

Adaptive and Slightly Deleterious Evolution in a Conifer, *Cryptomeria japonica*

Akihiro Fujimoto · Tomoyuki Kado ·
Hiroschi Yoshimaru · Yoshihiko Tsumura ·
Hidenori Tachida

Received: 4 May 2007 / Accepted: 23 June 2008
© Springer Science+Business Media, LLC 2008

Abstract In order to evaluate effects of the population structure and natural selection on organisms having long generation times, we surveyed DNA polymorphisms at five loci encoding 9-cis-epoxycarotenoid dioxygenase (*NCED*), ammonium transporter, calmodulin, aquaporin, and the second major allergen with polymethylgalacturonase enzyme activity in the pollen (*Cryj2*) in a conifer, *Cryptomeria japonica*. The average nucleotide diversity at silent sites across 12 loci including the previously analyzed seven loci was 0.0044. The population recombination rate ($4Nr$, where N and r are the effective population size and recombination rate per base per generation, respectively) was estimated as 0.00046 and a slow reduction in the population size was indicated, according to the maximum likelihood method implemented in LAMARC. At *NCED*, the McDonald-Kreitman (MK) test revealed an excess of replacement polymorphisms, suggesting contributions of slightly deleterious mutations. In contrast, the MK test revealed an excess of replacement divergence at *Cryj2* and a maximum likelihood approach using the PAML package revealed that certain amino acid sites had a nonsynonymous/synonymous substitution rate ratio (ω) > 4.0, indicating adaptive evolution at this locus. The overall analysis of the 12 loci suggested that adaptive, neutral, and slightly deleterious evolution played important roles in the evolution of *C. japonica*.

Keywords Conifer · DNA polymorphism · Slightly deleterious mutation · Adaptive evolution · Recombination rate · Population size

Introduction

In recent years, DNA sequence variation has been investigated in many organisms (Baum et al. 2003; Kvist et al. 2003; Yu et al. 2003). The development of analytical tools based on population genetics has enabled the expansion of our knowledge of the evolutionary forces that affect sequence variation (e.g., Felsenstein 2004). Because the effects of evolutionary forces are thought to differ depending on species-specific factors (e.g., generation time, mutation rate, recombination rate, population structure, mating system (Hamrick and Godt 1990)) and because the pattern of natural selection would in turn be affected by these factors, it is important to characterize variation in a wide variety of species. However, in order to evaluate the influences of such species-specific factors, it is necessary to analyze multilocus data, which is costly and time-consuming. Although multilocus surveys have been carried out to study herbaceous plants (Wright and Gaut 2005) such as *Arabidopsis* (Aguade 2001; Nordborg et al. 2005) and maize (Tenaillon et al. 2001), and species-specific characteristics have been shown to affect patterns of genetic variation (Bustamante et al. 2002), the corresponding data regarding trees, which characteristically have long generation times, have just recently begun to accumulate (Ingvarsson 2005; Neale and Savolinen 2004; González-Martínez et al. 2006). Although these studies indicated the importance of the population structure on genetic variation, and the levels of variation and patterns of linkage disequilibrium (LD) have been shown to differ among species,

A. Fujimoto · T. Kado · H. Tachida (✉)
Department of Biology, Faculty of Sciences, Kyushu University,
Ropponmatsu, Fukuoka 810-8560, Japan
e-mail: htachscb@mbox.nc.kyushu-u.ac.jp

H. Yoshimaru · Y. Tsumura
Forestry and Forest Products Research Institute, Kukizaki,
Ibaraki 305-8687, Japan

the details regarding patterns of variation and their evolutionary causes, especially as regards the role of selection in each species, are not well understood in the tree species.

Here we report sequence variation in a conifer, *Cryptomeria japonica*, which belongs to Cupressaceae sensu lato (Gadek et al. 2000; Kusumi et al. 2000). This species is a long-lived and predominantly outcrossing tree found in Japan (for more information see Takahashi et al. 2005). The average F_{IS} in this species based on data of 148 cleaved amplified sequence markers is negative and not significantly different from zero (Tsumura et al. 2007). Kusumi et al. (2002) reported the characteristics of molecular evolution using 11 nuclear genes in Cupressaceae. They found a significant correlation between synonymous substitution rates (d_s) and nonsynonymous substitution rates (d_n) across genes, and the estimate of the index of dispersion (Gillespie 1989) for nonsynonymous substitutions was low (≈ 1.0). Based on these observations, it was suggested that nonsynonymous substitutions in Cupressaceae were primarily driven by mutation, and not by diversifying selection. This pattern differs from that of mammals (>5) (Gillespie 1989; Ohta 1995) and may indicate that a larger portion of the nonsynonymous sites in Cupressaceae evolved neutrally, compared to that of mammals, although results based only on the index of dispersion may not be conclusive. Indeed, *Drosophila* data also show a low estimate of the index of dispersion (1.6; see Zeng et al. 1998), but other analyses show significant roles of positive and negative selection in protein evolution in *Drosophila* (Fay et al. 2002; Smith and Eyre-Walker 2002).

Kado et al. (2003) surveyed DNA sequence polymorphisms of *C. japonica* at seven loci. They found that the level of nucleotide variation at silent sites was moderate ($\pi_{sil} = 0.00384$) in this species, and the LD extended over a longer distance than in the case of either *Z. mays* or *Drosophila*. They also found a locus (*Acl5*) possibly affected by selective sweep and another locus (*Lcyb*) affected by slightly deleterious mutations. However, a few issues remained unresolved in that study. First, their analysis of the frequency spectrum for the combined data indicated a significant deviation from that of the standard neutral model, although neither Tajima's test nor other tests gave significant results. To account for the deviation found in the frequency spectrum, Kado et al. (2003) suggested that the population of *C. japonica* had previously undergone a reduction in size. However, their data were not sufficiently exhaustive to estimate previous population structures. Second, as the lengths of the regions analyzed by Kado et al. (2003) were for the most part relatively short, the population recombination rate, an important parameter in evolutionary analyses, could not be estimated. Third, many loci involved in disease resistance or sexual reproduction

are considered to be affected by adaptive evolution (Bergelson et al. 2001; Clark et al. 2006) but such loci were not included in the previous survey.

