Annals of Botany **99**: 461–475, 2007
doi:10.1093/aob/mcl288, available online at www.aob.oxfordjournals.org

OXFORD JOURNALS
OXFORD UNIVERSITY PRESS

Divergent Evolution of the Chloroplast Small Heat Shock Protein Gene in the Genera *Rhododendron* (Ericaceae) and *Machilus* (Lauraceae)

MIAO-LUN WU[1], TSAN-PIAO LIN[2], MIN-YI LIN[3], YU-PIN CHENG[4] and SHIH-YING HWANG[1,*]

[1]*Graduate Institute of Biotechnology, and* [3]*Department of Forestry and Natural Resources Conservation, Chinese Culture University, 55 Hwagang Road, Yangmingshan, Taipei 11114, Taiwan,* [2]*Institute of Plant Biology, National Taiwan University, 1 Roosevelt Road, Section 4, Taipei 10617, Taiwan and* [4]*Division of Forest Biology, Taiwan Forestry Research Institute, 53 Nanhai Road, Taipei 10066, Taiwan*

Received: 3 September 2006 Returned for revision: 8 November 2006 Accepted: 27 November 2006 Published electronically: 9 February 2007

• *Background and Aims* Evolutionary and ecological roles of the chloroplast small heat shock protein (CPsHSP) have been emphasized based on variations in protein contents; however, DNA sequence variations related to the evolutionary and ecological roles of this gene have not been investigated. In the present study, a basal angiosperm, *Machilus*, together with the eudicot *Rhododendron* were used to illustrate the evolutionary dynamics of gene divergence in CPsHSPs.
• *Methods* Degenerate primers were used to amplify CPsHSP-related sequences from 16 *Rhododendron* and eight *Machilus* species that occur in Taiwan. Manual DNA sequence alignment was carried out according to the deduced amino acid sequence alignment performed by CLUSTAL X. A neighbour-joining tree was generated in MEGA using conceptual translated amino acid sequences from consensus sequences of cloned CPsHSP genes from eight *Machilus* and 16 *Rhododendron* species as well as amino acid sequences of CPsHSPs from five monocots and seven other eudicots acquired from GenBank. CPsHSP amino acid sequences of *Funaria hygrometrica* were used as the outgroups. The aligned DNA and amino acid sequences were used to estimate several parameters of sequence divergence using the MEGA program. Separate Bayesian inference of DNA sequences of *Rhododendron* and *Machilus* species was analysed and the resulting gene trees were used for detection of putative positively selected amino acid sites by the Codeml program implemented in the PAML package. Mean hydrophobicity profile analysis was performed with representative amino acid sequences for both *Rhododendron* and *Machilus* species by the Bioedit program. The computer program SplitTester was used to examine whether CPsHSPs of *Rhododendron* lineages and duplicate copies of the *Machilus* CPsHSPs have evolved functional divergence based on the hydrophobicity distance matrix.
• *Key Results* Only one copy of the CPsHSP was found in *Rhododendron*. However, a higher evolutionary rate of amino acid substitutions in the *Hymenanthes* lineage of *Rhododendron* was inferred. Two positively selected amino acid sites may have resulted in higher hydrophobicity in the region of the α-crystallin domain (ACD) of the CPsHSP. By contrast, the basal angiosperm, *Machilus*, possessed duplicate copies of the CPsHSP, which also differed in their evolutionary rates of amino acid substitutions. However, no apparent relationship of ecological relevance toward the positively selected amino acid sites was found in *Machilus*.
• *Conclusions* Divergent evolution was found for both *Rhododendron* lineages and the paralogues of CPsHSP in *Machilus* that were directed to the shift in hydrophobicity in the ACD and/or methionine-rich region, which might have played important roles in molecular chaperone activity.

Key words: Adaptive evolution, chloroplast small heat shock protein, *Machilus*, *Rhododendron*.

INTRODUCTION

All small heat shock proteins (sHSPs) in plants are encoded by nuclear genes and are divided into six classes. Among the six classes of plant sHSPs, three classes were found to be localized in plastids, the endoplasmic reticulum and the mitochondria along with three which are localized in the cytosol (Sun *et al.*, 2002). It is thought that sHSPs function as molecular chaperones, preventing intracellular proteins from degradation and inappropriate aggregations (Sun *et al.*, 2002). Plant sHSPs show unusual abundance and diversity (Sun *et al.*, 2002). These diversified and abundant classes of sHSPs reflect the need by plants to adapt rapidly to continuously changing physical conditions such as temperature, light intensity and oxidative stress (Sun *et al.*, 2002; Sundby *et al.*, 2005). Organellar forms of sHSPs in plants were all derived from one nuclear-encoded cytosolic sHSP during the evolution of land plants (Waters and Vierling, 1999).

The nuclear-encoded chloroplast small heat shock protein (CPsHSP) exhibits a high degree of conservation in three consensus regions (Vierling, 1991; Waters, 1995). Near the carboxy-terminal end of the CPsHSP are consensus regions I and II that exhibit homology to consensus regions found in other families of low-molecular-weight HSPs and the α-crystallin domain (ACD) (Caspers *et al.*, 1995; Waters and Vierling, 1999; Sun *et al.*, 2002). In addition to the carboxy-terminal heat-shock domain, the CPsHSP has another highly conserved domain, consensus III, which is predicted to form a methionine-rich amphipathic α-helix (Chen and Vierling, 1991).

* For correspondence. E-mail hsy9347@ms34.hinet.net

The methionine-rich domain was the most conserved region within the predicted amino acid sequences of the CPsHSP in ten species of angiosperms analysed by Waters (1995). The methionine-rich amphipathic α-helix may be involved in substrate binding (Gustavsson *et al.*, 1999). The ACD is the hallmark of sHSPs. Both plant and mammalian sHSPs as well as α-crystallins in mammalian lens exhibit molecular chaperone activity to prevent thermal aggregation of proteins and to facilitate refolding of denatured proteins (Vierling, 1991). Various observations have suggested that the ACD is involved in subunit interactions of sHSPs and in binding partially unfolded proteins (Ganea, 2001; Sun *et al.*, 2002).

The hydrophobicity of amino acid residues in the ACD is thought to be important for polypeptide binding (Sharma *et al.*, 1998). Evidence has been reported that the substitution of the less hydrophobic residues of alanine or valine for the highly hydrophobic residue, leucine, in the conserved ACD of sHSP16·3 resulted in reduced chaperone-like activity in *Mycobacterium tuberculosis* (Mao and Chang, 2001). The hydrophobicity of leucine in the ACD is thought to play an important role in maintaining both the structural stability and chaperone-like activity. sHSPs usually form large oligomeric complexes and provide a means for rapidly exposing subunits, a process which offers hydrophobic surfaces onto which hydrophobic regions of partially denatured substrate proteins can bind to, thereby protecting them from aggregation (Ganea, 2001; Sun *et al.*, 2002).

Diversification following gene duplication plays an important role in the evolutionary processes of species (Ohta, 1991). If the selective constraints are reduced, non-synonymous mutations might have accumulated and consequently resulted in pseudogenization or neofunctionalization (Lynch and Conery, 2000; Zhang, 2003; Lynch and Katju, 2004). Therefore, it is possible that new functions of duplicated genes were derived as a result of advantageous mutations, while the original function was preserved in another copy. On the other hand, subfunctionalization also may occur after duplication. It is common to estimate the ratio of replacement substitutions (non-synonymous substitution, d_N) against synonymous substitutions (d_S) with the pairwise alignment of coding sequences (Li, 1997). For a d_N/d_S (=ω) ratio of <1, purifying selection which eliminates deleterious mutations is always implicated. Very low ω is often found in orthologous gene pair comparisons between species, indicating pervasive purifying selection (Seoighe *et al.*, 2003; Jordan *et al.*, 2004). When $d_N/d_S = 1$, non-synonymous mutations are neutral and are fixed at the same rate as synonymous mutations. Positive Darwinian selection occurs only when a higher fixation rate for non-synonymous mutations than synonymous mutations occurs. The higher rate of fixation in replacement substitutions is thought to be related to the fitness of individuals or species fitness to new habitats (Wagner, 2002). Although the d_N/d_S ratio is a commonly used estimate for proposing the actions of positive selection, it is very stringent in detecting adaptive evolution of protein sequences when the entire aligned sequences are taken into account. Potential adaptive evolution can statistically be identified by comparing substitution patterns in sequences and identifying individual sites with elevated ω ratios along specific lineages of a phylogenetic tree (Yang and Nielsen, 2002).

Plants are often exposed to drastically changing temperatures, daily fluctuations of low and high light intensities, and many other biotic and abiotic stresses. Maintenance of a robust photosynthetic system under heavy metal contamination is reported to be relevant to CPsHSPs (Heckathorn *et al.*, 2004). Recently, evolutionary and ecological roles of HSPs have been discussed (Sørensen *et al.*, 2003). Among eight closely related species from the genus *Ceanothus*, no correlation was found between levels of CPsHSP expression and the maximum temperature during the time of investigation within the species ranges; however, the expression of CPsHSP was found to be associated with photosynthetic thermal tolerance (Knight and Ackerly, 2001). Interestingly, Barua *et al.* (2003) reported that polymorphism in the expression levels of CPsHSPs have played a key role in the population fitness of *Chenopodium album*.

Taiwan hosts more than 4000 species of vascular plants, distributed from sea level to over 3900 m in elevation, and is rich in endemic species (Hsieh, 2002). High levels of endemism provide invaluable plant materials for evolutionary and ecological studies. *Machilus* is a genus classified as a basal angiosperm, while the genus *Rhododendron* (a eudicot) is taxonomically more advanced. Species in these two genera display both wide and limited distributions in a variety of diverse habitats from tropical, subtropical to temperate zones at different elevations. Habitat diversity may have acted as a selective factor in the evolution of genetic changes that resulted in new functions of genes (Wright and Gaut, 2005; Mitchell-Olds and Schmitt, 2006). However, functional divergence including neofunctionalization and subfunctionalization of retained duplicated copy of genes after genome doubling (polyploidy) is also a prominent and significant force in plant evolution (Adams and Wendel, 2005).

