ONLINE RESOURCES

Genetic relationship of a cucumber germplasm collection revealed by newly developed EST-SSR markers

JIANBIN HU, JIANWU LI*, FANGFANG LIANG, LIJUN LIU and SHENGWEI SI

College of Horticulture, Henan Agricultural University, Zhengzhou 450002, People's Republic of China

Introduction

Cucumber (*Cucumis sativus* L.) is one of the most widely cultivated vegetables, ranked fourth after tomato, onion and cabbage (Pitrat *et al*. 1999). It originates from India, and China is widely regarded as its secondary origin centre (Staub *et al*. 1997). There are a large number of cucumber germplasms spread around China and they are generally classified into six ecotypes based on the geographical distribution and ecological characters, i.e. south China ecotype, north China ecotype, Europe greenhouse ecotype, Europe open-field ecotype, occident processing ecotype and wild ecotype (Xu 1994). Although various ecotypes have been identified in cucumber, the fact that cucumber has a narrow genetic base is widely acknowledged (Staub *et al*. 2005), and this limits the possibilities for trait improvement using cross breeding. Genetic diversity is desirable for long-term crop improvement and reduction of vulnerability to important crop stresses. Many cucumber cultivars have been developed from closely related parents, but limited improvement in certain agricultural traits in recent years demands a more diverse range of germplasms.

Molecular markers have extensively been used for studying genetic diversity and genetic relationship in cucumber, especially PCR-based markers, such as random amplified polymorphic DNA (RAPD) (Horejsi and Staub 1999), amplification fragment length polymorphism (AFLP) (Li *et al*. 2004), inter-simple sequence repeat (ISSR) (Wang *et al*. 2007) and simple sequence repeat (SSR) (Danin-Poleg *et al*. 2001). From all these studies, different estimates for the degree of genetic variation were obtained, reflecting the differences in the selected sets of genotypes or marker systems.

PCR-based markers mentioned above amplify non-coding regions or the whole genome randomly and the loci obtained are generally far away from the genes of targeted traits. This limits the application of these molecular markers in breeding programmes; thus, genic markers, e.g, express sequence tag (EST) derived SSR markers are desirable. EST-SSR is a new marker system which is commonly developed by electronic search of EST databases (Yasodha *et al*. 2008). The polymorphism derived from EST-SSR is associated with the coding regions of the genome and reflects the genetic diversity available inside or adjacent to the genes (Varshney *et al*. 2005), which can provide more useful information on cross breeding. To date, EST-SSR markers have been developed for a wide range of plant species and used for genetic studies with multiple purposes.

In this study, we designed a set of novel EST-SSR markers from cucumber EST database and used them to analyse genetic relationship of cucumber germplasm collection containing a range of ecotypes distributed in China. Our study provides information regarding the amount of relevant genetic diversity of cucumber accessions, and would be helpful to select the most diverse genotypes for cross breeding and molecular mapping, as well as in germplasm conservation efforts.

Materials and methods

Plant materials and DNA extraction

A total of 29 *C. sativus* accessions, representing a major range of cucumber germplasms distributed in China, were used in the present study. These accessions included 12 north China (NC) accessions (L53, L57, L102, L107, L109, D0462, D01108, Naire, Jinchun30, Jinyou30, Shandong5 and Tianfengzhichun), nine south China (SC) accessions (Baipi, Tangshanqiugua, Wangzhongwang, Changfeng10, Chubaiyu, Jinzhengbanggua, Yichuanlian, Zaochunxinxiu and Chunbai2), three Europe greenhouse (EG) accessions

*For correspondence. E-mail: lijw555@sohu.com.

[Hu J., Li J., Liang F., Liu L. and Si S. 2010 Genetic relationship of a cucumber germplasm collection revealed by newly developed EST-SSR markers. *J. Genet.* **89**, e28–e32. Online only: http://www.ias.ac.in/jgenet/OnlineResources/89/e28.pdf]

Keywords. cucumber; cluster; dendrogram; EST-SSR; genetic relationship; cucumber (*Cucumis sativus* L.).

(L112, D0351 and D0442), two Europe open-field (EO) accessions (LEG1 and SW1), two occident processing (OP) accessions (D01105 and Ningjia1) and one wild species (*C. hystrix* Chakr.) (SHG). The NC, SC and wild accessions are Chinese native (CN) germplasms and the accessions of EG, EO and OP were introduced from Europe and America (EA) and are domesticated in China. Seeds of the cucumber accessions were donated by several research institutions or purchased from seed companies in China. Genomic DNA was extracted from all the accessions mentioned above using CTAB method described by Murray and Thompson (1980).

SSR marker development

Up to 10 August 2008, a total of 6344 ESTs were downloaded from GenBank and assembled with CAP 3 (Huang and Madan 1999), resulting in production of 4036 unigenes (containing 832 contigs and 3204 singletons). A web tool, SSRIT (http://www.gramene.org/db/markers/ssrtool), was used for searching SSRs in the unigenes with the criteria of minimum five repeats units for di-nucleotide to pentanucleotide. With primer Premier 5.0 program (PREMIER Biosoft International, California, USA), primers were designed flanking the SSRs and allowed to generate PCR products 100–300 bp in length and annealing temperature (T_a) of 50–60° C. Primers were synthesized by Sangon, Shanghai, China. Primer pairs were initially screened using the genomic DNA of accession L57 and those without obvious non-specific amplification were selected for further study.

SSR amplification and PCR product analysis

PCR amplification was performed in 15 μL reaction containing 1× PCR buffer, ~50 ng of sample DNA, 0.5 μM of each primer, 200 μM of each dNTP, 1.5 mM $MgCl_2$ and 0.75 unit of *Taq* DNA polymerase (TaKaRa, Dalian, China). All amplifications were carried out in a PTC-200 thermal cycler (MJ Research, Waterdown, USA) as follows: 5 min at 94°C, followed by 30 cycles of 40 s at 94°C, 40 s at annealing temperature (T_a) and 1 min at 72°C, and 8 min at 72°C for final extension. Amplified products were electrophoresed in 6% non-denaturing polyacrylamide gels (19:1 acrylamide:bis) and the gels were silver-stained according to the method of Creste *et al*. (2001). The size of the amplified bands for each microsatellite was estimated by reference to a DNA ladder (pUC19 DNA/*Msp*I marker, Sangon, Shanghai, China).

Data analysis

Each fragment amplified with each primer pairs was visually scored as either present (1) or absent (0) for each accession. With the marker data, a binary matrix was established and used to calculate Jaccard's similarity coefficient (Jaccard 1908). The accessions were clustered based on the similarity matrix data using the UPGMA (unweighted pair group method using arithmetic averages). A dendrogram was then constructed with SHAN program in the software NTSYSpc-2.10 (Rohlf 2000). Also, principal coordinate analysis (PCA) was performed based on the variance–covariance matrix calculated from the marker data. To evaluate diverse level of each EST-SSR marker, polymorphism information content (PIC) value of each EST-SSR marker was calculated using the formula (Anderson *et al*. 1993): PIC = $1 - \sum P_{ij}^2$, where P_{ij} is the frequency of the jth allele (marker) for the ith EST-SSR locus.

Results and discussion

EST-SSR marker selection

A total of 533 SSRs between di-nucleotide and pentanucleotide repeats were identified from 4036 unigenes which were assembled from 6344 cucumber ESTs. Thirty SSR-containing ESTs were randomly selected and used for primer design. Most of the sequences with significant protein homology were related to energy metabolism, transcription, protein synthesis, translational control and photosystem (annotated with BLASTx; e-value $< 10^{-7}$). Since 20 ESTs were used for SSR marker development by Kong *et al*. (2006), development of new EST-SSR primers in present study did not involve these sequences. After repeated screening with L57 DNA template, 21 primer pairs were chosen for further study due to presence of clear and distinct band patterns.

EST-SSR polymorphism in cucumber

Twenty-one well-developed markers were used for amplification of 29 cucumber accessions containing various cucumber ecotypes in China. Totally 73 alleles were obtained and their size ranged from 110 to 489 bp. Most EST-SSR primer pairs amplified bands with their expected band sizes or around them, suggesting that EST-SSR polymorphism mainly derived from the length variation of the SSRs in the amplified bands. Unlike the most, several makers (EC11, EC31, EC28 and EC52) amplified multiple homoeoloci of the genome within at least one accession. Each EST-SSR marker detected two (EC12 and EC41) to five (EC11 and EC34) alleles with an average of 3.48. The allelic diversity of the polymorphic markers was from low to medium for which PIC values range from 0.185 (EC41) to 0.642 (EC34) with an average of 0.374 (table 1). The level of polymorphism observed herein is similar to that reported by Kong *et al*. (2006), but lower than that detected by genomic SSRs (Watcharawongpaiboon and Chunwongse 2008). Despite the fact that low level of polymorphism has been observed in EST-SSR markers, their origin from coding sequence and high interspecific transferability make them more attractive in genetic study (Varshney *et al*. 2005).