In the present study, we surveyed five new nuclear loci including one possibly involved in sexual reproduction in *C. japonica*. A combined analysis of the present data with those of Kado et al. (2003) using the maximum likelihood (ML) method indicated a very slow reduction in the population size of this species. The population recombination rates were also estimated by the ML method and were compared with those in two other species, *Pinus taeda* and *Z. mays*. Neutrality tests identified loci affected by slightly deleterious mutations and positive selection. Interestingly, positive selection was observed at a locus expressed at the pollen surface, and this locus is thus considered to play a role in sexual reproduction.

Materials and Methods

Plant Materials and Sequencing of Alleles

We used the same 48 seed samples of *C. japonica* as those used in the study by Kado et al. (2003). Sixteen samples were collected from multiple artificial forests in each of three areas, Kantou-Toukai (KT), Hokuriku (HK), and Iwate (IW), covering a broad area in Japan (for their locations see Kado et al. 2003). Because the genetic composition of the artificial forests is very similar to that of the natural populations in *C. japonica* as a whole (Kado et al. 2006), we think those samples represent most of the current genetic variation in *C. japonica*. We also used a seed sample of *Taxodium distichum*, which is among the closest relatives of *C. japonica*. The Forestry and Forest Products Research Institute in Japan provided us the seed samples, and haploid DNA was extracted from the megagametophyte according to the SDS method. To examine molecular evolution at one locus (*Cryj2*), we also employed a seed sample of *Chamaecyparis obtusa*, which was obtained from the Forestry and Forest Products Research Institute in Japan, and a seed sample of *Metasequoia glyptostroboides* sampled in Fukuoka prefecture in Japan. Both species belong to Cupressaceae sensu lato (Gadek et al. 2000; Kusumi et al. 2000) but are more distantly related to *C. japonica* than *T. distichum* is. The haploid DNAs of these samples were extracted as described above.

Five nuclear loci were amplified by PCR and directly sequenced by using haploid megagametophyte DNA. On the basis of their sequence homology, four of the loci were considered to encode 9-cis-epoxycarotenoid dioxygenase (*NCED*), ammonium transporter (*AMT*), calmodulin (*Cal*), and aquaporin (*AQU*), respectively. Another locus codes

for the second major allergen of Japanese cedar pollinosis (*Cryj2*) (Namba et al. 1994) with polymethylgalacturonase enzyme activity (Ohtsuki et al. 1995). The PCR primers for the former four loci were designed based on data in the Sugi (*C. japonica*) Genome Database (Forestry and Forest Products Research Institute of Japan). The PCR primers for *Cryj2* were designed based on sequences in the NCBI database under accession numbers D37765 and AB084068. The amplification was carried out for 10 min at 95°C, followed by 35 cycles of 30 s at 95°C, 30 s at 50°C, and 1–4 min at 72°C, with a final 10-min incubation at 72°C. In order to increase the specificity of amplification, we carried out nested PCR using two primer sets for each gene. The PCR primers and the duration of extension are listed in Table 1.

After amplification, the PCR products were sequenced on an ABI Prism 3100 automatic sequencer by using the BigDye terminator cycle sequencing ready reaction kit ver. 3 or ver. 3.1 (PE Biosystems, Foster City, CA). Additional primers were designed in order to sequence both strands. All sequences were assembled and aligned manually with no ambiguity.

Data Analysis

The polymorphism data were analyzed by DnaSP ver. 4.00.0 (Rozas et al. 2003). Nucleotide diversity, π (Nei and Li 1979), and θ_w (Watterson 1975) were estimated. F_{ST} (Hudson et al. 1992) and S_{nn} (Hudson 2000) were used as indexes of population subdivision. The neutral hypothesis was tested by Tajima's (1989) D , Fu and Li's (1993) D^* , Fu and Li's

(1993) F^* , Fay and Wu's (2000) H , R_2 (Ramos-Onsins and Rozas 2002), the Hudson-Kreitman-Aguade (HKA) test (Hudson et al. 1987), and the McDonald-Kreitman (MK; 1991) test. We also carried out the multilocus HKA test for all 12 loci including those analyzed by Kado et al. (2003) by using a program HKA developed by J. Hey (<http://lifesci.rutgers.edu/~heylab/HeylabSoftware.htm#HKA>). The ratio r/u , where u is the mutation rate per site per generation and r is the recombination rate per site per generation, was estimated by using coalescent-likelihood programs in the LAMARC package (Kuhner et al. 2004) along with the past exponential population growth rate g (Kuhner et al. 2000). The exponential population growth rate g is defined by $\theta(t) = \theta(0) \times \exp(-gt)$, where t is the time measured backward from the present in units of u^{-1} and $\theta(t) = 4N_tu$, with N_t designating population size at t (Kuhner et al. 1998).

As regards *Cryj2*, because we obtained indication of action of positive selection as described in the result, we used the PAML v. 3.14 (Yang 1997) to test for adaptive molecular evolution and to identify amino acid sites that had undergone diversifying selection. ML estimates of the nonsynonymous/synonymous substitution rate ratio (ω) and the posterior probabilities of sites belonging to particular classes were obtained according to various models, i.e., M0 (one-ratio), M1a (nearly neutral), M2a (positive selection), and M3 (discrete), as well as M7 (β) and M8 (β & ω). We carried out likelihood-ratio tests in order to examine the statistical significance of the fits of the models (M0 vs. M3, M1a vs. M2a, M1a vs. M3, and M7 vs. M8). We use the notations described in the PAML manual for the estimated parameters.

