ISSN 1022-7954, Russian Journal of Genetics, 2013, Vol. 49, No. 8, pp. 827–838. © Pleiades Publishing, Inc., 2013.

PLANT GENETICS

Large-Scale Development of PIP and SSR Markers and Their Complementary Applied in *Nicotiana*[1]

L. Huang[b, *], H. Cao[a, *], L. Yang[a,], Y. Yu[b], and Y. Wang[a]**

[a] *Tobacco Laboratory, College of Plant Protection, Shandong Agricultural University, Daizong Road no. 61, Tai'an, Shandong, 271018 China*

e-mail: lyang@sdau.edu.cn; chcsnd@163.com

[b] *College of Agronomy, Shandong Agricultural University, Daizong Road no. 61, Tai'an, Shandong, 271018 China*

e-mail: huang2002003@gmail.com

Received October 15, 2012

Abstract—PIP (Potential Intron Polymorphism) and SSR (Simple Sequence Repeats) were used in many species, but large-scale development and combined use of these two markers have not been reported in tobacco. In this study, a total of 12,388 PIP and 76,848 SSR markers were designed and uploaded to a web-accessible database (http://yancao.sdau.edu.cn/tgb/). E-PCR analysis showed that PIP and SSR rarely overlapped and were strongly complementary in the tobacco genome. The density of markers was 3.07 PIP and 1.72 SSR per 10 kb of the known sequences. A total of 153 and 166 alleles were detected by 22 PIP and 22 SSR markers in 64 *Nicotiana* accessions. SSR produced higher PIC (polymorphism information content) values and identified more alleles than PIP, whereas PIP could identify larger numbers of rare alleles. Mantel testing demonstrated a high correlation coefficient ($r = 0.949$, $P < 0.001$) between PIP and SSR. The UPGMA dendrogram created from the combined PIP and SSR markers was clearer and more reliable than the individual PIP or SSR dendrograms. It suggested that PIP and SSR can make up the deficiency of molecular markers not only in tobacco but other plant.

DOI: 10.1134/S1022795413070077

INTRODUCTION

Tobacco (*Nicotiana tabacum* L.), an important species in the genus *Nicotiana*, is a raw material for the cigarette industry and is widely cultivated in many different countries [1]. It is well known that cultivated tobacco is a natural amphidiploid ($2n = 24$II) through the hybridization of wild progenitor species and *N. sylvestris* ($n = 12$) and *N. tomentosiformis* ($n = 12$) [2]. The recent taxonomic revision of *Nicotiana* produced a well defined group of species which was composed of 76 naturally occurring *Nicotiana* species within 13 sections [3]. The chromosome numbers of these 76 species belong to eleven types: $2n = 9$II, 10II, 12II, 16II, 18II, 19II, 20II, 21II, 22II, 23II and 24II [4]. Species with chromosome numbers of $2n = 12$II and 24II are euploid while the rest are aneuploid due to them losing one or more chromosome formations during the evolutionary process [4, 5]. Chromosome number $2n = 24$II (33 species) was the most frequent type [6]. The highly polymorphic chromosome numbers have provided a large gene reservoir that can be used to study genetic diversity in tobacco.

Unlike other major crop species, research progress of tobacco genetic markers has been relatively slow [7]. Although many DNA markers, such as RAPD (Random Amplified Polymorphic DNA), ISSR (Inter Simple Sequence Repeat) and AFLP (Amplified Fragment Length Polymorphism), have been utilized in the genus *Nicotiana* for genetic diversity analysis [1, 8–10], the disadvantages of these markers, such as a small number, randomly amplification and poor reproducibility, have limited their wide application. In contrast, SSR (Simple Sequence Repeats) are ubiquitous throughout the eukaryotic genomes [11]. SSR has some positive attributes, such as having a high level of polymorphism, are co-dominant and are easily detected by PCR detection [12]. Since Bindler et al. [7] reported the first linkage map for tobacco, there has been much research into using SSR markers to analyze genetic relationships in *Nicotiana* [13–16]. With the increase in the numerous functional sequences and genome sequences for tobacco being submitted to the public database (PiantGDB, http://www.plantgdb.org/ and TGI, tobacco genome initiative, http://www.pngg.org/tgi/), large scale mining for SSR markers using these known sequences became possible. Recently, Bindler et al. [17] developed 5,119 SSR markers by exploiting TGI sequences

[1] The article is published in the original.

[*] These authors contributed equally to this work.

[**] Corresponding author.

but the number was still not large enough to be used in the study of tobacco.

ILP (Intron Length Polymorphism) has attracted more and more attention as a novel molecular marker, since it not only has similar advantages to SSR, but also some particular characteristics, such as directly reflecting variation within specific genes and subspecies [18]. Before ILP markers could be designed, the whole genome sequence and position of introns needed to be identified, as happened with the successful use of ILP in the sequenced model plant, rice [19, 20]. Apart from a few model plants, most crop plants or economically important plants have been not sequenced yet. Fortunately, a database of expressed sequence tags (ESTs) has been constructed for plants [21], which meant that the positions of the introns could be annotated by a comparative genomics approach. So cross species amplification became possible when primers were designed in flanking exons in order to amplify introns by PCR [22]. Yang et al. [22] developed a database of PIP (Potential Intron Polymorphism) markers based on the cross species predicted position of introns. PIP markers have been used in several plants [23–25], but not in tobacco.

PIP and SSR have their own advantages. SSR markers have the highest expected heterozygosity compared with other markers [12]. The advantages of PIP markers include: subspecies specificity, neutrality (no phenotypic effect) and an ability to directly identify variation within genes. These advantages can complement SSR markers [18]. Thus, PIP markers, in combination with SSR markers, could be used to estimate genetic diversity. The combined use of these two marker systems in rice has been successful, but their use in tobacco has not been reported.

This study undertook the large-scale development of PIP and SSR markers and analyzed their distribution in the tobacco genome. In addition, a set of PIP and SSR markers were selected in order to undertake phylogenetic analysis and to analyze the prospects of the combined use of PIP and SSR in tobacco.

MATERIALS AND METHODS

Plant Material and DNA Isolation. A set of 64 accessions, including 12 wild species and 52 lines of cultivated tobacco (*N. tabacum* L.), were selected for this study (Table 1). Twelve wild *Nicotiana* species were chosen from three subgenera, including *Petunioides, Rustica* and *Tabacum*, and 52 cultivars, representing the five market classes of air-cured, burley, flue-cured, oriental and sun-cured tobacco, were also selected. Most materials were collected from China and America. These seeds were germinated and grown under greenhouse conditions. One leaf of young fresh material from one seedling per accession was used for DNA extraction. Genomic DNAs were extracted using the CTAB/NaCl method [26].

Development of PIP and SSR Markers. The EST sequences of four tobacco species (*N. tabacum, N. sylvestris, N. benthamiana* and *N. langsdorffi* × *N. sanderae*) and the ORF (Open Read Frames) of *N. tabacum* were downloaded from PlantGDB and TGI, respectively.