Diversity analysis of cucumber accessions

Pair-wise comparison was performed among all the accessions included in this study. Jaccard's similarity coefficients calculated from EST-SSR data varied from 0.542 (L109 and D01105) to 0.941 (Yichuanlian and SHG) with a mean of 0.792. This value is similar to that calculated from ISSR data

Table 1. Details of 21 cucumber EST-SSR markers developed from GenBank. The GenBank accession number, repeat motif, primer sequence, annealing temperature (T_a) are indicated. The number of alleles and PIC detected in 29 cucumber accessions are presented.

Locus name	GenBank accession no.	Repeat motif	Primer sequence (5′–3′)	T_a (°C)	No. of alleles	PIC
EC11	DN910157	$(TCT)_{10}$	F: TCTTCGCAGTCACCATTTC R: CCTTCCTCTGTTTCTGTTCC	55	5	0.578
EC12	DN910504	$(TCTT)_6$	F: TATTCTTCTTTGCTACCCAT R: TAAGTTTATTTTCTGCTGTCTA	53	2	0.279
EC13	DN909514	$(TTC)_9$	F: GCAATGAATCATGACCTCCA R: CTGAGAATTGGGAAGGGACA	55	3	0.426
EC15	DN909840	$(AGA)_6$	F: ACCAAAAACAGACCCCTATG R: GAAAGGGAAAACAAACGAGG	55	4	0.354
EC18	DN910067	$(CCT)_6$	F: TGCCATTTCATCGACTCTTC R: GCATTTCTGCTGTGGCTTAG	58	3	0.285
EC19	CK700725	$(TCC)_7$	F: TTTCTCTCCAACTCACCCTG R: TACATCGGCTTTGCTCATCT	55	3	0.347
EC20	DN910665	$(CT)_8$	F: AAAGTTGCTCTTGTTTGTCC R: GAGGTGAATGGTGGTGGCT	55	3	0.247
EC22	DN910469	$(AAG)_7$	F: CAGCAGAGAAACTCAATCCA R: CTGTTTACGGCTGCATTGGT	55	3	0.328
EC24	DN910437	$(TTA)_7$	F: ACAACACAACCGCTTCTCGT R: TGAGCCCAAGCACATAACAG	58	4	0.438
EC27	DN909941	$(ATC)_7$	F: GTTGGAAGGCACACAAAGTC R: CGAGATGATTGGAGGATGATG	55	3	0.461
EC28	DN910959	$(CAA)_8$	F: CTGAGTTATGGGGAAAGCAA R: TGTTAGTGATGTTGTTGGACC	58	4	0.441
EC31	CK758579	$(TTC)_7$	F: CTAACCAGCAGAACCCAATG R: GTATCCTGTTTCCAGCGAGA	55	4	0.238
EC34	AF202378	$(TA)_{15}$	F: GATCCCCATCATAATCACCC R: CAAAGGGCTACAATAACAAAC	55	5	0.642
EC35	CV000928	$(TTTTC)_5$	F: ATCCACAACACAAAAACCAC R: AAGAAGAACAGCCAAGAATG	55	3	0.392
EC39	BI740103	$(TCA)_7(CTT)_5$	F: CCAAGTTTAAGTTATTTAGGAG R: GAAGAGGACGATAAAGATGA	51	4	0.404
EC41	EW968287	$(AGA)_5$	F: AGCATGTGGAGGAGAAAGCA R: TTCATCATCGAGTGGGTCTG	58	2	0.185
EC47	AY942801	$(CT)_8$	F: CGATCTTTGTCATCCGACCT R: AGAACGAGCACGTTTTGAGC	60	4	0.485
EC49	CK758649	$(TCTTTC)_6$	F: CGTGTTTTCTCAGATTTCCCA R: CACTTCCCTTATCAACCCCA	58	4	0.540
EC50	CK758649	$(GAA)_5$	F: GGAACAGGGAAATCCACCAT R: TCGCTTCATCTCCCTCCTCC	58	3	0.247
EC52	CO998209	$(TAA)_6$	F: TCAAACACGAACCCGAAACG R: CAAGAAATTGCCAGGACGAG	58	4	0.300
EC56	AB112672	$(AGA)_{16}$	F: TTTTTGGGGGTTTTTGAGAG R: AGCTTTGTTCCCTATCTTCC	53	3	0.245

by Wang *et al*. (2007) but much lower than diversity value of melon (mean ≈0.64) (Xu *et al*. 2008), again confirming low genetic diversity in cucumber.

A dendrogram based on the similarity coefficients of the 29 accessions was constructed (figure 1), which clustered all the accessions into two major groups (groups I and II) at the similarity level of 0.753. This division is in line with geographical origins of the accessions. Almost all of the CN germplasms (NC and SC accessions and a wild accession SHG) were clustered into group I except for Tangshanqiugua and Ningjia1; all EA germplasms (EG, EO and OP accessions) were included in group II. From these, CN germplasms were distinctly separated from EA germplasms, suggesting different genetic backgrounds between them. EA accessions had higher diverse level than CN accessions; a mean similarity level of 0.785 in EA accessions versus a mean value of 0.847 in CN accessions was observed. High diverse level was observed within EA accessions corresponded

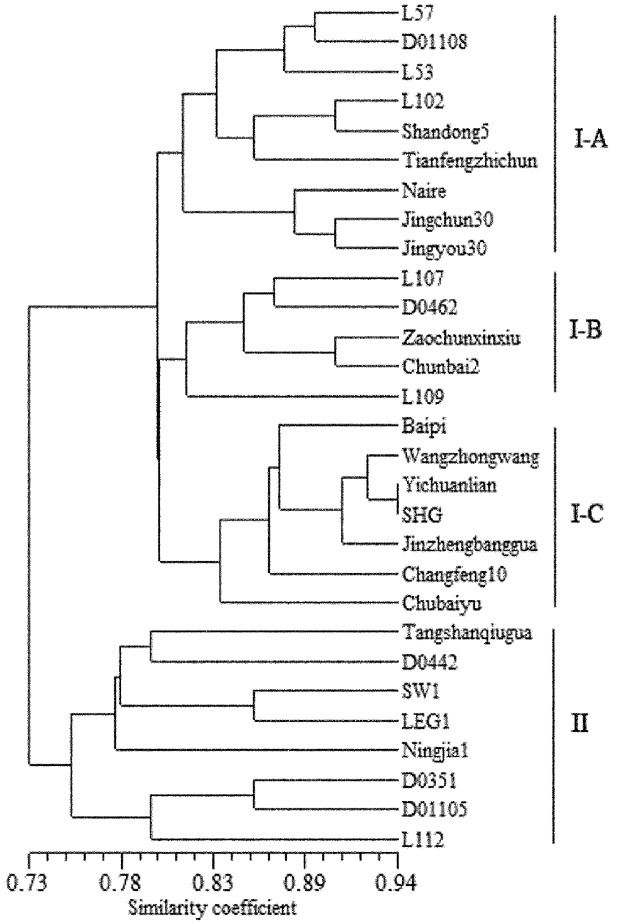

Figure 1. An unweighted pair-group method with arithmetic averages (UPGMA) dendrogram of genetic relationships among 29 cucumber accessions based on Jaccard's similarity coefficients from the EST-SSR data.

with their multiple origins: Holland, France and USA. Nevertheless, our results contrasted with the finding reported by Staub *et al*. (1999) that Chinese cucumber germplasms had higher genetic variation than from non-Chinese germplasms, and this difference could be due to different experimental materials or marker systems adopted in the latter case. As for the two exceptions, Ningjia1, usually being used for pickling, has an ancestral lineage of SW1 according to its breeding record; Tangshanqiugua may have a lineage of EA germplasm but there is no clear evidence at this time. At similarity level of 0.798, group I could be distinctly clustered into three subgroups (I-A, I-B and I-C), which accorded with the various ecotypes of the accessions on the whole. Cluster I-A contained nine NC accessions and cluster I-C consisted of six SC accessions and one wild accession (SHG). The wild accession SHG, characterized by prickless pericarp and short and club-shaped fruit, is Chinese native species and distributed in Yunnan province. It was closely clustered with few SC accessions (e.g., highest similarity level of 0.941 found between SHG and SC accession Yichuanlian), indicating that it could be an ancestry of Chinese cucumber, especially the SC cucumber. In our previous study with chloroplast SSRs (Hu *et al*. 2009), SHG shared a haplotype with some CN accessions, again supporting this opinion. At similarity level of 0.800, three NC accessions (L107, L109 and D0462) and two SC accessions (Zaochunxinxiu and Chunbai2) formed a group (I-B). A suggested explanation for this division is that these accessions probably had common genomic components that originated from other cucumber germplasm (not NC or SC germplasm). From the above results, the relationships among various types of cucumber accessions could be well resolved by the newly-developed markers, strongly verifying their effectiveness in identification of cucumber germplasm. Possibly, this finding implies a potential relation between the genic markers and cucumber phenotypes.