Table 1 Primers used to amplify the five loci in the first and second PCR analyses

Locus	First primer sequence (5'–3')	t^a	Second primer sequence (5'–3')	t^a
<i>NCED</i>	TGTCCTCAACTTCATCACCATCTT AGCAGATAGCAGAGAATGAGAA	2 min	TGTCCTCAACTTCATCACCATCTT ^b	1.5 min
			CGGCACCACGACATACTTG ^b	
			GAAGCAAGGAATCGCCAGTT ^c AGCAGATAGCAGAGAATGAGAA ^c	1.5 min
<i>AMT</i>	CTACCAAATCCAGGCGACAA CCCTGGCTGAATGTAAAGTAA	3 min	CTACCAAATCCAGGCGACAA ^b	1.5 min
			CTCCATAGTTAGTCTCCAG ^b	
			TTCTCATCAAGCAAGCCAAAGT ^c CCCTGGCTGAATGTAAAGTAA ^c	1.5 min
<i>Cal</i>	TATATTACATCAAAGCCATTAG TGCCAGAATGCCACCAAAATAA	4 min	GCCATCTCACAACAATCAGT GCCACCAAAATAAAAGACAATG	4 min
<i>AQU</i>	TTTTAGGGAATTTTGTCAAGG AGGGGGATTGCTGGTTCTAA	2 min	GCTAAGGAAGAAGGCAAGGA GCCCTGCTGCTACAGAGAATA	2 min
<i>Cryj2</i>	CTCGATCCCACTAGAATTCA TGCCTTAGAAACATTTATGATT	4 min	TACCTTCTACTAGCCTAATTT TTCCTAGCACATACAGTTTCA	4 min

^a Extension time

^b Primer used to amplify the 5' portion of the region

^c Primer used to amplify the 3' portion of the region

Results

Diversity in *C. japonica* at Five Newly Sequenced Loci

All exons and introns were sequenced (AB211598–AB211844) except for the first eight codons of *AQU* (Fig. 1). In total, we obtained DNA sequence data of ~ 670 kb (13,914 bp × 48 samples). Altogether, 1.99% of the synonymous sites, 0.63% of the nonsynonymous sites, and 2.76% of the silent sites were polymorphic, and 3.32% of the total region contained indel polymorphisms. The lengths of sequenced regions, excluding those involving indels, and the estimates of nucleotide variation are summarized in Table 2. The nucleotide diversity at silent sites, π_{sil} , varied from 0.00171 at *AMT* to 0.00948 at *Cryj2*. The unweighted average π_{sil} over the five loci was 0.0046. This value was slightly higher than that over the seven loci (Kado et al. 2003). The unweighted average of the combined 12 loci was 0.0044.

There was no nonsynonymous polymorphic site at *AMT*, *AQU*, or *Cal*. However, the *NCED* locus and *Cryj2* had 12 and 17 nonsynonymous polymorphic sites, respectively. The ratios of nucleotide diversity at nonsynonymous sites, π_{rep} , to π_{sil} were 0.96 at *NCED* and 0.37 at *Cryj2*. Therefore, the ratio differed greatly among loci. In particular, in the signal sequence region (1238–1399 bp) and the C-terminal cleaved region (3881–4244 bp) of *Cryj2* and in the putative signal sequence region (44–110 bp) of *NCED*, more nonsynonymous polymorphisms were found than in the other regions. The putative promoter region of *Cryj2*

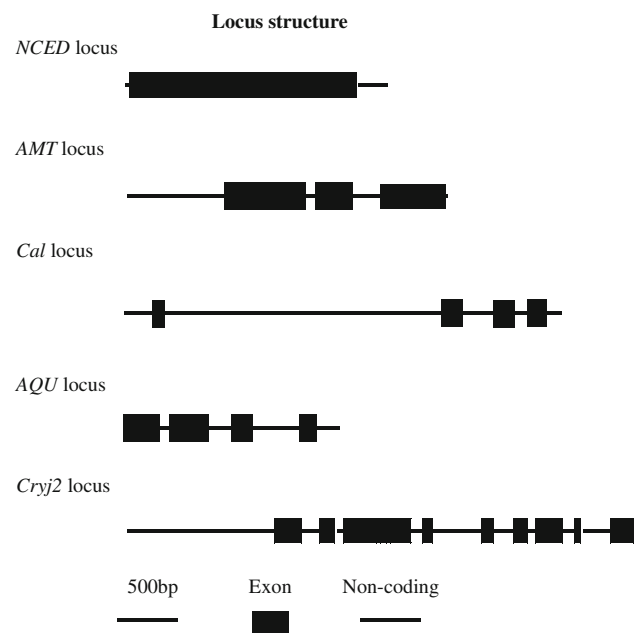


Fig. 1 A schematic presentation of the structure of the five loci analyzed

Table 2 Nucleotide variation at five loci

Locus	Nonsynonymous	Synonymous	Silent	Total
<i>NCED</i>				
Length	1427.87	447.13	690.13	2118
S^a	12	5	8	20
Π^b	0.00259	0.00222	0.00265	0.00261
θ_w^c	0.00189	0.00252	0.00261	0.00213
<i>AMT</i>				
Length	1109.33	348.67	1218	2678
S^a	0	1	11	11
Π^b	0	0.00097	0.00171	0.001
θ_w^c	0	0.00065	0.00158	0.00093
<i>Cal</i>				
Length	356.33	90.67	3092.67	3449
S^a	0	3	71	71
Π^b	0	0.00731	0.0037	0.00332
θ_w^c	0	0.00746	0.00517	0.00464
<i>AQU</i>				
Length	526.28	181.72	1250.72	1777
S^a	0	7	38	38
Π^b	0	0.00289	0.00536	0.0037
θ_w^c	0	0.00868	0.00685	0.00482
<i>Cryj2</i>				
Length	1206.08	335.92	2685.92	3892
S^a	17	12	119	136
Π^b	0.00353	0.00846	0.00948	0.00763
θ_w^c	0.00317	0.00804	0.00998	0.00787

^a S : the number of segregating sites

^b π : nucleotide diversity

^c Watterson's estimator of $\theta = 4Nu$

had an indel polymorphism of approximately 300 bp. Mononucleotide repeats of 10–13T were found in the introns of *Cal* and *Cryj2*.

Divergences at silent sites, K_{sil} , and replacement sites, K_{rep} , between *C. japonica* and *T. distichum*, at the five loci are listed in Table 3.

Population Structure and Recombination

The five loci sequenced here showed generally low F_{ST} values, and permutation tests of S_{nn} (Hudson 2000) did not reveal any significant population differentiation (data not shown). This is consistent with the previous results showing low differentiation among populations in this species (Tomaru et al. 1994; Tsumura and Tomaru 1999; Takahashi et al. 2005; Tsumura et al. 2007). The frequency spectrum of the 12 loci is shown in Fig. 2. The resulting frequency spectrum did not lead to a rejection of the standard neutral model ($p < 0.1941$ by the same test of the frequency spectrum described by Kado et al. [2003]).