A pipeline in Perl script was developed to design the PIP and SSR markers. The method of designing PIP markers was taken from the PIP database and used 120 base pair (bp) sequences cut from the tobacco EST sequences [22]. Primer pairs were designed for each side of the alternative splicing joint positions or the SSR loci using ePrimer3 [27]. The designed primers were tested using e-PCR (electronic PCR) [28] in all the tobacco sequences. Finally, suitable and reliable PIP and SSR markers were identified.

Selection of PIP and SSR Markers. All the PIP and SSR markers were tested by e-PCR, and then 44 markers (22 PIP and 22 SSR markers, Table 2) were selected out for real PCR. We selected those markers because of the 22 PIP markers could identify the different amplification products in two or more *Nicotiana* species, and the 22 SSR markers could amplify polymorphism in different *Nicotiana* species or the products had a polymorphic length of >5 bp in the same species from e-PCR.

PCR Conditions and Allele Detection. PCR reactions were performed in 15 μL volumes containing 20–25 ng DNA, 1.5 μL of 10× PCR buffer, 2.5 mmol L^{-1} $MgCl_2$, 0.25 mmol L^{-1} dNTPs, 0.36 μmol L^{-1} forward primers, 0.36 μmol L^{-1} reverse primers and 1 U *Taq* DNA polymerase. Thermocycling conditions started with extension for 4 min initial denaturation at 94°C, followed by 30 cycles of 30 s at 94°C, 1 min at 55°C, 30 s at 72°C and a final extension at 72°C for 10 min. 6% non-denaturing PAGE (320 V, 2.5 h) was used to separate the PCR products and silver staining was used to visualize the DNA bands. The PIC (polymorphism information content) value was calculated as follows [29]: $PIC = 1 - \sum_{i=1}^{k} p_i^2$, where p_i, is the frequency of the ith allele, and k is the total number of different alleles for the locus [30].

Construction of a Phylogenetic Tree. All distinct bands from the PCR products were scored as present (1) or absent (0). The binary matrix was used to generate a similarity coefficient matrix according to the Dice coefficient [31]. A dendrogram was constructed from the computed similarity matrix using the algorithm from UPGMA. To evaluate the robustness of the phylogenetic tree, Mantel testing [32] was used to compare co-phenetic matrices among the genetic similarities based on PIP, SSR and combined markers (in this paper, combined markers means PIP markers and SSR markers). Genetic similarity, dendrogram and co-phenetic correlations were calculated using NTSYS-pc version 2.1 [33]. The reliability of the dendrogram branches was checked using the Bootstrap

Table 1. Pedigree and origin of 64 *Nicotiana* accessions used in this study

Name	Type	Subgenus	Origin	2*n*	Pedigree/Background information
N. sylvestris	Wild	*Petunioides*	South America	24	
N. repanda	Wild	*Petunioides*	North America	48	
N. nesophila	Wild	*Petunioides*	North America	48	
N. alata	Wild	*Petunioides*	South America	18	
N. stocktonii	Wild	*Petunioides*	South America	48	
N. acuminata	Wild	*Petunioides*	North America	24	
N. undulata	Wild	*Petunioides*	South America	24	
N. bethanriana	Wild	*Petunioides*	Australia and South Pacific	38	
N. rustica	Wild	*Rustica*	South America	48	
N. glauca	Wild	*Rustica*	South America	24	
N. panianlata	Wild	*Rustica*	South America	24	
N. otophora	Wild	*Tabacum*	South America	24	
Pipaliu	Sun-cured	*Tabacum*	China	48	Landrace
Jinying	Sun-cured	*Tabacum*	China	48	Landrace
Tuanyuke	Sun-cured	*Tabacum*	China	48	Landrace
Tengchongxiaoliuye	Sun-cured	*Tabacum*	China	48	Landrace
Niuerduomaoyan	Sun-cured	*Tabacum*	South America	48	Landrace
Baihuaxiang	Sun-cured	*Tabacum*	China	48	Landrace
Baihuabantiecang	Sun-cured	*Tabacum*	China	48	Landrace
Huboxiang	Sun-cured	*Tabacum*	China	48	Landrace
Wushiye	Sun-cured	*Tabacum*	China	48	Landrace
Yanshai No. 2	Sun-cured	*Tabacum*	China	48	No pedigree information
GAT-2	Sun-cured	*Tabacum*	Japan	48	No pedigree information
Shandongdaye	Sun-cured	*Tabacum*	China	48	Landrace
Lanhanyan	Sun-cured	*Tabacum*	China	48	Landrace
Lianeryan	Sun-cured	*Tabacum*	China	48	Landrace
Dalianer	Sun-cured	*Tabacum*	China	48	Landrace
Dawanlei	Sun-cured	*Tabacum*	China	48	Landrace
Dabailei	Sun-cured	*Tabacum*	China	48	Landrace
Maoyan	Sun-cured	*Tabacum*	China	48	Landrace
Xiangyan	Sun-cured	*Tabacum*	China	48	Landrace
Samsun	Oriental	*Tabacum*	Albania	48	No pedigree information
KOKUBU	Oriental	*Tabacum*	Zimbabwe	48	No pedigree information
IRABOURBONNe	Oriental	*Tabacum*	Zimbabwe	48	No pedigree information
Xanthi Basma	Oriental	*Tabacum*	Turkey	48	No pedigree information
G140	Flue-cured	*Tabacum*	China	48	No pedigree information
NC82	Flue-cured	*Tabacum*	South America	48	6129/Coker319
Coker86	Flue-cured	*Tabacum*	South America	48	Coker 258/// 65-16R/ Coker 319 //175L
Guiyan 11	Flue-cured	*Tabacum*	China	48	76D-1/77E-1, then F6/NC89
DELCREST E.24	Flue-cured	*Tabacum*	United States	48	No pedigree information
Delhi-76	Flue-cured	*Tabacum*	Canada	48	No pedigree information
NC95	Flue-cured	*Tabacum*	South America	48	Coker 139/Bell 4-30// Coker139/Hicks
Dahuangjin 5210	Flue-cured	*Tabacum*	China	48	No pedigree information
Panyuanhuang	Flue-cured	*Tabacum*	China	48	Selection from Xujing No. 2
Cuibi No. 1	Flue-cured	*Tabacum*	China	48	No pedigree information
Hicks 55	Flue-cured	*Tabacum*	Zimbabwe	48	No pedigree information
TI1112	Flue-cured	*Tabacum*	South America	48	No pedigree information
Yuanyan No. 2	Flue-cured	*Tabacum*	China	48	Honghuadajinyuan/G-28

Table 1. (Contd.)