There are eight *Machilus* species found in Taiwan, of which two varieties each were classified for *M. japonica* Sieb & Zucc. and *M. zuihoensis* Hay. The two varieties of *M. japonica* Sieb & Zucc. are *M. japonica* var. *japonica* (*M. japonica*) and *M. japonica* var. *kusanoi* Hay. (*M. kusanoi*). The two varieties of *M. zuihoensis* Hay. are *M. zuihoensis* var. *zuihoensis* (*M. zuihoensis*) and *M. zuihoensis* var. *mushaensis* (*M. mushaensis*). Other species are *M. konishii* Hay., *M. obovatifolia* Hay., *M. philippinensis* Merr. and *M. thunbergii* Sieb. & Zucc. Distributions of some *Machilus* species within Taiwan are restricted, whereas others are found throughout large parts of the island. *Machilus thunbergii*, *M. japonica* and *M. zuihoensis* are widespread from subtropical to temperate zones; *Machilus kusanoi* is also widely distributed but mainly in the low lands near rivers. *Machilus konishii* is restricted to the subtropical zone in central and southern parts of Taiwan west of the Central Mountain Range (CMR). *Machilus obovatifolia* is only found on the tropical Hengchun Peninsula at the southern tip of the island. *Machilus philippinensis* is only found in the subtropical

southern part west of the CMR. These species have adapted to different edaphic and environmental conditions. *Rhododendron* in Taiwan comprises *R. rubropunctatum* Hay., *R. morii* Hay., *R. pseudochrysanthum* Hay., *R. hyperythrum* Hay., *R. formosanum* Hemsl., *R. ellipticum* Maxim., *R. kawakamii* Hay., *R. ovatum* Planch., *R. mariesi* Hemsl. and Wilson, *R. noriakianum* Suzuki, *R*, *nakaharae* Hay., *R. simsii* Planch., *R. kanehirai* Wilson, *R. breviperulatum* Hay., *R. rubropilosum* Hay. and *R. oldhamii* Maxim. These *Rhododendron* species were classified into the subgenera *Azaleasrum*, *Hymenanthes*, *Rhododendron* and *Tsutsusi*. Among these *Rhododendron* species, two monophyletic clades, i.e. the *Hymenanthes* clade (*R. rubropunctatum*, *R. morii*, *R. pseudochrysanthum*, *R. hyperythrum* and *R. formosanum*) and the *Tsutsusi* clade (*R. mariesii*, *R. noriakianum*, *R. nakaharae*, *R. simsii*, *R. kanehirai*, *R. breviperulatum*, *R. rubropilosum* and *R. oldhamii*) were resolved based on chloroplast DNA and nuclear intron data (Hwang *et al.*, 2006). In the *Hymenanthes* clade, *R. rubropunctatum*, *R. morii*, *R. pseudochrysanthum* and *R. hyperythrum* are grouped into a species complex termed the *R. pseudochrysanthum* complex (Chung *et al.*, 2007). Results from studies of Hwang and Hsu (2001), Tsai *et al.* (2003) and Chung *et al.* (2007) suggested the derivation of *R. pseudochrysanthum*, *R. morii*, *R. rubropunctatum* and *R. hyperythrum* from *R. formosanum*. Endemic *Rhododendron* species in the subgenus *Hymenanthes* are mainly distributed on high peaks in northern, central and southern Taiwan with similar habitat in the temperate zone. Morphological differences are small for the species in the subgenus *Hymenanthes*. Endemic *Rhododendron* species in the subgenus *Tsutsusi* are found in specific habitats: from *R. kanehirai* limited to river banks in northern Taiwan to *R. rubropilosum* distributed on sunny mountain slopes of the CMR. These vary morphologically from small shrubs (*R. breviperulatum*) to 3- to approx. 4-m-tall shrubs (*R. oldhamii*). Among these Taiwanese *Rhododendron* species, most have a limited areal distribution but others are widespread. The elevational distribution of Taiwanese *Rhododendron* from lowlands to the peak of the Yushan massif at 3950 m represents high adaptability of this genus. Morphologically, species in the subgenus *Hymenanthes* have large and coriaceous leaves in contrast to the small, chartaceous leaves of species in the subgenus *Tsutsusi*.

Genetic changes might have occurred in CPsHSPs of both *Machilus* and *Rhododendron* as a result of these diversified distributions from subtropical, tropical to temperate zones given that CPsHSP content has been found to be correlated with thermal tolerance of plants in different ecological niches (Barua *et al.*, 2003). Understanding whether diversified gene duplicates or genetic changes specific to the *Rhododendron* species in the subgenus *Hymenanthes* is of particular interest because dramatic daily fluctuations in temperature and exposure to high light intensity of these species may prompt more efficient molecular chaperone activities in the protective photosynthetic system. The study of CPsHSP evolution in the basal angiosperm *Machilus* in comparison with eudicots may aid in understanding the evolutionary dynamics of CPsHSP in plants. In the present study, a set of PCR primers were designed according to the methionine-rich domain as well as the sequences from the consensus carboxy-terminal end of the ACD using the concept of CODEHOPE (Rose *et al.*, 1998) based on the amino acid alignment of CPsHSPs (Waters and Vierling, 1999) to amplify related sequences from *Rhododendron* and *Machilus* species that occur in Taiwan. The primary objective of this study was to determine whether one or more copy of CPsHSP is harboured in Taiwanese *Rhododendron* and *Machilus* as determined using the designed primer set, as variations in species distributions suggest a possible variation in CPsHSP evolution that may be associated with different environments in which species have evolved. The second objective was to examine whether selective forces dictated evolution and are related to the functional divergence of CPsHSPs in *Rhododendron* and *Machilus*.

METHODS AND MATERIALS

Taxon sampling and DNA purification

The plant species included in this investigation for CPsHSP gene cloning were eight *Machilus* and 16 *Rhododendron* species. CPsHSP amino acid sequences from five monocots and seven other eudicots acquired from GenBank were used for comparisons. CPsHSP amino acid sequences of *Funaria hygrometrica* were used as the outgroups. Total DNA was extracted from ground leaf tissue according to a modified hexadecetyltrimethyl ammonium bromide (CTAB) procedure (Doyle and Doyle, 1987). DNA was precipitated with ethanol, and after washing with 70% ethanol, was dissolved in 200 μL TE buffer (pH 8·0) and stored at −20 °C. The DNA concentration was determined for each sample using the GeneQuant II RNA–DNA Calculator (Amersham Biosciences, Little Chalfont, UK).

Primers, PCR amplification and DNA sequencing

PCR amplification of partial sequences of the CPsHSP was performed with degenerate primers derived from the amino acid alignment (5′-MGNCARATGYTNGAYAC NATGGAY-3′ and 5′-NARNACNCCRTTYTTNARYTC NGC-3′) (Waters and Viering, 1999). The concept developed in CODEHOPE was used to design the degenerate primers (Rose *et al.*, 1998) by taking advantage of the conserved methionine-rich domain and a 3′ conserved region in the ACD. PCR products were cloned with a yT&A cloning kit (Yeastern Biotech) following the manufacturer's protocol. Plasmids containing PCR product were further screened using a colony PCR and purified with a QiaGen kit (QIAGEN). Subsequently, plasmid clones were sequenced in both directions using a *Taq* Dye Dideoxy Terminator Cycle Sequencing Kit (Applied Biosystems) and a model ABI373A automated sequencer (Applied Biosystems). For sequencing, the M13 forward and reverse primers were used for amplification.

All sequence polymorphisms were visually rechecked from the chromatograms.

Sequence alignment and variation

Consensus sequences were generated through the alignment of all cloned CPsHSP partial sequences for each species of *Rhododendron* and for each species and different types of *Machilus* using the program CLUSTAL X (Thompson *et al.*, 1997). The consensus sequences were subsequently translated into amino acid sequences with the aid of the BCM search launcher in six open reading frames, with the +1 frame resulting in amino acid sequences with no intron or stop codon. The deduced amino acid sequences were used in a BLAST search, which resulted in high similarities to CPsHSP sequences with very low *E* values. The deduced amino acid sequences of *Rhododendron* and *Machilus* as well as those of CPsHSP sequences (only those sequences containing the specific methionine-rich domain were recognized as CPsHSPs) acquired from GenBank were aligned for the phylogenetic analysis. The aligned amino acids in *Rhododendron* and *Machilus* were then separately used for the manual alignment of DNA sequences for the codeml analysis using the PAML package (Yang, 1997). DNA sequence divergence and amino acid sequence divergence were estimated based on the K2P (Kimura's two-parameter) model for DNA (Kimura, 1980) and the JTT (Jones, Taylor and Thornton) distance matrix for amino acids (Jones *et al.*, 1992) using MEGA (Kumar *et al.*, 2004). The numbers of synonymous and non-synonymous substitutions were also estimated using the MEGA program.

Gene tree analysis

A neighbour-joining (NJ) tree was generated by the MEGA 3·0 software (Kumar *et al.*, 2004) using the aligned amino acids of CPsHSP sequences from *Rhododendron* and *Machilus* as well as CPsHSP amino acid sequences of other species acquired from GenBank. For individual trees of *Rhododendron* and *Machilus*, aligned DNA sequences based on the aligned amino acids were first generated. Subsequently, the best-fitting substitution model was estimated by ModelTest version 3·7 (Posada and Crandall, 1998, 2001). The estimated parameters were then incorporated into MrBayes 3·1·2 (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003) to generate a Bayesian gene tree. For the Bayesian analysis, the Metropolis coupled Markov chain Monte Carlo (MCMC) process was run with four simultaneous chains (three 'heated') for $1{\cdot}5 \times 10^6$ generations with the first 3×10^5 generations set as the burn-in after chain stationarity had been identified from plots of the likelihood against generation.

PAML analysis

The Bayesian tree of CPsHSP nucleotide sequences in *Rhododendron* and *Machilus* was separately subjected to analysis by the codeml program of PAML (Yang, 1997). The aligned nucleotide sequences for *Machilus* and *Rhododendron* are available online as Supplementary Material. To assess if the evolutionary rates of the CPsHSP were the same between lineages of *Rhododendron* (the *Hymenanthes* lineage vs. the other *Rhododendron* species) and between duplicate copies of the CPsHSP in *Machilus*, a likelihood ratio test (LRT) was applied (Yang and Nielsen, 1998). To test the hypothesis of equal evolutionary rates between the *Rhododendron* lineages at the amino-acid level, a one-ratio and a two-ratio model were compared using the *Hymenanthes* lineage as foreground branch against all other *Rhododendron* species under investigation. The two models differed in that the former assumed the same evolutionary rate on the branches leading to all *Rhododendron* taxa, whereas the latter did not impose equal rates between *Hymenanthes* and other *Rhododendron* species. A similar codeml analysis for the evolutionary rates between the two types of CPsHSPs of *Machilus* was also performed. The lineage of type II CPsHSP was set as the foreground branch against the type I CPsHSP lineage. The codeml program using sequence type = 2 for amino-acid sequences in the PAML package was run for each of the two models.