Table 3 Divergence between *C. japonica* and *T. distichum*

	<i>NCED</i>	<i>AMT</i>	<i>Cal</i>	<i>AQU</i>	<i>Cryj2</i>
Total					
Length	2088	2610	3189	1763	2590
K^a	0.0612	0.0334	0.1189	0.0493	0.0757
Silent					
Length	660.06	1499.99	2832.66	1149.00	1502.22
K_{sil}^b	0.1281	0.0533	0.1353	0.0708	0.0872
Nonsynonymous					
Length	1427.94	1104.01	356.34	611.00	1087.78
K_{rep}^c	0.0322	0.0073	0	0.0107	0.0599

^a Total divergence per site between *C. japonica* and *T. distichum*

^b Silent divergence per site between *C. japonica* and *T. distichum*

^c Replacement divergence per site between *C. japonica* and *T. distichum*

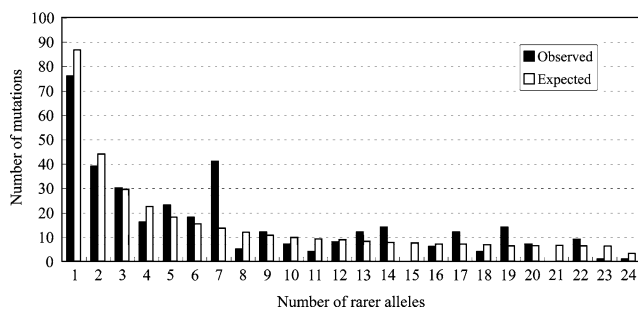


Fig. 2 The frequency spectrum of the nucleotide polymorphisms obtained from all polymorphic sites at the 12 loci. Expected values are based on a coalescent simulation assuming the neutrality, random mating, and constant size

The population growth rate and recombination rate were estimated using the LAMARC package. For this estimation, we excluded three loci, *Acl5*, *HemA*, and *Ferr*, which were suspected to have been affected by recent selective sweeps or balancing selection based on one or more neutrality tests (Kado et al. 2003, 2006), and the polymorphism data of the remaining nine loci were used. The ML estimates of θ (0), g , and r/u were 0.003894, -202.38 , and 0.1182, respectively. The minimum number of recombination events varied from one at *AQU* to six at *Cryj2*. Because the estimate of θ (0) was 0.003894, the present population recombination rate, $4Nr$, is estimated as 0.0004603. We also estimated r/u in *P. taeda* by LAMARC using the data for 19 loci of Brown et al. (2004) for comparison because they used different estimation methods. The ratio was 0.1308, and the estimate of θ in this species was 0.006322. Thus, the population recombination rate was estimated as 0.0008270. This estimate did not differ greatly from those estimated according to two other methods (0.00175 and 0.00115) by Brown et al. (2004). The ML estimates of r/u , θ , and $4Nr$ in *Z. mays* using the

data for 21 loci of Tenaillon et al. (2001) were 0.1992, 0.01491, and 0.002970, respectively. Therefore, the population recombination rate in *C. japonica* is approximately one-sixth that of *Z. mays* and one-half that of *P. taeda*.

Tests of Neutrality

Tajima's D , Fu and Li's D^* , and Fu and Li's F^* were calculated for each locus, but none of these values deviated significantly from the expectations of the standard neutral model at the 5% level (Table 4). In addition, there were no clear tendencies. Fu and Li's D , Fu and Li's F , Fay and Wu's H , and R_2 also showed no significant deviations from the neutral expectations (data not shown).

The results of the pairwise HKA test were not significant at the 5% level for the five loci sequenced here (data not shown). The multilocus HKA test for the 12 loci including those analyzed by Kado et al. (2003) gave a highly significant result ($p < 0.00198$) but this was due to *Acl5*, for which Kado et al. (2003) found indications of recent selective sweep, and the test gave a nonsignificant result ($p < 0.05513$) when we removed this locus.

However, the MK test detected a significant excess of nonsynonymous polymorphisms at *NCED* ($p < 0.0172$) and an almost-significant excess of nonsynonymous divergences at *Cryj2* ($p < 0.0544$) compared to the silent variation, although the result was not significant compared to the synonymous variation (Table 5). The use of silent variation rather than the synonymous variation in the MK test is not appropriate if the times to the most recent common ancestor in the coding and adjacent noncoding regions are different. However, the ratios of the nucleotide diversity to the divergence (π/K) are 0.091 at the synonymous sites and 0.092 at the noncoding sites used for the MK test. Therefore, there seems to be an excess of nonsynonymous divergences at *Cryj2*. In fact, the case of *Cryj2* was rather complex as revealed by the sliding plot analysis of the π_{rep}/π_{sil} and K_{rep}/K_{sil} ratios in the coding region (Fig. 3). There are higher amino acid polymorphisms than amino acid divergence in the 5' portion of the region (mostly the signal peptide region), but the reverse tendency was observed in the remaining portion (Fig. 3). If the signal peptide region was excluded at *Cryj2*, then the excess amount of nonsynonymous divergences became

Table 4 Summary of the results of the neutrality tests at the five loci

	<i>NCED</i>	<i>AMT</i>	<i>Cal</i>	<i>AQU</i>	<i>Cryj2</i>
Tajima's D	0.724	0.233	-1.072	-0.799	-0.082
Fu and Li's D^*	0.220	-0.266	-1.12	-0.922	1.165
Fu and Li's F^*	0.466	-0.120	-1.316	-1.045	0.846

Table 5 Summary of the McDonald-Kreitman test results

Locus	Nonsyn.	Syn.	Silent	p (syn.)	p (silent)	NI ^a (syn.)	NI (silent)
<i>NCED</i>							
Polymorphic	12	5	6	0.0150*	0.0172*	4.27	3.81
Fixed	41	73	78				
<i>AMT</i>							
Polymorphic	0	1	11	1.0000	0.5891	0.00	0.00
Fixed	8	17	73				
<i>Cal</i>							
Polymorphic	0	3	63	—	—	—	—
Fixed	0	16	342				
<i>AQU</i>							
Polymorphic	0	8	38	0.1477	0.1752	0.00	0.00
Fixed	6	15	73				
<i>Cryj2</i>							
Polymorphic ^b	13 (10)	11 (11)	50 (50)	0.336 (0.125)	0.054 (0.013)*	0.59 (0.42)	0.51 (0.38)
Fixed ^b	56 (56)	28 (26)	109 (107)				

