Gene
October 2016, Volume 591 Issue 1 Pages 97-107
https://doi.org/10.1016/j.gene.2016.06.061
https://archimer.ifremer.fr/doc/00382/49377/

Diversity of cytosolic HSP70 Heat Shock Protein from decapods and their phylogenetic placement within Arthropoda

Baringou Stephane [1, *], Rouault Jacques-Deric [2, 5], Koken Marcel [3], Hardivillier Yann [1], Hurtado Luis [4], Leignel Vincent [1]

[1] Univ Maine, Lab Mer Mol Sante, CNRS, FR IUML 3473, Ave Olivier Messiaen, F-72085 Le Mans, France.
[2] CNRS, Lab Evolut Genomes & Speciat, UPR9034, F-91198 Gif Sur Yvette, France.
[3] LABOCEA CNRS, F-29280 Plouzane, France.
[4] Texas A&M Univ, Dept Wildlife & Fisheries, Old Heep Bldg 110E, Tamu College Stn, TX 77843 USA.
[5] Univ Paris 11, F-91406 Orsay, France.

* Corresponding author : Stephane Baringou, email address : stephane.baringou@univ-lemans.fr ;

jacques-deric.rouault@orange.fr ; mhmkoken@gmail.com ; yann.hardivillier@univ-lemans.fr ; lhurtado@tamu.edu ; vincent.leignel@univ-lemans.fr

Abstract :

The 70 kDa heat shock proteins (HSP70) are considered the most conserved members of the HSP family. These proteins are primordial to the cell, because of their implications in many cellular pathways (e.g., development, immunity) and also because they minimize the effects of multiple stresses (e.g., temperature, pollutants, salinity, radiations). In the cytosol, two ubiquitous HSP70s with either a constitutive (HSC70) or an inducible (HSP70) expression pattern are found in all metazoan species, encoded by 5 or 6 genes (Drosophila melanogaster or yeast and human respectively). The cytosolic HSP70 protein family is considered a major actor in environmental adaptation, and widely used in ecology as an important biomarker of environmental stress. Nevertheless, the diversity of cytosolic HSP70 remains unclear amongst the Athropoda phylum, especially within decapods.

Using 122 new and 311 available sequences, we carried out analyses of the overall cytosolic HSP70 diversity in arthropods (with a focus on decapods) and inferred molecular phylogenies. Overall structural and phylogenetic analyses showed a surprisingly high diversity in cytosolic HSP70 and revealed the existence of several unrecognised groups. All crustacean HSP70 sequences present signature motifs and molecular weights characteristic of non-organellar HSP70, with multiple specific substitutions in the protein sequence. The cytosolic HSP70 family in arthropods appears to be constituted of at least three distinct groups (annotated as A, B and C), which comprise several subdivisions, including both constitutive and inducible forms. Group A is constituted by several classes of Arthropods, while group B and C seem to be specific to Malacostraca and Hexapoda/ Chelicerata, respectively. The HSP70 organization appeared much more complex than previously suggested, and far beyond a simple differentiation

according to their expression pattern (HSC70 versus HSP70). This study proposes a new classification of cytosolic HSP70 and an evolutionary model of the distinct forms amongst the Arthropoda phylum. The observed differences between HSP70 groups will probably have to be linked to distinct interactions with co-chaperones or other co-factors.

Highlights

► A large dataset of cytosolic HSP70 genes fragments was sequenced from decapods. ► Phylogenetic analyses reveal multiple clusters with structural specificities. ► Clusters of HSP70s found in this study include both consitutive and inducible forms. ► An evolutionary model of cytosolic HSP70 within the Arthropoda phylum is proposed.

Keywords : HSP70 family, Molecular evolution, Phylogeny, Arthropods, Crustacea

Abbreviations

HSP70	heat shock protein 70 kDa
HSC70	cognate heat shock protein 70 kDa
Cyt-HSP70	cytosolic heat shock protein 70 kDa
ER-HSP70	endoplasmic reticulum heat shock protein 70 kDa
HSP90	heat shock protein 90 kDa
NCBI	National Center of Biotechnology Information
NEF	nucleotide exchange factor
gDNA	genomic DNA
cDNA	complementary DNA
PCR	Polymerase Chain Reaction
ML	Maximum Likelihood
MP	Maximum Parsimony
NJ	Neighbour-Joining
NBD	Nucleotide-Binding Domain
SBD	Substrate-Binding Domain
CTD	C-Terminal Domain

1. Introduction

Heat Shock Proteins (HSPs) are known as major actors of the heat shock response (Ritossa, 1962; Tissiéres et al., 1974), and as proteins, generally involved in eukaryotic development and stress response (Craig and Schlesinger, 1985; Lindquist, 1986). Many HSPs types have been described and constitute a superfamily of proteins ranging from 15 to 110 kDa. Their functions include mainly the correct folding of newly synthesized or translocated proteins, denatured protein refolding, and aggregate solubilisation (Feder and Hofmann, 1999; Fink, 1999; Lindquist and Craig, 1988).

The 70 kDa Heat Shock Protein (HSP70) is considered one of the most ubiquitous and conserved proteins ever described (Gupta and Singh, 1994). Within eukaryotes, the HSP70

family includes distinct members, located in various cellular compartments (Craig et al., 1993; Karlin and Brocchieri, 1998). The cytosolic, endoplasmic reticulum (ER), mitochondrial and plastid HSP70s share identical signature patterns, but can be discriminated by distinct C-terminal motifs (Boorstein et al., 1994; Rensing and Maier, 1994). Even prokaryotic and eukaryotic HSP70s appear to have evolved from a common ancestor, and have at least 47 % amino acid sequence identity (Boorstein et al., 1994).

HSP70s are constituted of an N-terminal ATPase domain (Nucleotide Binding Domain, NBD, 44 kDa), a Substrate Binding Domain (SBD, 18 kDa) and a variable C-Terminal Domain (CTD, 10 kDa) (Kiang and Tsokos, 1998). The interactions of the SBD with the substrate are regulated by the NBD status: the HSP70 ATP hydrolysis activity converts the ATP-bound state to the ADP-bound state. In an ATP-bound state, the SBD substrate affinity is low, with a high association-dissociation rate. At the opposite, the ADP-bound or nucleotide-free state induces high substrate affinity and a low association-dissociation rate (Arakawa et al., 2010; Bukau and Horwich, 1998; Palleros et al., 1991; Swain et al., 2007). HSP70 chaperone activity is regulated through ATPase cycle modulations, by interactions with co-chaperones and Nucleotide Exchange Factors (NEFs) on specific sites, principally located in the NBD and the CTD (Demand et al., 1998; Mayer, 2010). Furthermore, it was recently demonstrated that a conserved fragment of 10-15 amino acids, named interdomain-linker (located between the NBD and the SBD), also appears to be involved in ATPase activity stimulation, chaperone complex stabilization and HSP70 oligomerization (Alderson et al., 2014; Aprile et al., 2013; Swain et al., 2007; Vogel et al., 2006).

Besides being involved in basic cellular processes such as correct protein folding, HSP70s are also implicated in cell proliferation and development, apoptosis, senescence and the immune response (Feder and Hofmann, 1999; Kiang and Tsokos, 1998). These fundamental functions led medical research to consider HSP70s as targets for cancer therapies (Meng et al., 2014;

Rodina et al., 2013; Schlecht et al., 2013; Sherman and Gabai, 2015) or for neurodegenerative disease (Abisambra et al., 2011; Fontaine et al., 2015; Lim and Yue, 2015; Manna, 2014). HSP70s are also considered as major ecological and physiological biomarkers for stress, owing to their involvement in adaptation and survival of organisms to varying environmental conditions (Bedulina et al., 2013; Colson-Proch et al., 2010; De Pomerai, 1996; Franzellitti and Fabbri, 2005; Kelley et al., 2011). Within decapods, over-expression of cytosolic HSP70 genes was observed facing various stressors, such as heat and cold shocks (Fu et al., 2013; Luan et al., 2010; Madeira et al., 2012; Qian et al., 2012; Yang et al., 2013), salinity and pH variations (Fu et al., 2013; Qian et al., 2012; Spees et al., 2002), metal exposure (Luan et al., 2010; Qian et al., 2012) and bacterial challenges (Cui et al., 2010; Xiu et al., 2014; Zhang et al., 2013; Zhou et al., 2010). Nevertheless, different expression profiles of cytosolic HSP70 members led to suspect multiple functions for different members of the subfamily. The cyt-HSP70 multigene family is well known in model organisms, including 6 homologues in human and yeast (Daugaard et al., 2007; Werner-Washburne and Craig, 1989) and 5 genes in *Drosophila melanogaster* (Bettencourt and Feder, 2001). Basal expression levels of each cyt-HSP70 form vary in different types of tissues (Daugaard et al., 2007) and present distinct sensitivities to heat stress leading to repression or induction of expression (Boorstein et al., 1994; Werner-Washburne and Craig, 1989). In this way, a distinction was made between the inducible HSP70, which in general presents very low basal levels and high stress-induction; and the constitutive HSC70 (heat shock cognate protein), evenly transcribed under normal or stress conditions (Ingolia and Craig, 1982; Krenek et al., 2013).

Within decapods, occasional HSC70 over-expression, however, was also described during environmental stresses (Lo et al., 2004; Qian et al., 2012), but at much lower induction levels (2-8 folds) than the inducible HSP70 (9-100 folds) (Luan et al., 2010). The complexity of HSC70 and HSP70 discrimination is reinforced by variable tissue-specific modality of

expression, depending on the stress (Liu et al., 2004; Lo et al., 2004; Xiu et al., 2014; Yamuna et al., 2000). Although physiological significance of these differences is not well understood, synthesis of HSP/HSC70 may be linked to cellular damages and tissues vulnerability facing to distinct stressors (e. g., temperature, contaminants). Diversity of cytosolic HSP70 was observed within several decapods species (Cottin et al., 2010; Leignel et al., 2007; Liu et al., 2004; Luan et al., 2010) and an especially high complexity was suspected in the brachyuran infra-order (Cui et al., 2010; Leignel et al., 2007). The links between different HSP70 subfamily members, and most importantly, their inducible or constitutive features remain a real point of discussion (Leignel et al., 2007; Liu et al., 2004; Mestre et al., 2015).

In this study, we characterize nucleotide sequences of cytosolic HSP70 from 23 brachyuran (16 families), 5 anomuran and 7 astacidean species to investigate the HSP70 diversity and phylogeny. Screening of the international databases allowed us to also perform several phylogenetic and aggregative analyses within decapods, and extend our classification to the whole arthropod phylum. Our analyses present a new hypothesis for the cytosolic-type HSP70 subfamily organization, based on aggregative methodology and multiple phylogenetical analyses, and propose an evolutionary model amongst the Arthropoda phylum. Biological relevance of the HSP/HSC70 classification is also discussed in view of our results.

2. Materials and Methods

2.1. *HSP70* data collection and sequence analyses

2.1.1. International database searches:

Nucleotide collection, Whole-genome shotgun contigs (WGS), Expressed sequence tags (EST) and Transcriptome Shotgun Assembly (TSA) available at NCBI were screened with the

BLASTn program. Due to the large quantity of matches and in order to increase the clearness of the results, one to four representative species among hexapod and chelicerate orders were retained for phylogenetic analyses within Arthropoda. Sequences used for expression analyses from the literature were preferentially selected. Several preliminary phylogenetic analyses were performed to remove redundant sequences from the same species and appearing grouped together.

2.1.2. Biological materials:

In this study HSP70 genes or cDNAs were isolated from a large panel of organisms. The crabs, *Carcinus maenas, Necora puber, Pachygrapsus marmoratus* and *Xantho pilipes* were collected in Noirmoutier (GPS: 47.026073, -2.250086) and Saint-Gildas-de-Rhuys (GPS: 47.488567, -2.819185) whereas the anomurans (*Galathea squamifera, Galathea strigosa, Munida rugosa* and *Pisidia longicornis*) originated from *Saint-Enogat* (GPS: 48.642000, -2.074227). The crabs, *Cancer antennarius* and *C. productus* were found at Bolinas beach in California (GPS: 37.907478, -122.681650). Several species were also bought in fish shops (Astacidea: *Astacus astacus, Enoplometopus debelius*, *Homarus americanus*, *Homarus gammarus, Nephrops norvegicus, Procambarus aleni* and *Procambarus clarkii;* Brachyura: *Cancer pagurus*, *Maia squinado*, and *Perisesarma bidens*). Hydrothermal crabs *Bythograea thermydron*, *Cyanagraea praedator* and *Segonzacia mesatlantica,* as well as *Chaceon affinis*, were collected during the oceanographic campaigns BIOBAZ (36°N, 2013) and HOPE (9 and 13°N, 1999). Fresh samples were frozen in liquid nitrogen and stored at -80°C for genomic DNA (gDNA) or RNA extractions. Samples stored in ethanol were only used for gDNA isolation.

2.1.3. Nucleic acids extractions:

Tissues were ground using a sterile pestle and mortar in liquid nitrogen. DNA was isolated from 0.1 to 2 g of tissue (whole animal, muscle or gill) using a standard phenol-chloroform extraction procedure and stored at 4°C to avoid frosting-defrosting damage). Total RNA was isolated from 10 to 50 mg of gill tissue with Tri-Reagent® solution (Ambion®) according to the manufacturer's instructions. 1 µg of each RNA sample was treated with the RQ1 RNase-Free DNase kit (Promega®), according to the manufacture's recommendations, to remove a potential gDNA contamination. A total of 0.5 µL of random primers (252.5 µM) and 1 µL of oligo-dT_{18N} (100 µM) were added to DNAse-treated total RNA and reverse transcription was performed with the M-MLV Reverse Transcriptase kit (Promega®) according to the manufacturer's instructions. Resulting cDNA samples were stored at -20°C.

2.1.4. DNA amplifications, cloning and sequencing:

Specific primer pairs for each putative cyt-HSP70 subgroup (Table 1) were deduced from multiple sequence alignments (Genedoc version 2.7, Nicholas et al., 1997). Genomic DNA (gDNA) was PCR amplified using GoTaq® G2 flexi DNA polymerase (Promega®). Reaction mixtures contained dNTPs (0.2 mM), 1.25 µM of each primer, and 150 ng of DNA. Amplifications were performed with the following regime: 5' 94°C; 35x (45", 94°C; 45", 55 to 66°C (according to primer sequences); 2'30", 72°C) and an additional extension step for 5' at 72°C. The PCR products were cloned into the pGEM®-T vector (Promega) and transformed into competent *Escherichia coli* DH5α. Positive colonies were identified by a white/blue screen. Extracted plasmids from two clones (Wizard® *Plus* SV Minipreps DNA Purification System, Promega®) were systematically sequenced (Beckman Coulter Genomics, UK) using SP6, T7 and custom primers.

2.1.5. Sequences analyses:

HSP70 sequences were translated *in silico* with the Expasy translate tool (http://web.expasy.org/translate/). Multiple sequence alignment was performed using MAFFT (**M**ultiple **A**lignment using **F**ast **F**ourier **T**ransform, Katoh and Standley, 2013) to investigate the presence of the cytosolic HSP70-associated patterns reported in the literature (Rensing and Maier, 1994). Alignment inconsistencies led to the suspicion of the existence of chimeric sequences probably due to a specificity failure of some primers. Therefore, the "uchime_ref command" from the USEARCH program (Edgar et al., 2011) was launched on our dataset. Chimera detection was based on a reference dataset, which we constituted of well-represented sequences, confirmed with two more clones from two different PCR amplification of long sequences (1500 bp and more), according to their patterns and finding them reproducibly within multiple species. Suspicious sequences were removed from our final dataset. Introns were detected using alignments and *in silico* translations (http://web.expasy.org/translate/). Secondary structures and 3D conformation of complete HSP70s were predicted using Phyre2 (Protein Homology/analogy Recognition Engine V 2.0; Kelley et al., 2015). Models were visualized and analysed with UCSF Chimera (Pettersen et al., 2004).

Relevant sequences were named according to their nucleotide length, and introduced into our database to investigate cytosolic-type HSP70 relationships. Expasy tools were used to calculate molecular weights (http://web.expasy.org/compute_pi/).

2.2.Relationship analyses of cytosolic HSP70 within the arthropods:

2.2.1. Phylogenetic analyses:

The longest representative HSP70 sequences from each Malacostraca taxon (Amphipoda, Astacidea, Anomura, Brachyura, Caridea, Dendrobranchiata and Euphausiacea) were selected

through one-to-one alignments, and subsequently used for our phylogenetic analyses. 84 translated sequences ranging from 530 to 560 amino acids (on a total protein length of 640 to 670 amino-acids) were aligned with MAFFT using the default parameters and submitted to the ProtTest 2.4 server (http://darwin.uvigo.es/software/prottest2_server.html; Abascal et al., 2005) to define the best fit-model of protein evolution. Molecular evolutionary analyses were conducted using four approaches (Bayesian inference, Maximum Likelihood, Maximum Parsimony and Neighbour-Joining). Bayesian inferences were calculated using respectively MrBayes 3.2 (Ronquist et al., 2012). Support for Maximum Likelihood trees was calculated using the RAxML server (http://embnet.vital-it.ch/raxml-bb/; Stamatakis et al., 2008), with a bootstraps analysis (100 bootstrap pseudoreplicates) ; only node values equal to 100 are shown in the tree. Maximum Parsimony and Neighbour-Joining analyses were performed using MEGA, version 6 (Tamura et al., 2013) ; nodal support was evaluated by nonparametric bootstrap analysis with 1000 replicates (Figure 1).

A supplementary phylogenetic analysis was made to confirm the placement of cyt-HSP70 classification within Arthropoda (Supplementary Figure S1). Bayesian inference, Maximum Likelihood, Maximum Parsimony and Neighbour-Joining were performed on 101 representative HSP70 amino acid sequences from each group, as previously defined amongst Malacostraca (Figure 1), and presented by the aggregative analysis within Arthropoda (Figure 2).

2.2.2. Aggregative methodology:

Larger investigations on cytosolic HSP70 relationships, within the arthropods, required a classification procedure that allows the analysis of a large heterogeneous dataset with variable sequence lengths, located at different positions in the ORF. Therefore, we used the Automatic Classification Program developed by Rouault et al. (2009) based on aggregative methodology.

Analysis was performed using 428 cytosolic-HSP70 nucleotide sequences from Chelicerata, Crustacea, and Hexapoda, (with 2 ER-HSP70 and 2 HSP90 out-group sequences) of variable lengths ranging from 327 to 2007 nucleotides. Introns were removed from the sequences prior to analysis. Resulting clusters included full-length HSP70s, as well as shorter fragments, as there is only a marginal impact of sequence length in this type of approach (Figure 2). Groups were defined according to the results obtained from the phylogenetic analyses.

3. Results and discussion

3.1.Sequence analysis:

3.1.1. *HSP70 fragments isolated and characterized in this study*:

122 HSP70s fragments ranging from 634 to 1977 bp were amplified from genomic and complementary DNAs of 35 crustacean's species (23 brachyurans, 7 astacideans and 5 anomurans). Amplicons were located at various positions inside the complete gene sequences, highlighting mostly conserved blocks with few polymorphic loci of 4 to 10 bp and a GC-rich 3' end of highly variable length. USEARCH analyses for chimera detection revealed several chimerical PCR amplifications, probably caused by the high sequence similarities between different HSP70 homologs. For this reason, 16 suspicious sequences have been removed from the sequence dataset used in this study (data not shown). Accession numbers of the sequences isolated in this study are listed in "Supplementary Table S1" with NCBI annotation details (matrix used, intron presence, literature references), with group affiliations, protein lengths and molecular weights of the complete sequences. *In silico*-translated complete sequences are between 635 to 658 amino-acid long with molecular weights ranging from 69.6 to 72 kDa. All sequences present the highly HSP70-specific patterns: signature I: G-I-D-L-G-T-T-[YF]-S,

signature II: D-L-G-G-G-T-F-D, signature III: T-V-P-A-Y-F-N, signature IV: N-E-P-T-A-A, signature V: R-A-R-F-E-E-[LM] and the C-terminal end sequence G-P-[TN]-[IV]-E-E-V-D typical for the cytosolic HSP70s (Karlin and Brocchieri, 1998; Rensing and Maier, 1994).

3.1.2. *Sequences extracted from the international databases*:

310 cytosolic HSP70 sequences present in the Genbank™ database, were used to extend our dataset to 162 Arthropoda species: Hexapoda (58 species), Chelicerata (20 species), Crustacea (83 species) and Myriapoda (1 species). Each extracted sequence was carefully screened for the presence of the characteristic cytosolic HSP70 signatures. ER-HSP70 and HSP90 sequences were used as out-groups for the phylogenetic and aggregative analyses. All these accession numbers and annotation details are also presented in "Supplementary Table S1".

3.2.Cytosolic-HSP70 relationship analyses:

3.2.1. *Phylogenetic analyses and classification of cytosolic HSP70 within Malacostraca*:

Results are summarized in the phylogenetic tree shown in Figure 1. Analyses revealed two well-defined groups of cytosolic HSP70s (referred as A and B). This division within the cytosolic HSP70s (cyt-HSP70) have been described before in the literature (Cascella et al., 2015; Cui et al., 2010; Leignel et al., 2007; Luan et al., 2010; Ravaux et al., 2007). However, beside these two groups, our results suggest an even more intricate structure of cyt-HSP70 subfamily. Within cyt-HSP70 cluster A, five subgroups can be observed (noted A1 to A5), whereas two distinct subdivisions were detected in cluster B (noted B1 and B2). A third subdivision level was noticed in the brachyurans cytosolic HSP70s (B1a and B1b). Each group is represented by at least three distinct infraorders, except for group A2 (specific for astacideans). Amphipods sequences gathered in two groups named HSP70 A4 and A5,

according to the distribution of sequences from *Hyalella Azteca* and *Gammarus* sp.. Maximum parsimony analyses did not support the HSP70 A5 group, but similar topologies of all phylogenetic trees suggest that these sequences are related to HSP70 A. Nevertheless, phylogenetic analysis and variability of amino acid sequences lead to the suspicion of a more complex evolutionary scenario for HSP70 in amphipods, which may consist of supplementary subdivisions within the HSP70 A5 group, and could be resolved by increasing the datasets. Thus, our results revealed a completely unexpected presence of four cyt-HSP70 subtypes within the Astacidea and Brachyura, three in Caridea and Euphausiacea, and at least two in Amphipoda and Dendrobranchiata.

3.2.2. *Aggregative and phylogenetic analyses of cytosolic HSP70 within the Arthropoda*: In order to confirm these cyt-HSP70 affiliations and extend the analysis to the whole arthropods phylum, aggregative methods were performed on our nucleotide dataset. Results are presented in Figure 2 and the sequence-list in "Supplementary Table S1". The aggregative method approach confirmed the phylogenetic analysis within Malacostraca, suggesting two major cyt-HSP70 groups: A and B, each containing several subdivisions. The Malacostraca cyt-HSP70 group A is well delimited, in contrast to group B that is split in 2 parts. Branchiopoda HSP70s are separated in 2 groups (both annotated A / B), each including sequences from *Daphnia* sp., whereas the 2 subdivisions of sequences from maxillopods comprising distinct orders and could be explained by the distance between organisms. Amongst the other Arthropoda subphyla, a third well-delimited group, annotated as group C, gathers only HSP70 sequences from the chelicerates and hexapods. Other sequences from these two subphyla are also present amongst the A and B form of crustacean sequences (named HSP70 A/B). Complementary phylogenetic analyses (Supplementary Figure S1) confirmed the aggregative results, with well-defined HSP70 form A and B for Malacostraca

and a form C clustering sequences from chelicerates and hexapods. However, the A/B group remains ambiguous with no clear relation with either of HSP70 A or HSP70 B, observed within Malacostraca. These results suggest the hypothesis that HSP70 group A and B may have appeared after emergence of Malacostraca, and could derive from the same common HSP70 ancestor as the A/B group. The HSP70 group C is totally absent from Malacostraca and seems to have been lost before their emergence.

3.3.Comparison and discrimination determinants of different cytosolic-HSP70 forms:

3.3.1. *Intron characteristics*:

The cyt-HSP70 sequence details are presented in "Supplementary Table S1". A single intron can only be found in the HSP70 genes of form B1 in Decapoda, located around the nucleotide 174 to 180 from the ATG; whereas HSP70 form A and B2 seem to be intron-free. Amphipod HSP70 genes appear also intron-free. The relative lack of genomic data in Caridea and Euphausiacea did not allow any general conclusions. Genomic sequences from Branchiopoda and different orders within the Chelicerata and Hexapoda phyla present multiple introns (until 7) at various locations. However, no clear correlation was found between these forms, with or without intron, and HSP70 classification. In conclusion, intron presence in HSP70 genes seems particular to the B1 group in Anomura, Astacidea and Brachyura, and absence of an intron does not seem a valid criterion for HSP70 classification.

3.3.2. *Protein sequence comparison*:

Complete cyt-Hsp70 sequences from our data collection present ORF lengths ranging from 1869 to 2031 nucleotides, corresponding to translated sequences of 623 to 677 amino acids

and calculated molecular weights of 68.2 to 73.6 kDa (theoretical molecular weight, average). The rather variable length of the C-terminal region explains these differences. Comparison of ORFs helps HSP70 classification, and suggests that short sequences from anomurans (unsuitable for phylogenetical analyses) may belong to the cyt-HSP70 A1 and B1 group. Distinctive features in each of the HSP70 forms are reported in supplementary data (Table S2). Most discriminating patterns are located in the C-Terminal Domain (CTD), a few others in the Nucleotide Binding Domain (NBD) and a single one in the Substrate Binding Domain (SBD) that seems to be highly conserved between the different cyt-HSP70 groups (Figure 3). Four major amino acid variations can be observed in the discriminating patterns: presence/absence of amino acids (motifs 2 and 3), significant residue changes (aromatic to small, or modification of steric hindrance) (motifs 1, 7, 10 and 11), different charge repartitions (motifs 4, 5, 8 and 9), or presence/absence of tetrapeptide patterns GGAP, GGMP or GGFP (motifs 12). The leucine-rich pattern in the interdomain-linker, annotated as motif 6, presents an apparently slight residue change from leucine to valine, but could still be relevant due to the typically extreme conservation of this fragment. Discrimination of cyt-HSP70 A and B group sequences can be done in the region from amino acid 100 to 200, in the NBD and SBD with motif 1, 2 and 7, respectively. Motif 1 and 7 contain an aromatic residue in the HSP70 A group, whereas HSP70 B presents an asparagine (motif 1) or a leucine (motif 7) at the same position. The motif 2 in the NBD has an insertion of 2 amino acids (including a positively charge residue) in HSP70 B1, and 4 variable amino acids in the HSP70 B2 group. The HSP70 group A and B can also be discriminated through the interdomain-linker and the CTD: motifs 5, 9 and 10 present distinct charge repartitions in each HSP70 group, whereas motif 8 presents a functional significant hyper-charged run (Karlin and Brocchieri, 1998) in HSP70 A , and mostly small neutral residues in the HSP70 B group. One of the more discriminating patterns found between HSP70 group A and B is the $(GGXP)_n$ motif, which is

repeated 2 to 6 times in HSP70 A1, A2 and A3, and totally lacking in a large majority of sequences belonging to HSP70 B group. Comparison of sequences from HSP70 A/B and HSP70 C group from Chelicerata and Hexapoda revealed other discriminating motifs. The HSP70 C group presents an insertion of a serine (motif 2) and a change of the positively charged lysine to a neutral leucine (motif 3) or a change of valine/leucine to methionine (motif 11) in HSP70 A/B and HSP70 C respectively. According to motifs 1, 8 and 12, sequences from the HSP70 A/B share common features with HSP70 A group, whereas HSP70 C features seem to be related to the HSP70 B group. These results suggest the hypothesis of a convergent evolution of HSP70 sequences toward two specific structures, as previously formulated in bacterial and yeast models (Karlin and Brocchieri, 1998; Krenek et al., 2013). The NBD surface and the CTD are known as binding regions for a wide range of co-chaperones which modulate HSP70 activity (Arakawa et al., 2010; Brehmer et al., 2001; Demand et al., 1998). The NBD binds Nucleotide Exchange Factors (NEFs) that are involved in the ATP hydrolysis activity (Arakawa et al., 2010; Mayer, 2010; Palleros et al., 1991), whereas the CTD harbours binding sites for co-chaperones involved in HSP70-complex stability and oligomerization (Aprile et al., 2013; Demand et al., 1998). In this way, the $(GGXP)_n$ motif, is suspected to be a binding-site for the Hip-cofactor (Demand et al., 1998; Nollen et al., 2001; Prapapanich et al., 1998). The interdomain-linker ensures two distinct roles: it binds the NBD to stimulate ATPase activity, and is involved in HSP70 oligomerization and co-chaperone binding, to stabilize the substrate in the chaperone complex (Alderson et al., 2014; Aprile et al., 2013; Kumar et al., 2011; Swain et al., 2007; Vogel et al., 2006).

3.3.3. *Discussion of the correlation between phylogenetic position and the absence/presence of induction capacity:*

The relationship between the different HSP70 subgroups that we define in this study was examined in light of their constitutive (HSC70) or inducible (HSP70) expression reported in the literature and summarised in Table 2 and Figure 1. The assignment of a gene/cDNA/protein to either the HSC70 or HSP70 "group" appears to be often rather subjective. Some authors support a relation between expression pattern and some structure specificities (Liu et al., 2004; Luan et al., 2010; Qian et al., 2012; Ravaux et al., 2007), which they claim can be used for an *a priori* annotation (Mestre et al., 2015), whereas others remain sceptical about such a correlation (Bedulina et al., 2013; Cui et al., 2010; Leignel et al., 2007; Rhee et al., 2009).

In Malacostraca, the HSC70-annotated sequences isolated from shrimps (*Palaemonetes varians* and *Rimicaris exoculata*) and prawn (*Fenneropenaeus chinensis*), showed a constitutive expression pattern and were all clustered into the HSP70 A1 group. The HSP70-annotated sequences from the same species were falling into our HSP70 groups B1 (*F. chinensis*) and B2 (*P. varians*, *R. exoculata*), and showed a high and significant induction after heat shock (Cottin et al., 2010; Luan et al., 2010). However, HSP70 A1 also include heat-inducible sequences from the shrimp *Litopennaeus vannamei*, the prawn *Macrobrachium nippponense*, and the crab *Scylla paramamosain*, with 9.5 to 23.3-fold induction, compared to their respective control group (Qian et al., 2012; Xiu et al., 2014; Yang et al., 2013). In the same way, sequences gathered in HSP70 B1 and B2 isolated from crab, *Chaceon affinis,* and the krill, *Euphausia crystallorophias*, do not present significant heat-inducible profiles (Cascella et al., 2015; Mestre et al., 2015). Consequently, correlations between HSP70 sequence similarity and their expression features could lead to mistakes or annotation inconsistencies. Several debatable suggestions have been found in the literature: 11 to 12-fold induced genes are assigned to the constitutive form (HSC70) (Wu et al., 2008; Xiu et al.,

2014), whereas genes showing no significant induction are assigned to the inducible forms (HSP70) (Mestre et al., 2015) (Table 2A).

No obvious HSC70 or HSP70 features could be discerned in either HSP70 group A/B or C in chelicerates or hexapods. For example, studies on the red mite *Panonychus citri* revealed that two HSP70 forms that gathered both in the same A/B group, but present either a constitutive or an inducible expression profile facing heat shock (Yang et al., 2012). In the HSP70 C group, hexapods sequences also showed variable expression profiles after heat stress: constitutive in *Tribolium castaneum* (Liu et al., 2014) and inducible in *Plutella xylostella* (Zhang et al., 2015). Conversely, HSC70 and HSP70 sequences from the scale insect *Ericerus pela*, gathered in the A/B and C groups respectively, and present very variable gene expression profiles: with a high over-expression in females (55 to 145-fold) but only a slight induction in males (Liu et al., 2014) (Table 2B).

According to the literature, the HSP70 gene response closely depends on the type of tissue analysed (Liu et al., 2004; Wu et al., 2008; Xiu et al., 2014), the populations (Cottin et al., 2015), or between male and female within the same species (Liu et al., 2014). Furthermore, induction of HSP70s is tightly connected to the stress type, with distinct expression patterns in response to heat-shock, bacterial challenges (Altincicek et al., 2008; Yang et al., 2013), metal exposure (Luan et al., 2010; Qian et al., 2012), or osmotic stress (Fu et al., 2013; Yang et al., 2013); and in function of exposure time and recovery period (Liu et al., 2014; Luan et al., 2010; Qian et al., 2012).

4. Conclusions

This study presents 122 new HSP70 sequences from 35 decapods species. Comprehensive analyses of all these arthropod HSP70s (based on 432 sequences) highlight a thus far

unsuspected diversity of the cytosolic-HSP70 isoforms (A, B and C groups). We demonstrate the existence of, at least, four paralogous cyt-HSP70s in astacideans, brachyurans, and probably the anomurans. The cyt-HSP70 groups defined in this study do not seem to be directly correlated to any ecological feature. The organism's adaptation to stressful and variable environment may be more related to the level plasticity in the expression response of the gene (Papot et al., 2016).

According to our knowledge, HSP70 sequences in the B1 group are the only one to harbour a single intron. The lack of genomic data for Caridea and Dendrobranchiata did not allow us to draw conclusions about intron characteristics for the cyt-HSP70 types in these groups (Supplementary Table S1). The cyt-HSP70 A, B and C groups can be discriminated through eleven amino acid motifs, and the presence or absence of GGXP repeats (Supplementary Table S2). These differences appear to involve distinct co-chaperone interactions (Brehmer et al., 2001; Demand et al., 1998; Nollen et al., 2001), and may influence chaperone complex stability (Aprile et al., 2013; Swain et al., 2007), rather than directly regulate inducible or constitutive expression.

Indeed, the cyt-HSP70 classification proposed in our study is not at all consistent with the constitutive or inducible features detected for these genes. Therefore *a priori* annotation only based on protein structural features is clearly not possible, and should thus be avoided (Chuang et al., 2007; Mestre et al., 2015; Qian et al., 2012; Wu et al., 2008; Xiu et al., 2014). The differential expression of HSP70 genes is thus probably regulated through promoter elements like transcription factor recognition sites (Heat Shock Element, CCAT box: (Bienz, 1986; Chuang et al., 2007; Zhao et al., 2013)), AT richness (Chen et al., 2011) or transposable element insertion (Haney and Feder, 2009).

The existence of HSP70 A and B forms was clearly demonstrated in the Malacostraca class, but whereas this distinction was not quite possible in the other arthropods (HSP70 A/B

group). Thus, HSP70 forms belonging to A and B groups have probably resulted from several duplication events of an ancestral gene, common to the HSP70 A/B group. The last duplication seems to have occurred on the cyt-HSP70 B1 gene, during the Brachyura emergence (Porter et al., 2005; Tsang et al., 2014), leading to B1a and B1b formation (Figure 1). None of HSP70 C forms was found within the Malacostraca and this led to the hypothesis of a putative loss of this form before the formation of the Multicrustacea clade (Oakley et al., 2013; Regier et al., 2010; Shen et al., 2013). The classification supported in this study provides new insights in HSP70 organization in Arthropoda, and more particularly in Malacostraca. Regarding to these results, further investigations could improve understandings about molecular adaptation process, facing to environment stressors.

List of abbreviations

HSP70: Heat shock protein 70 kDa

HSC70: Cognate heat shock protein 70 kDa

Cyt-HSP70: cytosolic heat shock protein 70 kDa

ER-HSP70: endoplasmic reticulum heat shock protein 70 kDa

HSP90: Heat shock protein 90 kDa

NCBI: National Center of Biotechnology Information

NEF: Nucleotide exchange factor

gDNA: genomic DNA

cDNA: complementary DNA

PCR: Polymerase Chain Reaction

ML: Maximum Likelihood

MP: Maximum Parsimony

NJ: Neighbour-Joining

NBD: Nucleotide-Binding Domain

SBD: Substrate-Binding Domain

CTD: C-Terminal Domain

Acknowledgements

This work was supported by a doctoral grant provided by the French Ministry of Higher Education and Research. The authors thank François Lallier, chief scientist of the HOPE (9 and 13°N, 1999) and the BIOBAZ (36°N, 2013) cruises and Richard Cosson for allowing us to get samples of several hydrothermal crustaceans. We express thanks to the captain and crew of the R/V *Pourquoi pas*? and especially the *Victor 6000* ROV team (*Remotely Operated Vehicle*) for their assistance in this work.

References:

Abascal, F., Zardoya, R., Posada, D., 2005. ProtTest: selection of best-fit models of protein evolution. Bioinforma. Oxf. Engl. 21, 2104–2105. doi:10.1093/bioinformatics/bti263

Abisambra, J.F., Jinwal, U.K., Jones, J.R., Blair, L.J., Koren, J., Dickey, C.A., 2011. Exploiting the diversity of the heat-shock protein family for primary and secondary tauopathy therapeutics. Curr. Neuropharmacol. 9, 623–631. doi:10.2174/157015911798376226

Alderson, T.R., Kim, J.H., Cai, K., Frederick, R.O., Tonelli, M., Markley, J.L., 2014. The specialized Hsp70 (HscA) Interdomain Linker binds to its Nucleotide-Binding Domain and stimulates ATP hydrolysis in both *cis* and *trans* configurations. Biochemistry (Mosc.) 53, 7148–7159. doi:10.1021/bi5010552

Altincicek, B., Knorr, E., Vilcinskas, A., 2008. Beetle immunity: Identification of immune-inducible genes from the model insect *Tribolium castaneum*. Dev. Comp. Immunol. 32, 585–595. doi:10.1016/j.dci.2007.09.005

Aprile, F.A., Dhulesia, A., Stengel, F., Roodveldt, C., Benesch, J.L.P., Tortora, P., Robinson, C.V., Salvatella, X., Dobson, C.M., Cremades, N., 2013. Hsp70 oligomerization is mediated by an interaction between the Interdomain Linker and the Substrate-Binding Domain. PLoS ONE 8, e67961. doi:10.1371/journal.pone.0067961

Arakawa, A., Handa, N., Ohsawa, N., Shida, M., Kigawa, T., Hayashi, F., Shirouzu, M., Yokoyama, S., 2010. The C-Terminal BAG Domain of BAG5 induces conformational changes of the Hsp70 Nucleotide- Binding Domain for ADP-ATP exchange. Structure 18, 309–319. doi:10.1016/j.str.2010.01.004

Bedulina, D.S., Evgen'ev, M.B., Timofeyev, M.A., Protopopova, M.V., Garbuz, D.G., Pavlichenko, V.V., Luckenbach, T., Shatilina, Z.M., Axenov-Gribanov, D.V., Gurkov, A.N., Sokolova, I.M., Zatsepina, O.G., 2013. Expression patterns and organization of the *hsp70* genes correlate with thermotolerance in two congener endemic amphipod species (*Eulimnogammarus cyaneus* and *E. verrucosus*) from Lake Baikal. Mol. Ecol. 22, 1416–1430. doi:10.1111/mec.12136

Bettencourt, B.R., Feder, M.E., 2001. Hsp70 duplication in the *Drosophila melanogaster* species group: how and when did two become five? Mol. Biol. Evol. 18, 1272–1282.

Bienz, M., 1986. A CCAAT box confers cell-type-specific regulation on the Xenopus hsp70 gene in oocytes. Cell 46, 1037–1042. doi:10.1016/0092-8674(86)90703-8

Boorstein, W.R., Ziegelhoffer, T., Craig, E.A., 1994. Molecular evolution of the HSP70 multigene family. J. Mol. Evol. 38, 1–17. doi:10.1007/BF00175490

Brehmer, D., Rüdiger, S., Gässler, C.S., Klostermeier, D., Packschies, L., Reinstein, J., Mayer, M.P., Bukau, B., 2001. Tuning of chaperone activity of Hsp70 proteins by

modulation of nucleotide exchange. Nat. Struct. Mol. Biol. 8, 427–432. doi:10.1038/87588

Bukau, B., Horwich, A.L., 1998. The Hsp70 and Hsp60 chaperone machines. Cell 92, 351–366. doi:10.1016/S0092-8674(00)80928-9

Cascella, K., Jollivet, D., Papot, C., Léger, N., Corre, E., Ravaux, J., Clark, M.S., Toullec, J.-Y., 2015. Diversification, evolution and sub-functionalization of 70kDa Heat-Shock proteins in two sister species of Antarctic krill: differences in thermal habitats, responses and implications under climate change. PLoS ONE 10, e0121642. doi:10.1371/journal.pone.0121642

Chen, B., Jia, T., Ma, R., Zhang, B., Kang, L., 2011. Evolution of hsp70 gene expression: A role for changes in AT-richness within promoters. PLoS ONE 6, e20308. doi:10.1371/journal.pone.0020308

Chuang, K.-H., Ho, S.-H., Song, Y.-L., 2007. Cloning and expression analysis of heat shock cognate 70 gene promoter in tiger shrimp (*Penaeus monodon*). Gene 405, 10–18. doi:10.1016/j.gene.2007.08.016

Colson-Proch, C., Morales, A., Hervant, F., Konecny, L., Moulin, C., Douady, C.J., 2010. First cellular approach of the effects of global warming on groundwater organisms: a study of the HSP70 gene expression. Cell Stress Chaperones 15, 259–270. doi:10.1007/s12192-009-0139-4

Cottin, D., Foucreau, N., Hervant, F., Piscart, C., 2015. Differential regulation of hsp70 genes in the freshwater key species *Gammarus pulex* (Crustacea, Amphipoda) exposed to thermal stress: effects of latitude and ontogeny. J. Comp. Physiol. B 185, 303–313. doi:10.1007/s00360-014-0885-1

Cottin, D., Shillito, B., Chertemps, T., Thatje, S., Léger, N., Ravaux, J., 2010. Comparison of heat-shock responses between the hydrothermal vent shrimp *Rimicaris exoculata* and

the related coastal shrimp *Palaemonetes varians*. J. Exp. Mar. Biol. Ecol. 393, 9–16. doi:10.1016/j.jembe.2010.06.008

Craig, E.A., Gambill, B.D., Nelson, R.J., 1993. Heat shock proteins: molecular chaperones of protein biogenesis. Microbiol. Rev. 57, 402–414.

Craig, E.A., Schlesinger, M.J., 1985. The Heat Shock Respons. CRC Crit. Rev. Biochem. 18, 239–280. doi:10.3109/10409238509085135

Cui, Z., Liu, Y., Luan, W., Li, Q., Wu, D., Wang, S., 2010. Molecular cloning and characterization of a heat shock protein 70 gene in swimming crab (*Portunus trituberculatus*). Fish Shellfish Immunol. 28, 56–64. doi:10.1016/j.fsi.2009.09.018

Daugaard, M., Rohde, M., Jäättelä, M., 2007. The heat shock protein 70 family: Highly homologous proteins with overlapping and distinct functions. FEBS Lett. 581, 3702–3710. doi:10.1016/j.febslet.2007.05.039

Demand, J., Lüders, J., Höhfeld, J., 1998. The carboxy-terminal domain of Hsc70 provides binding sites for a distinct set of chaperone cofactors. Mol. Cell. Biol. 18, 2023–2028. doi:10.1128/MCB.18.4.2023

De Pomerai, D.I., 1996. Review: Heat-shock proteins as biomarkers of pollution. Hum. Exp. Toxicol. 15, 279–285.

Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C., Knight, R., 2011. UCHIME improves sensitivity and speed of chimera detection. Bioinforma. Oxf. Engl. 27, 2194–2200. doi:10.1093/bioinformatics/btr381

Feder, M.E., Hofmann, G.E., 1999. Heat-shock proteins, molecular chaperones, and the stress response: evolutionary and ecological physiology. Annu. Rev. Physiol. 61, 243–282. doi:10.1146/annurev.physiol.61.1.243

Fink, A.L., 1999. Chaperone-mediated protein folding. Physiol. Rev. 79, 425–449.

Fontaine, S.N., Rauch, J.N., Nordhues, B.A., Assimon, V.A., Stothert, A.R., Jinwal, U.K., Sabbagh, J.J., Chang, L., Stevens, S.M., Zuiderweg, E.R.P., Gestwicki, J.E., Dickey, C.A., 2015. Isoform-selective genetic inhibition of constitutive cytosolic Hsp70 activity promotes client Tau degradation using an altered co-chaperone complement. J. Biol. Chem. 290, 13115–13127. doi:10.1074/jbc.M115.637595

Franzellitti, S., Fabbri, E., 2005. Differential HSP70 gene expression in the Mediterranean mussel exposed to various stressors. Biochem. Biophys. Res. Commun. 336, 1157–1163. doi:10.1016/j.bbrc.2005.08.244

Fu, W., Zhang, F., Liao, M., Liu, M., Zheng, B., Yang, H., Zhong, M., 2013. Molecular cloning and expression analysis of a cytosolic heat shock protein 70 gene from mud crab *Scylla serrata*. Fish Shellfish Immunol. 34, 1306–1314. doi:10.1016/j.fsi.2013.02.027

Gupta, R.S., Singh, B., 1994. Phylogenetic analysis of 70 kD heat shock protein sequences suggests a chimeric origin for the eukaryotic cell nucleus. Curr. Biol. 4, 1104–1114. doi:10.1016/S0960-9822(00)00249-9

Haney, R.A., Feder, M.E., 2009. Contrasting patterns of transposable element insertions in Drosophila Heat-Shock promoters. PLoS ONE 4, e8486. doi:10.1371/journal.pone.0008486

Ingolia, T.D., Craig, E.A., 1982. Drosophila gene related to the major heat shock-induced gene is transcribed at normal temperatures and not induced by heat shock. Proc. Natl. Acad. Sci. 79, 525–529.

Karlin, S., Brocchieri, L., 1998. Heat shock protein 70 family: multiple sequence comparisons, function, and evolution. J. Mol. Evol. 47, 565–577. doi:10.1007/PL00006413

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772–780. doi:10.1093/molbev/mst010

Kelley, A.L., de Rivera, C.E., Buckley, B.A., 2011. Intraspecific variation in thermotolerance and morphology of the invasive European green crab, *Carcinus maenas*, on the west coast of North America. J. Exp. Mar. Biol. Ecol. 409, 70–78. doi:10.1016/j.jembe.2011.08.005

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., Sternberg, M.J.E., 2015. The Phyre2 web portal for protein modeling, prediction and analysis. Nat. Protoc. 10, 845–858. doi:10.1038/nprot.2015.053

Kiang, J.G., Tsokos, G.C., 1998. Heat shock protein 70 kDa: molecular biology, biochemistry, and physiology. Pharmacol. Ther. 80, 183–201. doi:10.1016/S0163-7258(98)00028-X

Krenek, S., Schlegel, M., Berendonk, T.U., 2013. Convergent evolution of heat-inducibility during subfunctionalization of the Hsp70 gene family. BMC Evol. Biol. 13, 49. doi:10.1186/1471-2148-13-49

Kumar, D.P., Vorvis, C., Sarbeng, E.B., Cabra Ledesma, V.C., Willis, J.E., Liu, Q., 2011. The four hydrophobic residues on the Hsp70 Inter-domain Linker have two distinct roles. J. Mol. Biol. 411, 1099–1113. doi:10.1016/j.jmb.2011.07.001

Leignel, V., Cibois, M., Moreau, B., Chénais, B., 2007. Identification of new subgroup of HSP70 in Bythograeidae (hydrothermal crabs) and Xanthidae. Gene 396, 84–92. doi:10.1016/j.gene.2007.02.025

Lim, J., Yue, Z., 2015. Neuronal Aggregates: Formation, clearance, and spreading. Dev. Cell 32, 491–501. doi:10.1016/j.devcel.2015.02.002

Lindquist, S., 1986. The heat-shock response. Annu. Rev. Biochem. 55, 1151–1191. doi:10.1146/annurev.bi.55.070186.005443

Lindquist, S., Craig, E.A., 1988. The Heat-Shock Proteins. Annu. Rev. Genet. 22, 631–677. doi:10.1146/annurev.ge.22.120188.003215

Liu, J., Yang, W.-J., Zhu, X.-J., Karouna-Renier, N.K., Rao, R.K., 2004. Molecular cloning and expression of two HSP70 genes in the prawn, *Macrobrachium rosenbergii*. Cell Stress Chaperones 9, 313. doi:10.1379/CSC-40R.1

Liu, W.-W., Yang, P., Chen, X.-M., Xu, D.-L., Hu, Y.-H., 2014. Cloning and expression analysis of four eat Shock Protein genes in *Ericerus pela* (Homoptera: Coccidae). J. Insect Sci. 14, 142. doi:10.1093/jisesa/ieu032

Lo, W.-Y., Liu, K.-F., Liao, I.-C., Song, Y.-L., 2004. Cloning and molecular characterization of heat shock cognate 70 from tiger shrimp (*Penaeus monodon*). Cell Stress Chaperones 9, 332. doi:10.1379/CSC-47R.1

Luan, W., Li, F., Zhang, J., Wen, R., Li, Y., Xiang, J., 2010. 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. doi:10.1007/s12192-009-0124-y

Madeira, D., Narciso, L., Cabral, H.N., Vinagre, C., Diniz, M.S., 2012. HSP70 production patterns in coastal and estuarine organisms facing increasing temperatures. J. Sea Res. 73, 137–147. doi:10.1016/j.seares.2012.07.003

Manna, J., 2014. Activation of Heat Shock Proteins by Nanocurcumin to Prevent Neurodegenerative Diseases. Brain Disord. Ther. 03. doi:10.4172/2168-975X.1000139

Mayer, M.P., 2010. Gymnastics of molecular chaperones. Mol. Cell 39, 321–331. doi:10.1016/j.molcel.2010.07.012

Meng, C., Zhu, H., Song, H., Wang, Z., Huang, G., Li, D., Ma, Z., Ma, J., Qin, Q., Sun, X., others, 2014. Targets and molecular mechanisms of triptolide in cancer therapy. Chin. J. Cancer Res. 26, 622. doi:10.3978/j.issn.1000-9604.2014.09.01

Mestre, N.C., Cottin, D., Bettencourt, R., Colaço, A., Correia, S.P.C., Shillito, B., Thatje, S., Ravaux, J., 2015. Is the deep-sea crab *Chaceon affinis* able to induce a thermal stress response? Comp. Biochem. Physiol. A. Mol. Integr. Physiol. 181, 54–61. doi:10.1016/j.cbpa.2014.11.015

Nicholas, K.B., Nicholas, H.B.J., Deerfield, D.W., 1997. GeneDoc: analysis and visualization of genetic variation. Embnew News 4, 1–4.

Nollen, E.A.A., Kabakov, A.E., Brunsting, J.F., Kanon, B., Hohfeld, J., Kampinga, H.H., 2001. Modulation of *in vivo* HSP70 chaperone activity by Hip and Bag-1. J. Biol. Chem. 276, 4677–4682. doi:10.1074/jbc.M009745200

Oakley, T.H., Wolfe, J.M., Lindgren, A.R., Zaharoff, A.K., 2013. Phylotranscriptomics to bring the understudied into the fold: Monophyletic Ostracoda, fossil placement, and Pancrustacean phylogeny. Mol. Biol. Evol. 30, 215–233. doi:10.1093/molbev/mss216

Palleros, D.R., Welch, W.J., Fink, A.L., 1991. Interaction of hsp70 with unfolded proteins: effects of temperature and nucleotides on the kinetics of binding. Proc. Natl. Acad. Sci. U. S. A. 88, 5719–5723.

Papot, C., Cascella, K., Toullec, J.-Y., Jollivet, D., 2016. Divergent ecological histories of two sister Antarctic krill species led to contrasted patterns of genetic diversity in their heat-shock protein (hsp70) arsenal. Ecol. Evol. 6, 1555-1575. doi:10.1002/ece3.1989

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., 2004. UCSF Chimera--a visualization system for exploratory research and analysis. J. Comput. Chem. 25, 1605–1612. doi:10.1002/jcc.20084

Porter, M.L., Pérez-Losada, M., Crandall, K.A., 2005. Model-based multi-locus estimation of decapod phylogeny and divergence times. Mol. Phylogenet. Evol. 37, 355–369. doi:10.1016/j.ympev.2005.06.021

Prapapanich, V., Chen, S., Smith, D.F., 1998. Mutation of Hip's carboxy-terminal region inhibits a transitional stage of progesterone receptor assembly. Mol. Cell. Biol. 18, 944–952. doi:10.1128/MCB.18.2.944

Qian, Z., Liu, X., Wang, L., Wang, X., Li, Y., Xiang, J., Wang, P., 2012. Gene expression profiles of four heat shock proteins in response to different acute stresses in shrimp, *Litopenaeus vannamei*. Comp. Biochem. Physiol. Part C Toxicol. Pharmacol. 156, 211–220. doi:10.1016/j.cbpc.2012.06.001

Ravaux, J., Toullec, J.-Y., Léger, N., Lopez, P., Gaill, F., Shillito, B., 2007. First hsp70 from two hydrothermal vent shrimps, *Mirocaris fortunata* and *Rimicaris exoculata*: Characterization and sequence analysis. Gene 386, 162–172. doi:10.1016/j.gene.2006.09.001

Regier, J.C., Shultz, J.W., Zwick, A., Hussey, A., Ball, B., Wetzer, R., Martin, J.W., Cunningham, C.W., 2010. Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences. Nature 463, 1079–1083. doi:10.1038/nature08742

Rensing, S.A., Maier, U.-G., 1994. Phylogenetic analysis of the stress-70 protein family. J. Mol. Evol. 39, 80–86. doi:10.1007/BF00178252

Rhee, J.-S., Raisuddin, S., Lee, K.-W., Seo, J.S., Ki, J.-S., Kim, I.-C., Park, H.G., Lee, J.-S., 2009. Heat shock protein (Hsp) gene responses of the intertidal copepod *Tigriopus japonicus* to environmental toxicants. Comp. Biochem. Physiol. Part C Toxicol. Pharmacol. 149, 104–112. doi:10.1016/j.cbpc.2008.07.009

Ritossa, F., 1962. A new puffing pattern induced by temperature shock and DNP in drosophila. Experientia 18, 571–573. doi:10.1007/BF02172188

Rodina, A., Patel, P.D., Kang, Y., Patel, Y., Baaklini, I., Wong, M.J.H., Taldone, T., Yan, P., Yang, C., Maharaj, R., Gozman, A., Patel, M.R., Patel, H.J., Chirico, W., Erdjument-Bromage, H., Talele, T.T., Young, J.C., Chiosis, G., 2013. Identification of an allosteric pocket on human Hsp70 reveals a mode of inhibition of this therapeutically important protein. Chem. Biol. 20, 1469–1480. doi:10.1016/j.chembiol.2013.10.008

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61, 539–542. doi:10.1093/sysbio/sys029

Rouault, J.-D., Casse, N., Chénais, B., Hua-Van, A., Filée, J., Capy, P., 2009. Automatic classification within families of transposable elements: Application to the mariner Family. Gene 448, 227–232. doi:10.1016/j.gene.2009.08.009

Schlecht, R., Scholz, S.R., Dahmen, H., Wegener, A., Sirrenberg, C., Musil, D., Bomke, J., Eggenweiler, H.-M., Mayer, M.P., Bukau, B., 2013. Functional analysis of Hsp70 inhibitors. PLoS ONE 8, e78443. doi:10.1371/annotation/5a7961d9-a7ea-4b10-9b48-5b106c405b02

Shen, H., Braband, A., Scholtz, G., 2013. Mitogenomic analysis of decapod crustacean phylogeny corroborates traditional views on their relationships. Mol. Phylogenet. Evol. 66, 776–789. doi:10.1016/j.ympev.2012.11.002

Sherman, M.Y., Gabai, V.L., 2015. Hsp70 in cancer: back to the future. Oncogene 34, 4153–4161. doi:10.1038/onc.2014.349

Spees, J.L., Chang, S.A., Snyder, M.J., Chang, E.S., 2002. Osmotic induction of stress-responsive gene expression in the lobster *Homarus americanus*. Biol. Bull. 203, 331–337. doi:10.2307/1543575

Stamatakis, A., Hoover, P., Rougemont, J., 2008. A rapid bootstrap algorithm for the RAxML Web servers. Syst. Biol. 57, 758–771. doi:10.1080/10635150802429642

Swain, J.F., Dinler, G., Sivendran, R., Montgomery, D.L., Stotz, M., Gierasch, L.M., 2007. Hsp70 chaperone ligands control domain association via an allosteric mechanism mediated by the interdomain linker. Mol. Cell 26, 27–39. doi:10.1016/j.molcel.2007.02.020

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol. Biol. Evol. 30, 2725–2729. doi:10.1093/molbev/mst197

Tissiéres, A., Mitchell, H.K., Tracy, U.M., 1974. Protein synthesis in salivary glands of *Drosophila melanogaster*: Relation to chromosome puffs. J. Mol. Biol. 84, 389–398. doi:10.1016/0022-2836(74)90447-1

Tsang, L.M., Schubart, C.D., Ahyong, S.T., Lai, J.C.Y., Au, E.Y.C., Chan, T.-Y., Ng, P.K.L., Chu, K.H., 2014. Evolutionary History of True Crabs (Crustacea: Decapoda: Brachyura) and the Origin of Freshwater Crabs. Mol. Biol. Evol. 31, 1173–1187. doi:10.1093/molbev/msu068

Vogel, M., Mayer, M.P., Bukau, B., 2006. Allosteric regulation of Hsp70 chaperones involves a conserved interdomain linker. J. Biol. Chem. 281, 38705–38711. doi:10.1074/jbc.M609020200

Werner-Washburne, M., Craig, E.A., 1989. Expression of members of the *Saccharomyces cerevisiae* hsp70 multigene family. Genome Natl. Res. Counc. Can. Génome Cons. Natl. Rech. Can. 31, 684–689. doi:10.1139/g89-125

Wu, R., Sun, Y., Lei, L.M., Xie, S.T., 2008. Molecular identification and expression of heat shock cognate 70 (HSC70) in the pacific white shrimp *Litopenaeus vannamei*. Mol. Biol. 42, 234–242. doi:10.1134/S002689330802009X

Xiu, Y., Feng, J., Lu, W., Liu, D., Wu, T., Zhu, H., Liu, P., Li, W., Ren, Q., Gu, W., Meng, Q., Wang, W., 2014. Identification of a novel cognate cytosolic Hsp70 gene (MnHsc70-2) from oriental river prawn *Macrobrachium nipponense* and comparison of its expressions with the first cognate Hsc70 (MnHsc70-1) under different stresses. Cell Stress Chaperones 19, 949–961. doi:10.1007/s12192-014-0519-2

Yamuna, A., Kabila, V., Geraldine, P., 2000. Expression of heat shock protein 70 in freshwater prawn *Macrobrachium malcolmsonii* (H. Milne Edwards) following exposure to Hg and Cu. Indian J. Exp. Biol. 38, 921–925.

Yang, L.-H., Jiang, H.-B., Liu, Y.-H., Dou, W., Wang, J.-J., 2012. Molecular characterization of three heat shock protein 70 genes and their expression profiles under thermal stress in the citrus red mite. Mol. Biol. Rep. 39, 3585–3596. doi:10.1007/s11033-011-1132-4

Yang, Y., Ye, H., Huang, H., Li, S., Liu, X., Zeng, X., Gong, J., 2013. Expression of Hsp70 in the mud crab, *Scylla paramamosain* in response to bacterial, osmotic, and thermal stress. Cell Stress Chaperones 18, 475–482. doi:10.1007/s12192-013-0402-6

Zhang, F., Jiang, K., Sun, M., Zhang, D., Ma, L., 2013. Multiplex immune-related genes expression analysis response to bacterial challenge in mud crab, *Scylla paramamosain*. