Molecular analysis of almond germplasm in China

Ma R., Xie H., Xu Y., Ma Y., Jiang Y.Q., Cao M.Q.

in

Oliveira M.M. (ed.), Cordeiro V. (ed.).
XIII GREMPA Meeting on Almonds and Pistachios

Zaragoza : CIHEAM
Options Méditerranéennes : Série A. Séminaires Méditerranéens; n. 63

2005
pages 281-290

Molecular analysis of almond germplasm in China

R. Ma, H. Xie, Y. Xu, Y. Ma, Y.-Q. Jiang and M.-Q. Cao[1]
Beijing AgroBiotechnology Research Centre, Beijing Academy of Agriculture and Forestry Sciences, Ban-Jing Road, Hai-Dian District, Beijing100089, China
rcma1@yahoo.com

SUMMARY – Chinese almond is represented by six species and over 100 cultivars/genotypes of common almond (*Prunus dulcis* Mill.) have been identified in China. To analyse this diverse germplasm we have used SSRs derived from both genomic and cDNA clones, and the self-incompatibility alleles. Almond germplasm available from other countries was also introduced in the analyses. Recently, we have obtained over 1000 ESTs from almond pistil, and 723 unique sequences were assembled. These ESTs have been used in our EST-SSR analysis. The SSR markers obtained showed high cross-species transferability and allowed us to construct the phylogenetic tree of common almond, wild almond and their relative species. In this work, 54 SSR pairs were newly developed and 27 were used for the phylogenetic analysis and fingerprinting. A phylogenetic tree was then constructed based in the 178 polymorphic alleles distinguished among the 45 accessions. Species-specific SSR alleles were easily identified in our study. The results showed that almond cultivars from China and the Mediterranean region countries were classified in two independent groups. The two groups were differentiated by specific EST-SSR alleles, strongly suggesting that geographic isolation or founder effect contributed to the evolution of these alleles. Thirty-six new *S*-RNase gene sequences have been amplified by PCR and sequenced from the 28 almond cultivars and other almond species of China. The *S*-genotypes have been identified in some of the cultivars. Phylogenetic analysis indicated that the *S*-RNase sequences formed a relatively independent subgroup from those of almond cultivars originated from the Mediterranean region, suggesting that the *S*-genes have different evolutionary history after almond was introduced from the Central and Western Asia to the Mediterranean region, which might have been related to their geographic origin. The almond pistil EST analysis provides a first picture of the numerous almond genes potentially involved in the pistil development, and it not only contributes to the understanding of gene expression patterns in pistils but also provides an extensive reservoir for the gene cloning and genetic mapping in almond and other related fruit trees.

Key words: *Prunus dulcis* Mill., EST, SSR, *S*-RNase, almond.

RESUME – *"Analyse moléculaire du germoplasme d'amandier en Chine". L'amandier chinois est représenté par six espèces, et plus de 100 cultivars/génotypes de l'amandier commun (*Prunus dulcis *Mill.) ont été identifiés en Chine. Afin d'analyser ce germoplasme divers, nous avons utilisé des SSR dérivés de clones génomiques et de clones cDNA, et d'allèles d'auto-incompatibilité. Le germoplasme d'amandier disponible à partir d'autres pays a également été introduit dans les analyses. Récemment, nous avons obtenu plus de 1000 EST de pistils d'amandiers, et 723 séquences uniques ont été assemblées. Ces EST ont été utilisés dans notre analyse EST-SSR. Les marqueurs SSR obtenus ont montré une forte transférabilité à travers les espèces et nous ont permis de construire l'arbre phylogénétique de l'amandier commun, de l'amandier sauvage et des espèces apparentées. Dans ce travail, 54 paires de SSR ont été nouvellement développées et 27 ont été utilisées pour l'analyse phylogénétique et le "fingerprinting". Un arbre phylogénétique a été construit basé sur les 178 allèles polymorphes distingués parmi les 45 accessions. Des allèles SSR spécifiques à l'espèce ont été facilement identifiés dans notre étude. Les résultats ont montré que les cultivars d'amandiers de Chine et des pays de la région méditerranéenne ont été classifiés en deux groupes indépendants. Ces deux groupes se différenciaient par des allèles spécifiques EST-SSR, ce qui suggère fortement que leur isolation géographique ou l'effet des fondateurs ont contribué à l'évolution de ces allèles. Trente-six nouvelles séquences de gènes* S*-RNase ont été amplifiées par PCR et séquencées à partir des 28 cultivars d'amandier et d'autres espèces d'amandiers de Chine. Les* S*-génotypes ont été identifiés chez certains des cultivars. L'analyse phylogénétique a indiqué que les séquences* S*-RNase formaient un sous-groupe relativement indépendant par rapport aux cultivars d'amandiers originaires de la région méditerranéenne, suggérant que les* S*-gènes ont une histoire évolutive différente après que l'amandier ait été introduit dans la région méditerranéenne en provenance d'Asie Centrale et Occidentale, ce qui pourrait être mis en liaison avec leur origine géographique. L'analyse EST des pistils d'amandiers nous apporte une première vision des nombreux gènes d'amandiers potentiellement impliqués dans le développement du pistil, et ceci non seulement contribue à la compréhension des modes d'expression des gènes dans les pistils mais nous apporte également un réservoir extensif pour le clonage de gènes et la cartographie génétique de l'amandier et autres arbres fruitiers apparentés.*

Mots-clés : Prunus dulcis *Mill., EST, SSR,* S*-RNase, amandier.*

[1]Paper submitted, although the authors could not attend the meeting.

Introduction

Almond (*Prunus dulcis* Mill. syn. *Amygdalus communis* L. var. *dulcis*) is an important commercial fruit tree in the world. Twenty-six almond species form a distinct and easily identified taxonomic group in the world (Browicz *et al.,* 1996). Common almond (*P. communis* Fritsch.) is one of the major tree crops of the world, and an ancient cultivar supposedly originated in the Central Asia (Socias i Company, 1998). Traditional almond culture primarily utilized open-pollinated seedlings (Socias i Company, 1998). This fact, together with self-incompatibility, has created a very high heterozygosity in this species, which is one of the most polymorphic fruit species.

Studies at the molecular level for the identification, characterisation and relatedness analysis of almond cultivars of the Mediterranean area have been performed using molecular markers, such as isozymes (Arulsekar *et al*., 1986; Hauagge *et al.,* 1987; Cerezo *et al*., 1989; Jackson and Clarke, 1991; Arús *et al.,* 1994; Viruel *et al*., 1995), RFLP (Viruel *et al.,* 1995), RAPD (Joobeur *et al.,* 1998; Resta *et al*., 1997; Bartolozzi *et al*., 1998; Martins *et al*., 2001 and 2003; Ryan *et al*., 2001), ISSR and AFLP (Martins *et al*., 2001; 2003). SSRs, however, have not been fully explored in almond (Joobeur *et al.,* 2000).

Most almond cultivars show 100% self-incompatibility, a character controlled by a single multi-allelic *S* locus encoding specific ribonucleases (*S*-RNase). Based on the conserved structures of *S*- RNase, PCR has been performed in many plants to clone *S*-genes, study evolutionary relationships, and identify S-genotypes. More than twenty *S*-alleles have been reported among European almond cultivars (Batlle *et al*.,1997; Bošković *et al*.,1999; Ma *et al*., 2002).

Several species of almond have been described in China, such as the sextaploid species, longstalk almond (*Prunus pedunculata* Maxim.) and the diploid species, common almond (*P. communis* Fritsch.), Mongolic almond (*P. mongolica* Maxim.), wild almond (*P. ledebouriana* Schleche.), Tangut almond (*P. tangutica* Batal.) and the flowering almond (*P. triloba* Lindl.) (Zhu, 1983; Shang and Su, 1985). All the almond species have adapted to the arid, cold and dry climate and, therefore, potentially represent an important gene pool. It is therefore, of great importance to characterize Chinese almond genetic resources, assess their role for almond breeding and production, and effectively preserve and utilize them.

More than 40 almond cultivars and varieties have been used in the almond production in China (Zhu, 1983). So far, over 100 genotypes have been identified in Xin Jiang, China. We still have very little knowledge about the taxonomic status of almond species at the molecular level, and little is known about the genetic diversity and relatedness between the two eco-geographic types of cultivated almond represented by Chinese and the Mediterranean almond, respectively. Up to now, there is little knowledge about the *S*- phenotypes and genotypes of almond cultivated in China; therefore, the almond tree planting is solely based on practical experience, resulting in low yields.

Here, SSRs derived from both the genomic and EST sequences, and self-incompatibility genotyping have been used in the analysis of Chinese almond germplasm and those introduced from other countries.

Materials and methods

Plant material

The 44 accessions including 36 of common almond, two hybrids of almond and peach, two peaches and four related wild species were used for SSR analysis (Table 1). The 26 accessions of Chinese almond, including 22 cultivars of common almond and other 4 almond species were used for the isolation of *S*-RNase gene homologs.

Plant genomic DNA Isolation

Genomic DNA of all the accessions studied was extracted from young leaves according to the CTAB method described by Eldredge *et al.* (1992).

Table 1. Plant materials used in this work

Acc. No.	Plant species/cultivars	Collections	Origin
Peach [*Prunus persica* (L.) Batsch.]			
1	f. *aganopersica*	Beijing, China	China
2	var. *nucipersica* f. *aganonucipersica*	Beijing, China	China
Common almond (*P. communis* Fritsch.)			
3	'Ayuehunzibadan'	Xin-Jiang, China	China
4	'Bianzuihe'	Xin-Jiang, China	China
5	'Jianzuihuang'	Xin-Jiang, China	China
6	'Zhipi'	Xin-Jiang, China	China
7	'Shuangren'	Xin-Jiang, China	China
8	'Kexi'	Xin-Jiang, China	China
9	'Gongbadan'	Xin-Jiang, China	China
10	'Dabadan'	Xin-Jiang, China	China
11	'Hanfeng'	Xin-Jiang, China	China
12	'Shuangguo'	Xin-Jiang, China	China
13	'Yeerqiang'	Xin-Jiang, China	China
14	'Badanwang'	Xin-Jiang, China	China
15	'Huangshuang'	Xin-Jiang, China	China
16	'Make'	Xin-Jiang, China	China
17	'Duoguo'	Xin-Jiang, China	China
18	'Amannisha'	Xin-Jiang, China	China
19	'Baibadan'	Xin-Jiang, China	China
20	'Ao1'	Hebei Province, China	Unknown†
21	'Ao2'	Hebei Province, China	Unknown†
22	'Ao3'	Hebei Province, China	Unknown†
23	'NonPareil'	Shan Xi Province, China	USA
24	'Jeffery'	Shan Xi Province, China	USA
25	'Texas'	Shan Xi Province, China	USA
26	'Titan'	Shan Xi Province, China	USA
27	'I.X.L.'	Shan Xi Province, China	USA
28	'Mission'	Shan Xi Province, China	USA
29	'Ne Plus Ultra'	Shan Xi Province, China	USA
30	'Ferragnès'	Shan Xi Province, China	France
31	'Ferraduel'	Shan Xi Province, China	France
32	'Supernova'	Shan Xi Province, China	Italy
33	'$H_3$6'	Shan Xi Province, China	Unknown††
34	'$H_1$4'	Shan Xi Province, China	Unknown††
35	'Fura Saco'	Trás-os-Montes, Portugal	Portugal
36	'Tianrentaobadan'	Xinjiang, China	China
37	'Taobadan'	Xinjiang, China	China
38	'Kubadan'	Xinjiang, China	China
Peach and almond hybrid			
39	'Dulcis Pioneer'	Shan Xi Province, China	USA
40	'Liquan 1'	Shan Xi Province, China	Unknown
Wild almond species			
41	*Prunus triloba* Lindl	Shan Xi Province, China	China
42	*P. ledebouriana* Schleche.	Xinjang, China	China
43	*P. mongolica* Maxim.	Inner Mongolia, China	China
44	*P. tangutica* Batal.	Xinjiang, China	China

