Molecular Evolutionary Mechanisms Driving Functional Diversification of the HSP90A Family of Heat Shock Proteins in Eukaryotes

Lorenzo Carretero-Paulet,*[1,2] Victor A. Albert,[2] and Mario A. Fares*[1,3]

[1]Institute for Plant Molecular and Cell Biology – IBMCP (CSIC-UPV), Integrative Systems Biology Group, Valencia, Spain

[2]Department of Biological Sciences, SUNY-University at Buffalo

[3]Department of Genetics, School of Genetics and Microbiology, University of Dublin, Trinity College, Dublin, Ireland

***Corresponding author:** E-mail: lorenzoc@buffalo.edu; mfares@ibmcp.upv.es.

Associate Editor: Helen Piontkivska

Abstract

The ubiquitous and conserved cytosolic heat-shock proteins 90 (HSP90A) perform essential functions in the cell. To understand the evolutionary origin of HSP90A functional diversification, we analyzed the distribution of HSP90A family from 54 species representing the main eukaryotic lineages. Three independent HSP90A duplications led to the paralog subfamilies HSP90AA (heat-stress inducible) and HSP90AB (constitutive) and trace back to key time points during vertebrate, seed plant, and yeast evolution. HSP90AA and HSP90AB present divergent selection pressures, positive selection (PS), and signatures of functional divergence (FD) after duplication. The differential evolutionary patterns support different mechanisms for HSP90A functional diversification in vertebrates and seed plants. Mapping of PS and FD residues onto the HSP90A structure suggests the acquisition of novel and/or specialized client protein and/or cochaperone binding functions. We propose these residues as targets for further experimental studies of HSP90A proteins, reported to be capacitors of rapid evolutionary change, and targets for anticancer therapeutics.

Key words: **gene duplication, functional specialization, heat shock proteins, positive selection, divergent selection pressures.**

Letter

Heat-shock protein 90 (HSP90), a ubiquitous molecular chaperone, assists in the folding, degradation, transport, and subsequent activation of a highly select but functionally and structurally diverse set of client proteins, including steroid hormone receptors, protein kinases, and transcription factors. HSP90 is directly involved in the regulation and signaling of a wide range of essential processes mainly related to thermotolerance and stress responses, but also hormone signalling, cell cycle control, development, and apoptosis (Young et al. 2001; McClellan et al. 2007). In eukaryotes, HSP90 proteins can be found in different subcellular locations (cytosol, endoplasmic reticulum, chloroplast, and mitochondrion). Although cytosolic HSP90s, called HSP90As, are essential for viability under all conditions in eukaryotes, the bacterial homolog HTPG is dispensable under nonheat stress conditions (Bardwell and Craig 1988).

Functional promiscuity in HSP90A is promoted by its history of duplication events, which provides raw genetic material to evolve novel gene and gene functions (Ohno 1970; Zhang 2003; Conant and Wolfe 2008). HSP90As appear in two major subfamilies that originated through gene duplication: HSP90AA (HSP90α) and HSP90AB (HSP90β). Expression divergence after HSP90A gene duplication has been reported previously; in a number of animal species, HSP90AA is strongly upregulated in response to elevated temperatures, whereas HSP90AB is not induced during heat shock (Meng et al. 1993; Krone and Sass 1994; Pepin et al. 2001). In *Arabidopsis thaliana*, of the four *HSP90A* genes reported, one is heat inducible, whereas the rest are constitutively expressed (Yabe et al. 1994). Similar evidence of expression divergence between HSP82 and HSC82 has been found in *Saccharomyces cerevisiae* (Borkovich et al. 1989). However, unlike expression divergence, functional divergence (FD) of HSP90A at the level of coding regions has not been explored before.

Understanding how natural selection shapes functional diversification of HSP90A, the mutational dynamics supporting this diversification and how this can be used in protein engineering is paramount to evolutionary biology, biochemistry, and biotechnology due to its roles in signal transduction and regulation of diverse essential biological processes, its ability to canalize evolution, and its essentiality under heat-stress conditions. To understand and test FD of HSP90As after gene duplication, we performed a comprehensive survey of the HSP90A family in 54 representative eukaryotic genomes, dated HSP90A duplication events, and characterized the amino acid changes that likely led to HSP90A FD. Our results set the ground for further experimental research aimed at characterizing the diversification of HSP90A functions. Also, HSP90A provides a good model to understand eukaryotic diversification since, as we show in this article, this gene has undergone a history of repeated duplication events concomitant with the emergence of major eukaryotic innovations.

Evolutionary Origin and Diversification of the Eukaryotic HSP90A Family

We sampled 95 HSP90A sequences from the entire sequenced genomes of 55 species representing a wide variety of eukaryotic lineages plus a prokaryotic homolog (supplementary table S1, Supplementary Material online), all of which bear functional HSP90A sequence signatures (Chen et al. 2006). Of the 55 species, 22 showed *HSP90A* multigene families formed by up to six members. Phylogenetic analysis confirmed the occurrence of independent HSP90A duplication events in higher order eukaryotic lineages, such as vertebrates, fungi species from the *Saccharomyces* group, and seed plants (fig. 1). Moreover, additional gene duplications occurred in specific lineages, such as the reptile *Anolis carolinensis* and the basal chordate *Branchiostoma floridae*, which, according to our phylogenetic analysis, emerged from independent lineage-specific duplication events (fig. 1). Searches through the genomes of the basal vertebrate species, the cartilaginous fish *Callorhinchus milii* and the jawless fish *Petromyzon marinus*, retrieved 11 contig sequences, clustering with both HSP90AA and HSP90AB subfamilies and two contig sequences, branching within subfamily HSP90AA, respectively, (supplementary table S2, Supplementary Material online; unpublished results). Assuming no *HSP90AB* gene was lost or missing in the current version of the *P. marinus* genome, the origin of the duplication yielding HSP90AA and HSP90AB might be well placed before the split between cartilaginous and bony vertebrates, approximately 525 Ma, but after the divergence between jawless and jawed vertebrates, dating 652 Ma (Blair and Hedges 2005). Therefore, this duplication might correspond to the second round of vertebrate whole-genome duplication (WGD) (Putnam et al. 2008).