Molecular adaptive evolution at individual sites along the CPsHSP sequences was investigated using the codeml program implemented in PAML version 3·15 (Yang, 1997). A range of codon models for measuring selection were evaluated and nested models were compared by the LRT statistic, $2\delta L$, against the χ^2 distribution with degrees of freedom equal to the difference in the number of parameters. To perform these analyses, different models implemented in the codeml program of the PAML package were applied. To investigate different selective pressures along specific lineages, a model was tested that allows only a single ω for all branches in the tree (M0) against a two-ratio model that allows an additional ω for specific branches in the tree. The settings of foreground and background branches were the same as for the analysis at the amino acid level. Site models that allow for heterogeneous non-synonymous/synonymous substitution rate ratios ($\omega = d_N/d_S$) among sites (models M0, M1, M2, M3, M7 and M8; Yang *et al.*, 2000) were used to test for diversifying selection at individual sites. The LRT was used to evaluate whether allowing sites to have $\omega > 1$ significantly improved the fit of the model to the data (M1–M2, M0–M3, and M7–M8, where M2, M3 and M7 can accommodate positively selected sites).

Hydrophobicity profile and functional divergence analyses

Mean hydrophobicity profiles of representative sequences of CPsHSPs in *Rhododendron* and *Machilus* were generated using the software Bioedit (Hall, 1999). Functional divergence of protein domains through sliding-window comparisons of the CPsHSP sequences of the different *Rhododendron* lineages and the different CPsHSP types in *Machilus* were generated based on a hydrophobicity distance matrix using the computer software SplitTester (Gao *et al.*, 2005).

RESULTS

Sequence variation of CPsHSPs from Rhododendron *and* Machilus

Forward and reverse degenerate primer sequences were designed according to an alignment of the amino acid sequences of CPsHSPs (Waters and Vierling, 1999). This set of degenerate primers worked for both *Rhododendron* and *Machilus* species. Among the 12 cloned sequences obtained from each of the 16 *Rhododendron* species, only one type of CPsHSP gene was found. In contrast, two types of CPsHSP genes were found among the 12 clones each from the eight individual *Machilus* species. All cloned sequences from both *Rhododendron* and *Machilus* were found to be the CPsHSP after an open reading frame search (ORF Finder) in the NCBI with no intron found with the genomic amplification using the designed primers. Lack of an intron was also confirmed with the GENSCAN prediction (http://genes.mit.edu/GENSCAN.html) of coding sequences. The 16 *Rhododendron* and 16 *Machilus* amino acid sequences were aligned and are shown in Fig. 1.

The type I CPsHSP in *Machilus* had the largest K2P divergence for DNA (4·5 %) and JTT divergence for amino acid (5·8 %), while *Rhododendron* CPsHSP had the smallest K2P divergence for DNA (1·7 %) and JTT divergence for amino acid (1·8 %) (Table 1). The DNA sequence of type II CPsHSP in *Machilus* was more similar to the CPsHSP in *Rhododendron* (39·3 % K2P divergence) than that between the two types of CPsHSP in *Machilus* (50·9 % K2P divergence). The amino acid divergence based on the JTT matrix was 29·5 % between type II CPsHSP and *Rhododendron* CPsHSP, which is smaller than that of the two types of CPsHSP in *Machilus* (71·5 %). *Rhododendron* had a smaller within-genus amino acid divergence of 1·7 %, while the type II CPsHSP in *Machilus* had the highest within-genus divergence at 5·8 % and the 2·3 % divergence in the middle for the type II CPsHSP in *Machilus*. The number of variable sites for each comparison is also shown in Table 1. Lower numbers of synonymous and non-synonymous substitutions in CPsHSP were found for pairwise comparisons of *Rhododendron* species in contrast to higher numbers of synonymous and non-synonymous substitutions found in pairwise comparisons in type I and type II CPsHSPs of *Machilus* (Table 2). Type I CPsHSP had higher numbers of pairwise non-synonymous substitutions in comparison with that of the type II CpsHSP in *Machilus*.

Gene tree

Conceptual translated amino acid sequences from consensus sequences of cloned CPsHSP genes (eight *Machilus* and 16 *Rhododendron* species) as well as amino acid sequences acquired from GenBank were subjected to a phylogenetic analysis, and an NJ tree generated by MEGA is shown in Fig. 2. In this phylogenetic analysis, two amino acid sequences of the CPsHSP gene of moss (*Funaria hygrometrica*) (Waters and Vierling, 1999) were used as the outgroups. An interesting result was revealed by the NJ topology. The NJ tree was divided into three clades. Two types of CPsHSP that differed mainly in the ACD were found in *Machilus*. By contrast, only one type of CPsHSP gene sequence was found in *Rhododendron* species. In clade A, *Rhododendron* CPsHSP was sister to CPsHSPs of other eudicots including *Arabidopsis thaliana*, *Pisum sativum*, *Glycine max*, *Solanum esculentum*, *Capsicum annuum*, *Petunia hybrida*, *Nicotiana tabacum* and *N. sylvestris*. In clade B, the duplicated copy of CPsHSP (type II) found in *Machilus* was sister to the CPsHSPs of monocot species including *Hordeum vulgare*, *Agrostis stolonifera*, *Triticum aestivum*, *Oryza sativa* and *Zea mays*. The type I CPsHSP of *Machilus* was sister to another type of CPsHSP in *S. esculentum* and formed clade C. The CPsHSP of clade C was the common ancestor of the type II CPsHSP of *Machilus*, the CPsHSp of monocots and the CPsHSP of eudicots (including *S. esculentum* and *Rhododendron*). In clade A, the CPsHSP of the other eudicots was sister to the CPsHSP of *Rhododendron*. The type II CPsHSP of *Machilus* was sister to the CPsHSPs of monocots in clade B. It is apparent that paralogues of *Machilus* CPsHSP were displayed (Fig. 2), which implies a gene duplication event. Gene divergence was also found in the CPsHSPs of *Rhododendron* since the splitting of the subgenera *Hymenanthes* and *Tsutsusi*, although the level of variation was small. Nevertheless, it is probable that only one copy of CPsHSP exists in *Rhododendron*, but two copies of CPsHSP exist in the basal angiosperm *Machilus*.

Separate phylogenetic analyses were also performed for *Rhododendron* and *Machilus* using aligned DNA sequences according to the alignment of deduced amino acids. The aligned DNA sequences were first estimated for the best-fitting substitution model by Modeltest. Akaike's information criterion (AIC) estimations indicated that better substitution models were HKY+I (Hasegawa *et al.*, 1985; and invariant sites, $-\ln L = 639{\cdot}1492$) for *Rhododendron* and GTR + I+G (general time-reversible, invariant sites, and gamma distribution model, $-\ln L = 1251{\cdot}9891$) for *Machilus*. The parameters estimated for *Rhododendron* were invariant sites (I = 0·8430) with base frequencies for A of 0·3205, C 0·1936, G 0·2647 and T 0·2212. The parameters estimated for *Machilus* were invariant sites (I = 0·4256) plus gamma shape (G, a gamma distribution shape parameter α of 0·2108) model with a base frequency for A of 0·3057, C 0·1592, G 0·3586 and T 0·1764. The tree topologies of the Bayesian inferences for *Rhododendron* and *Machilus* are shown in Fig. 3.

LRT of evolutionary rate variation and positive selection in Rhododendron *and* Machilus *CPsHSPs*

A null model without adaptive evolution can be compared with an alternative model with adaptive evolution using the LRT that distinguishes a significant fit of the sequence data to the alternative model compared with the nested null model using the codeml program implemented in the computer package PAML (Yang, 1997). First, it was determined whether the amino acid substitution rate of CPsHSPs in the *Hymenanthes* lineage differed from those of other *Rhododendron* species. This comparison

