NTSL 2	NTSL 3	NTSL 4	NTSL 5	NTSL 6	NTSL 7	NTSL 8
I	Hicks	C	C	C	C	C	AT	T	C
	BY4	C	C	C	C	C	AT	T	C
	Walker's Broadleaf	C	C	C	C	C	AT	T	C
	Lizard Tail Turtlefoot	C	C	C	C	C	AT	T	C
	Shirey	C	C	C	C	C	AT	T	C
	Greenwood Dark	C	C	C	C	C	AT	T	C
	Little Crittenden	C	C	C	C	C	AT	T	C
II	PV-08	C	C	C	C	C	-	T	C
	MSKF0102-37	C	C	C	C	C	-	T	C
	Xanthi	C	C	C	C	C	-	T	C
	Goose Creek Red	C	C	C	C	C	-	T	C
III	Mokgicho	A	C	C	C	C	-	T	C
	Hyangcho	A	C	C	C	C	-	T	C
IV	Orinoco Tobacco	C	T	C	C	C	-	T	C
V	KB103	C	T	C	G	C	-	T	C
	VA528	C	T	C	G	C	-	T	C
	White Burley	C	T	C	G	C	-	T	C
VI	Monte Calme Tobacco	A	C	C	G	C	-	T	C
VII	Chilean Tobacco	A	C	C	C	C	-	T	T
	Perique Tobacco	A	C	C	C	C	-	T	T
	Iranian Tobacco	A	C	C	C	C	-	T	T
VIII	Basma	A	C	C	C	T	-	C	C
IX	Havana 608	A	C	A	C	T	-	T	C
X	Ismir	A	C	C	C	T	-	T	T

Burley, were grouped together (group V) and South American varieties, Chilean tobacco and Perique tobacco, were grouped together with Iranian tobacco (group VII). Interestingly, Oriental varieties, Xanthi, Basma, and Ismir, were not grouped together (Table 5).

In comparison with other species, SNP frequency of tobacco, 1/8,380 bp, is very low (Table 6). SNP frequency of plant species is usually between 1/46 bp and 1/336 bp (Ching *et al.,* 2002; Kolkman *et al.,* 2007; Nasu *et al.,* 2002; Schmid *et al.,* 2003; Zhu *et al.,* 2003). SNP frequencies are sensitive to the regions tested and the number of genotypes sampled. SNP frequency of the noncoding region, 1/32 bp, is higher than that of the coding region, 1/63 bp, from sunflower inbred lines (Kolkman *et al.,* 2007) and 1/32 bp in noncoding region and 1/124 bp in the coding region from maize inbred lines (Ching *et al.,* 2002). Larger samples have a greater probability of capturing rare SNPs (Kolkman *et al.,* 2007). In tobacco, two SNPs, NTS1 and NTS4, were found from coding regions and the others were identified from noncoding regions. We think it is not meaningful to calculate the SNP frequency from

Table 6. SNP frequency from plant species

Species	SNP frequency	Reference
Tobacco (*Nicotiana tabacum* L.)	1/8,380 bp	This study
Arabidopsis (*Arabidopsis thaliana* (L.) Heynh.)	1/336 bp	Schmid *et al.,* 2003
Soybean (*Glycine max* L. Merr.)	1/273 bp	Zhu *et al.,* 2003
Rice (*Oryza sativa* L.)	1/89 bp	Nasu *et al.,* 2002
Maize (*Zea mays* L.)	1/61 bp	Ching *et al.,* 2002
Sunflower (*Helianthus annuus* L.)	1/46 bp	Kolkman *et al.,* 2007

each region of tobacco because the number of tobacco SNPs is too small.

Tobacco SNPs are of low frequency because tobacco is an autogamous, domesticated, and inbred plant. Typically, SNPs are less abundant in autogamous than allogamous organisms, less in domesticated than wild genotypes, and less in inbred than outbred plant types (Kolkman *et al*., 2007).

In this study, the genotyping was done by direct sequencing of PCR products. For further genotyping of tobacco varieties, a more efficient system should be designed. A single base extension (SBE) assay would be a good candidate (Werbrouck *et al.,* 2008) and new techniques, such as MassARRAY, could be considered for high-throughput assay of genotypes (Jones *et al.,* 2007).

CONCLUSION

Identification of tobacco cultivars is crucial for cigarette manufacturers. They have to know the exact cultivars of dried tobacco leaves that are used in blending the leaves. In this study, we discovered 7 significant SNPs and 1 INDEL that can be used for cultivar identification. This study is the first attempt to identify tobacco cultivars using SNP markers. However, the number of discovered SNPs or INDELs is too small for exact identification of tobacco cultivars. Therefore, a high-throughput SNP discovery system will be needed for further identification of markers (Lijavetzky *et al.,* 2007).

ACKNOWLEDGEMENT

This work was supported by a Chungbuk National University Research Grant in 2007.

REFERENCES

Arumuganathan, K. and Earle, E.D. (1991) Nuclear DNA content of some important plant species. *Plant Mol. Biol. Rep.* 9: 208-218.

