



OPEN

The complex evolution of the metazoan HSP70 gene family

Er-meng Yu^{1,2}, Tatsuki Yoshinaga³, Frank L. Jalufka¹, Hashimul Ehsan¹, David B. Mark Welch⁴✉ & Gen Kaneko¹✉

The metazoan 70-kDa heat shock protein (HSP70) family contains several members localized in different subcellular compartments. The cytosolic members have been classified into inducible HSP70s and constitutive heat shock cognates (HSC70s), but their distinction and evolutionary relationship remain unclear because of occasional reports of “constitutive HSP70s” and the lack of cross-phylum comparisons. Here we provide novel insights into the evolution of these important molecular chaperones. Phylogenetic analyses of 125 full-length HSP70s from a broad range of phyla revealed an ancient duplication that gave rise to two lineages from which all metazoan cytosolic HSP70s descend. One lineage (A) contains a relatively small number of genes from many invertebrate phyla, none of which have been shown to be constitutively expressed (i.e., either inducible or unknown). The other lineage (B) included both inducible and constitutive genes from diverse phyla. Species-specific duplications are present in both lineages, and Lineage B contains well-supported phylum-specific clades for Platyhelminthes, Rotifera, Nematoda, Porifera/Cnidaria, and Chordata. Some genes in Lineage B have likely independently acquired inducibility, which may explain the sporadic distribution of “HSP70” or “HSC70” in previous phylogenetic analyses. Consistent with the diversification history within each group, inducible members show lower purifying selection pressure compared to constitutive members. These results illustrate the evolutionary history of the HSP70 family, encouraging us to propose a new nomenclature: “HSP70 + subcellular localization + lineage + copy number in the organism + inducible or constitutive, if known.” e.g., HSP70cA1i for cytosolic Lineage A, copy 1, inducible.

The 70-kDa heat shock protein (HSP70) family members play important roles in various cellular processes including heat shock response, folding of newly synthesized proteins, protein transport, and protein degradation. These apparently diverse functions are attributed to their chaperone activity^{1,2}, by which they prevent the aggregation and misfolding of target proteins. Typically, HSP70 family members bind to denaturing or newly synthesized proteins by recognizing up to ten hydrophobic amino acid residues exposed to the protein surface because of the misfolding. The release of the HSP70 members, which is triggered by ATP hydrolysis, facilitates the proper folding of the target proteins³. The HSP70 family contains organelle-specific members localized in cytosol, endoplasmic reticulum (ER), mitochondria, and chloroplasts⁴, and these organelle-specific types not only perform chaperoning functions in the organelles but are also known to contribute to protein transport across organelle membranes.

HSP70 family members are often upregulated upon various stresses that disrupt protein folding, such as heat treatment, exposure to toxic materials, ultraviolet irradiation, and pathogen attack^{5,6}. The upregulation is primarily regulated at the transcription level by transcription factors called heat shock factors (HSFs), particularly HSF1. Under unstressed conditions, HSF1 remains in a monomeric state. Various stresses lead to the formation of HSF1 trimers, which in turn bind to heat shock elements (HSEs) in the promoter region of HSP70 family member genes and promote their transcription. The HSF1-dependent transactivation system is generally conserved among eukaryotes, although species-specific differences in HSF1 function have been reported

¹School of Arts and Sciences, University of Houston-Victoria, Victoria, TX, USA. ²Key Laboratory of Tropical and Subtropical Fishery Resource Application & Cultivation, Key Laboratory of Aquatic Animal Immune Technology of Guangdong Province, Pearl River Fisheries Research Institute of CAFS, Guangzhou, China. ³School of Marine Biosciences, Kitasato University, Kanagawa, Japan. ⁴Josephine Bay Paul Center for Comparative Molecular Biology and Evolution, Marine Biological Laboratory, Woods Hole, MA, USA. ✉email: dmarkwelch@mbi.edu; kanekog@uhv.edu

in many organisms such as the fruit fly *Drosophila melanogaster*^{7,8} and yeast *Saccharomyces cerevisiae*⁹. The expression of HSP70 family member is also regulated by many other mechanisms such as the unfolded protein response, chromatin modification, and other transcription factors depending on the type of stress and subcellular localization^{10,11}. HSP70 family member genes have therefore been used as biomarkers of environmental stresses both in laboratory and field experiments^{12–15}.

On the other hand, several cytosolic HSP70 family members show constitutive expression patterns. These proteins are involved in the folding of newly synthesized proteins and are traditionally called 70-kDa heat shock cognates (HSC70). However, the distinction between HSP70s and HSC70s remains unclear because of the occasional reports of “constitutive HSP70s,” especially from invertebrates^{16–18}. These proteins were named “HSP70” because of their higher sequence similarity to known HSP70s than to HSC70s, but subsequent expression analyses failed to demonstrate their stress inducibility, resulting in assignment of the apparently self-contradicting protein names. Similarly, “stress-inducible HSC70s” have been reported in several animals including shrimps¹⁹, snails²⁰, and fish²¹. This confusion is likely attributable to the poorly understood evolutionary relationship between metazoan HSP70s and HSC70s, which is due to the lack of cross-phylum comparisons. A previous phylogenetic analysis indicated that metazoan HSP70 family members can be classified into invertebrate HSP70s, vertebrate HSP70s, and HSC70s from both vertebrates and invertebrates²², but this study comprises only a few phyla. In-depth phylogenetic analysis of HSP70 family members with a broad sampling across phyla would provide further insight into the classification of this important group of molecular chaperones.

In this study, we sought to trace the evolutionary history of metazoan HSP70 family members, with a particular attention to their stress inducibility, using phylogenetic analyses that included Porifera, Cnidaria, Gnathifera, Lophotrochozoa, Ecdysozoa, and Chordata. The specific hypothesis was that this broad sampling would provide essential information necessary to understand the evolutionary relationship between metazoan HSP70s and HSC70s. Our results add novel groups of stress-inducible HSP70, providing important insight into the evolutionary history of the metazoan HSP70 family.

Results

Phylogenetic analysis of HSP70 family members. We assembled a set of 125 HSP70 family members, including members associated with the mitochondria and ER, and many sequences known only from automated genome annotation as “HSP70” (Supplementary Table S1). Taxa were chosen to represent diverse phyla and included representatives of Porifera (1 species); Cnidaria (3 species from 1 class); Platyhelminthes (4 species from 3 classes); Gnathifera (6 species from two classes of rotifers); Lophotrochozoa (4 species from two classes of annelids, 1 brachiopod, and 6 species from 3 classes of molluscs); Ecdysozoa (6 species from 4 classes of arthropods, 6 species from 2 classes of nematodes, and 1 priapulid); and Chordata (9 species from 3 classes), with sequences from the yeast *Saccharomyces cerevisiae*. When possible, we chose reference or commonly recognized taxa (e.g., *Drosophila melanogaster*, *Caenorhabditis elegans*, *Nematostella vectensis*), and taxa with multiple HSP70 family members; we also included all 8 human Heat Shock Protein Family A members for reference (HSPA1A and HSPA1B are identical at the amino acid level).