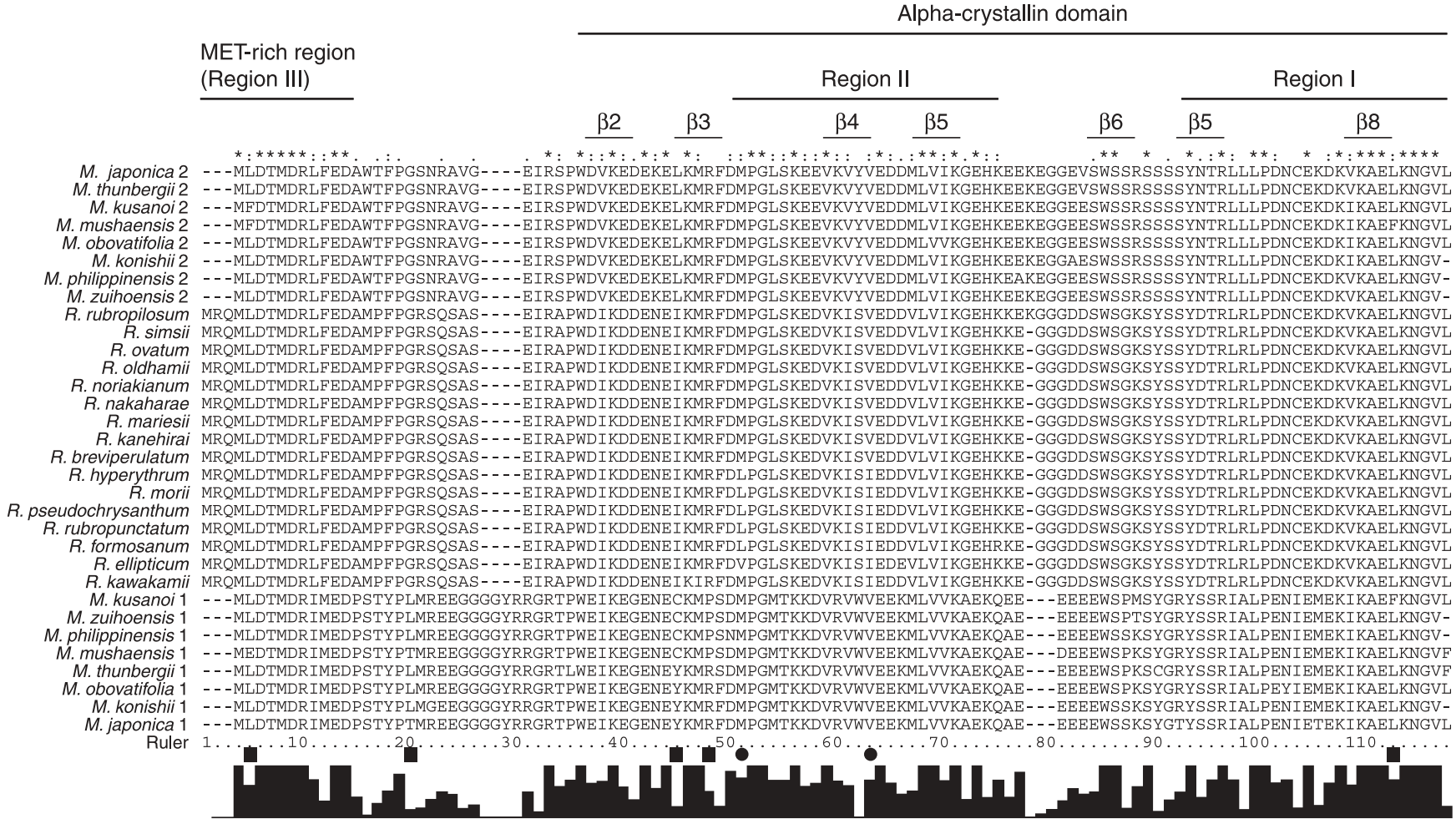

FIG. 1. Multiple sequence alignment of chloroplast small heat shock proteins (CPsHSPs). The amino acid sequences were conceptually translated from DNA sequences and aligned using CLUSTAL X. Locations of probable β-strands were predicted using the PredictProtein system (http://www.predictprotein.org/). The putative positively selected sites detected by the codeml program of the PAML package are indicated by closed circles (*Rhododendron*) and closed squares (*Machilus*) at the bottom of the aligned residue number. The shaded bars at the bottom indicate the level of similarity for each aligned amino acid. The symbols at the top of the aligned amino acids are as follows: –, no residue; *, identical residue;:, conserved substitution; ., semi-conserved substitution.

TABLE 1. *DNA and amino acid variation between chloroplast small heat shock proteins from* Rhododendron *and* Machilus

	DNA			Amino acid		
	Rhododendron	*Machilus* type I	*Machilus type* II	*Rhododendron*	*Machilus* type I	*Machilus* type II
Within-genus/type divergence	1·7	4·5	2·5	1·8	5·8	2·1
Rhododendron	–			–		
Machilus type I	37·1 (124/352)	–		78·6 (69/117)	–	
Machilus type II	56·6 (151/352)	50·8 (135/352)	–	29·5 (38/117)	71·5 (59/117)	–

Numbers in parentheses are (variable sites/aligned length). The alignment included CPsHSP sequences (DNA or amino acids) of 16 *Rhododendron* and 16 *Machilus* CPsHSPs.

TABLE 2. *Number of synonymous (below diagonal) and non-synonymous (above diagonal) differences of the chloroplast small heat shock protein gene (a) within* Rhododendron *and (b) within each type of* Machilus

(a) *Rhododendron**

Species	RM	RO	RB	RN	RNA	ROL	RR	RK	RS	RMO	RP	RRU	RH	RF	RE	RKA
RM		0	0	1	0	0	0	0	0	3	3	3	3	4	4	5
RO	0		0	1	0	0	0	0	0	3	3	3	3	4	4	5
RB	0	0		1	0	0	0	0	0	3	3	3	3	4	4	5
RN	1	1	1		1	1	1	1	1	4	4	4	4	5	4	6
RNA	2	2	2	1		0	0	0	0	3	3	3	3	4	4	5
ROL	2	2	2	1	0		0	0	0	3	3	3	3	4	4	5
RR	2	2	2	1	0	0		0	0	3	3	3	3	4	4	5
RK	2	2	2	1	2	2	2		0	3	3	3	3	4	4	5
RS	2	2	2	1	2	2	2	0		3	3	3	3	4	4	5
RMO	3	3	3	2	3	3	3	3	3		0	0	0	3	4	6
RP	3	3	3	2	3	3	3	3	3	0		0	0	3	4	6
RRU	3	3	3	2	3	3	3	3	3	0	0		0	3	4	6
RH	4	4	4	3	4	4	4	4	4	1	1	1		3	4	6
RF	4	4	4	3	4	4	4	4	4	3	3	3	2		5	7
RE	4	4	4	5	6	6	6	4	4	7	7	7	8	8		7
RKA	2	2	2	3	4	4	4	2	2	5	5	5	6	6	4	

(b) *Machilus*†

	Type I								Type II							
Species	MJ	MP	MK	MM	MZ	MT	MKU	MO	MJ	MP	MK	MM	MZ	MT	MKU	MO
MJ		9	6	9	10	9	11	8								
MP	6		5	6	3	6	5	7								
MK	7	1		9	6	5	8	4								
MM	10	8	9		5	8	6	8								
MZ	10	4	5	6		5	3	6								
MT	10	9	10	9	5		7	5								
MKU	11	6	7	6	2	5		6								
MO	12	9	8	6	7	11	8									
MJ										2	3	4	1	0	3	3
MP									6		3	4	1	2	3	3
MK									5	1		3	2	3	2	2
MM									6	6	5		3	4	1	3
MZ									5	1	0	5		1	2	2
MT									6	6	5	8	5		3	3
MKU									8	8	7	8	7	8		2
MO									9	8	7	9	7	7	7	

*Species codes: RM (*R. mariesii*); RO (*R. ovatum*); RB (*R. breviperulatum*); RN (*R. noriakianum*); RNA (*R. nakaharae*); ROL (*R. oldhamii*); RR (*R. rubropilosum*); RK (*R. kanehirai*); RS (*R. simsii*); RMO (*R. morii*); RP (*R. pseudochrysanthum*); RRU (*R. rubropunctatum*); RH (*R. hyperythrum*); RF (*R. formosanum*); RE (*R. ellipticum*); RKA (*R. kawakamii*).

†Species codes: MJ (*M. japonica*); MP (*M. philippinensis*); MK (*M. konishii*); MM (*M. mushaensis*); MZ (*M. zuihoensis*); MT (*M. thunbergii*); MKU (*M. kusanoi*); MO (*M. obovatifolia*).

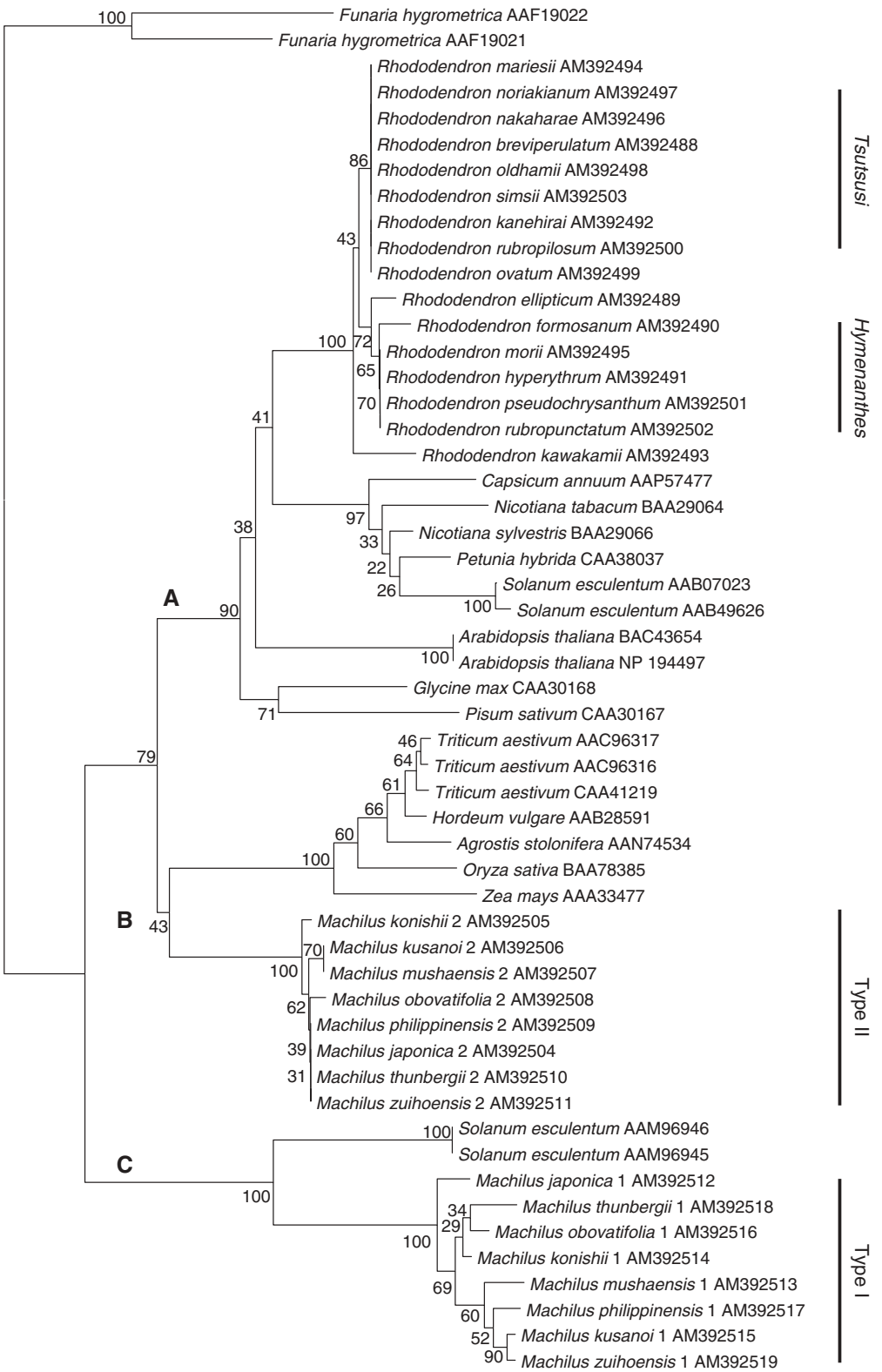

FIG. 2. Phylogenetic tree of chloroplast small heat shock proteins (CPsHSPs). The neighbour-joining tree was generated according to the amino acid JTT matrix (Jones *et al.*, 1992) using the MEGA program. Bootstrap values are shown. The tree is based on an analysis of an amino acid alignment of conceptually translated sequences of cloned CPsHSP genes from eight *Machilus* and 16 *Rhododendron* species as well as amino acid sequences of CPsHSPs from five monocots and seven other eudicots acquired from GenBank. The corresponding nucleotide sequences of *Machilus* and *Rhododendron* have been deposited in the EMBL database and their accession numbers are provided in the tree. GenBank accession numbers of acquired CPsHSP amino acid sequences are also provided in the tree. The bold letters A, B and C represent CPsHSP clusterings of eudicots, monocots and type II CPsHSP of *Machilus*, and type I CPsHSP of *Machilus*, respectively.