In plants, the duplication event leading to HSP90AA and HSP90AB emerged before the divergence between monocot and eudicot species (140–150 Ma) (fig. 1) but after the divergence of vascular plants from the basal moss species *Physcomitrella patens*, estimated to have occurred 450 Ma (Chaw et al. 2004; Rensing et al. 2008). Correspondingly, when the 2-Mb genomic regions containing *A. thaliana* and *Oryza sativa HSP90A* genes were compared using CoGE/GEvo, a high degree of synteny and collinearity was observed (http://genomevolution.org/r/5zk8), supporting their emergence before the divergence of the monocot and dicot lineages. Moreover, we found homologs showing distinctive sequence signatures for HSP90AA and HSP90AB in the genomes of the gymnosperm species *Pinus taeda*, *Pin. pinaster*, and *Chamaecyparis obtuse* (supplementary table S2, Supplementary Material online), which showed consistent branching when subjected to phylogenetic analyses with plant HSP90AA and HSP90AB sequences (unpublished results). This supports that the duplication leading to both HSP90A subfamilies traces back to the ancient ζ WGD shared by all extant seed plants 310 Ma (Jiao et al. 2011). These results, together with the HSP90A duplication in the WGD of *Saccharomyces* species 100 Ma (Wolfe and Shields 1997; Kellis et al. 2004), trace main *HSP90A* gene duplications back to polyploidy events occurring at eukaryote evolutionary landmarks, perhaps reflecting key roles for HSP90A gene duplicates in eukaryote evolutionary diversification.

Molecular Evolutionary Analysis of HSP90A Functional Specialization After Gene Duplication

Heat-induced activation of *HSP90A* genes is evolutionarily conserved across eukaryotes (Wu 1995; Zou et al. 1998; Yamada et al. 2007; Chan-Schaminet et al. 2009; Leach et al. 2012). Furthermore, HTPG, the homolog of HSP90A in prokaryotes, is also heat inducible (Nadeau et al. 1993), suggesting that fast and high accumulation of HSP90A protein during heat shock responses, and not constitutive expression, might represent the ancestral function. Expression divergence of HSP90A after duplication suggests that diversification in gene expression regulation may play an important role in maintaining the functional homeostasis of the eukaryote cell. However, little is known about functional diversification of HSP90A paralogs acting at the coding region level.

To understand the evolutionary basis of functional diversification of HSP90AA and HSP90AB, we estimated the nonsynonymous-to-synonymous rates ratio ($\omega = dN/dS$) (table 1) under different codon substitution-based evolutionary models. The mean ω values were equally low for vertebrate ($\omega = 0.0222$) and seed plants ($\omega = 0.0221$) HSP90As, reflecting strong purifying selection during most HSP90A evolution. The two-ratio branch model indicates asymmetrical evolution after gene duplication for all the four foreground branches. A branch-site model allowing heterogeneous ω values across sequences and branches leading to each HSP90A paralog subfamily, strongly indicated divergent selection pressures (DSP) at the sequence level in seed plant HSP90AA and HSP90AB and vertebrate HSP90AA but not along the vertebrate HSP90AB branch.

Furthermore, the branch-site models for positive selection (PS) were found to fit significantly better to the data in both vertebrate and seed plant HSP90AA and HSP90AB lineages, and identify amino acid changes fixed by PS (table 1; supplementary table S3, Supplementary Material online). Examination of the location of PS sites in the corresponding alignments (supplementary figs. S1 and S2, Supplementary Material online), and the 3D structure of yeast HSP82 (Ali et al. 2006) (fig. 2), revealed their distribution in the middle and COOH-terminal domains. Many PS sites corresponded to positions conserved in the sequences selected as foreground branch, whereas variable in the background branches (78, 480, and 530 in supplementary fig. S2, Supplementary Material online) or vice versa (640 in supplementary fig. S1, Supplementary Material online; 560 in supplementary fig. S2, Supplementary Material online). Some others corresponded to highly conserved positions involving amino acid changes between vertebrate (426 and 500, supplementary fig. S1, Supplementary Material online) or seed plant (77, supplementary fig. S2, Supplementary Material online) HSP90AA and HSP90AB subfamilies, respectively. Eight out of 12 as well as three out of four PS sites found in the analyses of