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

Brinkmann, B., Klintschar, M., Neuhuber, F., Huhne, J. and Rolf, B. (1998) Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. *Am. J. Hum. Genet.* 62: 1408-1415.

Ching, A., Caldwell, K.S., Jung, M., Dolan, M., Smith, O.S.,Tingey, S., Morgante, M. and Rafalski, A.J. (2002) SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines.

BMC Genet. 3: 19.

Grubwieser, P., Zimmermann, B., Niederstatter, H., Pavlic, M., Steinlechner, M. and Parson, W. (2007) Evaluation of an extended set of 15 candidate STR loci for paternity and kinship analysis in an Austrian population sample. *Int. J. Legal Med.* 121: 85-89.

Huang, X., Wang, J., Aluru, S., Yang, S.P. and Hillier, L. (2003) PCAP: a whole-genome assembly program. *Genome Res.* 13: 2164-2170.

Jacewicz, R., Berent, J., Prosniak, A., Dobosz, T., Kowalczyk, E. and Szram, S. (2004) Paternity determination of the deceased defendant in STR against RFLP analysis. *Int. Congress Series* 1261: 523-525.

Jones, E.S., Sullivan, H., Bhattramakki, D. and Smith, J.S. (2007) A comparison of simple sequence repeat and single nucleotide polymorphism marker technologies for the genotypic analysis of maize (*Zea mays* L.). *Theor. Appl. Genet.* 115: 361-371.

Julio, E., Verrier, J.L. and Dorlhac de Borne, F. (2006) 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, A.S., Gleba, Y.Y. and Bennett, M.D. (1993) Characterization of the *Nicotiana tabacum* L. genome by molecular cytogenetics. *Mol. Gen. Genet.* 240: 159-169.

Kolkman, J.M., Berry, S.T., Leon, A.J., Slabaugh, M.B., Tang, S., Gao, W., Shintani, D.K., Burke, J.M. and Knapp, S.J. (2007) Single nucleotide polymorphisms and linkage disequilibrium in sunflower. *Genetics* 177: 457-468.

Kondrashov, A.S. (2003) Direct estimates of human per nucleotide mutation rates at 20 loci causing Mendelian diseases. *Hum. Mutat.* 21: 12-27.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., *et al.,* (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23: 2947.

Lijavetzky, D., Cabezas, J.A., Ibanez, A., Rodriguez, V. and Martinez-Zapater, J.M. (2007) High throughput SNP discovery and genotyping in grapevine (*Vitis vinifera*L.) by combining a re-sequencing approach and SNPlex technology. *BMC Genomics* 8: 424.

Lim, K.Y., Matyasek, R., Kovarik, A. and Leitch, A.R. (2004) Genome evolution in allotetraploid *Nicotiana*. *Biol. J. Linn. Soc.* 82: 599-606.

Lim, K.Y., Souckova-Skalicka, K., Sarasan, V., Clarkson, J.J., Chase, M.W., Kovarik, A. and Leitch, A.R. (2006) A genetic appraisal of a new synthetic *Nicotiana tabacum* (Solanaceae) and the Kostoff synthetic tobacco. *Am. J. Bot.* 93: 875-883.

Nachman, M.W. and Crowell, S.L. (2000) Estimate of the mutation rate per nucleotide in humans. *Genetics* 156: 297-304.

Nasu, S., Suzuki, J., Ohta, R., Hasegawa, K., Yui, R., Kitazawa, N., Monna, L. and Minobe, Y. (2002) Search for and analysis of single nucleotide polymorphisms (SNPs) in rice (*Oryza sativa, Oryza rufipogon*) and establishment of SNP markers. *DNA Res.* 9: 163-171.

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

Rowold, D.J. and Herrera, R.J. (2003) Inferring recent human phylogenies using forensic STR technology. *Forensic Sci. Int.* 133: 260-265.

Schmid, K.J., Sorensen, T.R., Stracke, R., Torjek, O., Altmann, T., Mitchell-Olds, T. and Weisshaar, B. (2003) Large-scale identification and analysis of genome-wide

single-nucleotide polymorphisms for mapping in *Arabidopsis thaliana*. *Genome Res.* 13: 1250-1257.

Werbrouck, J., De Ruyck, K., Duprez, F., Van Eijkeren, M., Rietzschel, E., Bekaert, S., Vral, A., De Neve, W. and Thierens, H. (2008) Single-nucleotide polymorphisms in DNA double-strand break repair genes: Association with head and neck cancer and interaction with tobacco use and alcohol consumption. *Mutat. Res-Gen. Tox. En.* 656: 74-81.

Zhu, Y.L., Song, Q.J., Hyten, D.L., Van Tassell, C.P., Matukumalli, L.K., Grimm, D.R., Hyatt, S.M., Fickus, E.W., Young, N.D. and Cregan, P.B. (2003) Single-nucleotide polymorphisms in soybean. *Genetics* 163: 1123-1134.