To assess the robustness of the results we used two different alignment methods (T-Coffee and M-Coffee²³; M-Coffee itself combines multiple alignment algorithms) to construct consensus trees with maximum-likelihood (RAxML) and Bayesian (MrBayes) approaches. The four trees were in general agreement for major nodes; Fig. 1 shows the Bayesian consensus tree of the M-Coffee alignment. In all analyses HSP70 family members known to be mitochondria- and ER-specific formed two monophyletic groups (nodes 13 and 14 in Fig. 1) in agreement with previous reports^{4,24,25}. All family members known to be cytosolic formed a third group (node 1; clades 1–6 in Fig. 1), with yeast cytosolic HSP70s (clade 6) basal to all metazoan cytosolic HSP70s. The tree topology generally supported the species phylogeny. Within the metazoan cytosolic HSP70s (node 2; clades 1–5) there were two large clades (called lineages A and B): Lineage A was well supported and contained of genes from arthropods, annelids, brachiopods, rotifers, molluscs, priapulids, and cnidarians (node 3; clade 1); Lineage B was more poorly resolved and represented by many diverse vertebrate and invertebrate phyla (node 4; clades 2–5). There are two groups of vertebrate cytosolic HSP70s (nodes 10 and 11), which form a monophyletic group. Unexpectedly, we found three additional clusters of cytosolic HSP70 family composed of HSP70 genes from the phylum Platyhelminthes (node 5, clade 2), Rotifera (node 6; clade 3) and Nematoda (node 7; clade 4). This suggests the evolution of three unique subgroups of HSP70, distinct from conventional vertebrate and invertebrate HSP70s. Within the clade 5, Porifera and Cnidaria HSP70s were basal to other HSP70 family members, indicating that this clade has been conserved throughout metazoan evolution.

Figure 1 reveals that in many cases a species with multiple HSP70 family members has two closely related paralogs that are not informative to the deep history of the gene family. We therefore also constructed phylogenetic trees using a reduced set of 48 sequences, eliminating one of a paralog pair unless their inducibility had been previously determined (Supplementary Figs. S1 and S2). The overall topology and clustering patterns were retained, including both of the two HSP70 lineages as well as most of the nodes described above, although several clades were poorly resolved. Thus the analyses leading to the tree shown in Fig. 1 are unlikely to have been adversely effected by the large number of sequences included.

The phylogenetic breadth of our analyses provided the opportunity to search for sequence motifs characteristic of different groups within the HSP70 family. We identified a single region near the N terminus (but after signal peptide motifs) that discriminates between metazoan mitochondrial, ER, and cytosolic HSP70s (Fig. 2), which enables better discrimination for organelle-specific HSP70s together with the well-known C-terminal motifs for cytosolic (EEVD) and ER (HDEL/KDEL) HSP70s (Supplementary Fig. S3). We also found the RARFEEL motif, previously described as possibly cytosolic-specific in conjunction with the EEVD motif^{20,26,27}, in several invertebrate genes that are found with high confidence in the ER-associated node 14 (Fig. 1; *B. ibericus* HSC70,

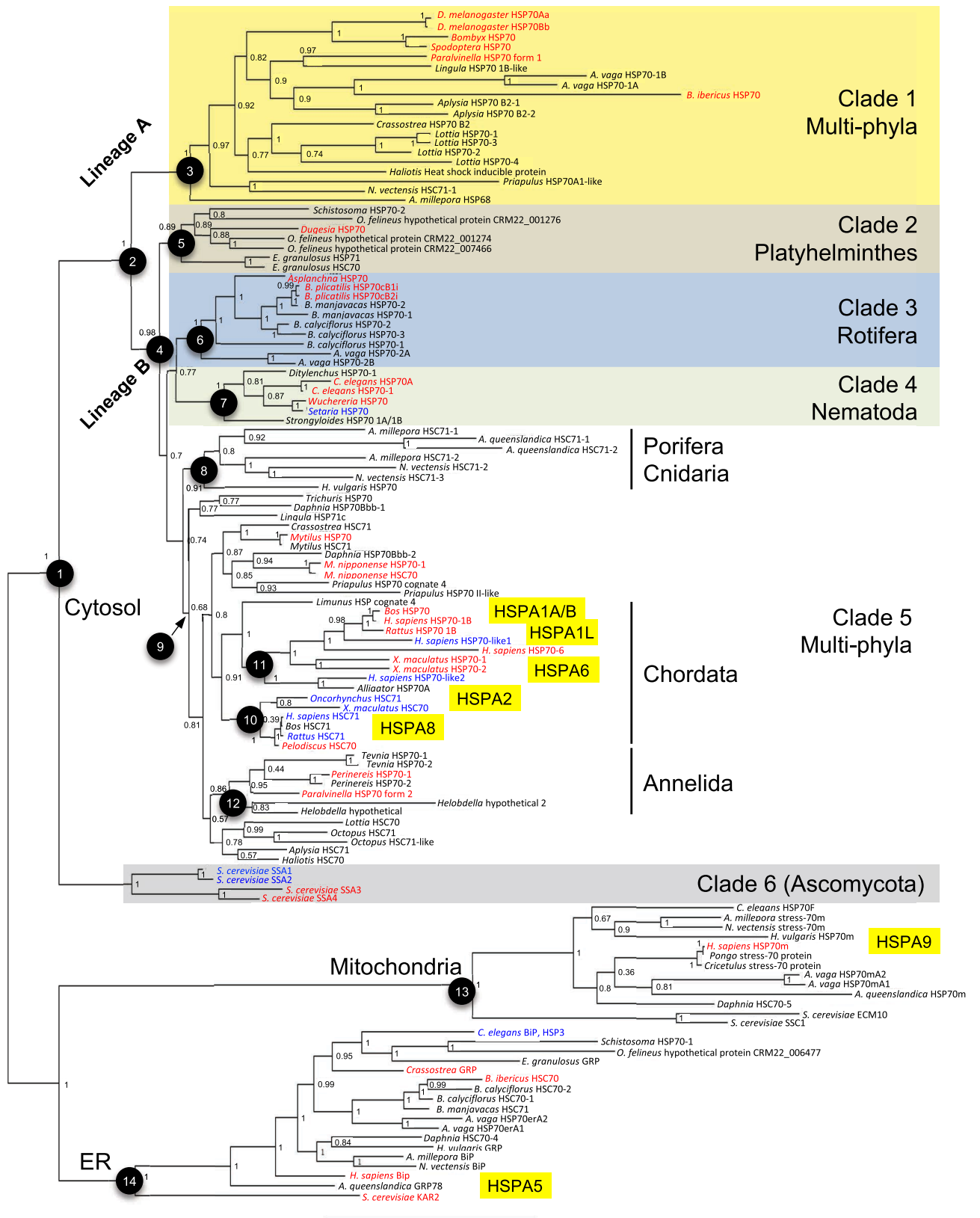
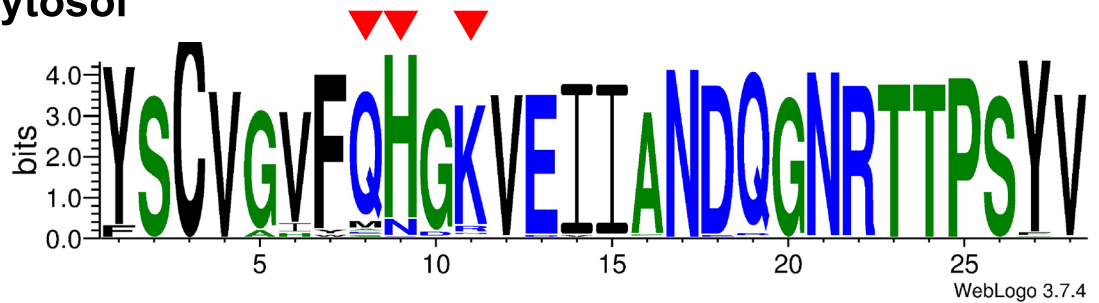
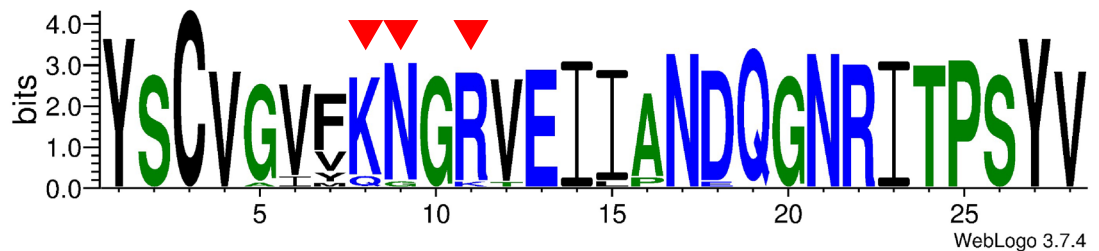