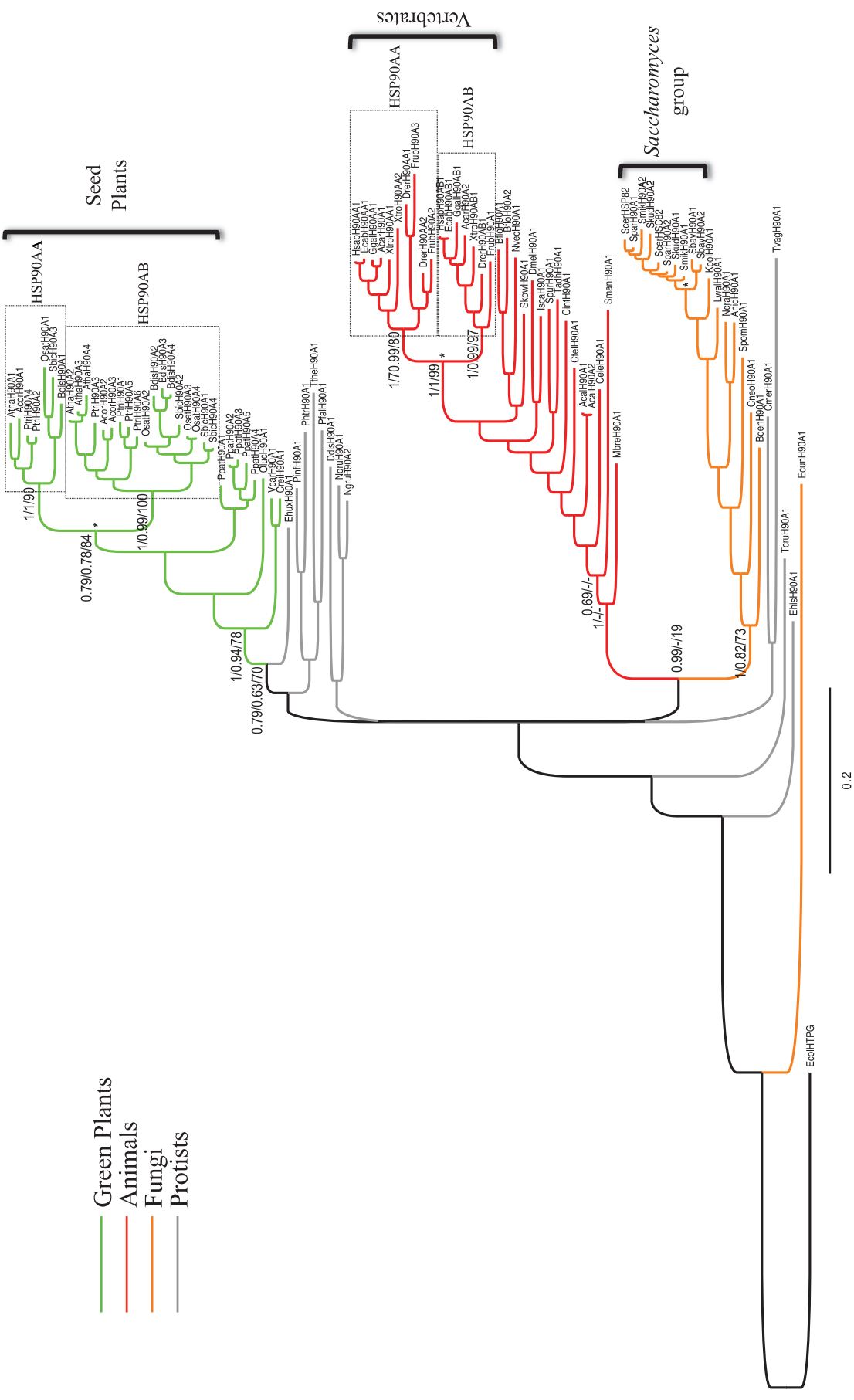

Fig. 1. Bayesian phylogeny of eukaryotic HSP90A. The BA tree depicts the evolutionary relationships among 94 HSP90A sequences from species representing distinct eukaryotic lineages, plus their homolog from *Escherichia coli*, HTPG, used as an outgroup. Statistical support corresponding to BA posterior probability, ML aLRT and bootstrap NJ values, respectively, are shown next to the corresponding nodes at relevant clades. Branch lengths in the tree are proportional to evolutionary distances between nodes, the scale bar indicating the number of inferred amino acid substitutions per site. The nodes defining the three independent duplications leading to subfamilies HSP90AA and HSP90AB in seed plant, vertebrate, and yeast HSP90A families, respectively, are indicated with asterisks. Clades clustering seed plant and vertebrate HSP90AA and HSP90AB subfamilies are enclosed within dashed boxes and were used as foreground branches in the molecular evolutionary and FD analyses.

Table 1. Parameter Estimates, Ln *L* Values and LRTs of Codon-Substitution Evolutionary Models.

Foreground Branch	Model	*P*[a]	Parameter Estimates (Frequency, *f* and ω Values)					Ln *L*	BEB[b]
Vertebrate HSP90AA	One-ratio model 0 ($\omega_0 = \omega_1$)	34	$\omega_0 = \omega_1 = 0.0222$					−16,366.77	NA
	Two-ratio model 2 (ω_0, ω_1)	35	$\omega_0 = 0.0212$; $\omega_1 = 999.0000$					−16,343.49	NA
	Model 3 (discrete)	38	—					−15,977.53	NA
	Clade model D ($K = 3$)	39	Site class	0	1	2		−15,944.42	NA
			f	0.761	0.021	0.217			
			ω_0	0.0025	0.433	0.069			
			ω_1	0.0025	0.433	2.891			
	Model Anull ($\omega_2 = 1$)	36	1					−16,173.79	NA
	Model A ($0 < \omega_0 < 1$)	37	Site class	0	1	2a	2b	−16,160.56	74, 78,* 281,* 426, 450,* 479,* 485,* 500,* 530, 572,* 575, 640*
			f	0.924	0.018	0.056	0.001		
			ω_0	0.017	1.000	0.017	1.000		
			ω_1	0.017	1.000	32.37	32.37		

LRT for Assymmetric Sequence Evolution: One-ratio model 0 ($\omega_0 = \omega_1$) vs. two-ratio model 2 (ω_0, ω_1) 2ΔLn *L* = 46.576; df = 1; *P* = 8.81E−12.

LRT for Divergent Selection: Model 3 (discrete) vs. Clade model D ($K = 3$) 2ΔLn *L* = 66.217; df = 1; *P* = 4.04E−16.

LRT for PS: Model Anull ($\omega_2 = 1$) vs. Model A ($0 < \omega_0 < 1$) 2ΔLn *L* = 26.457; df = 1; *P* = 2.69E−07.