†Introduced from Australia.

††Introduced from Hungary.

EST analysis

Total RNAs of Almond pistils before pollination were isolated using Trizol reagent (Invitrogen, USA), mRNA purification was done with the mRNA purification kit (Amersham Pharmacia, USA), and a cDNA

library was constructed using the SuperScript™ Lambda System (Invitrogen, USA). The phage library was converted to a plasmid library in the vector pZL1. The template DNAs for the sequencing reaction were prepared by the alkaline lysis method using Wizard SV 96 DNA purification system kit (Promega, USA). The 5'ends of the cDNA clones were sequenced with the M13/pUC reverse primer (5'-AGCGGATAACAATTTCACA-CAGG-3') on ABI Prism377 DNA sequencer (Applied Biosystem, USA). Contig analysis was performed with StackPACK v2.2 (Burke *et al*., 1999; Miller *et al*., 1999; http://www.sanbi.ac.za).

SSR Screening and PCR amplification

The 1005 EST above developed (Jiang and Ma, 2003) were used to assemble with other peach and almond EST sequences from NCBI (www.ncbi.nih.gov/dbEST) with SeqMan II (LASERGENE, DNASTAR, Inc.). MISA (http://pgrc.ipk.-gatersleben.de/misa) was used to detect EST-SSRs. When the repeat number was higher than 10 for mono-nucleotide repeat motif, 6 for di-nucleotide repeat motif, or 5 for tri-, tetra-, penta-, and hexa- nucleotide repeat motifs, the sequences were counted as SSRs.

SSR primer pairs were designed, with annealing temperatures of 50-60°C and expected sizes of PCR products ranging from 120 bp to 350 bp (Table 2).

SSR polymorphism detection

The samples were isolated on denatured PAGE electrophoresis (8M Urea). The DNA bands were visualized by silver staining. A permanent record of gels was made using Molecular Imager FX (BioRad, USA).

SSR data analysis

Information content of SSR loci was estimated by polymorphism information content (PIC) described by Botstein *et al*. (1980) and modified by Anderson *et al.* (1993) according to the formula $PIC = 1-\Sigma p_i^2$, where p_i is the frequency of the i^{th} allele, and by the power of discrimination (PD) (Kloostermen *et al*., 1993), using the formula as above, for which the alleles frequency was replaced by the genotype frequency. Accessions showing only one fragment at a locus were considered homozygous for that fragment.

Genetic relationships among the genotypes were calculated using UPGMA cluster analysis of the similarity matrix obtained from the proportion of shared fragments named Dice value performed with the programme NTSYS 2.10t (Exter Software, Stauket, New York, NY, USA).

The *S*-RNase cloning and sequencing

Four primers for *S* genotype identification were synthesized using the published *S*-RNase sequence of *Prunus* (Tamura *et al.,* 2000; Ma *et al*., 2002). After electrophoresis, PCR products were purified using a rapid E.Z.N.A. Gel extraction kit (Omega Bio-tek, USA) subsequently inserted into pUCm-T vector and transformed into DH5α for cloning. The presence of the insert in the plasmid was confirmed by restriction enzyme digestion with *Pst*I. DNA inserts were sequenced using the M13 forward (5'-GTAAAACGACGGCCAG-3') and reverse (5'-CAGGAAACAGCTATGAC-3') primer on ABI Prism377 DNA sequencer (Applied Biosystem, USA). The coding regions were obtained after comparison with published sequences according to the conserved exon-intron boundary sequences (CT-AG rule). The phylogenetic tree was formed based on the multiple alignment of the putative amino acid sequences of *S*-RNases.

Table 2. SSR primers used in this work

Primer Set	GenBank Acc. No.	Putative protein (organism)	SSR motif	Forward sequence (5'-3')/ Reverse sequence (5'-3')	Expected size of PCR products (bp)
EST-SSRs					
PSSR1	BI203163	No similarity	$(CAG)_6(CAA)_3(CAG)_6$	ACCCCACCACAAGTCCAATA TGCTGCGGGTACACCATCATCT	201
PSSR2	BI203162	Cysteine protease inhibitor (*Malus* x *domestica*)	$(CTT)_4$	GGCACGAGAACAAAAGCAGATGA GGCGAAGCGAGCGAGGT	233
PSSR3	BI203151	Unknown protein (*Arabidopsis thaliana*)	$(GAT)_4$	CAATGGCTGCACCGCTTCTGT CGGGGCATAGGGGTGGATAGTG	163
PSSR4	BI203129	Hexameric polyubiquitin (*Nicotiana sylvestris*)	$(AAAAAT)_{3imp}$	TCCCCAAGAGGCAGAGA CCGACTATAACATCCAGAAGG	196
PSSR17	CA853978	Auxin-repressed protein (*Malus* x *domestica*)	$(GA)_{15}$	GGACTGGACTGTGGATTGTTTTTG AGTCATCCACCGTGCCAGTTTTA	189
PSSR26	BQ703558	Transmembrane chemo-receptor family member (*Caenorhabditis elegans*)	$(TA)_{15}$	TCGTTCGCACAATAATAATG GCAACTCCAGCGCCTCCAAAACT	256
PSSR27	BQ703434	Cold stress protein (*Poncirus trifoliata*)	$(AAG)_7$	AGACGGCGATCACAAAGAGAAAG TGGGCGTAAACGGAAAGGAA	179 350
PSSR28	BQ641222	Ler glycine-rich protein (*Arabidopsis thaliana*)	$(AT)_{20}$	AAGCCCTTGGGTTTAGTCC CTTCCCAGAGCCATTCCACATT	250
PSSR39	BQ703684	Unknown protein (*Arabidopsis thaliana*)	$(GAA)_{10imp}$	TTCTTCAAAGCAAATCCATCT GCCACGCACCTCTCAAAG	186
PSSR40	BQ703480	Unknown protein (*Arabidopsis thaliana*)	$(AAG)_5$	TTACGTAAGAAGCCAAGAAGAGG TCCACTCCAACAAAATAAATAA	144
PSSR45	BQ703653	Cold stress protein (*Poncirus trifoliata*)	$(AAG)_7$	GAAAGACGGCGATCACAAAGAGA GCGCCCCAAACACCAGAATA	133
PSSR46	CA854086	No similarity	$(TA)_{12}$	AATGGTCTATGAACACCTCTC GAATACAAACGGGATTCAG	190
PSSR48	AF350937†	Expansin (EXP2) (*Prunus cerasus*)	$(CCT)_7$	AACACCTCCCCTGCATAAAAACCAA GCGCCTTCCCAGCCACCATAAT	177
PSSR51	AF0537699††	Homeodomain protein (Mdh1) (*Malus* x *domestica*)	$(GA)_{11}$	CAGAAGGGTTGAAGGTGATTG TCGAGGAGTTGCAGAGTATGAA	229
PSSR54	BU645586	Putative trehalose-phosphate phosphatase (*Arabidopsis thaliana*)	$(TA)_{12}$	CTTGGCTGGCTTCACTGC CTATGCATTTATAAACTTGCAGAG	133
PSSR57	CA854108	No similarity	$(T)_{16}$	TATCCCTCCTCATCCCTATCTATCATT ATCAAAACTGCACACTGTAACCCTAAG	132
PSSR60	BU645482	Formate dehydrogenase (*Solanum tuberosum*)	$(T)_{21}$	AGTCGGCTAGTTCGTGTTCTCTT TGTCCATCCAGGCAGCATAC	176
PSSR61	CA854205	Cold regulated LTCOR18 (LtCor18) (*Lavatera thuringiaca*)	$(GAA)_{9imp}$	GAAGATCTACGAGGAGCCAACT CTGCTGATCATCATACGAAGC	176
PSSR62	CA854147	CP12 (*Pisum sativum*)	$(TC)_{11imp}$	CAGCCATGCCCGCGACAAC GACAATCGACAACCTCAATCTAAG	153
PSSR63	CA854147	CP12 (*Pisum sativum*)	$(GAT)_5$	CACCAATTTATGTTGCAAGATTATATG GTTTTAGATTTCACAGTACTATG	154
Anonymous genomic SSRs					
Pchgms1†††			$(AC)_{12}(TA)_6$	GGGTAAATATGCCCATTGTGCAATC GGATCATTGAACTACGTCAATCCTC	194
Pchgms3†††			$(CT)_{14}$	ACGGTATGTCCGTACACTCTCCATG CAACCTGTGATTGCTCCTATTAAAC	163
Ps9f8††††				GGTTCTTGGTTATTATGA ACATTTCTATGCAGAGTA	156
Ps8e8††††				CCCAATGAACAACTGCAT CATATCAATCACTGGGATG	183
UDP98410†††††			$(AG)_{23}$	AATTTACCTATCAGCCTCAAA TTTATGGCAGTTTACAGACCG	146
UDP96008††††††			$(CA)_{23}$	TTGTACACACCCTCAGCCTG TGCTGAGGTTCAGGTGAGTG	165
PSSR11	BH023894		$(CT)_6$	ATGTCATGAGGTGTTTGTATCTTG TGTTTAGGTCTTCATGTTTTAGGA	129

†From cherry (*Prunus cerasus*).

††From apple (*Malus x domestica*).

†††Ref. Sosinski *et al.* (2000).

††††Ref. Joobeur *et al.* (2000).

†††††Ref. Testolini *et al.* (2000).

††††††Ref. Cipriani *et al.* (1999).

Results

Almond pistil EST analysis

In total, 1058 EST clones were sequenced. After vector sequences, ambiguous bases and excluding sequences without cDNA inserts were deleted, 1005 ESTs were obtained. 456 ESTs (45.4%) encode putative amino acid sequences with significant similarities to those registered in the NCBI non-redundant protein database. Sequences that didn't match with sequences in the protein databases were further analysed at the nucleotide level. In 716 unique sequences assembled, 318 (44%) having significant homology to previously identified genes, 162 (23%) matching hypothetical proteins, 61 having no significant similarity, and 175 showing low similarity and scores. The unique sequences whose functions could be predicted were then classified into 15 categories, and stress resistance and defence (14%), protein synthesis and processing (12%) and signal transduction (10%) were the three larger groups. The functionally unidentified sequences may either be new plant genes, or may be mostly composed of 5'-UTR sequences. This work provided some information of gene expression in almond pistils and new resources for developing almond SSR markers. These nucleotide sequence data appear in GenBank under the following accession No. BQ641013-BQ641227, BQ641229, BQ703418-BQ703709, BU645424-BU645594 and CA853933-CA854258.

SSR marker polymorphism and phylogenetic relationship of almond

Twenty-seven SSR primer pairs were designed from almond ESTs described above, only 18 of which produced repeatable and easily counted amplification products in silver-staining PAGE, and 2 EST-SSR primer pairs (PSSR48 and 51) were developed from apple and sweet cherry after mining the public database. Furthermore, 7 genomic SSR primer sets were also synthesized from published results. In total, 27 SSR primer pairs (20 EST-SSRs and 7 genomic SSRs) were used to evaluate the almond genetic resources.

The 20 EST-SSRs detected 79 alleles with an average 3.95 alleles per locus in the 36 almond cultivars, and the 7 genomic-SSRs detected 51 alleles with an average 7.29 alleles per locus. Among the 130 alleles detected, 72 alleles (55.38%) were shared by the two parts of almond cultivars originated from Central Asia and the Mediterranean region, and 30 (23.08%) only presented in Chinese cultivars and 28 (21.54%) only in other cultivars. Some alleles were shown unequally distributed in the two parts. For Example, PSSR4 detected two alleles, PSSR4-1 and PSSR4-2 in the 36 cultivated almonds. The PSSR4-1 occurred in all the 20 Chinese cultivars and in only 5 of the 16 foreign cultivars, and PSSR4-2 was present in only 2 of the Chinese cultivars and in most (14) of the foreign cultivars. Some of the cultivars could be easily discriminated from others by the EST-SSRs. The PSSR51 locus had the same allele in the 43 accessions including the different almond species and peach except for a Chinese almond cultivar, 'Make', which had a different allele. Similarly, the PSSR48 locus had only one allele in the 42 accessions except for the two French almond cultivars, Ferragnès and Ferraduel, both of which shared another allele.

The 27 SSR primer pairs were also used to assess their cross-transferability in other *Prunus* species, including 2 peach cultivars (*P. persica* L. var. *aganopersica* and var. *aganonucipersica*) and other 4 almond species (*P. mongolica*, *P. ledebouriana, P. tangutica*, and *P. triloba*). All SSRs were detected in the samples except for PSSR 26 which did not exist in 4 accessions, the 2 peach cultivars and the 2 almond species, *P. triloba* and *P. lebedouriana*. 24 of the 27 SSR loci (but PSSR3, PSSR39 and PSSR57) were polymorphic in peach and in the other 4 almond species. The 3 SSR loci, PSSR28, PSSR40, PSSR45 were monomorphic within cultivated almond, indicating they were highly conservative within the species, however, they displayed polymorphism in the 2 peach accessions and the 4 almond species.

The 44 accessions were divided into 5 groups based on UPGMA cluster analysis of the similarity data (Fig. 1). In the phylogenetic tree based on the 178 alleles detected by SSRs, almond cultivars formed two distinct subgroups which reflect their different geographic origin; one being mainly composed of Chinese cultivars and the other composed of the samples obtained from the Mediterranean region and USA, indicating that Chinese almond cultivars have a different evolutionary history from the Mediterranean almond.

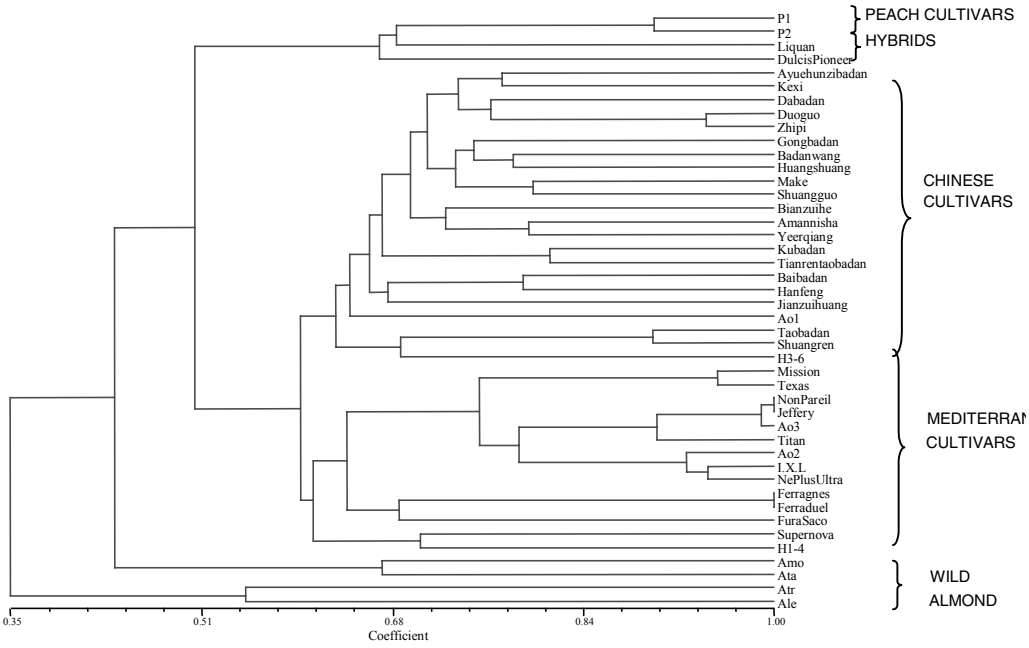

Fig. 1. Phylogenetic tree based on the 178 alleles detected by the 27 EST- and genomic SSRs in almond and other related species. The accessions are indicated in Table 1.

S-RNase cloning and sequence analysis from Chinese almond

In all the selected 26 almond samples of Chinese origin, 36 new *S*-RNase alleles (S50-85) were identified with 23 containing the five domains of *S*-RNases and 13 only C2 and C3 domains. The amplified sequences all harbour an intron ranging from 161 to 1525 bp. A phylogenetic tree was constructed based on the alignment of the deduced amino acid sequences of the 23 *S*-RNase gene homologues with those of the published *S*-RNase from almond and other *Prunus* species (Fig. 2).

Phylogenetic analysis indicated that the *S*-RNase sequences from Chinese almond formed relatively independent subgroups from those of almond cultivars originated from the Mediterranean region, suggesting that the *S*-genes have different evolutionary history, which might have been related to their geographic origin. Highly homologous *S*-RNase sequences have been cloned from wild almond species although so far, there is no knowledge about if they are self-compatible or incompatible.

Discussion

This is the first report on the molecular analysis of Chinese almond. The EST data from almond pistil cDNA library described here will be further used to isolate and identify candidate cDNAs with cell-type-specific or regulated expression patterns and to elucidate their biological functions in almond pistil development.

Almond was an ancient Mediterranean cultivar and differentiated into two ecotypes in the process of evolution and domestication, namely the Mediterranean and the Central Asia types (Socias i Company, 1998). In this work, we collected almond cultivars representing the two eco-geographic types, China and the Mediterranean, of common almond in the world. All the 36 cultivars were clustered into two subgroups, in a good agreement with their eco-geographic locations and historical pedigree records. The dendrogram can then be used to infer the genetic relatedness of the two eco-geographic types of almond cultivars.

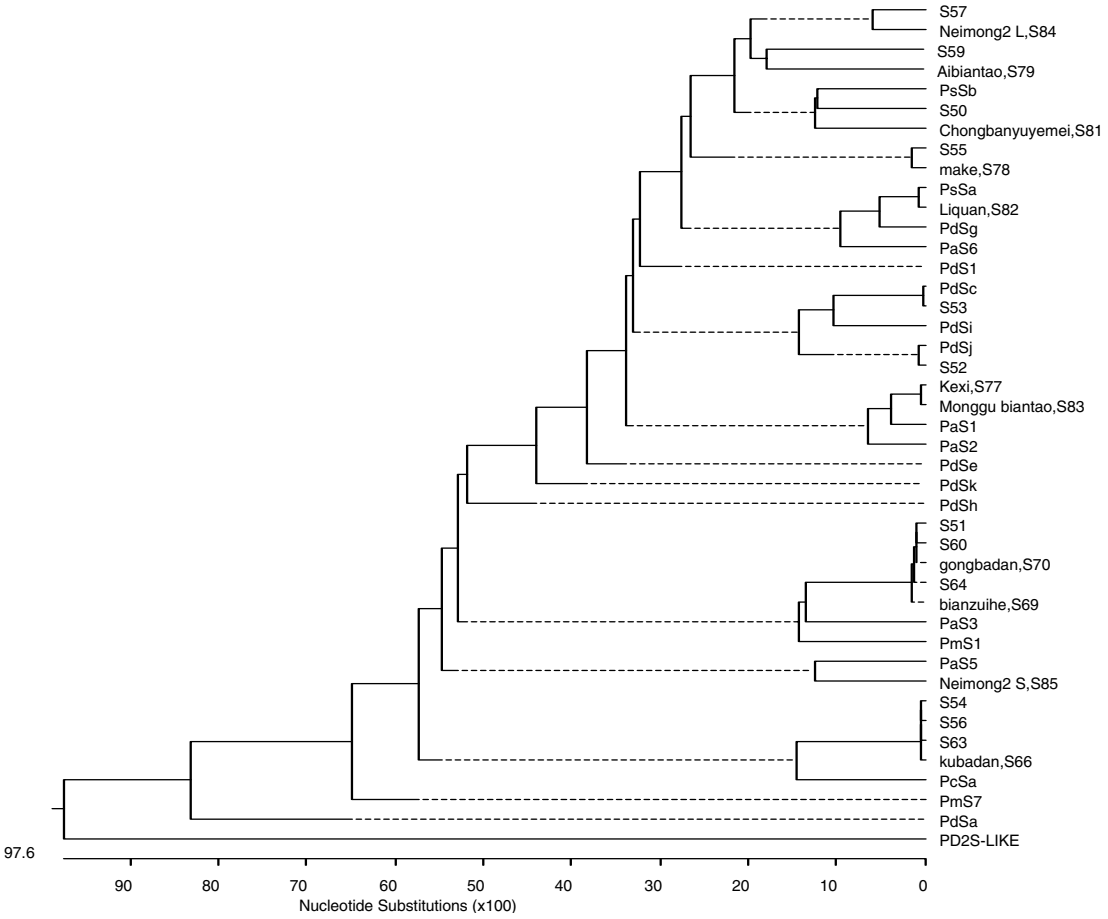

Fig. 2. The phylogenetic tree of the putative *S*-RNase amino acid sequences from almond and other *Prunus* species. Pa, *P. avium*; Pc, *P. cerasifera*; Pd. *P. dulcis*; Pm, *P. mume*; Ps, *P. salisina*; Neimong2, *P. mongolica* Acc. No. 2; Menggubiantao, *P. mongolica*; Changbaiyuyemei, *P. triloba*; Aibiantao, *P. ledebouriana.* Other *S*-RNase sequences are all from Chinese almond cultivars (*P. dulcis*).

Although some EST-SSRs showed low polymorphism in common almond (*P. communis* Fritsch.), they were highly polymorphic in other species of *Prunus*. This means that these SSRs should be useful for genetic analysis, marker assisted selection, and molecular mapping across taxa. An attempt was also made to mine EST data from other *Prunus* species and the relative *Malus × domestica* of the Rosaceae, and 2 (PSSR48 and PSSR51) were successfully used in this work, suggesting that more SSRs from the ESTs of different species could be developed for genetic analysis of different species in the genus *Prunus*, such as almond, apricot, cherry and peach.

It is well documented that within an SSR locus, length differences between alleles are not only due to a variation in the number of tandem repeats, and sequence mutations may cause the SSR variation (Grimaldi and Crouau-Roy, 1997; Orti *et al.*, 1997; Decroocq *et al.*, 2003). SSR length variation is inadequate to assess long-term evolutionary divergence between, for instance, the Maloideae and the Prunoideae of Rosaceae. However, it is useful to test the relationship between genetically close species of a subfamily, e.g. the Prunoideae, or cultivars within a species (Decroocq *et al.*, 2003). So far, there is little sequence information about SSR variation in the genus *Prunus*. The EST sequences analysed in this work provide the basis for further analysis, although the confirmation of the allelic relationship still needs extensive sequencing of these SSRs.

In this study, 36 new *S*-RNase sequences have been cloned from Chinese almond cultivars and related almond species. So far, cDNAs coding for these *S*-RNases have not been cloned, and further work is still required to elucidate their structures and functions. There is no previous knowledge regarding almond *S*-genotypes in China. Pollination tests are also needed to confirm if the newly cloned *S*-RNase sequences co-segregate with those identified in the breeding work.

References

Anderson, J.A., Churchill, G.A., Sutrique, J.E., Tanksley, S.D. and Sorrels, M.E. (1993). Optimizing parental election for genetic linkage maps. *Genome*, 36: 181-186.

Arulsekar, S., Parfitt, D.E. and Kester, D.E. (1986). Comparison of isozyme variability in peach and almond cultivars. *J. Hered.*, 77: 272-274.

Arús, P., Olarte, C., Romero, M. and Vargas, F. (1994). Linkage analysis of ten isozyme genes in F_1 segregating almond progenies. *J. Amer. Soc. Hort. Sci.*, 119: 339-34.

Bartolozzi, F., Warburton, M.L., Arulsekar, S. and Gradziel, T.M. (1998). Genetic characterization and relatedness among California almond cultivars and breeding lines detected by Randomly Amplified Polymorphic DNA (RAPD) analysis. *J. Amer. Soc. Hort. Sci.,* 123: 381-387.

Batlle, I., Ballester, J., Bošković, R., Romero, M.A., Tobutt, K.R. and Vargas, F.J. (1997). Use of stylar ribonucleases in almond breeding to design crosses and select self-compatible seedlings. *Nucis,* 6: 12-14.

Bošković, R., Tobutt, K.R., Duval, H., Batlle, I., Dicenta, F. and Vargas, F.J. (1999). A stylar ribonuclease assay to detect selfcompatible seedlings in almond progenies. *Theor. Appl. Genet.,* 99: 800-810.

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

Browicz, K. and Zohary, D. (1996). The genus *Prunus* L. (Rosaceae): Species relationships, distribution and evolution under domestication. *Genet. Resour. Crop Evol.*, 43: 229-247.

Burke, J., Davison, D. and Hide, W. (1999). d2_cluster: A validated method for clustering EST and full-length cDNA sequences. *Genome Res.,* 9: 1135-1142.

Cerezo, M., Socias i Company, R. and Vargas, F. (1989). Identification of almond cultivars by pollen isoenzymes. *J. Amer. Soc. Hort. Sci.*, 114: 164-169.

Cipriani, G., Lot, G., Huang, W.G., Marrazzo, M.T., Peterlunger, E. and Testolin, R. (1999). AC/GT and AG/CT microsatellites repeats in peach (*Prunus persica* (L.) Batsch): Isolation, characterization and cross-species amplification in *Prunus*. *Theor. Appl. Genet.*, 99: 65-72.

Decroocq, V., Fave, M.G., Hagen, L., Bordenave, L. and Decroocq, S. (2003). Development and transferability of apricot and grape EST microsatellite markers across taxa. *Theor. Appl. Genet.*, 106: 912-2.

Eldredge, L., Ballard, R., Baird, W.V., Abbott, A., Morgens, P., Callahan, A., Scorza, R. and Monet, R. (1992). Application of RFLP analysis to genetic mapping in peaches. *HortScience*, 27: 160-163.

Grimaldi, M.C. and Crouau-Roy, B. (1997). Microsatellite allelic homoplasy due to variable flanking sequences. *J. Mol. Evol.*, 44: 336-340.

Hauagge, R., Kester, D.E., Arulsekar, S., Parfitt, D.E. and Liu, L. (1987). Isozyme variation among California almond cultivars. II. Cultivar characterization and origins. *J. Amer. Soc. Hort. Sci.*, 112: 693-698.

Jackson, J.F. and Clarke, G.R. (1991). Gene flow in an almond orchard. *Theor. Appl. Genet.*, 82: 169-173.

Jiang, Y.Q. and Ma, R.C. (2003). Generation and analysis of expressed sequence tags from almond (*Prunus dulcis* Mill.) pistils. *Sex Plant Reprod.*,16: 197-207.

Joobeur, T., Periam, N., de Vicente, M.C., King, J. and Arus, P. (2000). Development of a second generation linkage map for almond using RAPD and SSR markers. *Genome,* 43(4): 649-55.

Joobeur, T., Viruel, M.A., de Vicente, M.C., Jauregui, B., Bellester, J., Dettori, M.T., Verde, I., Truco, M.J., Messeguer, R., Balester, J., Quarta, R., Dirlewanger, E. and Arús, P. (1998). Construction of a saturated linkage map for *Prunus* using an almond × peach F_2 progeny. *Theor. Appl. Genet.,* 97: 1034-1041.

Kloosterman, A.D., Boudowle, B. and Daselaar, P. (1993). PCR amplification and detection of the human D1S80 VNTR locus. Amplification conditions, population genetics and application in forensic analysis. *Int. J. Leg. Med.,* 105: 257-264.

Ma, R.C., Sánchez, A.M., Oliveira, M. (2002). Molecular evolution and distribution of S-RNase genes among almond cultivars. *Acta Horticul.*, 591: 213-216.

Martins, M., Farinha, A., Ferreira, E., Cordeiro, V., Monteiro, A., Tenreiro, R. and Oliveira, M. (2001). Molecular analysis of the genetic variability of Portuguese almond collections. *Acta Horticult.,* 546: 449-456.

Martins, M., Tenreiro, R. and Oliveira, M.M. (2003). Molecular diversity among Portuguese almond cultivars assessed by RAPD and ISSR markers. *Plant Cell Rep.*, 22: 71-78.

Miller, R.T., Christoffels, A.G., Gopalakrishnan, C., Burke, J., Ptitsyn, A.A., Broveak, T.R. and Hide, W. (1999). A comprehensive approach to clustering of expressed human gene sequence: The

sequence tag alignment and consensus knowledge base. *Genome Res.,* 9: 1143-1155.
Orti, G., Pearse, D.E. and Avise, J.C. (1997). Phylogenetic assessment of length variation at a microsattelite locus. *Proc. Natl. Acad. Sci. USA*, 94: 10745-10749.
Resta, P., Corona, M.G., Fanizza, G., Palasciano, M. and Godini, A. (1997). Random amplified DNA polymorphisms in *Amygdalus communis* L. and *A. webbii* Spach. *Acta Horticult.*, 470: 82-90.
Ryan, F.J., Ledbetter, C.A., Ramming, D.W., Palmquist, D., Bell, D.E. and Peterson, S.J. (2001). Challenges in developing molecular markers for almond (*Prunus*) and grape (*Vitis* species). *Acta Horticult.*, 546: 629-639.
Shang, Z.Y. and Su, G.X. (1985). Chromosomal numbers of six species in the genus *Amygdalus* from China. *J. Wuhan Bot. Res.*, 3: 363-366 (In Chinese).
Socias i Company, R. (1998). Fruit tree genetics at a turning point: The almond example. *Theor. Appl. Genet.* 96: 588-601.
Sosinski, B., Gannavarapu, M., Hager, L.D., Beck, L.E., King, G.J., Ryder, C.D., Rajapakse, S., Baird, W.V., Ballard, R.E. and Abbott, A.G. (2000). Characterization of microsatellite markers in peach (*Prunus persica* (L.) Bastch). *Theor. Appl. Genet.*, 101: 421-428.
Tamura, M., Ushijima, K., Sassa, H., Hirano, H., Tao, R., Gradziel, T.M. and Dandekar, A.M. (2000). Identification of self-incompatibility genotypes of almond by allele-specific PCR analysis. *Theor. Appl. Genet.*, 101: 344-349.
Testolin, R., Marrazzo, T., Cipriani, G., Quarta, R., Verde, I., Dettori, M.P., Pancaldi, M. and Sansavini, S. (2000). Microsatellite DNA in peach (*Prunus persica* L. Batsch) and its use in fingerprinting and testing the genetic origin of cultivars. *Genome*, 43: 512-520.
Viruel, M.A., Messeguer, R., de Vicente, M.C., Garcia-Mas, J., Puigdomenech, P., Vargas, F.J. and Arus, P. (1995). A linkage map with RFLP and isozyme markers for almond. *Theor. Appl. Genet.*, 91: 964-971.
Zhu, J.L. (1983). *Almond in Xin Jiang*. Xin Jiang Science Press, Xin Jiang, China (In Chinese).