Figure 1. Bayesian consensus tree of HSP70 family members. Nodes discussed in the text are indicated. Numbers on the branches indicate the posterior probability support for each node. Stress-inducible and constitutive genes are shown in red and blue, respectively. No data on the stress inducibility is available for genes shown in black. Approved HGNC names of human HSP70 family members are shown in yellow boxes for reference. The DDBJ/EMBL/GenBank accession numbers and other information are summarized in Supplementary Table S1. Clades 1 to 6 include only cytosolic members of the HSP70 family. Human HSPA1A and HSPA1B have identical sequences at the amino acid level and are shown together.

Cytosol



Endoplasmic reticulum



Mitochondria

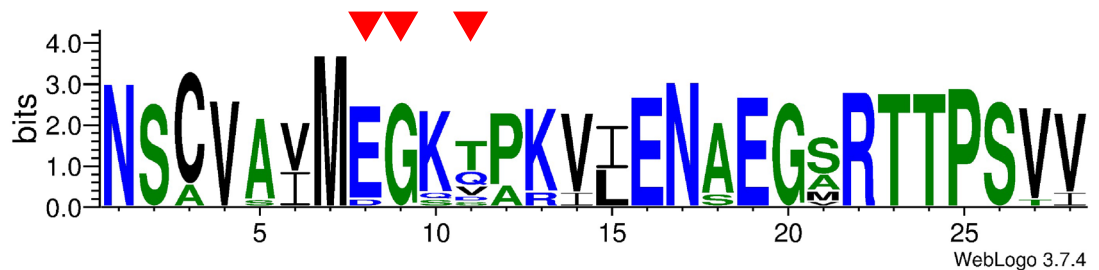


Figure 2. Logo visualization of cytosol-, endoplasmic reticulum (ER)-, and mitochondria-specific sequences near the N terminus of the HSP70 family members. All sequences in Fig. 1 were used to compute the sequence logos (93, 18, and 13 sequences for cytosol-, ER, and mitochondria-specific members, respectively). Compared to the cytosolic form, ER-specific forms have several conserved amino acid substitutions: Q8K, H9N, K11R, and T23I. Similarly, mitochondrial forms have Y1N, G5A, F7M, Q8E, H9G, G10K, K11T/Q/V, V12P/A, E13K/R, I14V, I15L, A16E, D17A, Q18E, and N21S/A/M. Residues that can be useful for discriminating organelle-specific members are shown with red arrows. Position 23 also discriminates ER-specific members from others.

C. gigas GRP, and *B. manjavacas* HSC71). This result suggests that RARFEEL is not a reliable predictor of cytosol association. Indeed, this motif was identified by a comparison of a small number sequences²⁸ and appears to have limited phylogenetic conservation.

Stress-inducible invertebrate HSP70s are reported to have an extra serine residue in their ATPase domain^{22,29}. In our analysis this insertion occurs only in Lineage A irrespective of the stress-inducibility, where it is present in all sequences except *B. ibericus* HSP70 and one of four *Lottia* HSP70s; in *Priapululus* HSP70-1 the serine has been replaced by alanine (Fig. 3). These results suggest that the extra serine residue is a characteristic of Lineage A and may be associated with the stress-inducibility of the ancestral HSP70, which will be tested by broader sampling focusing on this residue. The alignment result supports the idea that HSP70s from Nematoda and Rotifera evolved separately from those in Annelida, Mollusca, and Arthropoda, which have been generally recognized as “invertebrate HSP70s”.

Synonymous and nonsynonymous substitution rates. Lastly, we calculated the synonymous and nonsynonymous substitution rates of 34 representative cytosolic HSP70 member genes. We primarily selected genes with known expression patterns (stress-inducible or constitutive) from each clade, hypothesizing that constitutive HSP70 family members have been under stronger purifying selection compared to stress-inducible members. Because inducibility is only known for one member of Clade 2, we did not examine this group. The rotifer sequence *B. ibericus* HSP70 (GU574486) has an insertion at nucleotide position 403 and a nucleotide