Vertebrate HSP90AB	One-ratio model 0 ($\omega_0 = \omega_1$)	34	$\omega_0 = \omega_1 = 0.0222$					−16,366.77	NA
	Two-ratio model 2 (ω_0, ω_1)	35	$\omega_0 = 0.0212$; $\omega_1 = 999.0000$					−16,343.28	NA
	Model 3 (discrete)	38	—					−15,977.53	NA
	Clade model D ($K = 3$)	39	Site class	0	1	2		−15,977.53	NA
			f	0.761	0.022	0.217			
			ω_0	0.0025	0.433	0.069			
			ω_1	0.0025	0.433	999.0			
	Model Anull ($\omega_2 = 1$)	36	1					−16,173.79	NA
	Model A ($0 < \omega_0 < 1$)	37	Site class	0	1	2a	2b	−16,160.55	78, 479, 572
			f	0.924	0.018	0.056	0.001		
			ω_0	0.017	1.000	0.017	1.000		
			ω_1	0.017	1.000	999.0	999.0		

LRT for Assymmetric Sequence Evolution: One-ratio model 0 ($\omega_0 = \omega_1$) vs. two-ratio model 2 (ω_0, ω_1) 2ΔLn *L* = 46.983; df = 1; *P* = 7.16E−12.

LRT for Divergent Selection: Model 3 (discrete) vs. Clade model D ($K = 3$) 2ΔLn *L* = 0; df = 1; *P* = 1.

LRT for PS: Model Anull ($\omega_2 = 1$) vs Model A ($0 < \omega_0 < 1$) 2ΔLn *L* = 26.477; df = 1; *P* = 2.67E−07.

Seed Plant HSP90AA	One-ratio model 0 ($\omega_0 = \omega_1$)	50	$\omega_0 = \omega_1 = 0.0221$					−17,473.16	NA
	Two-ratio model 2 (ω_0, ω_1)	51	$\omega_0 = 0.0212$; $\omega_1 = 999.0000$					−17,453.07	NA
	Model 3 (discrete)	54	—					−17,039.52	NA
	Clade model D ($K = 3$)	55	Site class	0	1	2		−17,026.66	NA
			f	0.744	0.0434	0.213			
			ω_0	0.0023	0.230	0.064			
			ω_1	0.0023	0.230	999.0			
	Model Anull ($\omega_2 = 1$)	52	1					−17,369.96	NA
	Model A ($0 < \omega_0 < 1$)	53	Site class	0	1	2a	2b	−17,363.85	
			f	0.899	0.0149	0.0842	0.0014		77, 121, 231,* 560
			ω_0	0.0185	1.000	0.018	1.000		
			ω_1	0.0185	1.000	999.0	999.0		

LRT for Assymmetric Sequence Evolution: One-ratio model 0 ($\omega_0 = \omega_1$) vs. two-ratio model 2 (ω_0, ω_1) 2ΔLn *L* = 40.184; df = 1; *P* = 2.31E−10.

LRT for Divergent Selection: Model 3 (discrete) vs. Clade model D ($K = 3$) 2ΔLn L = 25.713; df = 1; *P* = 3.96E−07.

LRT for PS: Model Anull ($\omega_2 = 1$) vs. Model A ($0 < \omega_0 < 1$) 2ΔLn L = 12.228; df = 1; *P* = 4.71E−04.

Seed Plant HSP90AB	One-ratio model 0 ($\omega_0 = \omega_1$)	50	$\omega_0 = \omega_1 = 0.0221$					−17,473.16	NA
	Two-ratio model 2 (ω_0, ω_1)	51	$\omega_0 = 0.0212$; $\omega_1 = 999.0000$					−17,453.07	NA
	Model 3 (discrete)	54	—					−17,039.52	NA
	Clade model D ($K = 3$)	55	Site class	0	1	2		−17,028.97	NA
			f	0.743	0.043	0.214			
			ω_0	0.0023	0.231	0.064			
			ω_1	0.0025	0.433	2.89			

(continued)

Table 1. Continued

Foreground Branch	Model	P^a	Parameter Estimates (Frequency, f and ω Values)					Ln L	BEB^b
	Model Anull ($\omega_2 = 1$)	52	1					−17,369.96	NA
	Model A ($0 < \omega_0 < 1$)	53	Site class	0	1	2a	2b	−17,363.85	231
			f	0.899	0.0149	0.0842	0.0014		
			ω_0	0.0185	1.000	0.018	1.000		
			ω_1	0.0185	1.000	999.0	999.0		

LRT for Assymmetric Sequence Evolution: One-ratio model 0 ($\omega_0 = \omega_1$) vs. two-ratio model 2 (ω_0, ω_1) 2ΔLn $L = 40.184$; df = 1; $P = 2.31E{-}10$.

LRT for Divergent Selection: Model 3 (discrete) vs. Clade model D ($K = 3$) 2ΔLn $L = 21.082$; df = 1; $P = 4.40E{-}06$.

LRT for PS: Model Anull ($\omega_2 = 1$) vs. Model A ($0 < \omega_0 < 1$) 2ΔLn $L = 12.228$; df = 1; $P = 4.71E{-}04$.

NOTE.—Branches used as foreground branches (ω_1) are indicated. Numbering refers to the positions in the alignments of protein sequences from supplementary figures S1 and S2, Supplementary Material online, for plants and animals, respectively. Amino acid also identified in the analysis of FD is indicated with an underline.

[a]Number of parameters in the ω distribution.

[b]Amino acids detected in the BEB analysis as fixed by PS with posterior probabilities >95% (>99%, indicated with asterisks) are shown.

vertebrate and seed plant HSP90A, respectively, mapped to ordered regions of the protein forming alpha-helix or beta-sheet secondary structures (fig. 2*A* and *C*; supplementary table S3, Supplementary Material online).