Note: Nonsyn., nonsynonymous; syn.; synonymous

^a NI, neutrality index (Rand and Kann 1996)

^b Values in parentheses were calculated in the region, while excluding the signal peptide region

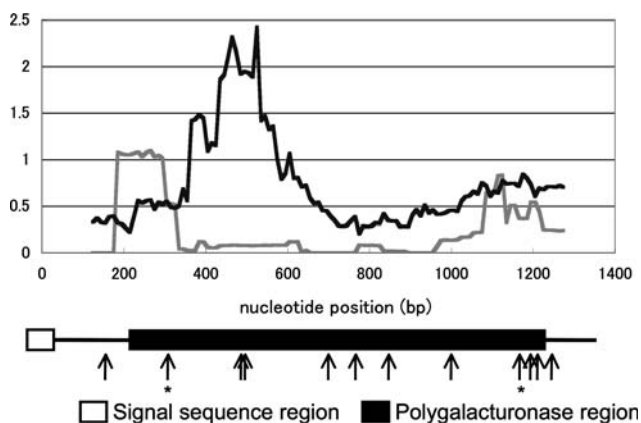


Fig. 3 Sliding window plots of the A/S ratio for the coding region of the *Cryj2* locus. The solid line and the dashed line show the ratios K_{rep}/K_{sil} and π_{rep}/π_{sil} , respectively. The window size is 250 bp, and the step size is 10 bp. The 12 sites belonging to the positively selected class based on the naive empirical Bayes (NEB) analysis are indicated by arrows. Sites with posterior probabilities >0.95 are marked with an asterisk

significant ($p < 0.0126$). When we carried out the MK test using only the signal peptide region, there were excess replacement polymorphisms although the result was not significant because of the low number of sites (Fisher's exact test, $p \leq 0.1$). Thus, at both *NCED* and *Cryj2*, excess amino acid polymorphisms were found in the signal peptide regions, as was reported at *Lcyb* by Kado et al. (2003). Although the number of cases is still small, signal peptide regions seem to have a tendency for excess

amino acid polymorphisms (see Discussion). In nuclear genes of Cupressaceae, codon biases are generally low (Kusumi et al. 2002). Therefore, we may assume selective neutrality of synonymous sites and thus the significant or nearly significant results of MK tests are suggestive of selection.

Molecular Evolutionary Analysis of *Cryj2*

Because the MK test for *Cryj2* indicated the effects of positive selection, we investigated the molecular evolution of *Cryj2* by applying the PAML package to the sequences of *C. japonica*, *T. distichum*, *Chamaecyparis obtusa*, and *M. glyptostroboides*. We were able to use only a portion of the coding region (from codon 31 to codon 449) for this analysis due to the difficulty of amplification and alignment of sequences from distantly related species. The results of the analysis carried out using the PAML package are reported in Table 6. The M2a, M3, and M8 models identified sites with ω values > 1.0 ($\omega_3 = 5.1222$ in M2a, $\omega_3 = 6.1214$ in M3, and $\omega_8 = 4.8212$ in M8), and their proportions were more than negligible. The results of the M1a versus M2a ($2\Delta l = 15.4833$, $p < 0.001$, $df = 2$), M0 versus M3 ($2\Delta l = 122.0963$, $p < 0.0001$, $df = 4$), M1a versus M3 ($2\Delta l = 16.2072$, $p < 0.001$, $df = 3$), and M7 versus M8 ($2\Delta l = 18.1357$, $p < 0.0003$, $df = 2$) model comparisons based on the likelihood-ratio tests were all highly significant. Several candidate target sites of natural selection were identified at *Cryj2* and they are dispersed throughout the gene (Table 7 and Fig. 3).

Table 6 Maximum likelihood estimates of ϖ among amino acid sites at *Cryj2* by PAML

Model	ϖ_1	ϖ_2	ϖ_3	log likelihood
M0 (one-ratio)	0.4852	–	–	–3959.93
M1a (nearly neutral)	0.07716	1		–3910.57
	$p_0 = 0.5753$	$p_1 = 0.4247$		
M2a (positive selection)	0.08395	1	5.1222	–3899.24
	$p_0 = 0.5590$	$p_1 = 0.4072$	$p_2 = 0.0338$	
M3 (discrete)	0.1167	1.2367	6.1214	–3898.88
	$p_0 = 0.6249$	$p_1 = 0.3513$	$p_2 = 0.0238$	
M7 (β)	$p = 0.1110$	$q = 0.1176$		–3909.09
M8 ($\beta + \varpi$)	$p = 0.1708$	$q = 0.2001$	$p_0 = 0.9609$	–3900.02
	$\varpi = 4.8212$	$p_1 = 0.03915$		

Note: For explanation of the models, see the text

Table 7 Posterior probabilities of sites belonging to the positively selected class based on the naive empirical Bayes (NEB) analysis

Codon position	M2a ($\varpi_3 = 5.1222$)	M8 ($\varpi = 4.8212$)
56		0.529
109	0.975*	0.982*
170	0.763	0.814
171	0.685	0.749
245	0.880	0.911
268	0.759	0.806
296	0.738	0.794
349	0.562	0.624
407	0.966*	0.974*
416	0.769	0.815
422		0.518
434		0.537

Note: Only sites having probabilities >50% by either model are shown. Significance level: * $p < 0.05$

Discussion

Past Population Structure of *C. japonica*

Kado et al. (2003) reported more intermediate-frequency variants than expected under the standard neutral model in the frequency spectrum over seven loci, and positive Tajima's D values were observed at five of seven loci. The remaining two loci showed low variation, which may well have been due to recent selective sweeps. Based on these observations, they suggested that there had been a reduction in population size in the recent past, although we should note that the number of loci was too small to draw any conclusive inference. This explanation was consistent with the past population structure that had been suggested based on fossil pollen analyses (Tsukada 1982; Takahara 1998). However, if we consider as well the data regarding the five loci analyzed here together with those for the seven loci described by Kado et al. (2003), the resulting

frequency spectrum (Fig. 2) does not lead to a rejection of the standard neutral model. Seven of 12 loci had a positive Tajima's D , and no clear tendency was observed (Table 4).

In order to better estimate the past population structure, we carried out a ML analysis of the data at nine loci using the LAMARC package. In the analysis, we excluded three loci (*Acl5*, *HemA*, and *Ferr*) that showed signs of selection based on the neutrality tests using polymorphism data (Kado et al. 2003, 2006). The exponential growth rate g was estimated to be -202 , with an approximate 95% confidence interval of $(-310.22, -90.75)$, and the estimate was significantly negative. Recall that g is measured in mutational units. In order to translate g into that per year, we first estimated the synonymous substitution rate per site per year as 1.935×10^{-9} from the estimate (0.387) of the synonymous divergence between Taxodioidae and Sequoioidea, which are considered to have separated 100 MYA (Kusumi et al. 2002). Because we used all sites of the nine loci for the estimation of g , we calibrated the mutation rate by using the K/K_{sil} ratio between *C. japonica* and *T. distichum* for the nine loci and obtained a mutation rate of 1.2831×10^{-9} per year. Therefore, $g = -202$ corresponds to -2.597×10^{-7} per year and thus a population size reduction of approximately 23% has taken place during the last 10^6 years.

Recombination Rate in *C. japonica*

The ML estimate of r/u in *C. japonica* was 0.1182. Because the mutation rate per generation in Cupressaceae is $1.283 \times 10^{-9} \times (50-100) \approx 65-130 \times 10^{-9}$ if we assume the generation time of 50–100 years, the recombination rate between neighboring nucleotides per generation is estimated to be $7.7-15.4 \times 10^{-9}$. If this value is extended to the whole genome, the total map distance would be 8500–17000 cM because the genome size of *C. japonica* is 11 Gbp (Hizume et al. 2001). This value is much larger than the 2000 cM obtained in a pseudo-test-cross experiment (Nikaido et al. 2000). Therefore, the regions we surveyed (mostly coding)

seem to have higher recombination rates than the other parts of the genome.

Nevertheless, the population recombination rate, $4Nr$, in *C. japonica* was low (0.0004603) and was approximately one-sixth that of *Z. mays* and one-half that of *P. taeda*. This is why LD was extended to more than 1 kb in this species (Kado et al. 2003). One possible reason for the differences of $4Nr$ among these three species is the difference among their population sizes. Estimates of r/u were 0.1182, 0.1308, and 0.1992 in *C. japonica*, *P. taeda*, and *Z. mays*, respectively, and they are similar. Therefore, if we assume that the mutation rates, u , per generation are similar, we can explain the differences in the estimates of $4Nr$ solely by differences in N . Under this hypothesis, the species with higher estimates of $4Nr$ would have higher nucleotide diversities. However, the patterns of LD in some tree species are not congruent with this simple pattern. For example, LD decays quickly in Norway spruce, whose nucleotide diversity at silent sites, π_{sil} , is comparable to that of *C. japonica* (Heurtz et al. 2006), while LD decays relatively slowly in Douglas fir, which has a relatively high π_{sil} (Krutovsky and Neale 2005). Therefore, differences in $4Nr$ among species may be caused not only by differences in N but also by differences in r . In any case, the extents of LD affect efficiencies of candidate gene-based association approaches (Neale and Savolainen 2004) and our current estimate makes *C. japonica* a suitable species for finding causative genes by this approach but not for finding causative nucleotides.

Slightly Deleterious Mutations in *C. japonica*

A significant excess of replacement polymorphisms found at *NCED* by the MK test indicated the contribution of slightly deleterious mutations (Ohta 1992) at this locus. However, the replacement variants were not rarer than the synonymous variants (Tajima's D was 1.09 at replacement sites and 0.04 at silent sites), as would typically be expected for slightly deleterious variants (Tachida 2000). Also, a marginally significant excess of replacement polymorphisms was observed at *Lcyb*, and those replacement variants were not rare (Tajima's D at both replacement and silent sites was 1.03 and 0.083, respectively) (Kado et al. 2003). The same pattern was observed in the signal peptide region of *Cryj2*. This contradiction may be resolved if we assume a recent relaxation of negative selection due to the population size reduction as suggested by the likelihood analysis above. If the population size had been reduced in the *C. japonica* lineage, thus rendering slightly deleterious mutations ($|N_s| \approx 1$) effectively neutral ($|N_s| < 1$), then the frequency of such mutations would not necessarily be low. Our estimate of size reduction was 23% per 10^6 years. If this trend has been continuing for the last 5×10^6 years,

the current population size would be one-fourth the size of the population 5 MYA. Such a change will cause significant differences in Tajima's D and the substitution rate if $4Ns$ in the larger population was between 5 and 10 (Tachida 2000). Therefore, nonsynonymous fixations might have been constrained until 5 MYA after *C. japonica* separated from the lineage of *T. distichum* 25 MYA. Of course, because the polymorphism data allow us to infer only recent changes of population size, this is only one possible explanation. Alternatively, negative selection at those loci might have been relaxed recently in *C. japonica* because of fixations of adaptive mutations at linked sites. This explanation may be more plausible for *Cryj2* because the analysis using PAML indicated multiple adaptive substitutions at this locus. However, there is no indication of adaptive substitutions at the other two loci. Currently it is difficult to determine which factor, size reduction or adaptive fixation at a linked site, caused the relaxation of negative selection at those loci. To clarify this, we need to increase the number of loci to infer the past population structure more accurately and to examine whether or not this pattern generally accompanies indication of adaptive substitutions at linked sites.

Adaptive Evolution at *Cryj2*

Although the results of the MK test were only marginally significant ($p < 0.0544$), they became significant ($p < 0.0125$) when data for the signal peptide region (codons 1–54) were excluded from the analysis. As noted above, all signal peptide regions investigated here indicated polymorphisms due to slightly deleterious mutations. Therefore, adaptive fixations and slightly deleterious mutations might have cancelled each other out at *Cryj2*. In order to pursue this possibility, we analyzed the *Cryj2* locus using the molecular evolutionary method of Yang (1997), which is based on the ML estimation. Because this method considers only interspecies differences, slightly deleterious mutations that rarely become fixed in the population will not affect the analysis. The model comparisons M1a versus M2a, M0 versus M3, M1a versus M3, and M7 versus M8 gave highly significant results, thus supporting adaptive evolution. The ML analysis identified that 3.38% of the sites had a ω value of 5.12 according to the positive selection model, M2a; 2.38% of the sites had a ω value of 6.12 according to the discrete model, M3; furthermore, 3.92% of the sites had a ω value of 4.82 according to the β & ω model, M8. Thus, approximately 2–4% of the sites (9–16 amino acid sites) would have a ω value $\gg 1$ and would be considered targets of adaptive evolution. When those candidate sites of positive selection were excluded from the analysis, and we performed the MK test again, the results were not significant. This finding shows that the candidate sites of adaptive evolution detected by the MK

test using closely related species (*C. japonica* and *T. distichum*) and those detected by the ML method using data from more distantly related species (including *C. obtusa* and *M. glyptostroboides*) are shared in common.