<i>Dm</i> HSP70Aa	ILTIID----EGSLFEVRSTAGD		<i>Trichuris</i>	ILTI---E-DG-IFEVKSTAGD	
<i>Dm</i> HSP70Bb	ILTIID----EGSLFEVRSTAGD		<i>Daphnia</i> _Bb1	MLTI---E-EG-IFEVKSTAGD	
<i>Bombyx</i>	ILTIID----EGSLFEVKSTAGD		<i>Lingula</i> _71c	ILTI---E-DG-IFEVKSTSGD	
<i>Spodoptera</i>	ILTIID----EGSLFEVRATAGD		<i>Crassostrea</i> C	ILTI---E-DG-IFEVKSTSGD	
<i>Paralvinella</i>	ILSIA----DGS L FEVKSTAGD		<i>Mytilus</i> _P	ILTI---K-DG-IFEVKSTSGD	
<i>Lingula</i> _1B	ILSIA----DGS L FEVRSTAGD		<i>Mytilus</i> _C	ILTI---E-DG-IFEVKSTSGD	
<i>Avaga</i> -1B	VLTISGGDVSGSVFEVKATAGD		<i>Daphnia</i> _Bb2	ILTI---E-DG-IFEVKSTAGD	
<i>Avaga</i> -1A	ILTISGGDAGGSVFEVRSTAGD		<i>Mn</i> _HSP70-1	ILTI---E-DD-IFEVKSTAGD	
<i>Bibericus</i>	VLSINDG-----LFEVLSTNGN		<i>Mn</i> _HSC70	ILTI---E-DG-IFEVKSTAGD	
<i>Aplysia</i> _B2-1	VLAIDEGS----IFEVKATAGD	3	<i>Priapulus</i> _C	ILTI---E-DG-IFEVKSTAGD	
<i>Aplysia</i> _B2-2	VLTIDEGS----MFEVKATAGD		<i>Priapulus</i> _P2	ILSI---E-EG-IFEVKSTAGD	
<i>Crassostrea</i> P	ILTIID----EGSIFEVRSTAGD		<i>Limulus</i> _C4	ILTI---E-DG-IFEVKSTAGD	
<i>Lottia</i> -1	ILTIID----EGSLFEVKSTAGD		<i>Bos</i> _HSP70	ILTI---D-DG-IFEVKATAGD	
<i>Lottia</i> -3	ILTIID----EGSLFEVKSTAGD		<i>Hs</i> _HSPA1AB	ILTI---D-DG-IFEVKATAGD	
<i>Lottia</i> -2	ILAIID----EGSLFEVKSTAGD		<i>Rattus</i> _1B	ILTI---D-DG-IFEVKATAGD	
<i>Lottia</i> -4	ILAIIDEG-----VFEVLATAGD		<i>Hs</i> _HSPA1L	ILTI---D-DG-IFEVKATAGD	
<i>Haliotis</i>	VLTID----EGSMFEVRSTAGD		<i>Hs</i> _HSPA6	VLSI---D-AG-VFEVKATAGD	
<i>Priapulus</i> _A1	VLTID----EGAMFEVRSTAGD		<i>Xm</i> _HSP70-1	ILTI---E-DG-IFEVKATAGD	
<i>Nvectensis</i> _1	ILTIID----DGS L FEVKSTAGD		<i>Xm</i> _HSP70-2	VLTI---E-DG-IFEVKATAGD	
<i>Am</i> _HSP68	VLTID----EGSLFEVRSTAGD		<i>Hs</i> _HSPA2	ILTI---E-DG-IFEVKSTAGD	9
<i>Schisto</i> _2	ILTI---E-DG-IFEVKSTAGD		<i>Alligator</i>	ILTI---E-DG-IFEVKSTAGD	
<i>Of</i> _001276	VLAI---E-EG-IFEVKSTAGD		<i>Oncorhynchus</i>	ILTI---E-DG-IFEVKSTAGD	
<i>Dugesia</i>	ILTI---E-DG-IFEVKSTAGD	5	<i>Xm</i> _HSC70	ILTI---E-DG-IFEVKATAGD	
<i>Of</i> _001274	ILTI---E-EG-IFEVKSTSGD		<i>Hs</i> _HSPA8	ILTI---E-DG-IFEVKSTAGD	
<i>Of</i> _007466	ILTI---D-DG-VFEVKSTAGD		<i>Bos</i> _HSC71	ILTI---E-DG-IFEVKSTAGD	
<i>Eg</i> _HSP70	ILSI---E-DG-IFEVKSTAGD		<i>Rattus</i> _HSC71	ILTI---E-DG-IFEVKSTAGD	
<i>Eg</i> _HSC70	ILSI---E-DG-IFEVKSTAGD		<i>Pelodiscus</i>	ILTI---E-DG-IFEVKSTAGD	
<i>Asplanchna</i>	ILTI---E-EG-IFEVKSTAGD		<i>Tevnia</i>	ILTI---E-DG-IFEVKSTAGD	
<i>Bp</i> _HSP70cB1i	ILTI---E-EG-IFEVKSTAGD		<i>Tevnia</i> 2	ILTI---E-DG-IFEVKSTAGD	
<i>Bp</i> _HSP70cB2i	ILTI---E-EG-IFEVKSTAGD		<i>Perinereis</i> 1	ILTI---E-DG-IFEVKSTAGD	
<i>Bm</i> HSP70-2	ILTI---E-EG-IFEVKSTAGD		<i>Perinereis</i> 2	ILTI---E-DG-IFEVKSTAGD	
<i>Bm</i> HSP70-1	ILTI---E-DG-IFEVKATAGD	6	<i>Paralvinella</i>	VLTI---E-DG-IFEVKSTSGD	
<i>Bc</i> HSP70-2	ILTI---E-EG-IFEVKSTAGD		<i>Helobdella</i> 2	ILTL---V-EG-IFEVKSTSGD	
<i>Bc</i> HSP70-3	ILTI---E-EG-IFEVKSTAGD		<i>Helobdella</i>	ILTI---E-DG-IFEVKSTSGD	
<i>Bc</i> HSP70-1	ILAI---E-EG-IFEVKSTAGD		<i>Lottia</i> _HSC70	ILTI---E-DG-IFEVKSTAGD	
<i>Avaga</i> -2A	VLKI---E-EG-IFEVKSTAGD		<i>Octopus</i> _C1	TLTI---E-DG-IFEVKSTAGD	
<i>Avaga</i> -2B	ILTI---E-EG-IFEVKSTAGD		<i>Octopus</i> _C2	ILAI---E-DG-IFEVKSTAGD	
<i>Ditylenchus</i>	ILTI---E-DG-IFEVKATAGD		<i>Aplysia</i>	ILTI---E-DG-IFEVKSTAGD	
<i>Ce</i> _HSP70A	ILTI---E-DG-IFEVKSTAGD	7	<i>Haliotis</i> _c	ILTI---E-DG-IFEVKSTAGD	
<i>Ce</i> _HSP70-1	ILTI---E-DG-IFEVKSTAGD		<i>Sc</i> _SSA1	LLSI---E-DG-IFEVKATAGD	
<i>Wuchereria</i>	ILTI---E-DG-IFEVKSTAGD		<i>Sc</i> _SSA2	LLSI---E-DG-IFEVKATAGD	12
<i>Setaria</i>	ILTI---E-DG-IFEVKSTAGD		<i>Sc</i> _SSA3	LLSI---D-EG-VFEVKATAGD	
<i>Strongyl</i>	VLTI---E-DG-IFEVKSTAGD		<i>Sc</i> _SSA4	LLSI---D-EG-VFEVKATAGD	
<i>Am</i> _HSC71-1	ILSI---E-DG-IFEVKSTAGD				
<i>Aq</i> _HSC71-1	ILTI---E-DG-IFEVRSTSGN				
<i>Aq</i> _HSC71-2	ILAI---E-NG-IFEVKSIIGD				
<i>Am</i> _HSC71-2	VLTI---D-DG-IFEVKATRGD	8			
<i>Nv</i> _HSC71-2	VLSI---E-DG-IFEVKSTHGD				
<i>Nv</i> _HSC71-3	VLTI---E-DG-IFEVKSTAGD				
<i>Hv</i> _HSP70	ILAI---E-DG-IFEVKSTAGD				

Figure 3. The serine residue specific to node 3. Cytosolic HSP70 family members in Fig. 1 were aligned by M-coffee, and sequences around the node 3-specific serine residue are shown in the same order as Fig. 1. Nodes in the phylogenetic tree (Fig. 1) are indicated on the right margin, and serine residues specific to node 3 are shown in bold with highlights. Stress-inducible and non-inducible genes are shown in red and blue, respectively. No data on the stress inducibility is available for genes shown in black. The accession numbers and other information are summarized in Supplementary Table S1.

deletion at position 563, causing a frame shift. Deduced amino acid sequences encoded by nucleotides 403–563 were markedly different from those of other HSP70 family members because of the frame shift, which resulted in the long branch to this sequence in phylogenetic trees (Fig. 1). Because the frame shift greatly affects the synonymous and nonsynonymous substitution rates, we excluded this sequence.

For the selected 34 sequences representing cytosolic members of the metazoan HSP70 family, we calculated synonymous and nonsynonymous substitution rates (Ks and Ka) from the well conserved regions of the alignment (Supplementary Fig. S4) for all combinations of sequences (Fig. 4A). Synonymous substitution rates (Ks) were saturated in many combinations and only a limited number of combinations produced meaningful results (229 out of 465 combinations, Supplementary Fig. S5). The Ka/Ks values calculated for the 229 combinations were averaged for each clade (Fig. 4B). Overall, Ka/Ks values were markedly lower than 1, indicating that cytosolic HSP70 family member genes have been under purifying selection. The purifying selection pressure appears to have been less in Lineage A (clade 1) than in Lineage B (clades 3–5). The effect size between these lineages ranged from 0.19 to 1.64 (Fig. 4C). Since the effect sizes larger than 0.8 indicate large difference in multiple comparisons, these results support our hypothesis that constitutive HSP70 family members [clade 4, clade 5, and clade 5 (node 10)] have been under stronger purifying selection.

Within Lineage B, the selection pressure was strongest in Nematode HSP70s (clade 3), invertebrate HSC70/HSP70s (clade 5), and vertebrate HSC70s (clade 5, node 10). In particular, vertebrate HSC70s (clade 5, node 10) and HSP70s (clade 5, node 11) showed significant difference in the Ka/Ks with a large effect size of 0.86 despite their close phylogenetic relationship. These results also support the above hypothesis. To confirm how much this finding depends on the sampling, we conducted the same analysis with a larger sample size, including most cytosolic genes used in the phylogenetic analysis (77 genes), and obtained consistent results (Supplementary Figs. S6, S7).

Discussion

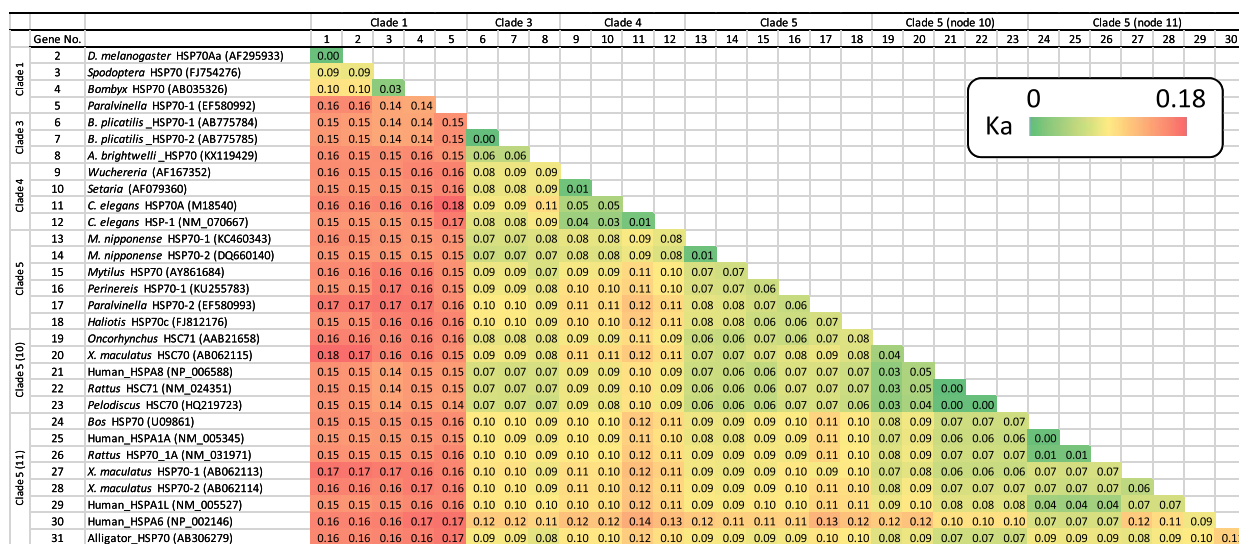
The monophyletic origin of HSC70s was proposed by Nikolaidis and Nei (2004) based on their analysis using *Drosophila* and nematode sequences²⁵. A subsequent phylogenetic analysis by Kourtidis et al. (2006) classified metazoan cytosolic HSP70s into three groups: invertebrate HSP70, vertebrate HSP70, and HSC70 including both vertebrate and invertebrate orthologues using data from multiple molluscs plus *D. melanogaster*, *C. elegans*, *Danio rerio*, and *H. sapiens*²². Surprisingly, to our knowledge, there has been no significant update on the evolution of metazoan HSP70 family after these small-scale studies despite the continuing expansion of available genomic data. Here we show that there are at least two types of ancestral cytosolic HSP70 family member genes in all analyzed phyla of metazoans, one (node 3; clade 1) giving birth to Lineage A of invertebrate HSP70s, and a second (node 4; clade 5) to Lineage B of both vertebrate and invertebrate HSP70s and all HSC70 genes. This second lineage has further diversified within diverse phyla: Platyhelminthes (node 5, clade 2), Rotifera (node 6; clade 3), Nematoda (node 7; clade 4), and Chordata (nodes 10 and 11; clade 5) (Fig. 1), which has not been identified in the previous reports. The tree topology is generally consistent with the species phylogeny, indicating that the effect of sequence contamination is very limited.

This newly identified evolutionary history of metazoan HSP70s demonstrates that stress inducibility has repeatedly evolved in many lineages of metazoans (i.e., convergent evolution). The multiple acquisition of stress inducibility has been observed in HSP70 family members duplicated within various taxa, such as those in a protist species and its close relative³⁰ and in mollusks²². To better understand this convergent evolution process, it is useful to discriminate between general stress inducibility and heat inducibility. For example, human mitochondrial HSP70 (mortalin/HSPA9) and ER-associated HSPA5 (BiP/GRP78) are not induced by heat but are induced by radiation³¹ and the unfolded protein response³², respectively. Comprehensive promoter analyses will add further mechanistic insights into the process by which HSP70 family members have acquired inducibility.

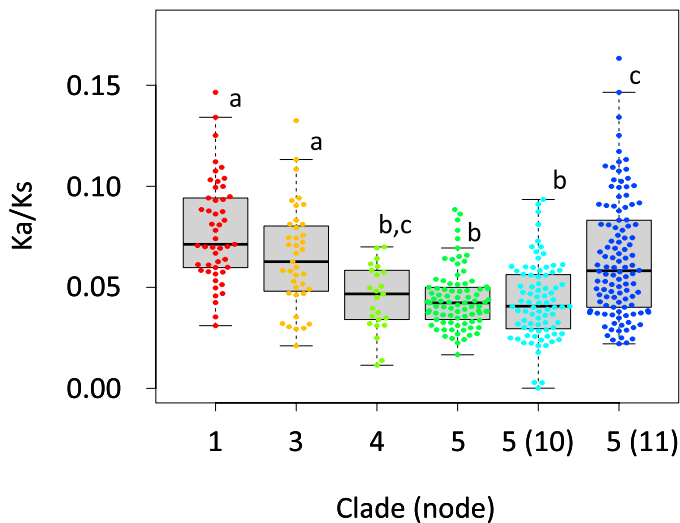
Another implication of our phylogenetic analysis is the nature of the ancestry HSP70 family members. Inter-clade comparison of the synonymous and nonsynonymous substitution rates shows that while all HSP70 family members are under strong purifying selection, the pressure to conserve amino acid sequence varies across clades. The presence of purifying selection has been reported for mammalian³³, nematode²⁵, molluscan²² and protist³⁰ HSP70 genes, and our multi-phylum analysis was consistent with these reports. Most dramatically, the canonical HSC70 genes of vertebrates (node 10; clade 5) showed stronger purifying selection pressure than the HSP70 family members in node 11 (clade 5). This concept can be further generalized by thoroughly investigating the saturation rate between each HSP70 gene³⁴. These results are in line with the concept that the role of HSC70 proteins, the folding of newly synthesized proteins, is more essential for cell survival compared to the role of the HSP70 cognates (node 11; clade 5). Indeed, cells deficient in the HSC70 gene are non-viable³⁵, whereas HSP70 knockout mice are mostly viable and fertile, although they are known to be sensitive to stress¹. The function of constitutive human HSPA8 is more diverse than other HSP70 family members and relates to many essential processes such as endocytosis and autophagy³⁶. From their relative importance it is possible to envision that constitutive expression may be ancestral in the metazoan cytosolic HSP70 family, but because of the presence of inducible and constitutive genes in many clades of metazoan cytosolic HSP70, it is difficult to predict the ancestral state with certainty. In addition, branch lengths between clades are not uniform, suggesting different evolutionary rates and gene conversion may occur between the paralogs found in many species. In this regard our data should be interpreted in a semiquantitative manner.

It is known that HSP70 paralogs in nematode, mammalian, and molluscan have experienced gene conversion events^{22,25,33}. In addition to purifying selection, gene conversion also likely contributes to the highly conserved nature of the HSP70 family members found in this study. The tree topology might have been affected by the gene conversion in closely related paralogs and orthologues such as human HSPA1A and HSPA1B³³. However, gene conversion generally functions to conserve the sequence in a species, and thus the overall topology of the

A



B



C

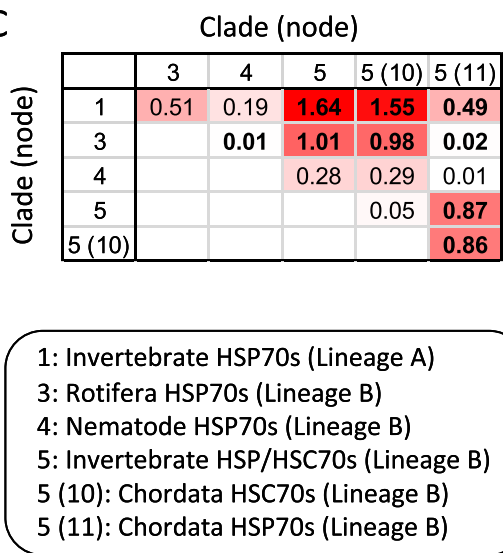


Figure 4. Selection pressure on cytosolic HSP70 family member genes. (A) Nonsynonymous substitution rates (Ka values) between 34 HSP70 family member genes. Gene number 1 is the fruit fly HSP70 Bb (AF295957) gene. The substitution rate of fruit fly HSP70 Aa and Bb gene was 00395526, which is shown as 0.00 in the figure. Accession numbers for other genes are indicated in the figure. Clades and nodes correspond to those in Fig. 2. (B) Bee swarm boxplots of Ka/Ks values for each clade. Only Ka/Ks values calculated from inter-cluster pairs were averaged. Statistical differences were calculated by Kruskal–Wallis test (chi-squared = 88.704, df = 5, P < 2.2e−16) followed by the non-parametric post-hoc tests (pairwise Wilcoxon test with P value adjustment by the Holm method). Clades sharing same letters are not significantly different at the 5% level of significance. One outlier in clade 3 (0.53) is not included in the plot (see supplementary material 1) although this value was used for all statistical analyses. (C) Effect sizes (Cohen’s d) for all comparisons. Combinations where significant differences were found in the pairwise Wilcoxon test are shown in bold.

metazoan tree is unlikely to be significantly affected. Gene conversion rate would need to be about 10 times higher than the mutation rate to significantly affect the topological distance according to a genomic simulation study³⁷, but the nucleotide identity between sequences is often lower than 70% in our phylogenetic analysis. Based on the largest-ever cross-phylum phylogenetic analysis of the HSP70 family, we propose a new nomenclature that reflects the phylogenetic relationship of family members rather than assumptions inferred about “cognate” or “constitutive” status (Fig. 5). The proposed name of an HSP70 family member protein consists of

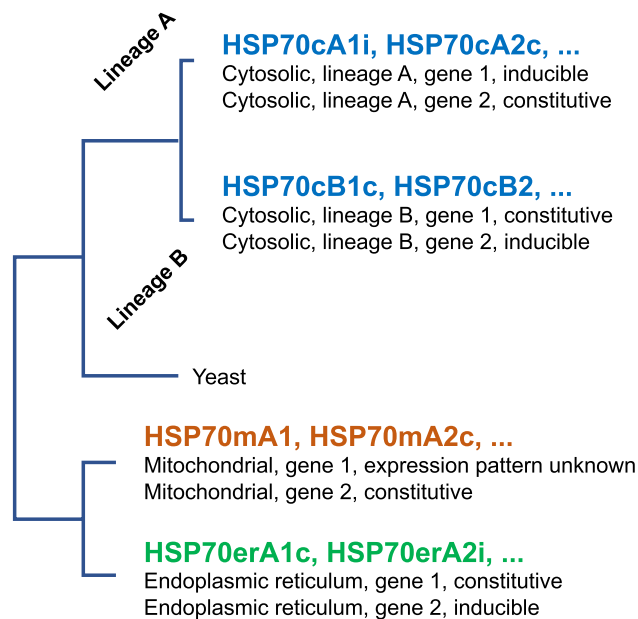


Figure 5. A proposed new nomenclature for HSP70 family members.

“HSP70 + subcellular localization + lineage + copy number found in the organism + inducible or constitutive, if known.” For example, HSP70cA1i represents a cytosolic HSP70 in Lineage A, copy 1, inducible; and HSP70mA1 for mitochondrial lineage of unknown inducibility. Although the current study found only one lineage in mitochondrial and ER members, the use of “A” in these organelle-specific isoforms would be beneficial. Systematic lineage names such as HSP70cD and HSP70mB can be assigned for novel lineages that may be discovered in future. Gene names are represented by lowercase italics of protein names (e.g., *hsp70cb2i* and *hsp70era1*).

In conclusion, the present study illustrated the evolutionary history of the metazoan HSP70 family, in which stress inducibility does not reflect evolutionary history. We also report the presence of novel signature sequences near the N terminus associated with cellular localization, and a previously unreported conserved serine residue. These findings, with the proposed nomenclature based on molecular evolution, will facilitate classification of new HSP70 family members and will help us understand the evolution and function of the HSP70 family.

Methods

Phylogenetic analysis. Metazoan HSP70 family genes were identified by BLAST searches to the NCBI nr database and to annotated and unannotated genome and transcriptome databases. Sequences were chosen to maximize phylogenetic diversity across Metazoa and where possible within phyla, and prioritized sequences where inducibility had been experimentally determined. Reference taxa and widely recognized taxa, and taxa with multiple annotated HSP70 family members were also prioritized. In some cases where there were multiple nearly-identical sequences in a species one representative sequence was chosen, although all human sequences were included for reference.

Alignments were generated with T-Coffee and M-Coffee using the tcoffee server at <http://tcoffee.org> with default parameters. The alignment revealed several sequences with extensive amino-terminal or carboxy-terminal extensions beyond the length of well characterized genes from model species. Because these are likely misannotated we excluded these regions from analysis and used the region of the alignment corresponding to human HSPA1A/HSPA1B positions 6-616.

Bayesian analyses were performed using MrBayes (v3.2.6) with two simultaneous runs of 4 chains each for 4×10^6 generations sampling every 100 generations. Evolutionary models were sampled during the run and for all runs with all alignments the WAG model was chosen with a posterior probability of 1. In all cases mcmc diagnostics found the average standard deviation of split frequencies between the runs to be < 0.05 , and ESS and PSRF convergence diagnostics indicated trees were sufficiently sampled. Consensus trees were generated by sampling the last 2×10^5 trees. Maximum likelihood analyses were performed using raxml-ng (v0.9.0) using the WAG model and combined ML tree search with 250 bootstrap trees, and default parameters. Trees were visualized with FigTree (v1.4.2).

Sequence analysis. Sequence logos were computed for the M-coffee alignments of cytosol-, ER, and mitochondria-specific members in Fig. 1 using WebLogo 3.7.4 (<http://weblogo.threeplusone.com>).

Synonymous and nonsynonymous substitution rates. R version 4.0.4 was used for the calculation of the synonymous and nonsynonymous substitution rates (the kaks function), Kruskal–Wallis test (`kruskal.test`

function), and post-hoc test (pairwise.wilcox.test function) on the Macintosh platform. Heat maps were created in Microsoft Excel 2016 (Redmond, MA).

Received: 29 April 2021; Accepted: 23 August 2021

Published online: 07 September 2021

References

1. Dugaard, M., Rohde, M. & Jäättelä, M. The heat shock protein 70 family: Highly homologous proteins with overlapping and distinct functions. *FEBS Lett.* **581**, 3702–3710 (2007).
2. Hartl, F. U., Bracher, A. & Hayer-Hartl, M. Molecular chaperones in protein folding and proteostasis. *Nature* **475**, 324–332 (2011).
3. Zuiderweg, E. R., Hightower, L. E. & Gestwicki, J. E. The remarkable multivalency of the Hsp70 chaperones. *Cell Stress Chaperones* **22**, 173–189 (2017).
4. Miemyk, J. A. The 70 kDa stress-related proteins as molecular chaperones. *Trends Plant Sci.* **2**, 180–187 (1997).
5. Sørensen, J. G., Kristensen, T. N. & Loeschcke, V. The evolutionary and ecological role of heat shock proteins. *Ecol. Lett.* **6**, 1025–1037 (2003).
6. Baird, N. A., Turnbull, D. W. & Johnson, E. A. Induction of the heat shock pathway during hypoxia requires regulation of heat shock factor by hypoxia-inducible factor-1. *J. Biol. Chem.* **281**, 38675–38681 (2006).
7. Jedlicka, P., Mortin, M. A. & Wu, C. Multiple functions of *Drosophila* heat shock transcription factor *in vivo*. *EMBO J.* **16**, 2452–2462 (1997).
8. Marchler, G. & Wu, C. Modulation of *Drosophila* heat shock transcription factor activity by the molecular chaperone DROJ1. *EMBO J.* **20**, 499–509 (2001).
9. Sorger, P. K. & Pelham, H. R. Yeast heat shock factor is an essential DNA-binding protein that exhibits temperature-dependent phosphorylation. *Cell* **54**, 855–864 (1988).
10. De Nadal, E., Ammerer, G. & Posas, F. Controlling gene expression in response to stress. *Nat. Rev. Genet.* **12**, 833–845 (2011).
11. Garbuz, D. Regulation of heat shock gene expression in response to stress. *Mol. Biol.* **51**, 352–367 (2017).
12. Sanders, B. M. Stress proteins in aquatic organisms—An environmental perspective. *Crit. Rev. Toxicol.* **23**, 49–75 (1993).
13. Ceyhan, S. B. *et al.* Deltamethrin attenuates antioxidant defense system and induces the expression of heat shock protein 70 in rainbow trout. *Comp. Biochem. Physiol. C Toxicol. Pharmacol.* **152**, 215–223 (2010).
14. Judge, M. L., Botton, M. L. & Hamilton, M. G. Physiological consequences of the supralittoral fringe: Microhabitat temperature profiles and stress protein levels in the tropical periwinkle *Cenchrithis muricatus* (Linnaeus, 1758). *Hydrobiologia* **675**, 143–156 (2011).
15. Kaneko, G., Kinoshita, S., Yoshinaga, T., Tsukamoto, K. & Watabe, S. Changes in expression patterns of stress protein genes during population growth of the rotifer *Brachionus plicatilis*. *Fish. Sci.* **68**, 1317–1323 (2002).
16. Jayasena, S., Chandrasekharan, N. & Karunanayake, E. H. Molecular characterisation of a hsp70 gene from the filarial parasite *Setaria digitata*. *Int. J. Parasitol.* **29**, 581–591 (1999).
17. Piano, A., Asirelli, C., Caselli, F. & Fabbri, E. Hsp70 expression in thermally stressed *Ostrea edulis*, a commercially important oyster in Europe. *Cell Stress Chaperones* **7**, 250 (2002).
18. Liu, L., Cheng, T.-Y. & Yang, Y. Cloning and expression pattern of a heat shock cognate protein 70 gene in ticks (*Haemaphysalis flava*). *Parasitol. Res.* **116**, 1695–1703 (2017).
19. Luan, W. *et al.* Identification of a novel inducible cytosolic Hsp70 gene in Chinese shrimp *Fenneropenaeus chinensis* and comparison of its expression with the cognate Hsc70 under different stresses. *Cell Stress Chaperones* **15**, 83–93 (2010).
20. Zheng, G., Dong, S., Hou, Y., Yang, K. & Yu, X. Molecular characteristics of HSC70 gene and its expression in the golden apple snails, *Pomacea canaliculata* (Mollusca: Gastropoda). *Aquaculture* **358–359**, 41–49 (2012).
21. Yabu, T. *et al.* Differential gene expression of HSC70/HSP70 in yellowtail cells in response to chaperone-mediated autophagy. *FEBS J.* **278**, 673–685 (2011).
22. Kourtidis, A. *et al.* Identification of several cytoplasmic HSP70 genes from the Mediterranean mussel (*Mytilus galloprovincialis*) and their long-term evolution in Mollusca and Metazoa. *J. Mol. Evol.* **62**, 446–459 (2006).
23. Wallace, I. M., O’Sullivan, O., Higgins, D. G. & Notredame, C. M-Coffee: Combining multiple sequence alignment methods with T-Coffee. *Nucleic Acids Res.* **34**, 1692–1699 (2006).
24. Boorstein, W. R., Ziegelhoffer, T. & Craig, E. A. Molecular evolution of the HSP70 multigene family. *J. Mol. Evol.* **38**, 1–17 (1994).
25. Nikolaidis, N. & Nei, M. Concerted and nonconcerted evolution of the Hsp70 gene superfamily in two sibling species of nematodes. *Mol. Biol. Evol.* **21**, 498–505 (2004).
26. Cottin, D. *et al.* Thermal biology of the deep-sea vent annelid *Paralvinella grasslei*: In vivo studies. *J. Exp. Biol.* **211**, 2196–2204 (2008).
27. Simoncelli, F., Morosi, L., Di Rosa, I., Pascolini, R. & Fagotti, A. Molecular characterization and expression of a heat-shock cognate 70 (Hsc70) and a heat-shock protein 70 (Hsp70) cDNAs in Rana (Pelophylax) lessonae embryos. *Comp. Biochem. Physiol. A Mol. Integr. Physiol.* **156**, 552–560 (2010).
28. Lo, W.-Y., Liu, K.-F., Liao, I.-C. & Song, Y.-L. Cloning and molecular characterization of heat shock cognate 70 from tiger shrimp (*Penaeus monodon*). *Cell Stress Chaperones* **9**, 332–343 (2004).
29. Garbuz, D. G. *et al.* Organization and evolution of hsp70 clusters strikingly differ in two species of Stratiomyidae (Diptera) inhabiting thermally contrasting environments. *BMC Evol. Biol.* **11**, 74 (2011).
30. Krenek, S., Schlegel, M. & Berendonk, T. U. Convergent evolution of heat-inducibility during subfunctionalization of the Hsp70 gene family. *BMC Evol. Biol.* **13**, 49 (2013).
31. Sadekova, S., Lehnert, S. & Chow, T. Induction of PBP74/mortalin/Grp75, a member of the hsp70 family, by low doses of ionizing radiation: A possible role in induced radioresistance. *Int. J. Radiat. Biol.* **72**, 653–660 (1997).
32. Wang, J., Lee, J., Liem, D. & Ping, P. HSPA5 Gene encoding Hsp70 chaperone BiP in the endoplasmic reticulum. *Gene* **618**, 14–23 (2017).
33. Hess, K. *et al.* Concurrent action of purifying selection and gene conversion results in extreme conservation of the major stress-inducible Hsp70 genes in mammals. *Sci. Rep.* **8**, 5082 (2018).
34. Pollock, D. D. & Larkin, J. C. Estimating the degree of saturation in mutant screens. *Genetics* **168**, 489–502 (2004).
35. Florin, L. *et al.* Nuclear translocation of papillomavirus minor capsid protein L2 requires Hsc70. *J. Virol.* **78**, 5546–5553 (2004).
36. Stricher, F., Macri, C., Ruff, M. & Muller, S. HSPA8/HSC70 chaperone protein: Structure, function, and chemical targeting. *Autophagy* **9**, 1937–1954 (2013).
37. Touchon, M. *et al.* Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLoS Genet.* **5**, e1000344 (2009).

Acknowledgements

We express our sincere thanks to Dr. Kristin Gribble, Marine Biological Laboratory, for helpful discussions and comments on the manuscript. The authors are grateful to Cynthia Rodriguez and Jamie Wilson, University of Houston-Victoria, for their help in gene expression analyses. GK was supported from M.G. and Lillie A. Johnson Foundation, Victoria, Texas. DMW was supported by the G. Unger Vetlesen Foundation.

Author contributions

Conceived and designed the experiments: T.Y., G.K. Performed experiments and data analyses: E.-m.Y., F.L.J., D.B.M.W., G.K. Contributed reagents/materials/analysis tools: H.E., D.B.M.W., G.K. Wrote the paper: D.B.M.W., G.K. Read and approved the final manuscript: E.-m.Y., T.Y., F.L.J., H.E., D.B.M.W., G.K.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-97192-9>.

Correspondence and requests for materials should be addressed to D.B.M.W. or G.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021