We examined FD after HSP90A duplications using type I (Gu 1999) and type II FD (Gu 2006). We found weak but significant FD I only in the lineage leading to seed plant HSP90AA and HSP90AB (table 2). Ten sites likely to be responsible for FD I were also identified (table 2; supplementary table S3 and fig. S2, Supplementary Material online). At these sites, amino acid residues were highly conserved in sequences from one of the subfamilies, but variable in the other one, likely reflecting a change in their functional roles between the two subfamilies. Some were apparently minor changes, as they involved amino acids with similar physicochemical features (e.g., 431, 549, and 632). Others involved radical amino acid changes, such as position 417, occupied by the highly conserved hydrophobic Phe in HSP90AB proteins but by polar hydrophilic amino acids, such as Asn and Ser, in HSP90AA proteins; or positions 563 and 634.

Mapping of the amino acid changes under FD I in seed-plant HSP90AA and HSP90AB onto the 3D structure of yeast HSP82 (Ali et al. 2006) (figs. 2*C* and *D*), located 8 of the 10 amino acids in alpha-helical secondary structures, revealing a significant structural preference (203 out of 709; Fisher exact test: $P = 0.041$). Moreover, 9 out of 10 residues were found on the surface of the protein, offering significantly larger solvent accessible surface area (ASA) than average (fig. 2*D*) (average ASA value for FD I amino acids: 50.5% compared with the average ASA for the remaining amino acids of the protein: 28.3%; t test: $t = 2.5919$, $P = 0.009$).

The different patterns of DSP, PS, and FD I identified after gene duplication permit sketching two distinct evolutionary scenarios for functional specialization of vertebrate and seed-plant HSP90AA and HSP90AB paralogs. In vertebrates, HSP90A duplication relaxed constraints on one copy (HSP90AA, which shows strong DSP signal) that explored and improved secondary HSP90A subfunctions encoded in the ancestral preduplication gene, whereas HSP90AB improved ancestral main function (a model known as "escape from adaptive conflict" [Conant and Wolfe 2008]). In seed plants, after a first phase of ancestral function partitioning between both subfamilies under a DSP model, both genes may have gained new functions through the fixation of adaptive amino acid changes under PS, a scenario in agreement with the sub-neo-functionalization model (He and Zhang 2005).

We mapped amino acid sites identified as PS or under FD onto functionally significant residues, as predicted by Evolutionary Trace (supplementary file S1, Supplementary Material online) (Mihalek et al. 2006) from the structure of *S. cerevisiae* HSP82 (Ali et al. 2006). None was found among the predicted residues ranking at the top 25% as involved in ATP binding and hydrolysis (Panaretou et al. 1998), p23/Sba1 cochaperone binding (Ali et al. 2006), interface between dimeric subunits, or possible novel functional surfaces. However, some of the amino acids are located very close to important functional residues. For example, amino acids 74, 77, and 78 are close to amino acid 79 (<6 Å distance), which is involved in ATP binding (Panaretou et al. 1998). These amino acids are also close (<8 Å distance) to amino acids 95–97, located to the interface of the two HSP90A monomers (Ali et al. 2006). These short distances between PS sites and functional sites were not expected by chance (e.g., resampling 1,000 sites randomly from the protein structure provides average Euclidean distances larger than 34 Å). Therefore, while functional specialization would likely not have affected essential residues involved in core HSP90A functional and/or structural features, adaptive changes at structurally proximal sites to functional ones may have strong implications on the optimization of ATP binding and hydrolysis and binding of client proteins or cofactors. Noticeably, amino acid changes fixed by PS and those related to FD I are preferentially distributed on ordered regions over the surface of the protein, most of them mapping to the middle and COOH domains, which have been suggested to provide independent binding sites for activation of specific client proteins and for recognition of specific cochaperones controlling HSP90A activity (Young et al. 2001; Meyer et al. 2003, 2004; Roe et al. 2004; Harst et al. 2005; Ali et al. 2006). Interestingly, some of the sites under PS (74, 426, 572, and 640) during vertebrate HSP90A functional