The product of the *Cryj2* gene is known as the second major allergen and was isolated from pollen (Namba et al. 1994). In addition, Ohtsuki et al. (1995) showed that the *Cryj2* gene product had enzyme activity toward pectin that is localized in plant cell walls; they suggested that *Cryj2* might play a role in the growth of pollen tubes in the pistil. Many loci involved in disease resistance or sexual reproduction are considered to be affected by adaptive evolution (Stahl et al. 1999; Bergelson et al. 2001; Clark et al. 2006). The detection of positive selection suggests that *Cryj2* has repeatedly undergone positive selection, due to its putative role in sexual reproduction. The physiological function of this locus has not yet been well characterized, but this locus is thought to be involved in species recognition by providing, for example, interspecies incompatibility.

Conclusion

By extending our previous analysis, we could characterize *C. japonica* as having a population slowly diminishing in size and longer extents of linkage disequilibrium than *P. taeda* or *Z. mays*. With this knowledge, we can now look for adaptively evolving genes (Wright et al. 2005). Although Kusumi et al. (2002) suggested that nonsynonymous substitutions in Cupressaceae might be primarily mutation-driven, we found evidence of positive selection for a gene (*Cryj2*) putatively involved in sexual reproduction. So prospects of finding such genes do not seem remote if we choose candidate genes appropriately. Also, it would be interesting to study effects of population size changes on weakly selected variants. In this respect, a comparison of polymorphisms at *NCED* and *Lcyb* in *C. japonica* with those in *T. distichum*, which seems to have a different population structure from that of *C. japonica* (Kado et al. 2006), would be informative.

Acknowledgments We thank Hitoshi Araki and two anonymous referees for helpful comments on the manuscript. This work was partially supported by Grants-in-Aid for Scientific Research from the Japan Society for the Promotion of Science (nos. 16370101, 19370099).

References

- Aguade M (2001) Nucleotide sequence variation at two genes of the phenylpropanoid pathway, the FAH1 and F3H genes, in *Arabidopsis thaliana*. *Mol Biol Evol* 18:1–9
- Baum J, Thomas AW, Conway DJ (2003) Evidence for diversifying selection on erythrocyte-binding antigens of *Plasmodium falciparum* and *P. vivax*. *Genetics* 163:1327–1336
- Bergelson J, Kreitman M, Stahl EA, Tian D (2001) Evolutionary dynamics of plant R-genes. *Science* 292:2281–2285
- Brown GR, Gill GP, Kuntz RJ, Langley CH, Neale DB (2004) Nucleotide diversity and linkage disequilibrium in loblolly pine. *Proc Natl Acad Sci USA* 101:15255–15260
- Bustamante CD, Nielsen R, Sawyer SA, Olsen KM, Purugganan MD, Hartl DL (2002) The cost of inbreeding in *Arabidopsis*. *Nature* 416:531–534
- Clark NL, Aagaard JE, Swanson WJ (2006) Evolution of reproductive protein from animals and plants. *Reproduction* 131:11–22
- Fay JC, Wu CI (2000) Hitchhiking under positive Darwinian selection. *Genetics* 155:1405–1413
- Fay JC, Wyckoff GJ, Wu CI (2002) Testing the neutral theory of molecular evolution with genomic data from *Drosophila*. *Nature* 415:1024–1026
- Felsenstein J (2004) Inferring phylogenies. Sinauer Associates, Sunderland, MA
- Fu YX, Li WH (1993) Statistical tests of neutrality of mutation. *Genetics* 133:693–709
- Gadek PA, Alpers DL, Heslewood MM, Quinn CJ (2000) Relationships within Cupressaceae sensu lato: a combined morphological and molecular approach. *Am J Bot* 87:1044–1057
- Gillespie JH (1989) Lineage effects and the index of dispersion of molecular evolution. *Mol Biol Evol* 6:636–647
- González-Martínez SC, Krutovsky KV, Neale DB (2006) Forest-tree population genomics and adaptive evolution. *New Phytol* 170:227–238
- Hamrick JL, Godt MJW (1990) Allozyme diversity in plant species. In: Brown AHD, Clegg MT, Kahler AL, Weir BS (eds) Plant population genetics, breeding and genetic resources. Sinauer Associates, Sunderland, MA, pp 43–63
- Heuertz M, De Paoli E, Kallman T, Larsson H, Jurman I, Morgante M, Lascoux M, Gyllenstrand N (2006) Multilocus patterns of nucleotide diversity, linkage disequilibrium and demographic history of Norway spruce [*Picea abies* (L.) Karst]. *Genetics* 174:2095–2105
- Hizume M, Kondo T, Shibata F, Ishizuka R (2001) Flow cytometric determination of genome size in the Taxodiaceae, Cupressaceae sensu stricto and Sciadopityaceae. *Cytologia* 66:307–311
- Hudson RR (2000) A new statistic for detecting genetic differentiation. *Genetics* 155:2011–2014
- Hudson RR, Kreitman M, Aguade M (1987) A test of neutral molecular evolution based on nucleotide data. *Genetics* 116:153–159
- Hudson RR, Slatkin M, Maddison WP (1992) Estimation of levels of gene flow from DNA sequence data. *Genetics* 132:583–589
- Ingvarsson PK (2005) Nucleotide polymorphism and linkage disequilibrium within and among natural populations of European aspen (*Populus tremula* L., Salicaceae). *Genetics* 169:945–953
- Kado T, Yoshimaru H, Tsumura Y, Tachida H (2003) DNA Variation in a conifer, *Cryptomeria japonica* (Cupressaceae sensu lato). *Genetics* 164:1547–1559
- Kado T, Ushio Y, Yoshimaru H, Tsumura Y, Tachida H (2006) Contrasting patterns of DNA variation in natural populations of closely related conifers, *Cryptomeria japonica* and *Taxodium distichum* (Cupressaceae sensu lato). *Genes Genet Syst* 81(2): 103–113
- Kuhner MK, Yamato J, Felsenstein J (1998) Maximum likelihood estimation of population growth rates based on the coalescent. *Genetics* 149:429–434
- Kuhner MK, Yamato J, Felsenstein J (2000) Maximum likelihood estimation of recombination rates from population data. *Genetics* 156:1393–1401
- Kuhner MK, Yamato J, Beerli P, Smith LP, Rynes E, Walkup E, Li C, Sloan J, Colacurcio P, Felsenstein J (2004) LAMARC v 1.2.1.