Name	Type	Subgenus	Origin	2*n*	Pedigree/Background information
Yuanyan No. 5	Flue-cured	*Tabacum*	China	48	No pedigree information
TT6	Flue-cured	*Tabacum*	China	48	Hicks BL/Holmes lines// Vamfen-Hicks / Kutsaga
K326	Flue-cured	*Tabacum*	South America	48	McNair225//McNair30/NC95
Yunyan 85	Flue-cured	*Tabacum*	China	48	Yunyan No. 2/K326
Taiwan No. 8	Flue-cured	*Tabacum*	China	48	Hicks BL/GAT-2
Jinxing6007	Flue-cured	*Tabacum*	China	48	Selection from Jinxingyan
NC27NF	Flue-cured	*Tabacum*	United States	48	Coker 319/NCTG21//Coker 319
Honghuadajingyuan	Flue-cured	*Tabacum*	China	48	Selection from Dajinyuan
K358	Flue-cured	*Tabacum*	United States	48	McN926/80241
Gexing No. 1	Flue-cured	*Tabacum*	China	48	Selection from Dahuangjin
Zhongyan 100	Flue-cured	*Tabacum*	China	48	NC82/920//NC82
Burley 37	Burley	*Tabacum*	South America	48	Gr.25/Gr.42
Ky12	Burley	*Tabacum*	South America	48	No pedigree information
Ky17	Burley	*Tabacum*	South America	48	Bx/By//BeI6611///B49, then F7/Va509
Diaozhiyan	Air-cured	*Tabacum*	China	48	Landrace
Maiyland 609	Air-cured	*Tabacum*	South America	48	MD Robinson/Florida301

method and 1000 replicates were checked using FreeTree [34].

RESULTS AND DISCUSSION

PIP and SSR Markers Were Widely Distributed and Covered the Whole Tobacco Genome from e-PCR

Based on the EST sequences of four *Nicotiana* species and the original read sequences of *N. tabacum*, a total of 12388 PIP, 10551 EST-SSR and 66297 genomic-SSR markers were obtained (Table 3). All of the information was uploaded to a web-accessible database (http://yancao.sdau.edu.cn/tgb).

The e-PCR results showed that PIP and SSR markers were widely distributed in the tobacco genome. Of all the 89236 markers, only 385 pairs (1.68%) of two types of markers overlapped the same position in the genome. This meant that 98.32% of PIP and SSR markers were distributed discretely. Furthermore, the frequency of the markers was 3.07 for PIP and l.72 for SSR per 10 kb sequence. These results indicated that PIP and SSR markers were discretely distributed in the tobacco sequences and almost covered the whole of the tobacco genome.

Characteristics of PIP and SSR Markers

In the real PCR, all the 22 PIP markers generated clear and stable bands in the 64 accessions. Among them, 21 out of 22 (all except NBP2307) were polymorphic (Table 4). A total of 153 alleles, with an average of 6.9 alleles per locus, were scored. The number of alleles ranged from 1 (NBP2307) to 17 (NTPOO 12). Allele length varied from 115 to 2617 bp with an average of 498.8 bp. PIP markers were designed for exon sequences with a length of 120 bp. As a result, the average length of the introns was about 378 bp. The PIC values of the PIP markers ranged from 0 (NBP2307) to 0.854 (NTP0012) with an average of 0.540.

With regards to the 22 SSR markers, a total of 166 alleles were detected. The number of alleles revealed

Table 2. Sequences of PIP and SSR markers

PIP	Forward primer	Reverse primer
NTP0003	TGATGCATTCGGTACAGCTC	AAACCTTTGAACCACCCACA
NTP0019	TCGTGATCGTGTTGTTGAGG	GCAGCATTCATAGCATTTGG
NTP0004	CTGCTGAGGCTGAATTACCA	AAGCTGCATAATCAAGGTGCTA
NTP0006	GGTGATTACGGATGGGACAC	CCACTTTCACCCATTTCTCAA
NTP0020	TTTGCCTCCTAATGAAGTTGG	CCATGGGTCTGAAAGATGCT
NTP0002	GCGACTGGGATAGAGCAAAT	CCCAGCTGTATGGCTAGTTCA
NTP0016	ACTTAACCTCCCGGTTCTCC	GCTCTCCATGAACCAAGACA
NTP0007	CCCTCATGGTGAGGATGTTT	GAGCAGTTTTGTTCCCAAGG
NTP0010	CATTCCTCATGGTGGAGGAG	TCGGCAAAATAAGTGCTGAA
NTP0013	CCATTTTCTTTGGGTTGGAG	TCCTCCCAGCTTTGTTATCAG
NTP0012	CCCTTATGCTGGTGGTGTTT	GATATGGTGAGGGCAGGACT
NTP0018	CGAAGAGAACCCAAGACGAG	GCCTCCACTAAGGGCAGTTT
NTP0009	CTGTTGCAAAAAGCATCGAA	CCGTCTAAGGTCCATTGCAT
NTP0008	GCATGACACAGTGGCTTACC	CCCTGAAGGAACAGACAAAGA
NTP0005	GGGTCTTACATCAGCAAATGG	ACTGTCTTGCATCGTTGCTG
NTP0011	TCAGGTTTTCAATCCAAGAATG	CATCATGTGTTGGTCCTTGC
NTP0014	TCGGGCACCTGAATTTTATT	TGTTCCATCTTGCATTGTCC
NTP0015	TGAGACACTGCCGAAGAATG	CGAAGATACCATGCAACTGG
NTP0017	TGGGCCTAAATTGGTCAGAG	CCTCTGGAATCAAAGCCATC
NTP0001	TGGCTAATGCATTGGAGTCT	TTGAGGTCTTCTCCGCTTGT
NBP2307	TTGGAAAGTCTGCCCTTGTC	GTGTGCTTTGACATTGATGC
NTP0357	TTGGTTATCATTGGAGCTGGT	CTCCAACAACTTCCCCTTCA
SSR	**Forward primer**	**Reverse primer**
NTGS66283	TTTTGCCTTTCGTGGAACTC	TGGTTGGTGGAATGGATTTT
NTGS66284	CAACAATCAAACCCCAAAGG	GTCAGCCATCCTAGCGACTC
NTES8708	GAAAGGTGCTTCTTGTTCTTCTTC	TGGGCTCACCAAAAGGATAG
NTGS66285	CACCAGGTGCACAGAGACAT	AGCTCAAACTCCCTCTCTTCC
NTES8709	CAGAGGACTGATACGAAGAAGAAGA	AAAAGCTGCTGCACCTCACT
NTGS66286	GGTAGGAGGAGGACGACCTTT	AAGACACAATCCCATGGCTA
NTES8710	AAGCTGCTGCCACCAATAAT	TCTCCACTTCCTTGCTCCAT
NTGS66287	ATCACGCAAAGCGAGTAGGT	ATCGGAGGCATCCTCAGTTA
NTES8711	GGACTGATACGAAGAAGAAGAAGAA	AAAGCTGCTGCACCTCACTT
NTES8712	AACCCCTATACATGAAGGACCA	GGGGTTTTTCAACTCCAGTG
NTGS66288	ATTCACCATGGGAGTTTTCG	TGGGCACTTGTGAATTTGAA
NTGS66289	GAACAGATTGCTGAAACTGAGAGA	GATGAGGTGGAGGCACAAG
NTGS66290	GCAGGAGTATGCCCAAGAAA	ATCAGTCATACGCACCCACA
NTGS66291	CACACAACCCACACCAACAT	AACGTGCAACACGCATAGAC
NTGS66292	CACACCAACAAGAATTCACACA	CTACGACACAAGATGGCTTGAT
NTGS66293	AAAGCTGCTGCACCTCACTT	CAGCTGTAGACGCCTAATTTAACA
NTGS66294	GAAGATTTTTGGGTGTGACCA	CTTGTCCTTGAAAGTTGAACCA
NTGS66295	AAGCACACGCAAGTGACAAG	AAGGCTGAAAGAGGGATTGC
NTGS66296	CCCAAAAGTCTTCGCTGAAT	TGCTAGGTTCTCACCCTCAGA
NTGS66297	ACAAACACCGAGAGCCTGTT	AAGCGTGATAAGCTATCCTTGG
NTES3747	CACGACAGATGGCCAGAATA	CCATCTTCTCGATCTCCTCCT
NTES0341	TTGTTCGAATAACAACAGAATATCG	CCAACACCAGCTCCAATGAT

