File S1

Materials and Methods

Plant materials: In the present study, we used 48 seed samples of *Cryptomeria japonica*. Genetic characterizations of natural populations of *C. japonica* are described in Tsumura et al. (2007). The samples were the same as those used in the study by Kado *et al.* (2003) and Fujimoto *et al.* (2008) and were collected from three areas in Japan (Kantou-Toukai, Hokuriku, and Iwate). The average estimates of F_{ST} between populations based on the data of seven nuclear loci were 0.01-0.04 (Kado *et al.* 2003). Hence, we regard the present samples of *C. japonica* as homogeneous and pool the data of the three populations. Also, although the samples were collected from artificial forests, their genetic compositions were very similar to those of the natural populations of *C. japonica* (Kado *et al.* 2006). We also used two seed samples of *Taxodium distichum*, one of the closest relatives of *C. japonica*. These samples were used as outgroups for the analysis of a putative gene found in the studied region. All seed samples were provided from The Forestry and Forest Products Research Institute in Japan, and haploid genome DNA was extracted from the megagametophytes using a modified SDS method. Briefly, the megagametophytes from seeds were homogenized in extraction buffer (0.1 M Tris-HCl pH 8.0, 10 mM EDTA, 0.5% SDS, and 0.1 mg/ml Proteinase K), and haploid DNA was extracted from it using phenol/chloroform and ethanol.

Sequence analysis of BAC clones: Eight BAC clones (BAC1-BAC8) were randomly chosen from the BAC library developed by The Forestry and Forest Products Research Institute in Japan and sequenced using a Roche-454 Genome Sequencer FLX Titanium (Roche, Indianapolis, IN, USA) by Takara Bio Inc. (Ohtsu, Japan). The sequences of the fragments of each BAC clone were assembled using GS De Novo Assembler version 2.0, with its default setting. We chose the largest contig in the assembled sequences of each BAC clone and used it as the sequence of the BAC for designing primers. The eight BAC sequences were named BAC1-BAC8. The average number of reads per base pair for those contigs ranged from 12.9 to 86.3. The full analysis of the BAC sequences will be published elsewhere (M. Tamura, A. Watanabe, K. Uchiyama, N. Futamura, K. Shinohara, Y. Tsumura, H. Tachida, unpublished results).

Loci and primers: The PCR primers were based on the BAC sequences of *C. japonica*. The BAC sequences contained putative transposable elements and many other repetitive sequences (M. Tamura, A. Watanabe, K. Uchiyama, N. Futamura, K. Shinohara, Y. Tsumura, H. Tachida, unpublished results). In the design of primers, we avoided the repetitive regions. The accuracy of the distances on the BAC sequences for which genetic diversity was investigated (BAC3, BAC6 and BAC7) was confirmed using long PCR and electrophoresis. To increase the specificity of the amplification, we carried out nested PCR at all loci. First, PCR primers were designed to generate fragments of approximately 3 to 15 kbp in size. Then, a second set of PCR primers was designed to amplify one or more regions within the first PCR product so that the lengths of the fragments

for which sequences were determined would be approximately 700 to 2000 bp. BAC6 contained a partial coding gene that was homologous to the calcium-dependent protein kinase of *Populus trichocarpa* (XP_002322129.1). The size of one of the introns was greater than 70 kb. We designed primer pairs so that all exons contained in BAC6 could be sequenced. The positions of the regions used for the diversity analysis and structure of the coding region are shown in Figure 1. If necessary, we designed internal primers for sequencing. All primers were designed using Primer3 software (http://primer3.sourceforge.net/). The primer sequences used are listed in Table S1 (supporting information).

