Plant Breeding
© 2011 Blackwell Verlag GmbH
doi:10.1111/j.1439-0523.2011.01872.x

iSNAP: a small RNA-based molecular marker technique

Yijie Gui[1], Guanghao Yan[1], Shiping Bo[1,2], Zhijun Tong[1], Yu Wang[1], Bingguang Xiao[3], Xiuping Lu[3], Yongping Li[3], Weiren Wu[1] and Longjiang Fan[1,2,4]

[1]Department of Agronomy, Zhejiang University, Hangzhou 310029, Zhejiang, China; [2]James D. Watson Institute of Genome Sciences, Zhejiang University, Hangzhou 310029, Zhejiang, China; [3]Yunnan Academy of Tobacco Agricultural Sciences and China Tobacco Breeding Research Center at Yunnan, Yuxi 653100, Yunnan, China; [4]Corresponding author, E-mail: fanlj@zju.edu.cn

With 4 figures and 3 tables

Received December 22, 2010|Accepted April 6, 2011
Communicated by C. Jung

Abstract

Endogenous non-coding small RNAs usually consist of 20–24 nucleotides, are found ubiquitously throughout the genome and play important regulatory roles in most eucaryotes. In this study, a new kind of PCR-based marker approach using multi-mapped small RNA sequences, called *i*nter *s*mall R*NA p*olymorphism (iSNAP), was developed. iSNAP primers were designed based on multi-mapped small RNA and its two end conserved flanking sequences. To demonstrate the marker system, 10 rice and 30 tobacco iSNAP primers were developed and used to construct a rice genetic map and to fingerprint eight tobacco varieties. The results indicate that this technique can be effectively applied in the mapping of rice and fingerprinting of species (such as tobacco) with extremely low genetic diversity and thus provides a new option of molecular marker techniques.

Key words: small RNA — inter small RNA polymorphism — tobacco — molecular marker technique

Since Botstein et al. (1980) first used a molecular marker technique, restriction fragment length polymorphism, for their human genetic map in 1980, many different kinds of molecular markers have been developed for the detection of DNA polymorphism (Agarwal et al. 2008). Basic molecular marker techniques can be classified into two types: non-PCR-based and PCR-based techniques. PCR-based techniques can be further subdivided into two groups: arbitrarily primed (such as random amplified polymorphic DNA, RAPD, and amplified fragment length polymorphism, AFLP) (Williams et al. 1990, Vos et al. 1995) and sequence-targeted PCR-based techniques. In the sequence-targeted group, in addition to marker techniques based on the interspersed repeat sequences (such as simple sequence repeat or SSR) (Litt and Luty 1989, Tautz 1989, Meyer et al. 1993, Wang et al. 1994, Zietkiewicz et al. 1994) and single-nucleotide polymorphism (Syvanen 2001, Rafalski 2002), transposable element-based molecular markers have been developed. For these kinds of markers, PCR primers were designed based on conserved short sequences of widespread, abundant transposable elements in genomes. Examples include the inter-retrotransposon amplified polymorphism (IRAP) (Waugh et al. 1997, Kalendar et al. 1999) and inter-MITE polymorphism (IMP) (Chang et al. 2001) markers that use long terminal repeat (LTR) of retrotransposons and terminal inverted repeats (TIR) of miniature inverted-repeat transposable elements (MITEs) for their primer sequences, respectively. Recently, a universal method (iPBS) for DNA fingerprinting and retrotransposon isolation was also developed (Kalendar et al. 2010).

Endogenous non-protein-coding small RNAs usually contain 21–24 nucleotides (nt) and play an important role in post-transcriptional gene regulation in plants. In general, small RNAs can be divided into two major classes, microRNAs (miRNAs) and short-interfering RNAs (siRNAs), based on the mechanisms by which they are synthesized and their function (Dugas and Bartel 2004). Small RNA sequences, such as miRNAs or their precursors, are usually conserved. However, their flanking sequences are usually polymorphic, and their sequences contain structural variations (such as insertions/deletions) (e.g. in rice by Wang et al. 2010).

We previously generated over 5 million 18–35 nt small RNA reads from rice via high-throughput sequencing using an Illumina genome analysis system (Zhu et al. 2008) (GSE11014). We mapped these reads into the rice genome and found that some small RNAs could be mapped to multiple genomic positions in the genome. This encourages us to try a new kind of PCR-based marker approach using these multi-mapped small RNA sequences. Our investigations in rice and tobacco indicated that this new kind of marker, called *i*nter *s*mall R*NA p*olymorphism (iSNAP), works well in rice and in species with extremely low genetic diversity (e.g. tobacco) and thus provides a new option for molecular marker techniques.

Materials and Methods

Plant materials: The mapping population consisted of 131 recombination inbred lines (RILs) that were developed from a cross of two *indica* rice (*Oryza sativa*) cultivars ('H359' and 'Acc8558') (Tang et al. 2000). Eight tobacco (*Nicotania tabacum*) cultivars representing the main types of tobacco were used for fingerprinting (Table 1).