In a separate experiment, genetic relationship of the same set of cucumber accessions was analysed using genomic SSR markers reported by Danin-Poleg *et al*. (2001). By means of UPGMA, EA accessions were distinctly separated from CN accessions, however, subgroups of CN accessions did not agree with their ecotypes (data not shown). In comparison to EST-SSR analysis, there was a minor difference on clustering of CN accessions, which could be the result of the different regions of cucumber genome targeted by the two marker systems.

To better understand the genetic relationships of the cucumber accessions, PCA was performed based on the genetic similarity matrix. In this analysis, the first and second coordinate explained 35.8% and 11.7% of the variation, respectively and the first three Eigen vectors accounted for 53.2% of the variation observed. Figure 2 represents the distribution

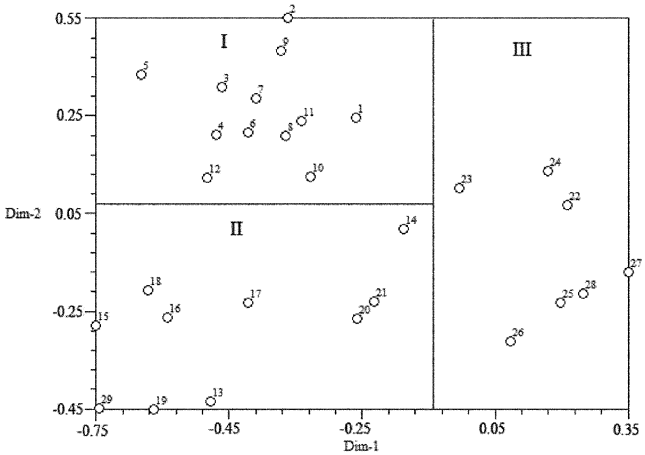

Figure 2. Diagram showing the relationships among 29 cucumber accessions based on principal coordinate analysis using EST-SSR markers. The accession numbers 1–29 represented by L53, L57, L102, L107, D01108, D0462, Naire, Jinchun30, Jinyou30, L109, Shandong5, Tianfengzhichun, Baipi, Tangshanqiugua, Wangzhongwang, Changfeng10, Chubaiyu, Jinzhengbanggua, Yichuanlian, Zaochunxinxiu, Chunbai2, D0351, L112, D0442, SW1, LEG1, D01105, Ningjia1 and SHG, respectively.

of different accessions according to the two principal axes of variation using PCA, which revealed classifications similar to the UPGMA analysis. NC accessions (region I), SC accessions (including SHG) (region II) and EA accessions (region III) were distinctly differentiated in the PCA diagram. Also, this division distinctly reflected the information of geographic origins and ecotypes of the accessions.

In conclusion, in this study, we developed 21 new EST-SSR markers for *C. sativus* and used them to analyse genetic variation of a set of cucumber accessions. Based on the EST-SSR data, different ecotypes of cucumber accessions were separated and the genetic relationships among them were well elucidated. Our results are of important reference value for parent selection, cross breeding or construction of mapping population.

Acknowledgements

This work is supported by grant from the National High Technology Research and Development Program of China (863 program) (No. 2007AA10Z177) and National Key Technology R&D Program (No. 2009BADB8B02). We are grateful to Prof. Jiancan Feng, College of Horticulture, Henan Agricultural University, for his assistance in the laboratory work.

References

Anderson J. A., Churchill G. A., Autrique J. E., Tanksley S. D. and Sorrells M. E. 1993 Optimizing parental selection for genetic linkage maps. *Genome* **36**, 181–186.

Creste S., Neto A. T. and Figueira A. 2001 Detection of single sequence repeat polymorphisms in denaturing polyacrylamide sequencing gels by silver staining. *Plant Mol. Biol. Rep*. **19**, 229–306.

Danin-Poleg Y., Reis N., Tzuri G. and Katzir N. 2001 Development and characterization of microsatellite markers in *Cucumis. Theor. Appl. Genet*. **102**, 61–72.

Horejsi T. and Staub J. E. 1999 Genetic variation in cucumber (*Cucumis sativus* L.) as assessed by random amplified polymorphic DNA. *Genet. Resour. Crop Evol*. **46**, 337–350.

Hu J. B., Li J. W. and Zhou X. Y. 2009 Analysis of cytoplasmic variation in a cucumber germplasm collection using chloroplast microsatellite markers. *Acta Physiol. Plant*. **31**, 1085–1089.

Huang X. and Madan A. 1999 CAP 3: a DNA sequence assembly program. *Genome Res*. **9**, 868–877.

Jaccard P. 1908 Nouvells recherchess sur la distribution florate. *Bull. Soc. Vaudosee Sci. Nat.* **44**, 223–270.

Kong Q., Xiang C. and Yu Z. 2006 Development of EST-SSRs in *Cucumis sativus* from sequence database. *Mol. Ecol. Notes* **6**, 1234–1236.

Li X. X., Zhu D. W., Du Y. C., Sheng D., Kong Q. S. and Song J. P. 2004 Studies on genetic diversity and phylogenetic relationship of cucumber (*Cucumis sativus L.*) germplasm by AFLP technique. *Acta Hortic. Sin*. **31**, 309–314 (in Chinese, with English abstract).

Murray M. G. and Thompson W. F. 1980 Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res*. **8**, 4321–4325.

Pitrat M., Chauvet M. and Foury C. 1999 Diversity, history, and production of cultivated cucurbits. *Acta Hortic*. **492**, 21–28.

Rohlf F. J. 2000 *NTSYS-pc: Numerical taxonomy and multivariate analysis system*, version 2.1. Exeter Publications, New York, USA.

Staub J. E., Chung S. M. and Fazio G. 2005 Conformity and genetic relatedness estimation in crop species having a narrow genetic base: the case of cucumber *(Cucumis sativus L.). Plant Breed*. **124**, 44–53.

Staub J. E., Serquen F. C. and McCreight J. D. 1997 Genetic diversity in cucumber (*Cucumis sativus* L.): III. An evaluation of Indian germplasm. *Genet. Resour. Crop Evol*. **44**, 315–326.

Staub J. E., Serquen F. C., Horejsi T. and Chen J. 1999 Genetic diversity in cucumber (*Cucumis sativus* L.): IV. An evaluation of Chinese germplasm. *Genet. Resour. Crop Evol*. **46**, 297–310.

Varshney R. K., Graner A. and Sorrells M. W. 2005 Genic microsatellite markers in plants: features and applications. *Trends Biotechnol*. **23**, 48–55.

Wang J., Xu Q., Miao M. M., Liang G. H., Zhang M. Z. and Chen X. H. 2007 Analysis of genetic relationship of cucumber (*Cucumis sativus* L.) germplasm by ISSR markers. *Mol. Plant Breed*. **5**, 677–682.

Watcharawongpaiboon N. and Chunwongse J. 2008 Development and characterization of microsatellite markers from an enriched genomic library of cucumber *(Cucumis sativus). Plant Breed*. **127**, 74–81.

Xu Y. 1994 Cucumber (*Cucumis sativus* L.). In *Vegetable germplasm resources* (ed. C. J. Zhou), pp. 163–171. Beijing Agricultural University Press, Beijing, P. R. China (in Chinese).

Xu Z. H., Xu Y. Y., Liu J. P. and Sun Z. Q. 2008 Studies on the genetic diversity and phylogenetic relationship of melon (*Cucumis melo* L.) germplasm. *J. Fruit Sci*. **25**, 552–558 (in Chinese, with English abstract).

Yasodha R., Sumathi R., Chezhian P., Kavitha S. and Ghosh M. 2008 Eucalyptus microsatellites mined *in silico:* survey and evaluation. *J. Genet*. **87**, 21–25.

Received 15 September 2009, in final revised form 30 December 2009; accepted 4 January 2010
Published on the Web: 30 June 2010