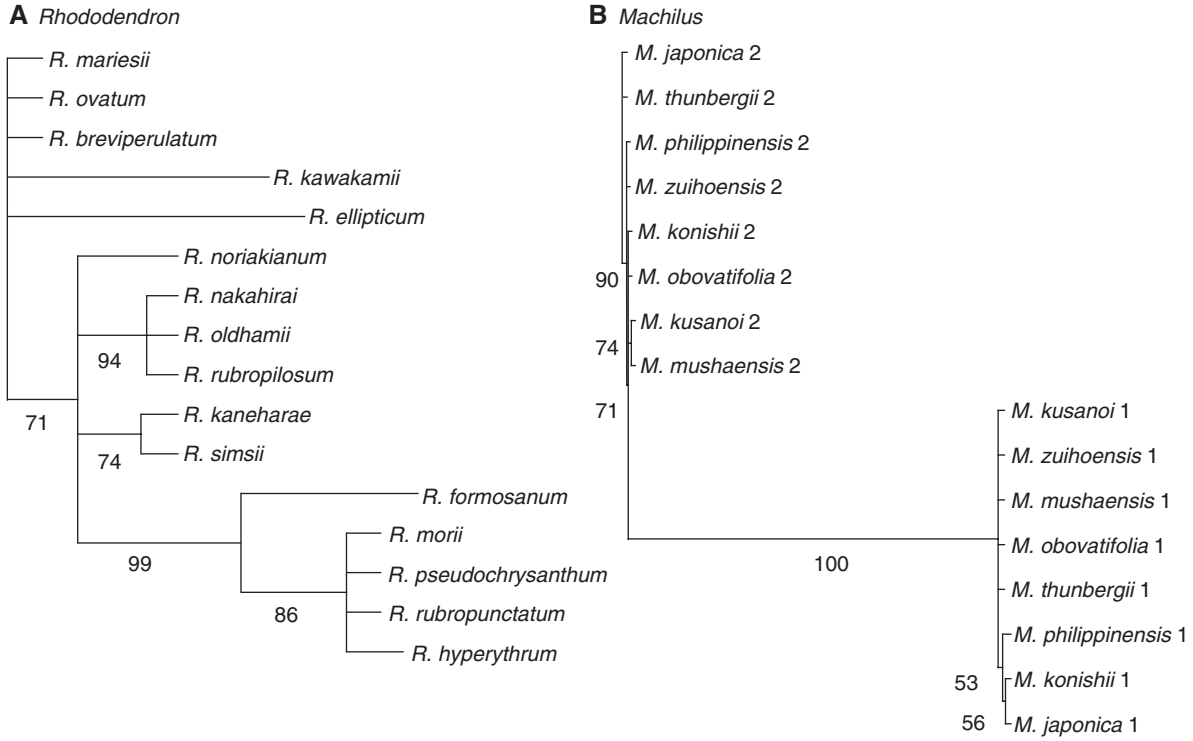

FIG. 3. Phylogenetic trees of the Bayesian analysis for (A) *Rhododendron* and (B) *Machilus* species. The Bayesian gene trees were generated by the MrBayes program based on the aligned DNA sequences according to the aligned amino acid sequences. Posterior probabilities of >50 % are shown. The Bayesian gene trees were used in the codeml analysis implemented in the PAML package. In *Rhododendron*, the *Hymenanthes* lineage was set as the foreground branch against the background branch, which including all other *Rhododendron* species investigated. In *Machilus*, the lineage of type II CPsHSP was set as the foreground branch and the lineage of type I CPsHSP was set as the background branch.

was also performed for the two duplicate copies of *Machilus*. Twice the difference of the maximum-likelihood values was compared with a chi-squares distribution. Both *Rhododendron* and *Machilus* displayed significant amino acid evolutionary rate differences when the *Rhododendron Hymenanthes* lineage was compared with the other *Rhododedron* species investigated ($2\delta L = 58{\cdot}79$, d.f. $= 1$, $P < 0{\cdot}0001$), when only species belonging to the subgenera *Hymenanthes* and *Tsutsusi* were compared ($2\delta L = 31{\cdot}73$, d.f. $= 1$, $P < 0{\cdot}0001$), and when the type II CPsHSP was compared with type I of *Machilus* ($2\delta L = 73{\cdot}12$, d.f. $= 1$, $P < 0{\cdot}0001$). The comparison in the nested model being significant suggests that the *Hymenanthes* lineage of *Rhododendron* has evolved faster than the others at the amino acid level. Unequal rates of evolution were also detected for the two types of CPsHSPs in *Machilus* at the amino acid level.

The LRT for positive selection compares the fit of the two nested models to the sequence data: a null model without adaptive evolution and an alternative model with adaptive evolution. Both models may invoke variation in ω among codons, but the null model is restricted to $\omega \leq 1$, whereas the alternative model allows adaptive evolution with $\omega > 1$. If the alternative model provides a significantly better fit to the data, then adaptive evolution is inferred. In *Rhododendron*, because interspecific sequence divergence of the CPsHSP is low, no significantly better model was detected in any of the comparisons (Tables 3 and 4). This suggests that, across the entire phylogenetic tree, most sites are evolving under absolute purifying selection. However, in all models that allow for positively selected site detection ($\omega > 1$), two amino acid sites (52M and 64V, Fig. 1, Table 3) were consistently targeted as being positively selected sites in the lineage leading to *R. formosanum* and the *R. pseudochrysanthum* complex (Table 3). One of the two sites that was detected as being a positively selected site (64V) had significant or high posterior probabilities under both the NEB and the BEB estimations. In *Machilus*, the two-ratio model that allows for a different ω for the type II CPsHSP compared with the type I CPsHSP (Table 3) was not a significantly better fit than the one-ratio model (M0) ($P = 0{\cdot}0803$) (Table 4). However, a 1·5-fold increase in ω leading to the type II CPsHSP relative to the type I CPsHSP lineage was found (Table 3). The results of the site models indicated that the selective pressure on the protein greatly varied among amino acid sites. All models allowing for positively selected sites (M2, M3 and M8) provided a significantly better fit to the data than their neutral counterparts (M1, M0 and M7, respectively) (Table 4). Four amino acid sites (5L, 21G, 49R and 114L) were consistently detected by the M2, M3 and M8 models as being positively selected sites, and some of them had significant or high posterior probabilities. Site 46L was detected as being a positively selected site by M3 (significant by NEB) and M8 (high posterior probability by NEB).

TABLE 3. *Results of the maximum-likelihood analysis of chloroplast small heat shock protein genes using a variety of methods*

Model	*P*	Likelihood	Parameter estimates	Positively selected sites: NEB	Positively selected sites: BEB
Rhododendron					
M0 (one-ratio)	1	−621·369412	$\omega = 0{\cdot}2430$	None	
Two-ratios	2	−621·368183	$\omega_0 = 0{\cdot}2451$	NA	
			$\omega_1 = 0{\cdot}2451$		
Site models					
M1: neutral	1	−618·220128	$p_0 = 0{\cdot}78109$	NA	
			$p_1 = 0{\cdot}21891$		
M2: selection	3	−617·542094	$p_0 = 0{\cdot}96391$	52M (0·895)	52M (0·620)
			$p_1 = 0{\cdot}0000$	64V (0·996**)	64V (0·845)
			$p_2 = 0{\cdot}03609$		
			$\omega_2 = 3{\cdot}97016$		
M3: discrete (k = 3)	5	−617·542214	$p_0 = 0{\cdot}47366$	52M (0·895)	
			$p_1 = 0{\cdot}49025$	64V (0·996**)	
			$p_2 = 0{\cdot}03609$		
			$\omega_0 = 0{\cdot}12453$		
			$\omega_1 = 0{\cdot}12453$		
			$\omega_2 = 3{\cdot}97017$		
M7: beta	2	−618·232193	$P = 0{\cdot}00876$	NA	
			$q = 0{\cdot}02865$		
M8: beta and ω	4	−617·544813	$p_0 = 0{\cdot}96408$	52M (0·889)	52M (0·744)
			$P = 14{\cdot}18467$	64V (0·995*)	64V (0·924)
			$q = 99{\cdot}00000$		
			$p_1 = 0{\cdot}03592$		
			$\omega = 3{\cdot}97668$		
Machilus					
M0 (one-ratio)	1	−1266·173004	$\omega = 0{\cdot}0831$	None	
Two-ratios	2	−1264·643353	$\omega_0 = 0{\cdot}0412$	NA	
			$\omega_1 = 0{\cdot}1020$		
Site models					
M1: neutral	1	−1241·605938	$p_0 = 0{\cdot}88034$	NA	
			$p_1 = 0{\cdot}11966$		
M2: selection	3	−1238·422465	$p_0 = 0{\cdot}88004$	5L (0·970*)	5L (0·968*)
			$p_1 = 0{\cdot}08398$	21G (0·883)	21G (0·8943)
			$p_2 = 0{\cdot}03598$	49R (0·692)	49R (0·739)
			$\omega_2 = 4{\cdot}47099$	114L (0·754)	114L (0·752)
M3: discrete (k = 3)	5	−1232·725612	$p_0 = 0{\cdot}57287$	5L (0·989*)	
			$p_1 = 0{\cdot}37527$	21G (1·000**)	
			$p_2 = 0{\cdot}05187$	46L (0·974*)	
			$\omega_0 = 0{\cdot}01236$	49R (0·994*)	
			$\omega_1 = 0{\cdot}14390$	114L (0·863)	
			$\omega_2 = 3{\cdot}40625$		
M7: beta	2	−1242·390531	$P = 0{\cdot}34373$	NA	
			$q = 2{\cdot}39091$		
M8: beta and ω	4	−1232·740909	$p_0 = 0{\cdot}94875$	5L (0·990*)	5L (0·981*)
			$P = 0{\cdot}51982$	21G (0·999**)	21G (0·978**)
			$q = 7{\cdot}00486$	46L (0·942)	46L (0·613)
			$p_1 = 0{\cdot}05125$	49R (0·989*)	49R (0·911)
			$\omega = 3{\cdot}46620$	114L (0·873)	114L (0·822)

p_0, p_1 and p_2 are the proportion of sites in categories 1, 2 and 3, respectively. ω is the d_N/d_S ratio in these categories of sites. *P* and *q* are beta estimates. The one-ratio (M0), two-ratios, M1 and M7 models do not allow positively selected sites. Asterisks indicate significance at the *0·95 and **0·99 levels.