Table 3. The number of PIP and SSR markers for the four species

Species	PIP	SSR
N. tabacum	8223	75009
N. sylvestris	699	81
N. benthamiana	2676	1524
N. langsdorffi × *N. sanderae*	790	234
The total number	12388	76848

by each marker ranged from 3 (NTGS66284) to 12 (NTES8710) with an average of 7.5. A set of fragments from 100 to 545 bp generated all primer pairs with an average of 159.1 bp. The PIC values for SSR markers varied from 0.238 (NTES8708) to 0.843 (NTES8710) with an average of 0.656, a little higher than that of the PIP markers. Rare alleles, with a frequency <0.05, accounted for approximately half of the total number of alleles observed with an average of 3.5 and 3.2 for PIP and SSR, respectively. The number of rare alleles with a frequency <0.01 detected by PIP (2.0) was much greater than the number detected by SSR (0.7).

The real PCR products were generally in agreement with the e-PCR predicted results. This meant that the methods for developing PIP used in this study were more feasible and practical than methods used in previous studies. Because of the different number of primers and species used in e-PCR and real PCR, the average allele number for PIP (6.9) and SSR (7.5) amplified by PCR were 1.28 and 1.44, respectively, which was more than that for e-PCR.

As we known that tobacco was a natural amphidiploid species, so genome duplication was consequent on frequent co-amplification of more than one locus and there was a clear correlation between number of alleles and chromosome number. In the 13 species, the total number of alleles observed by PIP was generally higher than SSR (Fig. 1). Furthermore, with the increase of chromosome numbers, total number of alleles also increased, PIP in particular. It suggested that PIP was more conserved than SSR between interspecies, and this advantage would be useful for developing genetic markers cross-species. Furthermore, it identified more rare alleles, which could be very useful as they were uniquely linked to particular genotypes [35].

Genetic Similarity among the 64 Accessions

Genetic similarity (GS) matrices derived from PIP and SSR were calculated (available on lyang@sdau.edu.cn). For PIP, GS values among the 64 accessions ranged from 0.22 to 1.0 with an average of 0.782. For SSR, GS values ranged from 0.13 to 0.98 with an average of 0.573. PIP showed a higher GS value than SSR, which indicated that PIP had a ability to reveal genetic diversity in tobacco.

In this study, the average GS value among the 52 cultivars was as high as 0.97 for PIP and 0.72 for SSR, which indicated a similar genetic relationship between them. This might have resulted from a genetic bottleneck where many accessions had a high genetic simi-

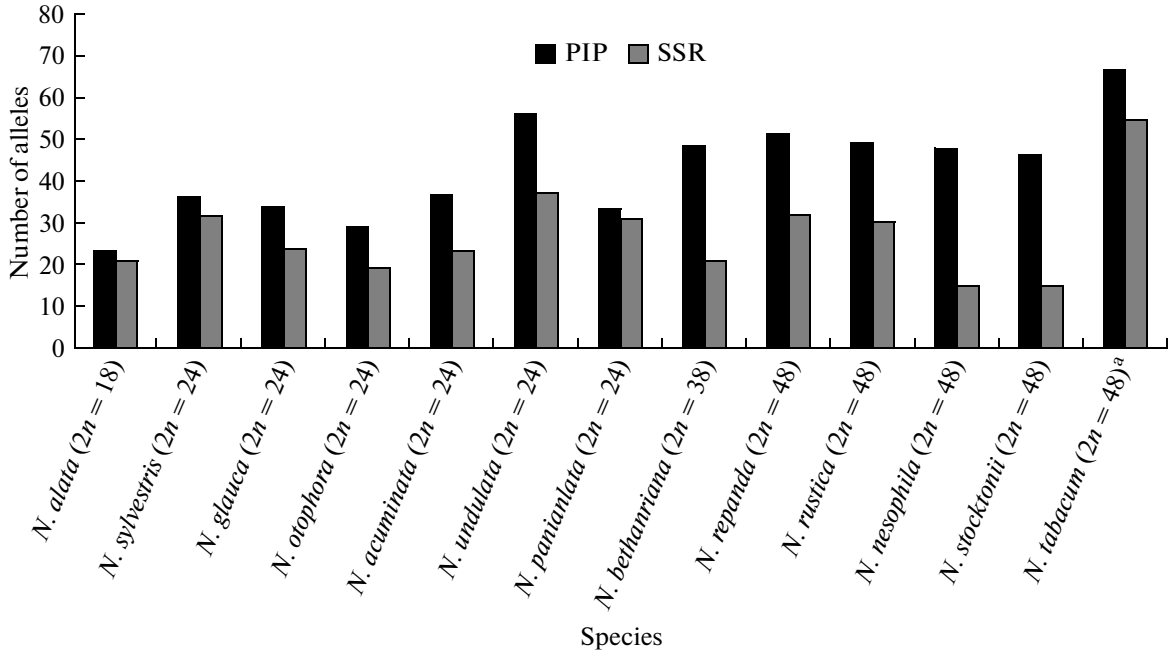

Fig. 1. Total number of alleles observed by PIP and SSR based on different species in *Nicotiana*. [a] average number of alleles observed in 52 cultivated tobacco.

Table 4. The statistical data of PIP and SSR markers used to genotype 64 varieties in *Nicotiana* genus

PIP	Alleles	Rare alleles		Total bands	Alleles sizes (bp)	Mean (bp)	PIC
		<0.05	<0.01				
NTP0003	10	6	5	235	464–1702	638	0.7316
NTP0019	14	10	6	249	260–1500	543	0.7595
NTP0004	5	1	0	77	164–381	299	0.2998
NTP0006	7	4	3	175	167–777	553	0.6321
NTP0020	7	5	2	116	212–380	321	0.4608
NTP0002	8	5	2	184	512–945	731	0.6663
NTP0016	12	5	3	386	240–1252	489	0.853
NTP0007	9	3	2	339	155–355	236	0.818
NTP0010	4	1	1	177	165–500	371	0.596
NTP0013	5	2	0	170	234–902	487	0.6342
NTP0012	17	10	10	411	251–1993	586	0.8541
NTP0018	4	2	0	69	199–433	224	0.1864
NTP0009	5	1	1	224	183–920	552	0.7067
NTP0008	7	3	2	201	258–540	381	0.7193
NTP0005	3	1	0	59	138–1043	875	0.1884
NTP0011	7	3	1	194	202–507	332	0.7164
NTP0014	11	6	4	286	376–1632	615	0.796
NTP0015	6	4	2	123	773–2617[b]	1001	0.4575
NTP0017	3	1	0	61	230–256	244	0.1993
NTP0001	4	2	1	108	482–639	551	0.4106
NBP2307	1	0	0	63	115[a]	115	0
NTP0357	4	2	0	56	285–927	830	0.1912
Mean	6.9	3.5	2	180.1	115[a]–2617[b]	498.8	0.540
Mean[c]	1.28				158[a]–3412[b]	270.8	
SSR	Alleles	Rare alleles		Total bands	Alleles sizes (bp)	Mean (bp)	PIC
		<0.05	<0.01				
NTGS66283	6	3	0	163	127–201	155	0.6811
NTGS66284	3	0	0	115	106–139	124	0.4786
NTES8708	4	1	0	64	100–112	103	0.2383
NTGS66285	6	2	1	227	100–210	148	0.7191
NTES8709	7	4	0	155	116–172	130	0.6934
NTGS66286	10	5	2	127	162–363	274	0.748
NTES8710	12	6	1	243	105–210	139	0.8429
NTGS66287	6	3	0	122	118–145	129	0.5713
NTES8711	5	2	0	58	113–159	129	0.5719
NTES8712	6	1	1	160	102–309	187	0.7275
NTGS 66288	7	3	1	208	100–153	120	0.7571
NTGS 66289	7	5	1	127	100–146	109	0.5094
NTGS 66290	11	4	1	241	114–195	146	0.842
NTGS 66291	9	5	1	114	152–277	217	0.7445
NTGS 66292	7	1	0	108	100[a]–173	129	0.7841
NTGS 66293	7	1	0	59	100[a]–200	149	0.6134
NTGS 66294	11	4	0	154	138–369	206	0.8003
NTGS 66295	10	5	4	150	101–474	277	0.7514
NTGS 66296	10	3	1	161	100[a]–177	133	0.8251
NTGS 66297	9	4	1	152	113–266	205	0.6481
NTES3747	7	5	0	73	115–545[b]	153	0.2662
NTES0341	6	3	1	184	107–190	139	0.6169
Mean	7.5	3.2	0.7	143.8	100[a]–545[b]	159.1	0.656
Mean[c]	1.44				100[a]–279[b]	131.3	

Note: [a, b] represent the shortest and longest bands, respectively, [c] e-PCR results.

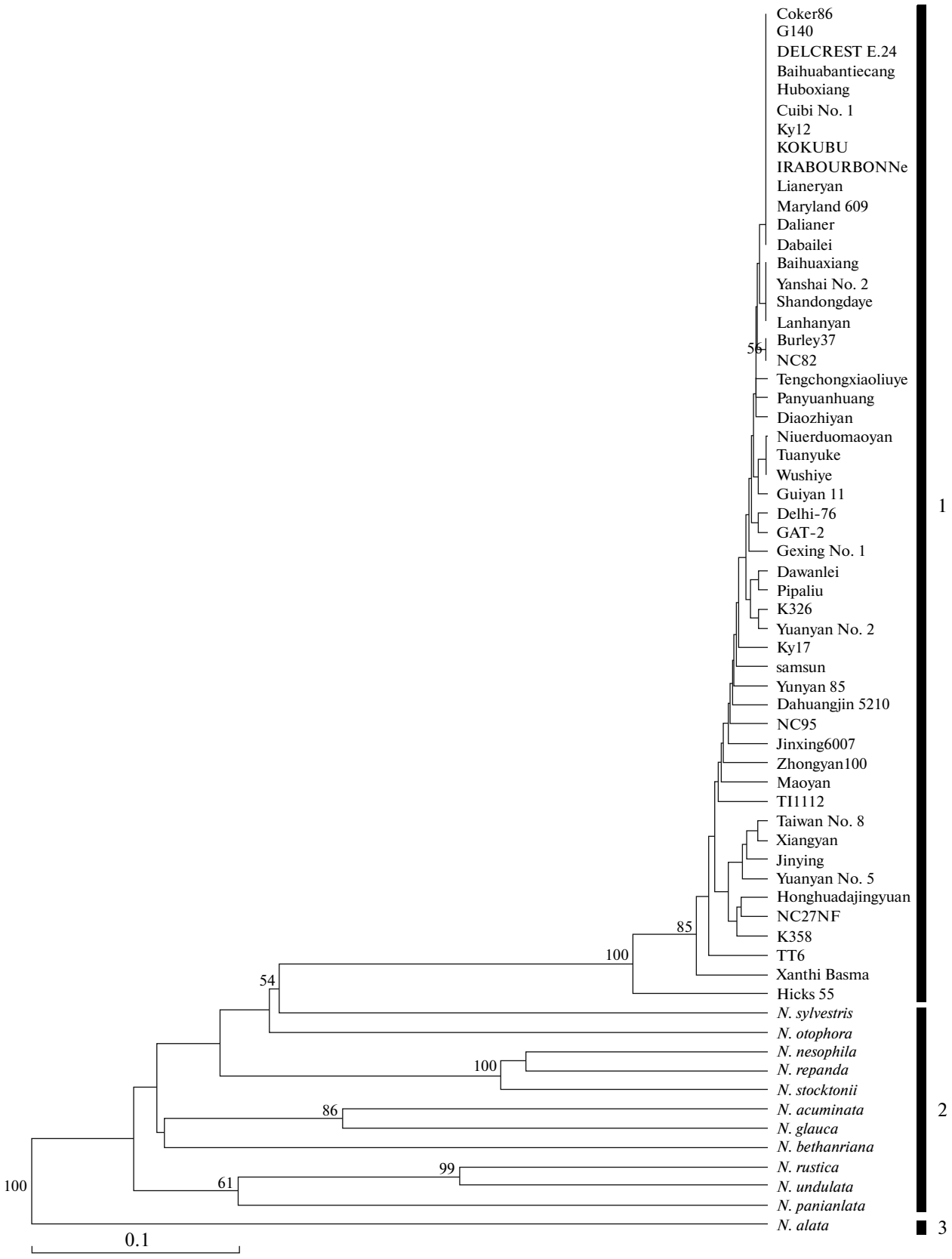

Fig. 2. UPGMA dendrogram derived from PIP markers. Bootstrap values greater than 50% were placed on the branches.

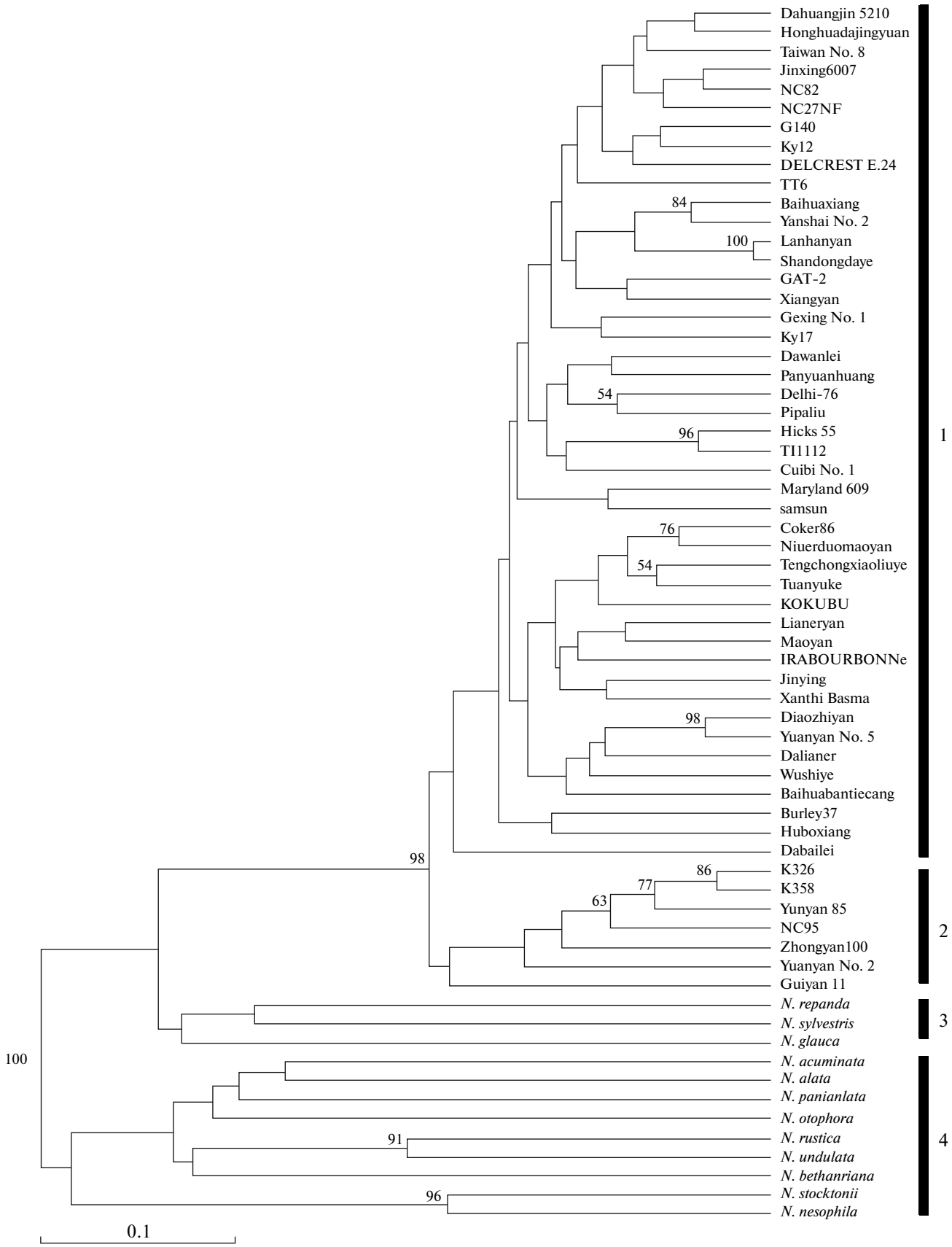

Fig. 3. UPGMA dendrogram derived from SSR markers. Bootstrap values greater than 50% were placed on the branches.

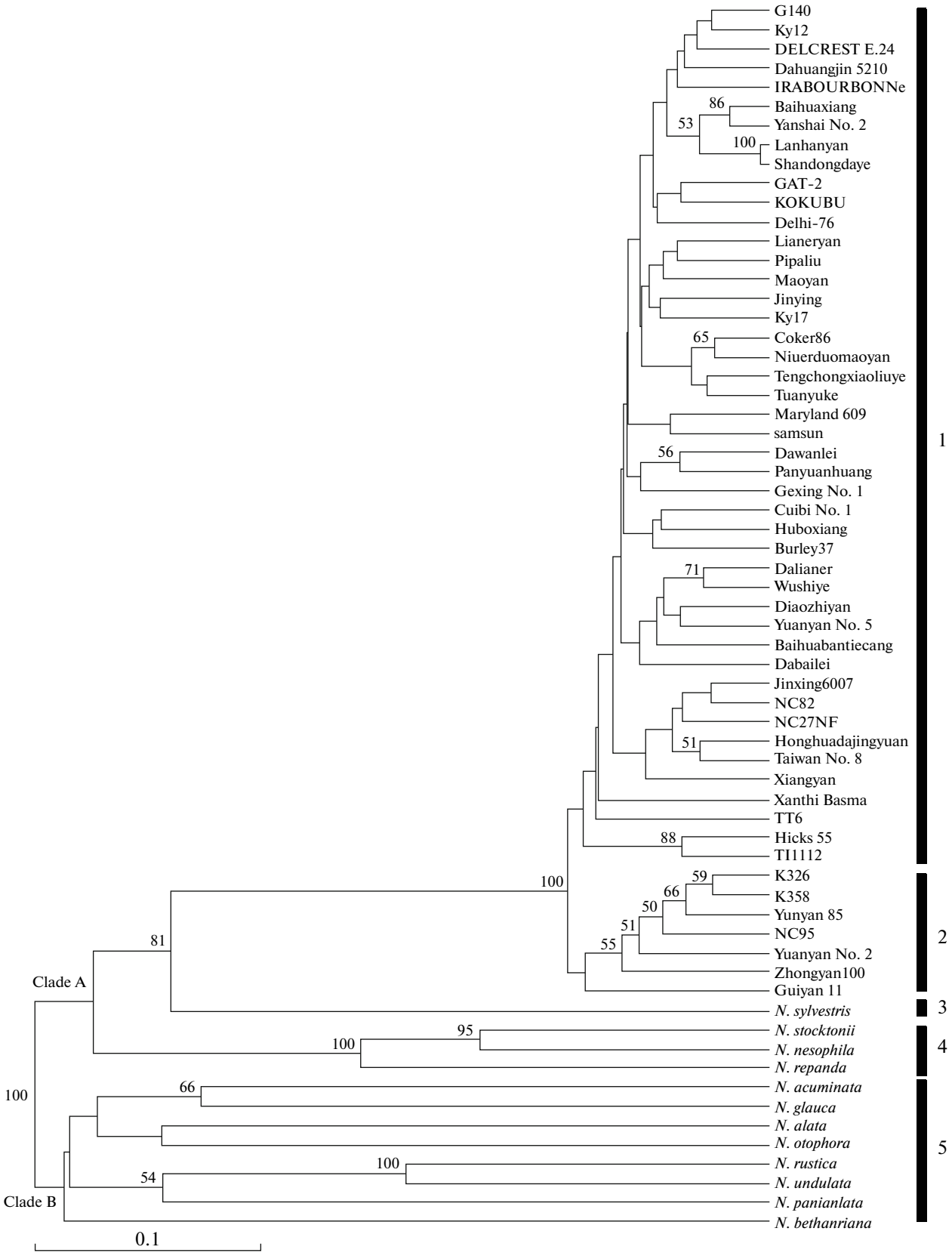

Fig. 4. UPGMA dendrogram derived from combined markers (PIP and SSR). Bootstrap values greater than 50% were placed on the branches.

larity. It was therefore difficult to find high-yielding, high-quality and high-resistance genes in the germplasms of *N. tabacum* L. However, the average GS value for PIP among the 12 wild species was 0.41, which was slightly higher than that of SSR (0.35). The relatively low GS values showed that there were obvious genetic variations among the wild species of *Nicotiana*. Therefore, it was necessary to find particular genes in the wild species that would broaden the genetic background of the commercial tobacco cultivars by means of a breeding program or through transgenic approaches.

Therefore, the ability of PIP to reveal polymorphism among cultivars was inferior to that of SSR which played an important role as it identified low level diversity among cultivated tobacco [1, 8, 10]. However, SSR was not suitable for estimating genetic similarities between relatively far taxa because of their rapid evolution and replication-slippage mechanisms [12, 36]. So, PIP and SSR had their own positive characteristics and could be complementary in the genome sequences. It is expected that the complete tobacco genome will confirm this conclusion. Furthermore, it is also presumed that PIP and SSR markers are distributed in the genomes of most plant species [19].

Comparison Analysis of the Cluster between PIP and SSR

Based on the genetic similarity coefficient derived from PIP and SSR, two UPGMA dendrograms were constructed (Figs. 2 and 3, respectively). Mantel testing demonstrated a high correlation coefficient ($r = 0.949$, $P < 0.001$) between PIP and SSR. Both of the dendrograms clearly differentiated wild species and cultivars. SSR (Fig. 3) effectively separated the cultivated tobacco even though genetic similarities were relatively high. However, for the pro-parent (*N. sylvestris*) of *N. tabacum* L., SSR was not as effective as PIP in gathering them together.

Cluster Analysis of Combined Markers

A dendrogram was constructed based on the combined markers (Fig. 4). The results showed that the dendrogram scattered 64 accessions into 5 groups. The 52 cultivars fell into 2 general groups while the 12 wild accessions fell into 3 groups. These 5 groups were divided into two main clades. In clade A (groups 1 to 4), members had the same chromosome number $2n = 48$ except *N. sylvestris* ($2n = 24$, group 4). Three wild species in group 4 (*N. repanda, N. nesophila* and *N. stocktonii*) belonged to *Nicotiana sect. Repandae*. In clade B (group 5), eight individuals were clustered together. Most of them contained less than 48 chromosomes except *N. rustica* ($2n = 48$).

There are five tobacco types and they can be distinguished by plant type. These are: leaf morphology, cured-leaf chemistry, senescence habit, and smoke flavor [37]. Therefore, the different tobacco types should have distinct genetic features and that produced the same type of tobacco tended to be grouped together. For instance, group 2 was composed of varieties that produced the same type of tobacco. However, in this study, the cluster results from the combined markers showed that 52 *N. tabacum* L. accessions did not completely group together in accordance with their tobacco types [7, 38]. Most of the accessions that belonged to different types of tobacco were found to be clustered in the same group, such as group 1.

In group 2, it was supported by high bootstrap values. Pedigree information showed that K326, K358, Yunyan 85 and Zhongyan 100 shared common genetic backgrounds with NC95. Some of them even had similar parental backgrounds. For example, K326 was a parent of K358 and Yunyan 85. Analysis of genetic diversity, using 702 *N. tabacum* accessions, revealed little association between clustering patterns [14]. Therefore, varieties belonging to different tobacco types may have had some unique genetic traits but did not show absolute separation. Overall, the final clustering tree absorbed the characteristics of both markers and showed more clearly the actual clustering results.

The combined use of PIP and SSR markers will serve as useful tools for future research into modern breeding, such as marker assisted selection, high density genetic map construction, quantitative trait locus (QTL) mapping and gene cloning.

ACKNOWLEDGMENTS

This work was supported by grant from the National Natural Science Foundation of China (project no. 30900780), China Postdoctoral Science Foundation (project nos. 20090461260 and 201104647), Postdoctoral Foundation of Shandong Agricultural University (project no. 76267), and Science and Innovation Foundation for Young Scientists of Shandong Agricultural University (project no. 23677). We are grateful to Dr. Jiajie Wu from College of Agronomy and Dr. Yalin Sang from College of Forestry, Shandong Agricultural University for helpful suggestions on the manuscript.

REFERENCES

1. Arslan, B. and Okumus, A., Genetic and geographic polymorphism of cultivated tobaccos (*Nicotiana tabacum*) in Turkey, *Russ. J. Genet.*, 2006, vol. 42, no. 6, pp. 667–671.
2. Lim, K.Y., Matyasek, R., Kovarik, A., and Leitch, A.R., Genome evolution in allotetraploid *Nicotiana*, *Biol. J. Linn. Soc.*, 2004, vol. 82, no. 4, pp. 599–606.
3. Knapp, S., Chase, M.W., and Clarkson, J.J., Nomenclatural changes and a new sectional classification in *Nicotiana* (Solanaceae), *Taxon*, 2004, vol. 53, no. 1, pp. 73–82.

4. Goodspeed, T.H., *The Genus Nicotiana*, Waltham: Mass. Chronica Botanica, 1954.
5. Arumuganathan, K. and Earle, E., Nuclear DNA content of some important plant species, *Plant Mol. Biol. Rep.*, 1991, vol. 9, no. 3, pp. 208–218.
6. Lewis, R.S., *Nicotiana Wild Crop Relatives: Genomic and Breeding Resources*, Berlin: Springer-Verlag, 2011.
7. Bindler, G., van der Hoeven, R., Gunduz, I., et al., A microsatellite marker based linkage map of tobacco, *Theor. Appl. Genet.*, 2007, vol. 114, no. 2, pp. 341–349.
8. Ren, N. and Timko, M.P., AFLP analysis of genetic polymorphism and evolutionary relationships among cultivated and wild *Nicotiana* species, *Genome*, 2001, vol. 44, no. 4, pp. 559–571.
9. Zhang, H.Y., Liu, X.Z., He, C.S., and Yang, Y.M., Genetic diversity among flue-cured tobacco cultivars based on RAPD and AFLP markers, *Braz. Arch. Biol. Technol.*, 2008, vol. 51, pp. 1097–1101.
10. Yang, B.C., Xiao, B.G., Chen, X.J., and Shi, C.H., Genetic diversity of flue-cured tobacco varieties based on ISSR markers, *Yi Chuan*, 2005, vol. 27, no. 5, pp. 753–758.
11. Tóth, G., Gáspári, Z., and Jurka, J., Microsatellites in different eukaryotic genomes: Survey and analysis, *Genome Res.*, 2000, vol. 10, no. 7, pp. 967–981.
12. Powell, W., Morgante, M., Andre, C., et al., The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis, *Mol. Breed.*, 1996, vol. 2, no. 3, pp. 225–238.
13. Davalieva, K., Maleva, I., Filiposki, K., et al., Genetic variability of Macedonian tobacco varieties determined by microsatellite marker analysis, *Diversity*, 2010, vol. 2, no. 4, pp. 439–449.
14. Moon, H.S., Nifong, J.M., Nicholson, J.S., et al., Microsatellite-based analysis of tobacco (*Nicotiana tabacum* L.) genetic resources, *Crop Sci.*, 2009, vol. 49, no. 6, pp. 2149–2159.
15. Moon, H.S., Nifong, J.M., Heineman, A., et al., Changes in genetic diversity of U.S. flue-cured tobacco germplasm over seven decades of cultivar development, *Crop Sci.*, 2009, vol. 49, no. 2, pp. 498–508.
16. Moon, H.S., Nicholson, J.S., and Lewis, R.S., Use of transferable *Nicotiana tabacum* L. microsatellite markers for investigating genetic diversity in the genus *Nicotiana*, *Genome*, 2008, vol. 51, no. 8, pp. 547–559.
17. Bindler, G., Plieske, J., Bakaher, N., et al., A high density genetic map of tobacco (*Nicotiana tabacum* L.) obtained from large scale microsatellite marker development, *Theor. Appl. Genet.*, 2011, vol. 123, no. 2, pp. 219–230.
18. Wang, X.S., Zhao, X.Q., Zhu, J., and Wu, W.R., Genome-wide investigation of intron length polymorphisms and their potential as molecular markers in rice (*Oryza sativa* L.), *DNA Res.*, 2006, vol. 12, no. 6, pp. 417–427.
19. Zhao, X.Q., Yang, L., Zheng, Y., et al., Subspecies-specific intron length polymorphism markers reveal clear genetic differentiation in common wild rice (*Oryza rufipogon* L.) in relation to the domestication of cultivated rice (*O. sativa* L.). *J. Genet. Genomics*, 2009, vol. 36, no. 7, pp. 435–442.
20. Huang, M., Xie, F.M., Chen, L.Y., et al., Comparative analysis of genetic diversity and structure in rice using 1LP and SSR markers, *Rice Sci.*, 2010, vol. 17, no. 4, pp. 257–268.
21. Dong, Q., Schlueter, S.D., and Brendel, V., PlantGDB, plant genome database and analysis tools, *Nucleic Acids Res.,* 2004, vol. 32, pp. D354–D359.
22. Yang, L., Jin, G., Zhao, X.Q., et al., PIP: A database of potential intron polymorphism markers, *Bioinformatics*, 2007, vol. 23, no. 16, pp. 2174–2177.
23. Wang, Y.Y., Chen, J., Francis, D., et al., Discovery of intron polymorphisms in cultivated tomato using both tomato and *Arabidopsis* genomic information, *Theor. Appl. Genet.*, 2010, vol. 121, no. 7, pp. 1199–1207.
24. Chen, X., Zhang, G., and Wu, W., Investigation and utilization of intron length polymorphisms in conifers, *New For.*, 2011, vol. 41, no. 3, pp. 379–388.
25. Liu, H.L., Lin, Y.A., Chen, Y., et al., Genome-scale identification of resistance gene analogs and the development of their intron length polymorphism markers in maize, *Mol. Breed.*, 2011, pp. 1–11.
26. Murray, M.G. and Thompson, W.F., Rapid isolation of high molecular weight plant DNA, *Nucleic Acids Res.*, 1980, vol. 8, no. 19, pp. 4321–4326.
27. Rozen, S. and Skaletsky, H., Primer3 on the WWW for general users and for biologist programmers, *Methods Mol. Biol.*, 2000, vol. 132, pp. 365–86.
28. Schuler, G.D., Sequence mapping by electronic PCR, *Genome Res.*, 1997, vol. 7, no. 5, pp. 541–550.
29. Saal, B. and Wricke, G., Development of simple sequence repeat markers in rye (*Secale cereale* L.), *Genome*, 1999, vol. 42, no. 5, pp. 964–972.
30. Nei, M., Analysis of gene diversity in subdivided populations, *Proc. Natl. Acad. Sci. U.S.A.*, 1973, vol. 70, no. 12, pp. 3321–3323.
31. Nei, M. and Li, W.H., Mathematical model for studying genetic variation in terms of restriction endonucleases, *Proc. Natl. Acad. Sci. U.S.A.*, 1979, vol. 76, no. 10, pp. 5269–5273.
32. Mantel, N., The detection of disease clustering and a generalized regression approach, *Cancer Res.*, 1967, vol. 27, no. 2, pp. 209–220.
33. Rohlf, J.F., *NTSYSpc: Numerical Taxonomy and Multivariate Analysis System*, Setauket: Exeter Software, 2000.
34. Hampl, V., Pavlicek, A., and Flegr, J., Construction and bootstrap analysis of DNA fingerprinting-based phylogenetic trees with the freeware program FreeTree: Application to trichomonad parasites, *Int. J. Syst. Evol. Microbiol.*, 2001, vol. 51, no. 3, pp. 731–735.
35. Ali, M., Rajewski, J., Baenziger, P., et al., Assessment of genetic diversity and relationship among a collection of U.S. sweet sorghum germplasm by SSR markers, *Mol. Breed.*, 2008, vol. 21, no. 4, pp. 497–509.
36. Bowcock, A.M., Ruizlinares, A., Tomfohrde, J., et al., High-resolution of human evolutionary trees with polymorphic microsatellites, *Nature*, 1994, vol. 368, no. 6470, pp. 455–457.
37. Lewis, R. and Nicholson, J., Aspects of the evolution of *Nicotiana tabacum* L. and the status of the United States *Nicotiana* germplasm collection, *Genet. Resour. Crop. Evol.*, 2007, vol. 54, no. 4, pp. 727–740.
38. Julio, E., Verrier, J.L., and de Borne, F., Development of SCAR markers linked to three disease resistances based on AFLP within *Nicotiana tabacum* L., *Theor. Appl. Genet.*, 2006, vol. 112, no. 2, pp. 335–346.