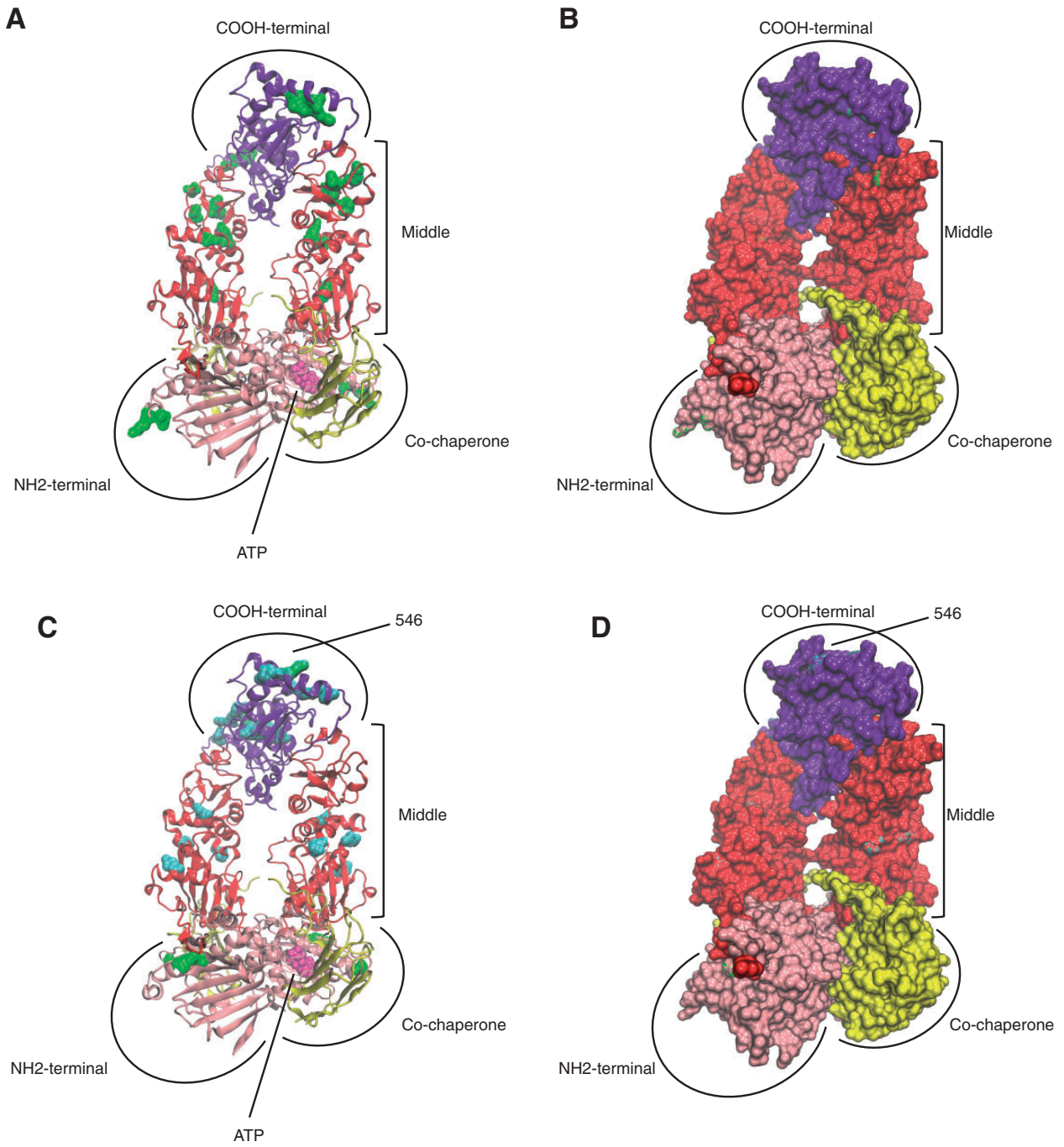

Fig. 2. 3D architecture of HSP90A/cochaperone complex showing PS and FD I sites. Cartoon backbone highlighting secondary structures (*A* and *C*) and molecular surface equivalents (*B* and *D*) representations of yeast HSP82 (Ali et al. 2006). PS residues identified in the analysis of functional diversification of vertebrate (*A* and *B*) and seed plant (*C* and *D*) HSP90AA and HSP90AB are shown in green. Residues identified as likely involved in FD of seed plant HSP90AA and HSP90AB are shown in cyan (*C* and *D*). Position 546, both related to FD and PS, is highlighted.

diversification or under PS (231) or FD (563) in seed plant HSP90As mediate HSP90A coevolution with its substrates or cofactors (Fares and Travers 2006; Travers and Fares 2007). In addition, site 231 belongs to the charged linker in HSP90A, which has been recently shown to mediate the interaction of HSP90A with its client proteins (Tsutsumi et al. 2012). It is tempting to speculate that the acquisition of new peptide-binding abilities, the modification of pre-existing ones, or their loss would drive adaptive functional shifts of HSP90A duplicates by increasing the number of key regulatory and signaling proteins dependent on its chaperone activity. Many other sites detected under PS or FD I had no defined role. However, their position in the structure close to sites for which a role has been defined in interacting with cofactors or clients support their possible compensatory effect of destabilizing functional mutations. Also, the close structural proximity of some of the amino acids under PS or FD I to one another (e.g., 417 is close in the structure to 426 and 431, which are close to 500, <8 Å distance) suggest their possible coordination in specific HSP90A functions.

Table 2. Analysis of FD.

FD	HSP90A Subfamilies	Coefficient θ ± SE (*P*)	Critical Amino Acid Sites
Type I	Vertebrate AA vs. AB	−0.0533 ± 0.0558	
	Seed Plant AA vs. AB	0.213 ± 0.0654 (1.27E−06)	303, 417,* 431, 446, 549, 560, 563,* 632, 634, 679
Type II	Vertebrate AA vs. AB	−0.998 ± 0.487	
	Seed Plant AA vs. AB	−1.115 ± 0.575	

NOTE.—Critical amino acid sites detected as related to FD with $P > 50\%$ (>70%, indicated with asterisks) are listed. Numbering refers to the positions in the alignments of protein sequences from supplementary figs. S1 and S2, Supplementary Material online, for seed plants and vertebrates, respectively. Amino acid also identified in the BEB analysis of PS is indicated with an underline.

A key role has been attributed to HSP90As in facilitating phenotypic diversity and rapid evolutionary change (Rutherford and Lindquist 1998; Queitsch et al. 2002). Moreover, the requirement of HSP90As for the function of key proteins controlling cell proliferation, including the tumor suppressor protein p53 (Hagn et al. 2011) and oncogenic protein kinases such as ErbB2, Cdk4, B-raf, and Akt/protein kinase B, makes them an attractive target for new cancer therapeutics (Ali et al. 2006). Our results identify residues involved in putative functional shifts during HSP90A evolution that can be used as targets for further studies to precisely determine their functional roles.

Materials and Methods

We used the *Homo sapiens* HsapHSP90AA1 sequence as a query to search for HSP90A sequences in full genomes of diverse eukaryote species. Phylogenetic analyses were performed on the basis of multiple alignments of HSP90A amino acid sequences obtained using the CLUSTALX 1.8 program (Thompson et al. 1997). Bayesian analysis (BA) was performed with MrBayes 3.1.2 (Huelsenbeck and Ronquist 2001), and a maximum likelihood (ML) tree was constructed using PhyML v3.0 (Guindon and Gascuel 2003). Neighbor Joining (NJ) phylogenetic analysis was conducted in MEGA 5.0 (Tamura et al. 2007). Resulting trees were represented and edited using FigTree v1.3.1.

