

Development of ten microsatellite markers for *Quercus mongolica* var. *crispula* by database mining

Saneyoshi Ueno · Yoshihiko Tsumura

Received: 30 October 2007 / Accepted: 4 November 2007 / Published online: 18 November 2007
© Springer Science+Business Media B.V. 2007

Abstract Simple sequence repeats (SSRs) in the NCBI dbEST database were surveyed to identify potential SSR markers for *Quercus mongolica*. In total, 2,691 gene sequences, mainly from expressed sequence tags (ESTs) for *Q. robur* and *Q. petraea* had been registered. Twenty-two PCR primers were designed for SSRs in these sequences and screened for polymorphisms in 16 *Q. mongolica* trees. Ten loci were easily genotyped and showed polymorphism, with numbers of alleles and expected heterozygosity ranging from 3 to 15 and 0.28 to 0.94, respectively. These EST-SSR markers should be useful for studying the genetic diversity of *Quercus* species.

Keywords Expressed sequence tag · Fagaceae · *Quercus petraea* · *Quercus robur* · SSR

Quercus is one of the most prominent genera of plants in the world, including about 500 species distributed throughout the Northern hemisphere (Nixon 1993). However, 35 species in the genus are recorded in the International Union for Conservation of Nature and Natural Resources Red List (<http://www.iucnredlist.org/>) as critically endangered, endangered or vulnerable. Genetic aspect of these threatened species should be carefully analyzed to develop effective conservation strategies.

One of the major obstacles for genetic studies of rare species is the general lack of genetic markers such as simple sequence repeats (SSRs), development of which is

costly and time-consuming. Fortunately, however, the massive accumulations of DNA sequences, especially expressed sequence tags (ESTs), in publicly available databases may facilitate the development of these markers (Pashley et al. 2006). Furthermore, since ESTs are gene sequence and conservative, primers designed for the target species are likely to work well in related species. The utility of DNA databases for developing SSR markers should be demonstrated to accelerate such a strategy.

Quercus mongolica var. *crispula* is one of the most silviculturally important broad-leaved deciduous tree species in Japan. Since increasing importance is being attached to multi-functional aspects of such trees, e.g. for watershed protection, recreational use and maintenance of biodiversity as well as timber production, the areas reforested with broad-leaved trees in Japan, especially *Quercus* species, have increased (The Ministry of Agriculture 2002). Ideally, the genetic composition and diversity of the populations used as sources of planting material should be carefully analyzed in such reforestation programs. Therefore, in the present study, we have developed SSR markers for *Q. mongolica*, by mining 2,691 publicly available sequences originating from mRNA.

There were 1,439 and 1,233 ESTs for *Q. robur* and *Q. petraea*, respectively, registered in dbEST (<http://www.ncbi.nlm.nih.gov/dbEST/>). These sequences, along with additional 19 mRNA sequences, were downloaded and processed using read2Marker (Fukuoka et al. 2005) under default settings. Locations of microsatellites (coding/non-coding region) and putative gene functions were also estimated using ESTScan (Iseli et al. 1999) and blastx (Altschul et al. 1990), respectively. The e-value cutoff for blastx was set at $1e-10$.

A total of 22 primer pairs were designed for sequences with at least nine and six repeats for di- and tri-SSRs,

S. Ueno (✉) · Y. Tsumura
Tree Genetics Laboratory, Department of Forest Genetics,
Forestry and Forest Products Research Institute, 1 Matsunosato,
Tsukuba, Ibaraki 305-8687, Japan
e-mail: saueno@ffpri.affrc.go.jp

Table 1 Characteristics of the ten polymorphic microsatellite markers in *Q. mongolica*

Accession no.	Primer sequence	Repeat motif	N	Size range (bp)	Na	H _o	H _e	SSR location	Blastx top hit ^a Description (E-value)
AI577265	ATTGAGCCGCAAAATCAATAAGG GAAACTGGTCCCTTCTCTTGGG	(GAA) ₆	16	374–380	3	0.50	0.66	Coding	Pheromone receptor-like protein (8E–77)
CN725667	GCTAAGCTCCAAAGCCATTGTGA CCGATGACGTGGATGTAATCTCC	(GAA) ₆	16	202–207	3	0.50	0.46	Coding	–
CR627959	GCTCCCTGGTAGTCGGCTAAAGA CAATTGGGACAAACATGGAAGCAT	(GCA) ₇	16	243–251	4	0.63	0.59	Coding	Cys-3-His zinc finger protein (3E–35)
DN949770	TCTCCTTCTGCACCTCACTACGA AACGTAAAAGGCAAAATTGATCGGG	(TC) ₉	14	277–337	15	0.86	0.94	Non-coding	–
DN949776	CAACCATCGAGGCCATTACGAA TCACCGATCTTGAAGGTCCTCAG	(AG) ₉	16	156–168	4	0.44	0.69	Non-coding	60S ribosomal protein L21 (1E–85)
DN950446	TCTCTTCTCCGTCCATTATCGC GTTTCTCCACAGACCCCAATTTC	(AG) ₁₉	15	149–171	13	0.87	0.91	Non-coding	–
DN950634	TCAAAACAATCTCAAAGGTCCCAA GCTTTTGAGAAACTTTGGCCACC	(TC) ₁₀	14	361–381	10	0.71	0.87	Non-coding	Putative carboxyl-terminal proteinase (2E–51)
DN950717	TAGTTTTCCCA GCGAATCCAACA CTTCTTGAAGGGACTGACCCCAT	(GGA) ₆	16	242–260	3	0.31	0.28	Coding	Salt tolerance protein 5-like protein (3E–58)
DN950726	GCAAGAAGCATGCAGATGGAGAT GCATGGCCGTCATTAGCATTAAAG	(GAT) ₆	15	314–319	3	0.60	0.54	Coding	–
DN950883	AGACCAGAGGAGAGCTGAGAGGTG AACCAACCTCCCTGGACAAAAT	(TC) ₉ TG(TC) ₄	15	394–401	6	0.33	0.36	Non-coding	–