TABLE 4. *Likelihood ratio tests for sites having different ω ratios along foreground branches*

Comparison	d.f.	*Rhododendron* 2δL	*Rhododendron* *P*	*Rhododendron* ω	*Machilus* 2δL	*Machilus* *P*	*Machilus* ω
One-ratio vs. two-ratio	1	0·0024	0·9609	0·2451	3·0593	0·0803	0·1020
M0 vs. M3	4	7·6543	0·1051	3·97017	66·8947	<0·0001	3·40625
M1 vs. M2	2	1·3560	0·5076	3·97016	6·3669	0·0414	4·47099
M7 vs. M8	2	1·3747	0·5029	3·97668	19·2992	<0·0001	3·46620

Hydrophobicity profile analyses

Hydrophobicity is important for the chaperone activity of sHSP in cells of various organisms (Shearstone and Baneyx, 1999; Liang *et al.*, 2000; Lindner *et al.*, 2000; Chowdary *et al.*, 2004). Therefore the mean hydrophobicity was calculated based on Kyte and Doolittle's (1982) method using the Bioedit program, which reveals the pattern of mean hydrophobicity along the aligned amino acid sequences (Fig. 4A, B). In *Rhododendron*, higher hydrophobicity was found for the CPsSHP of the representative sequence of the *Hymenanthes* lineage (*R. pseudochrysanthum*) in the ACD around amino acid site 40–60 in comparison with the representative sequence of the *Tsutsusi* lineage (*R. kanehirai*). In our data, changes in the hydrophobicity profile in the two CPsHSP paralogues of *Machilus mushaensis* also showed elevated hydrophobicity from the N-terminal into the ACD (Fig. 4B). By contrast, reduced hydrophobicity was found for the second half of the ACD.

The computer program SplitTester (Gao *et al.*, 2005) was used to examine also whether these *Rhododendron* lineages and duplicate copies of the *Machilus* have evolved CPsHSP functional divergence based on the hydrophobicity distance matrix. The SplitTester program compared the two groups of sequences based on the sequence alignment and produced phylogenetic trees in a series of sliding windows to identify whether two groups of sequences diverged in their functions. No difference was detected in the comparison between the CPsHSP sequences of the *Hymenanthes* lineage and those of other species in *Rhododendron* (Fig. 4C). By contrast, dramatic differences along the aligned sequences of the two types of CPsHSP in *Machilus* were observed, which indicates that the functions of the two types of CPsHSP in *Machilus* have greatly diverged (Fig. 4D).

DISCUSSION

A single phyletic line may diverge into a series of lineages that can adapt to rather different niches. Adaptive evolution can be a rapid process through which lineages derived from a recent common ancestor occur simultaneously. During this process, gene duplication or relaxed purifying selection of one orthologous gene may play an important role in functional divergence, which enables lineages to have better fitness in the rather different environments. This is exactly the case observed here for *Rhododendron*. The amino acid sequences of the CPsHSP are highly conserved across the four subgenera of *Rhododendron* in Taiwan with the 5–7 amino acid sites of *R. kawakamii* having diverged the most compared with those of the other CPsHSPs of *Rhododendron*. However, non-synonymous substitutions of two amino acids when comparing the *Hymenanthe*s lineage to the *Tsutsusi* lineage may have led to functional divergence.

No significant difference was found between the compared models, which probably resulted from the high sequence similarity of CPsHSPs in *Rhododendron*. However, a significant difference in the evolutionary rate at the amino acid level was detected, and along the ACD two putative positively selected amino acids were identified that were confined to the lineage leading to the five *Rhododendron* species of *R. formosanum*, *R. pseudochrysanthum*, *R. hyperythrum*, *R. morii* and *R. rubropunctatum* (classified in the subgenus *Hymenanthes*). Although no duplicated copy of CPsHSP was found in *Rhododendron*, diversifying evolution of the CPsHSP is apparent since the splitting of the subgenera *Hymenanthes* and *Tsutsusi*. It is likely that adaptively evolving mutations might be encountered in the ACD region of the CPsHSP that is associated with the evolution of the *Hymenanthes* lineage. Furthermore, relaxed purifying selection might have played an important role, resulting in the non-synonymous substitutions at sites 52 and 64 and related to oligomerization, substrate binding and the chaperone activity of CPsHSP. The major morphological difference in the five species in the subgenus *Hymenanthes* from the remaining *Rhododendron* species under investigation is the thick coriaceous leaves in the former and chartaceous or thin coriaceous leaves in the latter. *Rhododendron pseudochrysanthum*, *R. morii* and *R. hyperythrum* are believed to have originated from *R. rubropunctatum* and all are derivatives of *R. formosanum* (Chung *et al.*, 2007; Hwang *et al.*, unpubl. res.). Therefore, the derivation of the two positively selected non-synonymous substitutions is lineage-specific and probably related to the adaptation of these species to specific habitats. However, stochastic processes cannot be excluded. Determining whether changes in these two amino acids are related to the enhanced chaperone activity of the CPsHSP in *Rhododendron* requires further investigation.

The ecological and evolutionary roles of the CPsHSP in plants have thus far been investigated for the differential expression levels of CPsHSP in closely related species that occupy different habitats (Knight and Ackerly, 2001; Barua *et al.*, 2003). Differential CPsHSP content levels in closely related species or different ecotypes are considered the functional consequences of natural variation in the CPsHSP and are associated with evolutionary adaptations to specific habitats among closely related species or ecotypes. Sequence variation in the CPsHSP related to species radiation has not been reported. The data presented here demonstrate an example of DNA sequence variation in the ACD of the CPsHSP that might have had an important role in the adaptive evolution of the *Hymenanthes* lineage of *Rhododendron*, as its members possess contrasting morphological variations from the remaining *Rhododendron* species in Taiwan and are found in specific habitats.

Machilus demonstrates a history of duplication in the CPsHSP gene. Ancient copy of the CPsHSP (type I) in *Machilus* has accumulated more interspecific differences in non-synonymous substitutions than has the more-recent copy of the CPsHSP (type II). The level of intraspecific variation may be small given that selective constraints are usually great in the protein coding sequences, but the duplicates of CPsHSP might not have experienced similar selective forces (Kundsen and Miyamoto, 2001; Zhang *et al.*, 2002; Frydenberg *et al.*, 2003; Lynch and Katju, 2004; Yang *et al.*, 2005). The two types of CPsHSP in