- University of Washington. Available at: <http://evolution.gs.washington.edu/lamarc.html>
- Kusumi J, Tsumura Y, Yoshimaru H, Tachida H (2000) Phylogenetic relationships in Taxodiaceae and Cupressaceae sensu stricto based on *matK* gene, *chlL* gene, *trnL-trnF* IGS region, and *trnL* intron sequences. *Am J Bot* 87:1480–1488
- Kusumi J, Tsumura Y, Yoshimaru H, Tachida H (2002) Molecular evolution of nuclear genes in Cupressaceae, a group of conifer trees. *Mol Biol Evol* 19:736–747
- Krutovsky KV, Neale DB (2005) Nucleotide diversity and linkage disequilibrium in cold-hardiness- and wood quality-related candidate genes in Douglas fir. *Genetics* 171:2029–2041
- Kvist L, Martens J, Nazarenko AA, Orell M (2003) Paternal leakage of mitochondrial DNA in the great tit (*Parus major*). *Mol Biol Evol* 20:243–247
- McDonald JH, Kreitman M (1991) Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* 351:652–654
- Namba M, Kurose M, Torigoe K, Hino K, Taniguchi Y, Fukuda S, Usui M, Kutimoto M (1994) Molecular cloning of the second major allergen, *Cry jII*, from Japanese cedar pollen. *FEBS Lett* 353:124–128
- Neale DB, Savolinen O (2004) Association genetics of complex traits in conifers. *Trends Plant Sci* 9:325–330
- Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonuclease. *Proc Natl Acad Sci USA* 76:5269–5273
- Nikaido AM, Ujino T, Iwata H, Yoshimura K, Yoshimaru H, Suyama Y, Murai M, Nagasaka K, Tsumura Y (2000) AFLP and CAPS linkage maps of *Cryptomeria japonica*. *Theor Appl Genet* 100:825–831
- Nordborg M, Tina TH, Ishino Y et al (24 coauthors) (2005) The pattern of polymorphism in *Arabidopsis thaliana*. *PLoS Biol* 3(7):e196
- Ohta T (1992) The nearly neutral theory of molecular evolution. *Annu Rev Ecol Evol S* 23:263–286
- Ohta T (1995) Synonymous and nonsynonymous substitutions in mammalian genes and the nearly neutral theory. *J Mol Evol* 40:56–63
- Ohtsuki T, Taniguchi Y, Kohno K, Fukuda S, Usui M, Kurimoto M (1995) *Cry j 2*, a major allergen of Japanese cedar pollen, shows polymethylgalacturonase activity. *Allergy* 50:483–488
- Ramos-Onsins SE, Rozas J (2002) Statistical properties of new neutrality tests against population growth. *Mol Biol Evol* 19:2092–2100
- Rozas J, JSánchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19:2496–2497
- Smith NG, Eyre-Walker A (2002) Adaptive protein evolution in *Drosophila*. *Nature* 415:1022–1024
- Stahl EA, Dwyer G, Mauricio R, Kreitman M, Bergelson J (1999) Dynamics of disease resistance polymorphism at the *Rpm1* locus of Arabidopsis. *Nature* 400:667–671
- Tachida H (2000) Molecular evolution in a multisite nearly neutral mutation model. *J Mol Evol* 50:69–81
- Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595
- Takahara H (1998) Sugi-rin no henshen. In: Yasuda Y, Miyoshi N (eds) *Nippon rettou syokusei-si*. Asakura Shoten, Tokyo, pp 207–223 (in Japanese)
- Takahashi T, Tani N, Taira H, Tsumura Y (2005) Microsatellite markers reveal high allelic variation in natural populations of *Cryptomeria japonica* near refugial areas of the last glacial period. *J Plant Res* 118:83–90
- Tenaillon MI, Sawkins MC, Long AD, Gaut RL, Doebley JF, Gaut BS (2001) Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays ssp. mays* L.). *Proc Natl Acad Sci USA* 98:9161–9166
- Tomaru N, Tsumura Y, Ohba K (1994) Genetic variation and population differentiation in natural populations of *Cryptomeria japonica*. *Plant Species Biol* 9:191–199
- Tsukada M (1982) *Cryptomeria japonica*: glacial refugia and late-glacial and postglacial migration. *Ecology* 63:1091–1105
- Tsumura Y, Tomaru N (1999) Genetic diversity of *Cryptomeria japonica* using co-dominant markers based on sequenced-tagged sites. *Theor Appl Genet* 98:396–404
- Tsumura Y, Kado T, Takahashi T, Tani N, Ujino-Ihara T, Iwata H (2007) Genome scan to detect genetic structure and adaptive genes of natural populations of *Cryptomeria japonica*. *Genetics* 176:2393–2403
- Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* 13:555–556. Available at: <http://abacus.gene.ucl.ac.uk/software/paml.html>
- Yu N, Jensen-Seaman MI, Chemnick L, Kidd JR, Deinard AS, Ryder O, Kidd KK, Li WH (2003) Low nucleotide diversity in chimpanzees and bonobos. *Genetics* 164:1511–1518
- Watterson GA (1975) On the number of segregating sites in genetical models without recombination. *Theor Popul Biol* 7:256–276
- Wright SI, Gaut BS (2005) Molecular population genetics and the search for adaptive evolution. *Mol Biol Evol* 22(3):506–519
- Zeng LW, Comeron JM, Chen B, Kreitman M (1998) The molecular clock revisited: the rate of synonymous vs. replacement change in *Drosophila*. *Genetica* 102(103):369–382