Amplification and sequencing: We carried out long PCR using TaKaRa LA Taq (Takara, Ohtsu, Japan) for long fragments (> 10 kb), but for those with smaller sizes, we used TaKaRa ExTaq (Takara, Ohtsu, Japan). The first PCR conditions were 3 min at 98°, followed by 35 cycles of 10 sec at 98° and 15 min at 68°, and finally, 15 min at 68°. The products were directly used as templates for the second PCR. The second PCR conditions were 3 min at 95°; followed by 25 cycles of 30 sec at 95°, 30 sec at 50°, and 1-2 min at 72°; and a final extension of 7 min at 72°. The products were purified by PEG precipitation treatment to remove surplus primers and dNTP. The DNA fragments were directly sequenced on an ABI Prism 3100 automatic sequencer or Applied Biosystems 3730 DNA Analyzer using a Big Dye Terminator v3.1 Cyclesequence Kit (Applied Biosystems, Foster City, CA, USA). All sequences described in this article have been deposited in the GenBank/EMBL/DDBJ databases (accession nos. AB686666-AB687489).

Analyses of data: All sequences were assembled using the SeqMan package in Lasersene 7.1 (DNASTAR, Inc., Madison, WI, USA) and aligned manually using MEGA3 (Kumar *et al*. 2004). By comparing these sequences with those of the cDNA clones, we could determine the coding regions. Estimates of standard population genetics statistics, such as π, Tajima's *D* (Tajima 1989), Fu and Li's F^* and D^* (Fu and Li 1993) and the squared allelic correlation coefficient (r^2), between sites were calculated using DnaSP 5.0 (Rozas *et al.* 2003).

We also estimated the population recombination rate, $4N_er$, where N_e and r are the effective size and the recombination rate, respectively, using the composite likelihood method (Hudson 2001; McVean *et al.* 2002) implemented in LDhat 2.1 (www.stats.ox.ac.uk/~mcvean/LDhat.html). First, we made FASTA-like files in which undetermined nucleotides were represented by "?" and the program convert was run. We used all polymorphic sites as data. Next, the "locs" and "sites" files in the outputs of convert were used as inputs for the program pairwise. The program pairwise assumes a constant recombination rate across the region. We used the crossing-over model to estimate $4N_er$, and the parameter region searched was $4N_er$ = 0.0-100.0. The minimum number of recombination events (Hudson and Kaplan 1985) was also obtained by the program.

REFERENCES

Fu, Y. X., and W. H. Li, 1993 Statistical tests of neutrality of mutation. Genetics **133:** 693-709.

Fujimoto, A., T. Kado, H. Yoshimaru., Y. Tsumura and H. Tachida, 2008 Adaptive and slightly deleterious evolution in a conifer, *Cryptomeria japonica*. J. Mol. Evol. **67:** 201-210.

Hudson, R. R., 2001 Two-locus sampling distributions and their application. Genetics **159:** 1805-1817.

Hudson, R. R., and N. Kaplan, 1985 Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics **111:** 147-164.

Kado, T., H. Yoshimaru, Y. Tsumura and H. Tachida, 2003 DNA variation in a conifer, *Cryptomeria japonica* (Cupressaceae sensu lato). Genetics **164**: 1547-1559.

Kado, T., Y. Ushio, H. Yoshimaru, Y. Tsumura and H. Tachida, 2006 Contrasting patterns of DNA variation in natural populations of two related conifers, *Cryptomeria japonica* and *Taxodium distichum* (Cupressaceae sensu lato). Genes & Genetic Systems **81:** 103-113.

Kumar, S., K. Tamura and M. Nei, 2004 MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Briefings in Bioinformatics **5:** 150-163.

McVean, G., P. Awadalla and P. Fearnhead, 2002 A coalescent-based method for detecting and estimating recombination rates from gene sequences. Genetics **160:** 1231-1241.

Rozas, J., J. C. Sanchez-DelBarrio, X. Messeguer and R. Rozas, 2003 DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics **19:** 2496-2497.

Tajima, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123:** 585-595.

Tsumura, Y., T. Kado, T. Takahashi, N. Tani, T. Ujino-Ihara *et al.*, 2007 Genome scan to detect genetic structure and adaptive genes of natural populations of *Cryptomeria japonica*. Genetics **176:** 2393-2403.