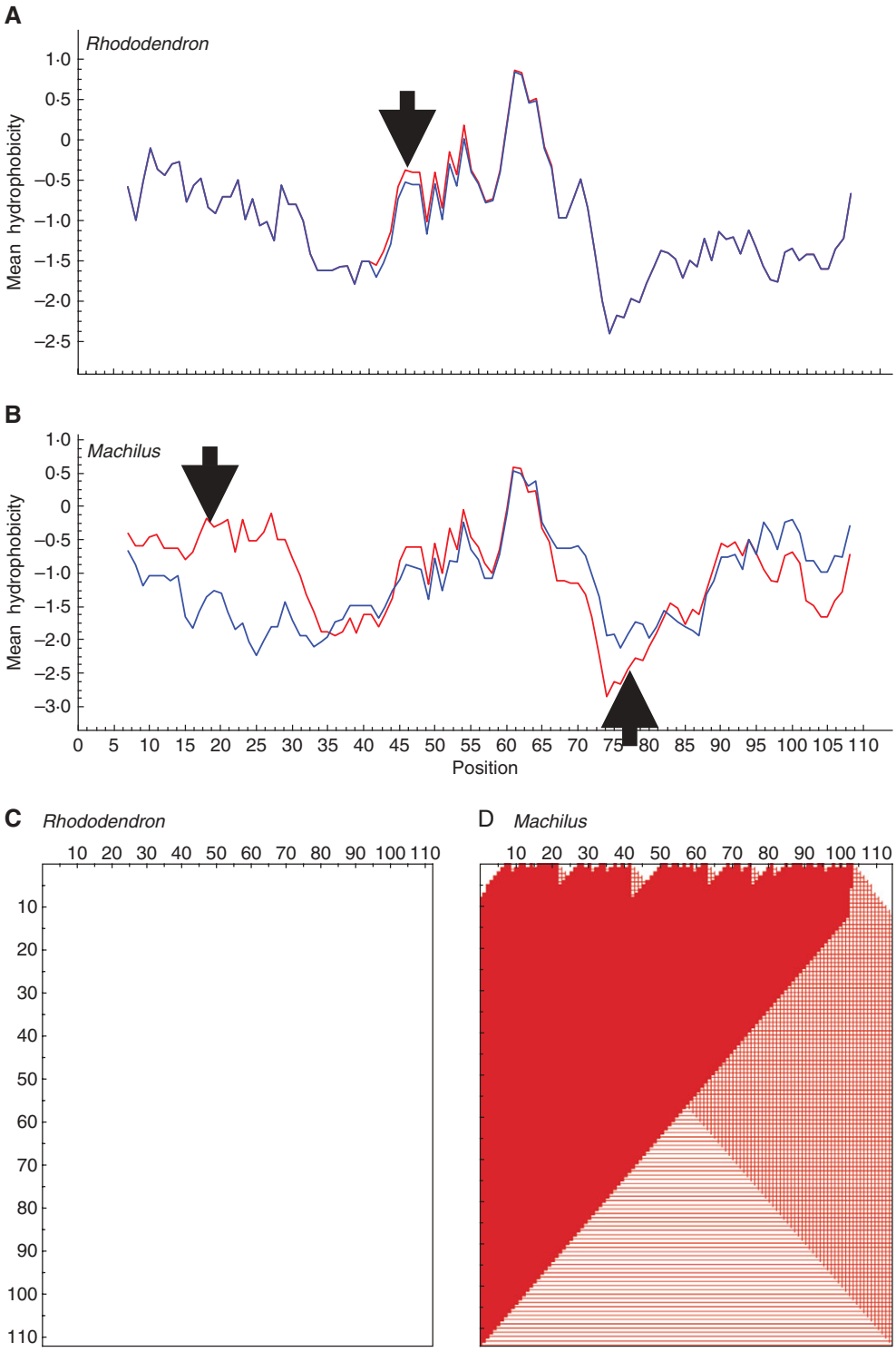

FIG. 4. Hydrophobicity profile and functional divergence analysis. The hydrophobicity profile were generated for: (A) *Rhododendron* – the arrow points to the amino acid sequences of *R. pseudochrysanthum* (red line), which had higher hydrophobicity than *R. kanehirai* (blue line); overlapping of two lines are seen due to sequence identity; and (B) *Machilus* – the arrow points to the amino acid sequences which had higher and lower hydrophobicities in different regions when comparing the type II (red line) with the type I CPsHSP (blue line) of *Machilus mushaensis*. Functional divergence of aligned amino acid sequences between different groups of (C) *Rhododendron* and (D) *Machilus* species was evaluated with the SplitTester program based on a hydrophobicity distance matrix (Gao *et al.*, 2005). The *x*-axis represents the length of the aligned sequences; the *y*-axis represents increasing sliding window size. SplitTester aims to identify specific domain sequences responsible for functional divergence of subgroups within a protein family. The red colour indicates where the degree of confidence supporting the functional split of domain sequences was 100 % across the entire sliding windows in the comparison between two types of CPsHSPs in *Machilus*. The comparison between the *Hymenanthes* and *Tsutsusi* lineages resulted in a blank output, indicating no significant functional split was found along the aligned amino acid sequences of the *Rhododendron* CPsHSPs.

Machilus and the CPsHSP in *Rhododendron* shared highly similar amino acid sequences in the methionine-rich region but diverged in the ACD. The high conservation of the methionine-rich region suggests that the selective constraints enforced on this region were great. The dramatic divergence in the amino acid sequences of the ACD indicates functional divergence. It is likely that during the course of evolution *Rhododendron* and *Machilus* shared a recent copy of the CPsHSP that probably evolved as a more efficient chaperonin molecule. The loss of the ancient copy of the CPsHSP is unique to *Rhododendron* as another eudicot, *Solanum esculentum*, possessed both types of CPsHSP. Furthermore, monocots apparently contain only one type of CPsHSP, as that was the only type found in the database.

The methionine-rich region was found to form an amphipathic α-helix (Chen and Vierling, 1991; Waters and Vierling, 1999). Selective constraints were found to be higher on the secondary structure than on the primary sequence of the sHSPs (Caspers *et al.*, 1995). The site 12 methionine, which is not one of the four conserved methionines mentioned by Waters and Vierling (1999) in the methionine-rich region, was found to have been substituted by phenylalanine in the more recent copy of the CPsHSP; this also contributed to elevation of the hydrophobicity profile in this region. However, this amino acid change was not detected as having been positively selected by the codeml analysis. This site corresponds to site 15 (phenylalanine) in the hydrophobic face of the amphipathic α-helix in *Arabidopsis thaliana* Hsp21 (Waters and Vierling, 1999). The four conserved methionine residues in the methionine-rich region of the CPsHSP were reported to be related to the oxidation-induced conformational changes in the oligomer and offer a possibility for dynamic formation of oligomeric structure for *A. thaliana* Hsp21 under oxidative environments (Gustavsson *et al.*, 1999). Replacement of the four most-conserved methionines in the amphipathic α-helix by leucine did not produce a response to oxidation resulting in loss of α-helical structure, but the chaperone-like activity was unaffected and was actually even better in the citrate synthase assay (Gustavsson *et al.*, 2001). The enhanced chaperone activity was thought to be due to increased conformational stability. Chloroplasts provide a safe reducing environment in which methionine can be kept in a reduced state and therefore substitution of methionine by leucine increased chaperone activity. Both leucine and phenylalanine have higher hydrophobicity than that for methionine. It is likely that conserving other methionines in the methionine-rich domain preserved the ability for dynamic oligomeric structure formation, and converting the methionine at site 12 (in the hydrophobic face of the amphipathic α-helix) to phenylalanine enhanced the hydrophobicity and therefore raised the stability of the oligomeric structure, consequently enhancing the molecular chaperone activity.

No apparent relationship of ecological relevance toward the positively selected amino acid sites was found in *Machilus*. This lack of correlation is possibly due to the overly divergent sequences of the duplicated CPsHSPs in *Machilus*, so it is not possible to deduce any meaningful inferences from the comparison. However, three of the five positively selected sites detected are located in the β-strand and are probably related to dimer formation (deJong *et al.*, 1998; Haslbeck *et al.*, 1999; Giese and Vierling, 2004). The high levels of divergence between the two types of CPsHSPs in *Machilus* may have resulted in the false positive detection of positively selected sites along the aligned sequences (Anisimova *et al.*, 2002). Nevertheless, a significant difference in evolutionary rates of the paralogues at the amino acid level was evident, which indicates functional divergence (Kundsen and Miyamoto, 2001; Gu *et al.*, 2002; Yang *et al.*, 2005). As the amino acid differences found were concentrated in the ACD, the domain-specific divergence in the CPsHSP paralogues of *Machilus* might have been an important clue for the identity of functional divergence between the paralogues. It is important to note that the *Rhododendron* CPsHSP and *Machilus* type II CPsHSP share relatively higher similarity in both the methionine-rich region and the ACD. From the results of the hydrophobicity profile analyses, it is likely that the majority of changes in the ACD have been directed at shifting hydrophobicity. From the data here, changes in the hydrophobicity profile in the CPsHSP paralogues of *Machilus* are interesting in that the type II CPsHSP had significantly elevated hydrophobicity from the methionine-rich region into the middle of the ACD (Fig. 4B). By contrast, reduced hydrophobicity was found for the second half of the ACD. A similar pattern of a hydrophobicity shift along the CPsHSP sequences was also observed with the comparison between the representative sequences of the two types of the CPsHSP in *Machilus*. The pattern of change was also similar when the CPsHSP of *Rhodoedndron* was compared with the type I CPsHSP of *Machilus* (data not shown). Consequently, it can be generalized that in angiosperms the recent copy of the CPsHSP has elevated hydrophobicity in the first half of the CPsHSP but decreased hydrophobicity in the second half, and these might have played important roles in the chaperone activity of the CPsHSP.

CONCLUSIONS

The evolutionary analysis herein suggests that the CPsHSP arose prior to the divergence of the major groups of angiosperms (Waters *et al.*, 1996). The CPsHSP was reported to have evolved through gene duplication from a nuclear-encoded cytosolic sHSP (Waters and Vierling, 1999). Furthermore, angiosperms gained a duplicate of the CPsHSP after its divergence from cytosolic sHSP and subsequently lost the ancient copy of CPsHSP probably in *Rhododendron* and monocots (Fig. 2). Plant sHSPs are heat-inducible and in some cases developmentally regulated; however, they are not significant components in non-stressed cells (Vierling, 1991; Waters *et al.*, 1996; Sun *et al.*, 2002). It is believed that the plant CPsHSP is functionally related to protection of photosystem II electron transport in the chloroplasts (Heckathorn *et al.*, 1998). Therefore, it is possible that genetic changes leading to the evolution of CPsHSPs have been necessary since their

divergence from the cytosolic sHSP. Selection pressures that might have been unique to plant lineages generated various types of sHSPs specific to plants. Conservation of the methionine-rich domain among the different types of CPsHSPs in *Rhododendron* and *Machilus* supports the hypothesis that selection constraints restricted the maintenance of the methionine-rich domain to ensure its functional divergence from other sHSPs such as mitochondrial, endoreticular and cytosolic sHSPs in plants (Waters and Vierling, 1999). Most importantly, it has been found in the present study that the evolving changes specific to the *Hymenanthes* lineage in *Rhododendron* and the selection favouring the shift in the hydrophobicity profile for *Rhododendron* and *Machilus* might have been related to the expanding substrate specificity required.

SUPPLEMENTARY MATERIAL

The aligned nucleotide sequences for *Machilus* and *Rhododendron* are available online at http://aob.oxfordjournals.org/.

ACKNOWLEDGEMENTS

This study was supported in part by grants from the National Science Council (NSC95-2313-B-034-001) and the Council of Agriculture (94AS-9·1·7-FB-e3), Executive Yuan, Taiwan. We are grateful to Jia-Bin Tsai, Taiwan Forest Research Institute, for assistance with some of the sample collection.

LITERATURE CITED

Adams KL, Wendel JF. 2005. Polyploidy and genome evolution in plants. *Current Opinion in Plant Biology* **8**: 135–141.

Anisimova M, Bielawski JP, Yang Z. 2002. Accuracy and power of Bayes prediction of amino acid sites under positive selection. *Molecular Biology and Evolution* **19**: 950–958.

Barua D, Heckathorn SA, Downs CA. 2003. Variation in chloroplast small heat-shock protein function is a major determinant of variation in thermotolerance of photosynthetic electron transport among ecotypes of *Chenopodium album*. *Functional Plant Biology* **30**: 1071–1079.

Caspers GJ, Leunissen JAM, de Jong WW. 1995. The expanding small heat-shock protein family, and structure predictions of the conserved 'α-crystallin domain'. *Journal of Molecular Evolution* **40**: 238–248.

Chen Q, Vierling E. 1991. Analysis of conserved domains identifies a unique structural feature of a chloroplast heat shock protein. *Molecular and General Genetics* **226**: 425–431.

Chowdary TK, Raman B, Ramakrishna T, Rao CM. 2004. Mammalian Hsp22 is a heat-inducible small heat-shock protein with chaperone-like activity. *Biochemistry Journal* **381**: 379–387.

Chung JD, Lin TP, Chen YL, Cheng YP, Hwang SY. 2007. Phylogeographic study reveals the origin and evolutionary history of a *Rhododendron* species complex in Taiwan. *Molecular Phylogenetics and Evolution* **42**: 14–24.

Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf material. *Phytochemistry Bulletin* **19**: 11–15.

Frydenberg J, Hoffmann AA, Loeschcke V. 2003. DNA sequence variation and latitudinal associations in hsp23, hsp26 and hsp27 from natural populations of *Drosophila melanogaster*. *Molecular Ecology* **12**: 2025–2032.

Ganea E. 2001. Chaperone-like activity of α-crystallin and other small heat shock proteins. *Current Protein and Peptide Science* **2**: 205–225.

Gao X, VanderVelden KA, Voytas DF, Gu X. 2005. SplitTester: software to identify domains responsible for functional divergence in protein family. *BioMed Central Bioinformatics* **6**: 137.

Giese KC, Vierling E. 2004. Mutants in a small heat shock protein that affect the oligomeric state. *The Journal of Biological Chemistry* **279**: 32674–32683.

Gu J, Wang Y, Gu X. 2002. Evolutionary analysis for functional divergence of Jak protein kinase domains and tissue-specific genes. *Journal of Molecular Evolution* **54**: 725–733

Gustavsson N, Härndahl U, Emanuelsson A, Roepstorff P, Sundby C. 1999. Methionine sulfoxidation of the chloroplast small heat shock protein and conformational changes in the oligomer. *Protein Science* **8**: 2506–2512.

Gustavsson N, Kokke BPA, Anzelius B, Boelens WC, Sundby C. 2001. Substitution of conserved methionines by leucines in chloroplast small heat shock protein results in loss of redox-response but retained chaperone-like activity. *Protein Science* **10**: 1785–1793.

Hall TA. 1999. BIOEDIT: a user-friendly biological sequence alignment, editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41**: 95–98.

Hasegawa M, Kishino H, Yano TA. 1985. Dating of the human–ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution* **22**: 160–174.

Haslbeck M, Walke S, Stromer T, Ehrnsperger M, White HE, Chen S, *et al.* 1999. Hsp26: a temperature-regulated chaperone. *The European Molecular Biology Organization Journal* **18**: 6744–6751.

Heckathorn SA, Downs CA, Sharkey TD, Coleman JS. 1998. The small, methionine-rich chloroplast heat-shock protein protects photosystem II electron transport during heat stress. *Plant Physiology* **116**: 439–444.

Heckathorn SA, Mueller JK, LaGuidice S, Zhu B, Barrett T, Blair B, Dong Y. 2004. Chloroplast small heat-shock proteins protect photosynthesis during heavy metal stress. *American Journal of Botany* **91**: 1312–1318.

Hsieh CF. 2002. Composition, endemism and phytogeographical affinities of the Taiwan flora. *Taiwania* **47**: 298–310.

Huelsenbeck JP, Ronquist F. 2001. MrBayes: bayesian inference of phylogenetic trees. *Bioinformatics* **17**: 754–755.

Hwang SY, Hsu KK. 2001. Molecular phylogeny of eight Taiwanese *Rhododendron* species based on chloroplast *trn*F-*trn*L DNA sequences. *Taiwan Journal of Forest Science* **16**: 153–160 (in Chinese with English summary).

Jones DT, Taylor WR, Thornton JM. 1992. A new approach to protein fold recognition. *Nature* **358**: 86–89.

deJong WW, Caspers GJ, Leunissen JA. 1998. Genealogy of the alpha-crystallin–small heat-shock protein superfamily. *International Journal of Biological Macromolecules* **22**: 151–162

Jordan IK, Wolf YI, Koonin EV. 2004. Duplicated genes evolve slower than singletons despite the initial rate increase. *BioMed Central Evolutionary Biology* **4**: 22.

Kimura M. 1980. A simple method for estimating evolutionary rate of base substitution through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* **16**: 111–120.

Knight CA, Ackerly DD. 2001. Correlated evolution of chloroplast heat shock protein expression in closely related plant species. *American Journal of Botany* **88**: 411–418.

Kumar S, Tamura K, Nei M. 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Briefings in Bioinformatics* **5**: 150–163.

Kundsen B, Miyamoto M. 2001. A likelihood ratio test for evolutionary rate shifts and functional divergence among proteins. *Proceedings of the National Academy of Sciences of the USA* **98**: 14512–14517.

Kyte J, Doolittle R. 1982. A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology* **157**: 105–132.

Li WH. 1997. *Molecular evolution.* Sunderland, MA: Sinauer Associates.

Liang JJ, Sun TX, Akhtar NJ. 2000. Heat-induced conformational change of human lens recombinant αA- and αB-crystallins. *Molecular Vision* **6**: 10–14.

Lindner RA, Carver JA, Ehrnsperger M, Buchner J, Esposito G, Behlke J, *et al.* 2000. Mouse Hsp25, a small heat shock protein: the role of its C-terminal extension in oligomerization and chaperone action. *European Journal of Biochemistry* **267**: 1923–1932.

Lynch L, Katju V. 2004. The altered evolutionary trajectories of gene duplicates. *Trends in Genetics* **20**: 544–549.

Lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. *Science* **290**: 1151–1154.

Mao Q, Chang Z. 2001. Site-directed mutation on the only universally conserved residue Leu122 of small heat shock protein Hsp16·3. *Biochemical and Biophysical Research Communications* **289**: 1257–1261.

Mitchell-Olds T, Schmitt J. 2006. Genetic mechanisms and evolutionary significance of natural variation in *Arabidopsis*. *Nature* **441**: 947–952.

Ohta T. 1991. Multigene families and the evolution of complexity. *Journal of Molecular Evolution* **33**: 34–41.

Posada D, Crandall KA. 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics* **14**: 817–818.

Posada D, Crandall KA. 2001. Selecting models of nucleotide substitution: an application to the human immunodeficiency virus 1 (HIV-1). *Molecular Biology and Evolution* **18**: 897–906.

Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**: 1572–1574.

Rose TM, Schultz ER, Henikoff JG, Pietrokovski S, McCallum CM, Henikoff S. 1998. Consensus–degenerate hybrid oligonucleotide primers for amplification of distantly related sequences. *Nucleic Acids Research* **26**: 1628–1635.

Seoighe C, Johnston CR, Shields DC. 2003. Significantly different patterns of amino acid replacement after gene duplication as compared to after speciation. *Molecular Biology and Evoution.* **20**: 484–490.

Sharma KK, Kaur H, Kumar GS, Kester K. 1998. Interaction of 1,1′Bi(4-anilino)naphthalene-5,5′-Disulfonic acid with α-crystallin. *Journal of Biological Chemistry* **273**: 8965–8970.

Shearstone J, Baneyx F. 1999. Biochemical characterization of the small heat shock proteins IbpB from *Escherichia coli*. *Journal of Biological Chemistry* **274**: 9937–9945.

Sørensen JG, Kristensen TN, Loeschcke V. 2003. The evolutionary and ecological role of heat shock proteins. *Ecology Letters* **6**: 1025–1037.

Sun W, VanMontagu M, Verbruggen N. 2002. Small heat shock proteins and stress tolerance in plants. *Biochimica et Biophysica Acta* **1577**: 1–9.

Sundby C, Härndahl U, Gustavsson N, Åhrman E, Murphy DJ. 2005. Conserved methionines in chloroplasts. *Biochimica et Biophysica Acta* **1703**: 191–202.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. 1997. The Clustal X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* **25**: 4876–4882.

Tsai CC, Huang SC, Chen CH, Tseng YH, Huang PL, Tsai SH, *et al.* 2003. Genetic relationships of *Rhododendron* (Ericaceae) in Taiwan based on the sequence of the internal transcribed spacer of ribosomal DNA. *Journal of Horticultural Science and Biotechnology* **78**: 234–240.

Vierling E. 1991. The roles of heat shock proteins in plants. *Annual Review of Plant Physiology and Plant Molecular Biology* **42**: 579–620

Wagner A. 2002. Asymmetric functional divergence of duplicate genes. *Molecular Biology and Evolution* **19**: 1760–1768.

Waters ER. 1995. The molecular evolution of the small heat-shock proteins in plants. *Genetics* **141**: 785–795.

Waters ER, Vierling E. 1999. Chloroplast small heart shock proteins: evidence for atypical evolution of an organelle-localized protein. *Proceedings of the National Academy of Sciences of the USA* **96**: 14394–14399.

Waters ER, Lee GJ, Vierling E. 1996. Evolution, structure and function of the small heat shock proteins in plants. *Journal of Experimental Botany* **47**: 325–338.

Wright SI, Gaut BS. 2005. Molecular population genetics and the search for adaptive evolution in plants. *Molecular Biology and Evolution* **22**: 506–519.

Yang J, Xie Z, Blover BJ. 2005. Asymmetric evolution of duplicate genes encoding the CCAAT-binding factor NF-Y in plant genomes. *New Phytologist* **165**: 623–631.

Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Computer Applications in Biosciences* **13**: 555–556.

Yang Z, Nielsen R. 1998. Synonymous and nonsynonymous rate variation in nuclear genes of mammals. *Journal of Molecular Evolution* **46**: 409–418.

Yang Z, Nielsen R. 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Molecular Biology and Evolution* **19**: 908–917.

Yang Z, Nielsen R, Goldman N, Pedersen AMK. 2003. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* **22**: 431–449.

Zhang J. 2003. Evolution by gene duplication: an update. *Trends in Ecology and Evolution* **18**: 292–298.

Zhang L, Peek AS, Dunams D, Gaut BS. 2002. Population genetics of duplicated disease-defense genes, *hm1* and *hm2*, in Maize (*Zea mays* ssp. *mays* L.) and its wild ancestor (*Zea mays* ssp. *parviglumis*). *Genetics* **162**: 851–860.