Estimation of ω values during HSP90A evolution was performed by means of the codeml program from the PAML v4.4 package (Yang 1997) on the basis of codon sequence alignments. Two different classes of models were implemented: 1) "branch-specific" models, which permit heterogeneity in the ω ratio among branches in the phylogeny previously defined as "foreground" branches (Yang 1998) and 2) "branch-site" models, which allow ω to vary in foreground branches, but also featuring heterogeneity in selective pressures throughout the sequences by defining different codon site-classes with different ω ratios (Yang and Nielsen 2002). These models perform ML estimates of ω ratios and attach a log-likelihood (ln *L*) value to each examined alignment and tree topology. Likelihood ratio tests (LRTs) permit comparison of the fit of two nested models and compute a *P* value for the fitting of the examined data set to the alternative model being tested. To test for asymmetric sequence evolution, we compared the branch-specific one-ratio model 0, constraining all branches in the tree to the same ω ratio, to a two-ratio model 2, allowing a different ω to be estimated for the foreground branch. To measure DSP acting on a significant number of codon sites, a test comparing the site-specific null discrete model 3 (which allows the ω ratio to vary among sites, but holding ω constant among branches in the tree) and the branch-site model D, allowing a class of sites to be under DSP between foreground branches and the rest of the tree, was used (Bielawski and Yang 2004). Finally, the comparison between the branch-site model A, with ω fixed at one for the examined branch as null model, and the model A, featuring an extra class of sites under PS with $\omega > 1$ in foreground branches, was used as a conservative test to detect PS as opposed to relaxed purifying selection affecting a few sites in the selected branch (Zhang et al. 2005). Model A also implements a Bayes Empirical Bayes (BEB) procedure that calculates posterior probabilities of a codon being subjected to PS (Yang et al. 2005).

The analysis of FD between proteins of paralog clades was performed using DIVERGE v2.0 software (Gu and Vander Velden 2002). DIVERGE performs the ML calculation of the theta (θ) type I and type II coefficients of FD (FD I and FD II), based on the occurrence of altered selective constraints or radical shifts of physiochemical properties, respectively (Gu 1999, 2006). The program also estimates the posterior probabilities of amino acid sites to be responsible for FD.

The Evolutionary Trace server was used for prediction of the active sites and functional interfaces of *S. cerevisiae* HSP82 protein structure (PDB id: 2CG9) (Ali et al. 2006; Mihalek et al. 2006). Protein 3D-structural models were displayed and edited using VMD 1.9.1 (Humphrey et al. 1996). ASA or solvent accessibility of amino acids in the 2CG9 protein 3D-structure was calculated using ASAView (Ahmad et al. 2004).

Supplementary Material

Supplementary file S1, tables S1–S3, and figures S1 and S2 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

References

Ahmad S, Gromiha M, Fawareh H, Sarai A. 2004. ASAView: database and tool for solvent accessibility representation in proteins. *BMC Bioinformatics* 5:51.

Ali MM, Roe SM, Vaughan CK, Meyer P, Panaretou B, Piper PW, Prodromou C, Pearl LH. 2006. Crystal structure of an Hsp90-nucleotide-p23/Sba1 closed chaperone complex. *Nature* 440:1013–1017.

Bardwell JC, Craig EA. 1988. Ancient heat shock gene is dispensable. *J Bacteriol.* 170:2977–2983.

Bielawski JP, Yang Z. 2004. A maximum likelihood method for detecting functional divergence at individual codon sites, with application to gene family evolution. *J Mol Evol.* 59:121–132.

Blair JE, Hedges SB. 2005. Molecular phylogeny and divergence times of deuterostome animals. *Mol Biol Evol.* 22:2275–2284.

Borkovich KA, Farrelly FW, Finkelstein DB, Taulien J, Lindquist S. 1989. hsp82 is an essential protein that is required in higher concentrations for growth of cells at higher temperatures. *Mol Cell Biol.* 9: 3919–3930.

Chan-Schaminet KY, Baniwal SK, Bublak D, Nover L, Scharf KD. 2009. Specific interaction between tomato HsfA1 and HsfA2 creates hetero-oligomeric superactivator complexes for synergistic activation of heat stress gene expression. *J Biol Chem.* 284:20848–20857.

Chaw S-M, Chang C-C, Chen H-L, Li W-H. 2004. Dating the monocot-dicot divergence and the origin of core eudicots using whole chloroplast genomes. *J Mol Evol.* 58:424–441.

Chen B, Zhong D, Monteiro A. 2006. Comparative genomics and evolution of the HSP90 family of genes across all kingdoms of organisms. *BMC Genomics* 7:156.

Conant GC, Wolfe KH. 2008. Turning a hobby into a job: how duplicated genes find new functions. *Nat Rev Genet.* 9:938–950.

Fares MA, Travers SA. 2006. A novel method for detecting intramolecular coevolution: adding a further dimension to selective constraints analyses. *Genetics* 173:9–23.

Gu X. 1999. Statistical methods for testing functional divergence after gene duplication. *Mol Biol Evol.* 16:1664–1674.

Gu X. 2006. A simple statistical method for estimating type-II (cluster-specific) functional divergence of protein sequences. *Mol Biol Evol.* 23:1937–1945.

Gu X, Vander Velden K. 2002. DIVERGE: phylogeny-based analysis for functional-structural divergence of a protein family. *Bioinformatics* 18:500–501.

Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol.* 52: 696–704.

Hagn F, Lagleder S, Retzlaff M, Rohrberg J, Demmer O, Richter K, Buchner J, Kessler H. 2011. Structural analysis of the interaction between Hsp90 and the tumor suppressor protein p53. *Nat Struct Mol Biol.* 18:1086–1093.

Harst A, Lin H, Obermann WM. 2005. Aha1 competes with Hop, p50 and p23 for binding to the molecular chaperone Hsp90 and contributes to kinase and hormone receptor activation. *Biochem J.* 387: 789–796.

He X, Zhang J. 2005. Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. *Genetics* 169:1157–1164.

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

Humphrey W, Dalke A, Schulten K. 1996. VMD: visual molecular dynamics. *J Mol Graph.* 14:33–38, 27–28.

Jiao Y, Wickett NJ, Ayyampalayam S, et al. (17 co-authors). 2011. Ancestral polyploidy in seed plants and angiosperms. *Nature* 473: 97–100.

Kellis M, Birren BW, Lander ES. 2004. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* 428:617–624.

Krone PH, Sass JB. 1994. HSP 90 alpha and HSP 90 beta genes are present in the zebrafish and are differentially regulated in developing embryos. *Biochem Biophys Res Commun.* 204: 746–752.

Leach MD, Klipp E, Cowen LE, Brown AJ. 2012. Fungal Hsp90: a biological transistor that tunes cellular outputs to thermal inputs. *Nat Rev Microbiol.* 10:693–704.

McClellan AJ, Xia Y, Deutschbauer AM, Davis RW, Gerstein M, Frydman J. 2007. Diverse cellular functions of the Hsp90 molecular chaperone uncovered using systems approaches. *Cell* 131:121–135.

Meng X, Jerome V, Devin J, Baulieu EE, Catelli MG. 1993. Cloning of chicken hsp90 beta: the only vertebrate hsp90 insensitive to heat shock. *Biochem Biophys Res Commun.* 190: 630–636.

Meyer P, Prodromou C, Hu B, Vaughan C, Roe SM, Panaretou B, Piper PW, Pearl LH. 2003. Structural and functional analysis of the middle segment of hsp90: implications for ATP hydrolysis and client protein and cochaperone interactions. *Mol Cell.* 11: 647–658.

Meyer P, Prodromou C, Liao C, Hu B, Roe SM, Vaughan CK, Vlasic I, Panaretou B, Piper PW, Pearl LH. 2004. Structural basis for recruitment of the ATPase activator Aha1 to the Hsp90 chaperone machinery. *EMBO J.* 23:1402–1410.

Mihalek I, Res I, Lichtarge O. 2006. Evolutionary trace report_maker: a new type of service for comparative analysis of proteins. *Bioinformatics* 22:1656–1657.

Nadeau K, Das A, Walsh CT. 1993. Hsp90 chaperonins possess ATPase activity and bind heat shock transcription factors and peptidyl prolyl isomerases. *J Biol Chem.* 268:1479–1487.

Ohno S. 1970. Evolution by gene duplication. New York (NY): Springer-Verlag.

Panaretou B, Prodromou C, Roe SM, O'Brien R, Ladbury JE, Piper PW, Pearl LH. 1998. ATP binding and hydrolysis are essential to the function of the Hsp90 molecular chaperone in vivo. *EMBO J.* 17: 4829–4836.

Pepin K, Momose F, Ishida N, Nagata K. 2001. Molecular cloning of horse Hsp90 cDNA and its comparative analysis with other vertebrate Hsp90 sequences. *J Vet Med Sci.* 63:115–124.

Putnam NH, Butts T, Ferrier DE, et al. (37 co-authors). 2008. The amphioxus genome and the evolution of the chordate karyotype. *Nature* 453:1064–1071.

Queitsch C, Sangster TA, Lindquist S. 2002. Hsp90 as a capacitor of phenotypic variation. *Nature* 417:618–624.

Rensing SA, Lang D, Zimmer AD, et al. (71 co-authors). 2008. The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. *Science* 319:64–69.

Roe SM, Ali MM, Meyer P, Vaughan CK, Panaretou B, Piper PW, Prodromou C, Pearl LH. 2004. The mechanism of Hsp90 regulation by the protein kinase-specific cochaperone p50(cdc37). *Cell* 116: 87–98.

Rutherford SL, Lindquist S. 1998. Hsp90 as a capacitor for morphological evolution. *Nature* 396:336–342.

Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol.* 24:1596–1599.

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 Res.* 25:4876–4882.

Travers SA, Fares MA. 2007. Functional coevolutionary networks of the Hsp70-Hop-Hsp90 system revealed through computational analyses. *Mol Biol Evol.* 24:1032–1044.

Tsutsumi S, Mollapour M, Prodromou C, Lee CT, Panaretou B, Yoshida S, Mayer MP, Neckers LM. 2012. Charged linker sequence modulates eukaryotic heat shock protein 90 (Hsp90) chaperone activity. *Proc Natl Acad Sci U S A.* 109:2937–2942.

Wolfe KH, Shields DC. 1997. Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* 387:708–713.

Wu C. 1995. Heat shock transcription factors: structure and regulation. *Annu Rev Cell Dev Biol.* 11:441–469.

Yabe N, Takahashi T, Komeda Y. 1994. Analysis of tissue-specific expression of *Arabidopsis thaliana* HSP90-family gene HSP81. *Plant Cell Physiol.* 35:1207–1219.

Yamada K, Fukao Y, Hayashi M, Fukazawa M, Suzuki I, Nishimura M. 2007. Cytosolic HSP90 regulates the heat shock response that is responsible for heat acclimation in *Arabidopsis thaliana*. *J Biol Chem.* 282:37794–37804.

Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci.* 13:555–556.

Yang Z. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol Biol Evol.* 15: 568–573.

Yang Z, Nielsen R. 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol Biol Evol.* 19:908–917.

Yang Z, Wong WS, Nielsen R. 2005. Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol Biol Evol.* 22: 1107–1118.

Young JC, Moarefi I, Hartl FU. 2001. Hsp90: a specialized but essential protein-folding tool. *J Cell Biol.* 154:267–273.

Zhang J. 2003. Evolution by gene duplication: an update. *Trends Ecol Evol.* 18:292–298.

Zhang J, Nielsen R, Yang Z. 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol Biol Evol.* 22:2472–2479.

Zou J, Guo Y, Guettouche T, Smith DF, Voellmy R. 1998. Repression of heat shock transcription factor HSF1 activation by HSP90 (HSP90 complex) that forms a stress-sensitive complex with HSF1. *Cell* 94:471–480.