Table 1: Tobacco varieties used for fingerprinting

Code	Variety	Type
T1	Honghuadajinyuan (HD)	Flue-cured
T2	K326	Flue-cured
T3	Hicks Broad Leaf	Flue-cured
T4	Florida301	Cigar
T5	Burley21	Burley
T6	TN86	Burley
T7	Samsun	Oriental
T8	Turkey Basma	Oriental

Small RNA and genome sequences: Rice genome sequences were downloaded from the Rice Genome Annotation Project (The Institute of Genomic Research, http://www.tigr.org, release 5.0), and over 12 million tobacco genomic survey sequences (GSS) determined by the Tobacco Genome Initiative (TGI) (http://www.tobaccogenome.org) were downloaded from GenBank. An unredundant assembling set based on the tobacco GSSs was generated by MIRA (Chevreux et al. 2004) with de novo assembling strategy which finally obtained nearly 300 000 contigs accumulating 243 mega bases with a maximum contig length of 71 127 nt and N50 of 942 nt. This data set was finally used as tobacco genome sequence. Small RNAs that were generated from the rice cultivar 'Nipponbare' (Zhu et al. 2008) (GSE11014) and tobacco cultivar Hicks Broad Leaf (D. Tang et al. unpublished results) were used to develop iSNAP markers.

iSNAP primers and PCR detection: Based on the multi-mapped small RNA sequences, eight small RNAs selected from rice and 30 small RNAs from tobacco were used to design iSNAP primers (Tables 2 and 3; for details see Results section). Their primer sequences are available from electronic supplementary Table S1.

PCR was performed in 10 μl volumes containing approximately 40–60 ng of template DNA, 1× PCR buffer (50 mM KCl, 10 mM Tris–HCl pH 8.8), 1.5 mM $MgCl_2$, 0.5 μM of each primer (1.0 μM of single primer), 0.2 mM dNTPs and 0.6 U of *Taq* polymerase (Dingguo,

Table 2: Rice iSNAP primer sequences used in this study

Primer_combination	Marker_name	Forward_primer	Reverse_primer	AT
RS1F-RS5R	iSNAP15	GGTGAAGACGTAGACGCCCAACA	AAGATGTACCCCTCTGAT	55
RS1F-RS6R	iSNAP16	GGTGAAGACGTAGACGCCCAACA	TAAAGATGTACACCTCTGAT	60
RS2F-RS6R	iSNAP26	TGGCATTAATTGACAAAAGTTC	TAAAGATGTACACCTCTGAT	60
RS2F-RS10R	iSNAP210	TGGCATTAATTGACAAAAGTTC	GGTCGGGTTGCCAGTTC	60
RS5F-RS4R	iSNAP54	TCTGATCAGAGGGGTACATCTT	TATTATGCCGCCGCTGTTGGG	60
RS5F-RS6R	iSNAP56	TCTGATCAGAGGGGTACATCTT	TAAAGATGTACACCTCTGAT	60
RS5F-RS7R	iSNAP57	TCTGATCAGAGGGGTACATCTT	TCCTACACGGCTGGCTAT	60
RS5F-RS9R	iSNAP59	TCTGATCAGAGGGGTACATCTT	TTGGGTTAAGGCAGGTTCGTGT	60
RS6F-RS5R	iSNAP65	TCTGATCAGAGGTGTACATCTTTA	AAGATGTACCCCTCTGAT	60
RS6F-RS7R	iSNAP67	TCTGATCAGAGGTGTACATCTTTA	TCCTACACGGCTGGCTAT	58
RS6F-RS9R	iSNAP69	TCTGATCAGAGGTGTACATCTTTA	TTGGGTTAAGGCAGGTTCGTGT	58
RS7F-RS9R	iSNAP79	ATAGGCCAGCCGTGTAGGA	TTGGGTTAAGGCAGGTTCGTGT	58
RS7F-RS10R	iSNAP710	ATAGGCCAGCCGTGTAGGA	GGTCGGGTTGCCAGTTC	55

Table 3: Tobacco iSNAP primer sequences used in this study

Primer	Forward_primer	Reverse_primer	AT
TS1F-TS2R	TTAAGTTCTAGGTCGCCCAC	TTCTAGGTCGCCCACCAGTATA	52
TS1F-TS15R	TTAAGTTCTAGGTCGCCCAC	GGCGTCCACCTGGAGAATGAGG	52
TS1F-TS17R	TTAAGTTCTAGGTCGCCCAC	GGGAGGTCCACGGGTTATGC	52
TS2F-TS27R	TATACTGGTGGGCGACCTA	GCTCAACTCAGGGATTGGAGCCCT	52
TS2F-TS28R	TATACTGGTGGGCGACCTA	TCATATCCTAAAGGTTCAACTCA	58
TS7F-TS18R	AGTTTAAATAGGCGCCCACCTGTA	ACCTTCTCCTAGGATCACAA	58
TS11F-TS6R	AGTCAGCATTAGGGTTGTAG	TTCATCCGCCCTCAAATAG	52
TS11F-TS18R	AGTCAGCATTAGGGTTGTAG	ACCTTCTCCTAGGATCACAA	58
TS15F-TS18R	CCTCATTCTCCAGGTGGACGCC	ACCTTCTCCTAGGATCACAA	58
TS16F-TS2R	AGAGATTGGAGCCCTAATGTCGG	TTCTAGGTCGCCCACCAGTATA	52
TS16F-TS3R	AGAGATTGGAGCCCTAATGTCGG	GCATTATACTGGTGGGCGAA	52
TS16F-TS7R	AGAGATTGGAGCCCTAATGTCGG	TACAGGTGCGCGCCTATTTAC	52
TS16F-TS9R	AGAGATTGGAGCCCTAATGTCGG	GTCTTAGTTTAAATAGGCGAC	52
TS16F-TS11R	AGAGATTGGAGCCCTAATGTCGG	GGTTTAAACCCTAATGCTGAC	52
TS16F-TS15R	AGAGATTGGAGCCCTAATGTCGG	GGCGTCCACCTGGAGAATGAGG	52
TS16F-TS17R	AGAGATTGGAGCCCTAATGTCGG	GGGAGGTCCACGGGTTATGC	52
TS18F-TS16R	TTTTGATCCTAGGAGAAGGT	GCCGACATTAGGGCTCCAATCCCT	52
TS24F-TS1R	AGGACATGGTGTCCTTAACTTTGA	GTGGGCGACCTAGAACTTA	52
TS27F-TS1R	AGGGCTCCAATCCCTGAGTTGAG	GTGGGCGACCTAGAACTTA	58
TS27F-TS7R	AGGGCTCCAATCCCTGAGTTGAG	TACAGGTGCGCGCCTATTTAC	58
TS27F-TS12R	AGGGCTCCAATCCCTGAGTTGAG	ATCCTCATTCTCCAGGTGGA	52
TS27F-TS16R	AGGGCTCCAATCCCTGAGTTGAG	GCCGACATTAGGGCTCCAATCCCT	52
TS27F-TS17R	AGGGCTCCAATCCCTGAGTTGAG	GGGAGGTCCACGGGTTATGC	52
TS27F-TS26R	AGGGCTCCAATCCCTGAGTTGAG	CTTTTGTGGATATGAGTTGAACCT	52
TS1F	TTAAGTTCTAGGTCGCCCAC		
TS3F	AGGTCGCCCACCAGTATAATGCGG		52
TS7F	AGTTTAAATAGGCGCCCACCTGTA		52
TS11F	AGTCAGCATTAGGGTTGTAG		52
TS15F	CCTCATTCTCCAGGTGGACGCC		52
TS27F	AGGGCTCCAATCCCTGAGTTGAGC		52
TS28F	ATGAGTTGAACCTTTAGGATATGA		52
TS3R		GCATTATACTGGTGGGCGAA	52
TS4R		CCAGTATAATGCGGGAATAC	52
TS28R		GCCGACATTAGGGCTCCAATCCCT	52

AT, annealing temperature.

Beijing, China). Thermocycling consisted of an initial denaturation step for 5 min at 95°C, 35 cycles of 45 s at 95°C, 45 s at 58°C and 1 min 30 s at 72°C, and a final extension step of 72°C for 10 min. PCR products were mixed with 10 µl of formamide loading dye (98% deionized formamide, 10 mM EDTA, 0.2% mg/ml bromophenol blue and 0.2% mg/ml Xylene Cyanole FF). The mixture was denatured for 5 min at 95°C and immediately chilled on ice for at least 10 min, and aliquots were separated by electrophoresis (DYCZ-20C; Beijing Liuyi Instrument Factory, Beijing, China) through 6% (w/v) polyacrylamide gels (Sambrook and Russell 2001) under the following conditions: 1800 V, 60 mA, 60 W, for 2 h.

Data analysis: A genetic map of rice consisting of 147 RFLP and 78 AFLP markers was previously constructed based on a recombination inbred population derived from a cross between two *indica* rice cultivars H359 × AC8558 (Tang et al. 2000). The map was then reconstructed by removing all the AFLP markers and adding 72 SSR markers (unpublished). Based on this map, iSNAP markers were mapped. The genetic mapping program MAPMAKER/EXP 3.0 (Lander et al. 1987) was used for data analysis. The order of RFLP and SSR markers in the original map was fixed, and the TRY command was then used to determine the locations of iSNAP markers. After the marker order was determined, the MAP command was used to calculate the distances between adjacent markers. The genetic map was drawn using the program MAPDRAW V2.1 (Liu and Meng 2003).

The matrix of polymorphic bands for each individual was scored as present (1), absent (0) or missing data (3). The data were then used to generate relative genetic similarity (GS) matrices using the method of Nei and Li (1979). The phylogenetic tree was constructed based on the GS matrices data using the neighbour-joining method. All statistical analyses were performed with NTSYSPC 2.10E software (Rohlf 2000). Weblogo was used to generate sequence logo (Crooks et al. 2004).

Reproducibility test: To assess the reproducibility of the iSNAP approach, each sample reaction was set internal repeat and amplified at three different annealing temperatures. The polymorphic bands were considered to be reproducible when they were present or absent in the internal repeat at least twice under the same conditions.

Results

Identification of multi-mapped small RNAs

Over 5 million 18–35 nt small RNA reads obtained via high-throughput sequencing from rice genome (va. Nipponbare) in our previous studies (Zhu et al. 2008) (GSE11014) were mapped into its genome (TIGR RELEASE 5.0). Many small RNAs could be perfectly mapped (i.e. without mismatch) to many different genomic positions in the genome. They usually were mapped to sense and antisense strands of chromosome sequences, respectively (Fig. 1 and Fig. S1). Eight small RNAs with the highest number of mapped positions are shown in Tables 2 and S1. Their genomic positions are well distributed across the whole genome. On average, chromosome 1–12 have 160, 146, 59, 249, 176, 138, 171, 176, 133, 83, 65 and 154 of the eight multi-mapped small RNAs, or 30, 33, 13, 56, 47, 35, 46, 50, 46, 29, 16 and 45 small RNAs per 1 Mb, respectively. Most of the small RNAs are located in inter-LTR regions of repetitive retrotransposons according to the Rice Genome Annotation Project.

To evaluate the presentation of iSNAP for species with low genetic diversity, we choose tobacco as an example. Tobacco is a highly homozygous plant based on molecular markers (such as AFLP) or genes (Ren and Timko 2001, Julio et al. 2006, Leng et al. 2010). Many crops have experienced a severe bottleneck in domestication, and this apparent lack of molecular diversity could be related to the strong domestication selection of tobacco, and/or perhaps only a few genotypes became the progenitors of most modern cultivars during the domestication of tobacco (Wernsman 1999). Small RNA populations from tobacco roots and leaves were sequenced using a high-throughput approach, and over 5 million reads were generated (D. Tang et al., unpublished results). We mapped these small RNAs into the tobacco genome, and some small RNAs could be perfectly mapped to many tobacco contigs. Thirty small RNAs with the different copy number of mapped contigs were selected for our further evaluation of iSNAP on tobacco (Table S1). Most of the selected small RNAs did not come from the repeat-related regions.

Identification of the high-copy number ($>10^3$) small RNAs may make it possible to develop iSNAP markers in the two crops.

Primer design of iSNAP marker

As mentioned earlier, structural variations at small RNA flanking sequences observed in the genome provide the genetic variations for iSNAP. The basic idea of iSNAP is to use the length polymorphism of PCR products by a primer pair from two flanking small RNAs or a single small RNA. Therefore, an iSNAP primer pair can be designed based on a single multi-mapped small RNA and/or its 5′ and 3′ conserved flanking sequences and used as an iSNAP primer pair (e.g. small RNA TS1F/R); a primer pair can also come from the combination of two multi-mapped small RNAs (e.g. RS5F/RS4R and TS2F/TS19R) (Table S1). The combination of iSNAP primers can lead to a large number of new iSNAP primer pairs. A similar annealing temperature for iSNAP primers can be expected with the use of the flanking sequences of small RNAs. Degenerated bases can be used based on the alignment of flanking sequence (see next paragraph).

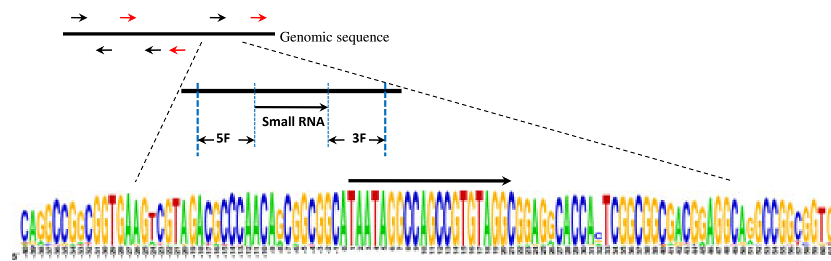

Fig. 1: Genomic distribution of multi-mapped small RNAs and the area for iSNAP primer design. Two different small RNAs (black and red arrows) and their directions were presented in the genomic sequence. Primers can be designed in the corresponding small RNA and its two ends conserved flanking sequences (5F and 3F). The sequence logo bases on a rice small RNA (Table 2) and its flanking sequences

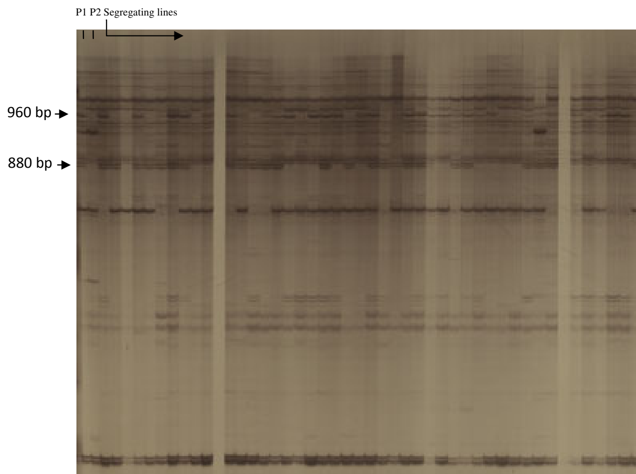

Fig. 2: A subset of the PCR results for a rice iSNAP primer (RS2F/10R) visualized on 6% polyacrylamide gel. The parent *indica* rice cultivars H359 (P1), Ac8558 (P2) and their segregating recombination inbred lines are shown at top line. Putative iSNAP markers are labelled with arrows in the left column

The flanking end sequences of small RNAs were conserved within a short distance. An example from a rice small RNA (zhu0607-87438-G1-1-G6-0) is shown in Fig. 1. On average, 44 and 43 bp of conserved 5′ and 3′ flanking sequences (90% identity) were observed in the eight rice small RNAs, respectively. A similar result was also obtained for tobacco small RNAs (data not shown).

Chromosomal localization of iSNAP markers

To evaluate the utility of small RNAs in genetic mapping, ten iSNAP primers were developed from rice small RNAs. Based on the 10 primers, 13 pairs of primer combinations were selected to construct a genetic map in an RIL population of 131 individuals from a cross between *indica* cultivars. Approximately 9–40 clear bands were detected within these primer combinations. The size range of the bands was approximately 100–1500 bp. Typical findings in polyacrylamide gel electrophoreses for the results of PCR amplification with an iSNAP primer pair (RS2F/RS10R) are shown in Fig. 2.

Thirty-nine polymorphic bands between the two parental lines were detected with the 13 pairs, and each primer pair showed one to seven such bands. All the polymorphic bands exhibited a presence/absence pattern between those of the two parental lines. Finally, 30 iSNAP markers were mapped into seven linkage groups (Fig. 3). The distribution of the mapping loci showed no significant clustering, suggesting that iSNAPs may be able to cover the entire genome using a limited set of iSNAP markers.

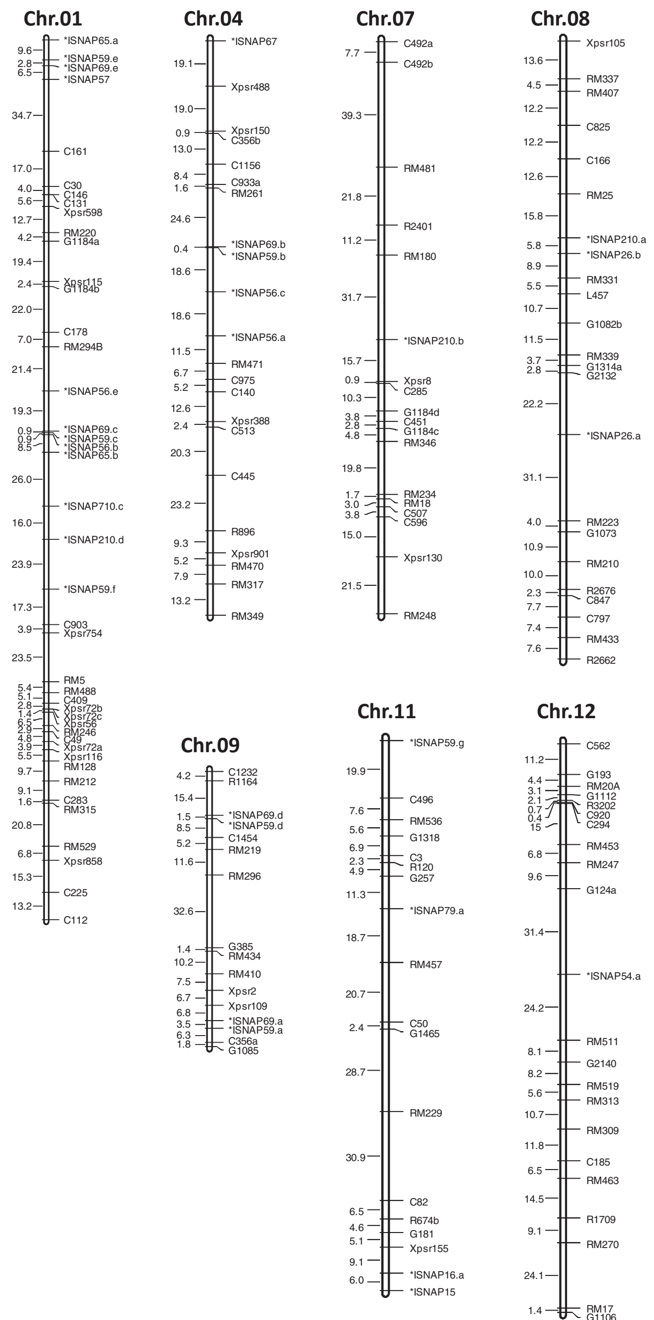

Fig. 3: Linkage map of the recombination inbred line population of *indica* rice cultivar H359 and Ac8338. The newly placed iSNAP markers are shown in asterisks

Development of tobacco iSNAP markers

To assess the function of iSNAPs for fingerprinting and GS analyses, a total of eight tobacco accessions, which included the main types of tobacco cultivars (Table 1), were collected and 30 iSNAP primers from tobacco small RNA populations were developed. A total of 41 polymorphic bands were scored using the 10 single primers and 24 pairs of primer combinations across the eight cultivars. Typical results of PCR are shown in Fig. 4. The 41 polymorphic bands were used to construct a neighbour-joining phylogenetic tree that separates the eight accessions into four tobacco types (Fig. S2).

Discussion

In this report, we have demonstrated that a small RNA could be mapped over 1000 genomic positions in crop genomes and could be used to anchor sequences in the development of molecular markers. iSNAP is unique and does not overlap other molecular marker systems, such as the retrotransposon-based IRAP (Waugh et al. 1997, Kalendar et al. 1999) and MITE-based IMP (Chang et al. 2001) markers that use LTR of retrotransposons and TIR of MITEs for their primer

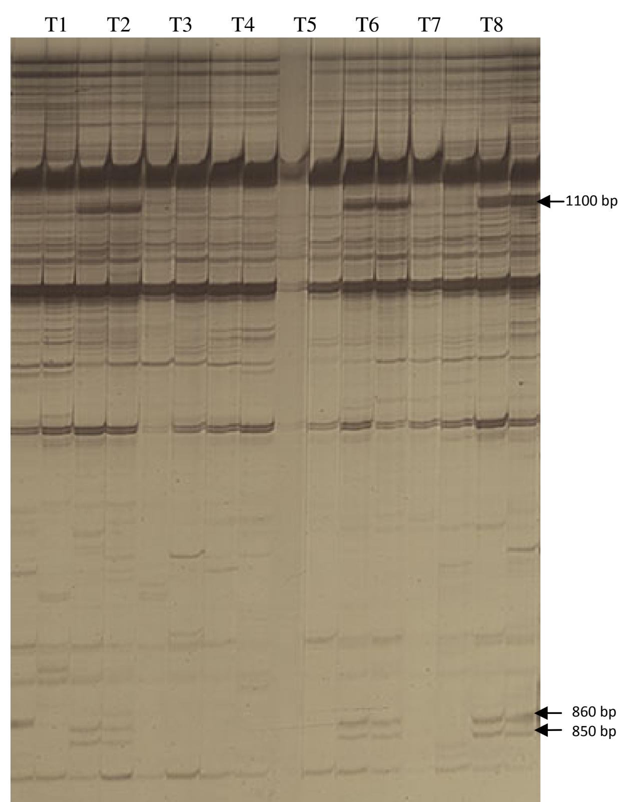

Fig. 4: Fingerprinting results obtained from eight tobacco varieties using an iSNAP primer pair. T1–T8 represent the eight tobacco cultivars (Table 1) in the top line, and two independent PCR products from the same cultivar are loaded in two adjacent lanes. Putative iSNAP markers are labelled with arrows in left column. Only a section of the polyacrylamide gel is shown

sequences, respectively. Most iSNAP used in rice is located in inter-LTR regions of retrotransposons. To our knowledge, iSNAP is the first effort to use small RNAs as molecular markers.

Three features of iSNAP marker are apparent. First, iSNAP is a high-throughput non-coding sequence-based marker system. There can be as many as 1000 copies of small RNAs in the rice and tobacco genomes, similar to MITEs (Oki et al. 2008, Kuang et al. 2009). Multiple bands, usually over 10, were usually generated by a single iSNAP primer pair in gels. Sensitive detection approaches, such as LI-COR4300 system, should further improve its efficiency. Moreover, the combination of different iSNAP primers, as in the IMP marker system, will generate many new iSNAP pairs. Although several new approaches to genotyping through high-throughput sequencing have recently been proposed, they are usually expensive. Second, iSNAP is a kind of functional (transcriptional sequences) marker. To our knowledge, small RNAs, even those of 18 nt or 30–40 nt, play functional roles in diverse organisms (Kasschau et al. 2007, Zhu et al. 2008). A recent study showed that a new class of RNAs, transcription initiation RNAs (tiRNAs), with a moderate length of 18 nucleotides that map within −60 to +120 of transcriptional start sites, exhibit regulatory function (Taft et al. 2009). The advent of high-throughput sequencing methods has enabled us to sequence many plant small RNA populations (Lister et al. 2009) and provides many candidate small RNAs for iSNAP development. Third, as mentioned above, small RNAs have been found in most organisms, which suggest that the iSNAP approach is widely applicable. This technique is promising even for plants (such as tobacco) with extremely low intraspecies genetic diversity. In brief, our results demonstrated a new marker system, a small RNA-based approach that has broad applicability.

However, the development of iSNAP strongly depends on the genome sequence and small RNA population. Although small RNAs are readily available via high-throughput sequencing, the efficiency of the markers will be limited if only these small RNA sequences are used and information to estimate their copy number is unavailable. Meanwhile, no significant conservation of those multiple-mapped small RNAs was observed among rice, tobacco and other plants, implying that iSNAP markers should be species specific. Moreover, iSNAP, IMP and other similar marker systems depend on length polymorphisms between short anchor/repeat sequences, and therefore, their present reproducibility is not as good as that of, for example, SSR. To confirm the reliability of the detected polymorphic bands, we set internal repeat under three different conditions. In our study, the primers designed from rice were more reproducible than those derived from tobacco. This is most likely due to the fact that rice iSNAP primers were designed based on a more accurate high copy number with a decoded reference genome. Generally, small RNAs with a moderate to high copy number in the genome promise good reproducibility and efficiency for iSNAP.

Acknowledgements

This study was supported by the National Basic Research Program of China (2011CB109306), National Tobacco Monopoly Bureau of China (110200701023, 08A05) and Yunnan Provincial Science and Technology Department (2008CD194).

References

Agarwal, M., N. Shrivastava, and H. Padh, 2008: Advances in molecular marker techniques and their applications in plant sciences. Plant Cell Rep. **27,** 617—631.

Botstein, D., R. L. White, M. Skolnick, and R. W. Davis, 1980: Construction of a genetic-linkage map in man using restriction fragment length polymorphisms. Am. J. Hum Genet **32,** 314—331.

Chang, R. Y., L. S. O'Donoughue, and T. E. Bureau, 2001: Inter-MITE polymorphisms (IMP): a high throughput transposon-based genome mapping and fingerprinting approach. Theor. Appl. Genet. **102,** 773—781.

Chevreux, B., T. Pfisterer, B. Drescher, A. J. Driesel, W. E. G. Muller, T. Wetter, and S. Suhai, 2004: Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. Genome Res. **14,** 1147—1159.

Crooks, G. E., G. Hon, J. M. Chandonia, and S. E. Brenner, 2004: WebLogo: a sequence logo generator. Genome Res. **14,** 1188—1190.

Dugas, D. V., and B. Bartel, 2004: MicroRNA regulation of gene expression in plants. Curr. Opin. Plant Biol. **7,** 512—520.

Julio, E., J. L. Verrier, and F. D. de Borne, 2006: Development of SCAR markers linked to three disease resistances based on AFLP within Nicotiana tabacum L. Theor. Appl. Genet. **112,** 335—346.

Kalendar, R., T. Grob, M. Regina, A. Suoniemi, and A. Schulman, 1999: IRAP and REMAP: two new retrotransposon-based DNA fingerprinting techniques. Theor. Appl. Genet. **98,** 704—711.

Kalendar, R., K. Antonius, P. Smykal, and A. H. Schulman, 2010: iPBS: a universal method for DNA fingerprinting and retrotransposon isolation. Theor. Appl. Genet. **121,** 1419—1430.

Kasschau, K. D., N. Fahlgren, E. J. Chapman, C. M. Sullivan, J. S. Cumbie, S. A. Givan, and J. C. Carrington, 2007: Genome-wide profiling and analysis of Arabidopsis siRNAs. PLoS Biol. **5,** 479—493.

Kuang, H. H., C. Padmanabhan, F. Li, A. Kamei, P. B. Bhaskar, O. Y. Shu, J. M. Jiang, C. R. Buell, and B. Baker, 2009: Identification of miniature inverted-repeat transposable elements (MITEs) and biogenesis of their siRNAs in the solanaceae: new functional implications for MITEs. Genome Res. **19,** 42—56.

Lander, E., J. Abrahamson, A. Barlow, M. Daly, S. Lincoln, L. Newburg, and P. Green, 1987: Mapmaker a computer package for constructing genetic-linkage maps. Cytogenet. Cell Genet. **46,** 642.

Leng, X. D., B. G. Xiao, S. Wang, Y. J. Gui, Y. Wang, X. P. Lu, J. H. Xie, Y. P. Li, and L. J. Fan, 2010: Identification of NBS-type resistance gene homologs in tobacco genome. Plant Mol. Biol. Rep. **28,** 152—161.

Lister, R., B. D. Gregory, and J. R. Ecker, 2009: Next is now: new technologies for sequencing of genomes, transcriptomes, and beyond. Curr. Opin. Plant Biol. **12,** 107—118.

Litt, M., and J. A. Luty, 1989: A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac-muscle actin gene. Am. J. Hum. Genet. **44,** 397—401.

Liu, R. H., and J. L. Meng, 2003: MapDraw: a Microsoft Excel macro for drawing genetic linkage maps based on given genetic linkage data. Hereditas **25,** 317—321.

Meyer, W., T. G. Mitchell, E. Z. Freedman, and R. Vilgalys, 1993: Hybridization probes for conventional DNA-fingerprinting used as single primers in the polymerase chain-reaction to distinguish strains of cryptococcus-neoformans. J. Clin. Microbiol. **31,** 2274—2280.

Nei, M., and W. H. Li, 1979: Mathematical-model for studying genetic-variation in terms of restriction endonucleases. Proc. Natl Acad. Sci. USA **76,** 5269—5273.

Oki, N., K. Yano, Y. Okumoto, T. Tsukiyama, M. Teraishi, and T. Tanisaka, 2008: A genome-wide view of miniature inverted-repeat transposable elements (MITEs) in rice, Oryza sativa ssp japonica. Genes Genet. Syst. **83,** 321—329.

Rafalski, A., 2002: Applications of single nucleotide polymorphisms in crop genetics. Curr. Opin. Plant Biol. **5,** 94—100.

Ren, N., and M. P. Timko, 2001: AFLP analysis of genetic polymorphism and evolutionary relationships among cultivated and wild Nicotiana species. Genome **44,** 559—571.

Rohlf, F. J., 2000: NTSYS-pc: Numerical Taxonomy and Multivariate Analysis System, 2.1 Ed. Exeter Software, Setauket, NY.

Sambrook, J., and D. W. Russell, 2001: Molecular Cloning a Laboratory Manual, 3rd edn. Cold Spring Harbor, New York.

Syvanen, A. C., 2001: Accessing genetic variation: genotyping single nucleotide polymorphisms. Nat. Rev. Gen. **2,** 930—942.

Taft, R. J., E. A. Glazov, N. Cloonan, C. Simons, S. Stephen, G. J. Faulkner, T. Lassmann, A. R. R. Forrest, S. M. Grimmond, K. Schroder, K. Irvine, T. Arakawa, M. Nakamura, A. Kubosaki, K. Hayashida, C. Kawazu, M. Murata, H. Nishiyori, S. Fukuda, J. Kawai, C. O. Daub, D. A. Hume, H. Suzuki, V. Orlando, P. Carninci, Y. Hayashizaki, and J. S. Mattick, 2009: Tiny RNAs associated with transcription start sites in animals. Nat. Genet. **41,** 572—578.

Tang, D., W. Wu, W. Li, H. Lu, and A. J. Worland, 2000: Mapping of QTLs conferring resistance to bacterial leaf streak in rice. Theor. Appl. Genet. **101,** 286—291.

Tautz, D., 1989: Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Acids Res. **17,** 6463—6471.

Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. Vandelee, M. Hornes, A. Frijters, J. Pot, J. Peleman, M. Kuiper, and M. Zabeau, 1995: AFLP – a new technique for DNA-fingerprinting. Nucleic Acids Res. **23,** 4407—4414.

Wang, Z., J. L. Weber, G. Zhong, and S. D. Tanksley, 1994: Survey of plant short tandem DNA repeats. Theor. Appl. Genet. **88,** 1—6.

Wang, Y., D. Shen, S. P. Bo, H. A. Chen, J. A. Zheng, Q. H. Zhu, D. G. Cai, C. Helliwell, and L. J. Fan, 2010: Sequence variation and selection of small RNAs in domesticated rice. BMC Evol. Biol. **10,** 119.

Waugh, R., K. McLean, A. J. Flavell, S. R. Pearce, A. Kumar, B. B. T. Thomas, and W. Powell, 1997: Genetic distribution of Bare-1-like retrotransposable elements in the barley genome revealed by sequence-specific amplification polymorphisms (S-SAP). Mol. Gen. Genet. **253,** 687—694.

Wernsman, E. A., 1999: An overview of tobacco breeding – past, present and future. Recent Adv. Tob. Sci. **25,** 5—35.

Williams, J. G. K., A. R. Kubelik, K. J. Livak, J. A. Rafalski, and S. V. Tingey, 1990: DNA polymorphisms amplified by arbitrary primers are useful as genetic-markers. Nucleic Acids Res. **18,** 6531—6535.

Zhu, Q. H., A. Spriggs, L. Matthew, L. J. Fan, G. Kennedy, F. Gubler, and C. Helliwell, 2008: A diverse set of microRNAs and microRNA-like small RNAs in developing rice grains. Genome Res. **18,** 1456—1465.

Zietkiewicz, E., A. Rafalski, and D. Labuda, 1994: Genome fingerprinting by simple sequence repeat (ssr)-anchored polymerase chain-reaction amplification. Genomics **20,** 176—183.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Figure S1. Distribution of three multi-mapped small RNAs in rice genome. A 100 kb window (bin) was used to scan sense and antisense strands of chromosome sequences, respectively. The three small RNAs (RS1, 2 and 5/6) see Table S1.

Figure S2. Phylogenetic tree of tobacco varieties based on 41 iSNAP markers. T1–T8 represent the eight tobacco cultivars in Table 1.

Table S1. Rice and tobacco small RNAs used for the development of iSNAP markers in this study.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.