N, Number of individuals genotyped; Na, Number of alleles per locus; H_o, Observed heterozygosity; H_e, Expected heterozygosity

^a No hits found

respectively. PCR was carried out in 8 μ l reaction mixtures containing ca. 10 ng genomic DNA, 1 \times PCR buffer, 200 μ M of each dNTP, 1.5 mM MgCl₂, 0.2 μ M of each primer designed in the present study and 0.4 U of *Taq* polymerase (Promega), using the following program: 94°C for 3 min, then 40 cycles of 94°C for 45 s, 55°C for 45 s and 72°C for 45 s, followed by a final extension at 72°C for 7 min. The PCR products were electrophoretically separated on 2% agarose gels and stained with ethidium bromide. Sixteen individuals were used for screening polymorphisms, by PCR carried out in 6 μ l reaction mixtures under the conditions described above, except that the amount of *Taq* polymerase was reduced to 0.15 U. PCR products were labeled with ChromaTide Rhodamine Green-5-dUTP (Molecular Probes) according to the method of Kondo et al (2000), and analyzed using a 3100 Genetic Analyzer with GeneScan software. For each locus, the number of alleles (N_a), observed heterozygosity (H_o) and expected heterozygosity (H_e) were calculated and Hardy-Weinberg (HW) and genotypic equilibrium were tested using FSTAT software (Goudet 1995). Ten loci showed clear polymorphic patterns, with H_o and H_e ranging from 0.31 to 0.87 and 0.28 to 0.94, respectively (Table 1). The locus DN949776 showed significant deviation from HW equilibrium ($P = 0.0265$). However, no loci showed HW disequilibrium after the significance levels were adjusted for multiple-tests. No significant genotypic disequilibrium was detected for any pair of loci. All of the di-SSR loci were estimated to be in non-coding regions, with N_a ranging from 4 to 15, while all tri-SSR loci were estimated to be in coding regions with N_a of 3 or 4. Half of the ten loci in Table 1 showed similarity with other proteins and putative functions were proposed for them. In summary, we have efficiently developed 10 EST-SSR markers for *Q. mongolica* from 2,691 *Q. robur* and *Q.*

petraea gene (EST) sequences. The study shows that developing SSR markers by database mining is rapid, convenient and cost-effective, and we hope that the ten markers developed in it will be useful for analyzing the genetic diversity of *Quercus* species.

Acknowledgements The authors are grateful to M. Kanno, A. Kanazashi, H. Yoshimaru, K. Ishida, Y. Koyama and K. Koono for sample collection and to Y. Taguchi for laboratory work. This research was supported by a grant for Research on Genetic Guideline for Restoration Programs using Genetic Diversity Information from the Ministry of Environment, Japan.

References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410
- Fukuoka H, Nunome T, Minamiyama Y, Kono I, Namiki N, Kojima A (2005) Read2Marker: a data processing tool for microsatellite marker development from a large data set. *Biotechniques* 39:472, 474, 476
- Goudet J (1995) FSTAT (Version 1.2): a computer program to calculate F-statistics. *J Hered* 86:485–486
- Iseli C, Jongeneel CV, Bucher P (1999) ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. In: Proceedings of the International Conference on Intelligent Systems for Molecular Biology, AAAI press, Menlo Park, California, pp 138–148
- Kondo H, Tahira T, Hayashi H, Oshima K, Hayashi K (2000) Microsatellite genotyping of post-PCR fluorescently labeled markers. *Biotechniques* 29:868–872
- Nixon KC (1993) Infrageneric classification of *Quercus* (Fagaceae) and typification of sectional names. *Ann Sci For* 50:25s–34s
- Pashley CH, Ellis JR, McCauley DE, Burke JM (2006) EST databases as a source for molecular markers: lessons from *Helianthus*. *J Hered* 97:381–388
- The Ministry of Agriculture, Forestry and Fisheries of Japan (2002) Ringyo Chi-iki Tyousa Houkokusyo (Forestry District Survey Report). Tokyo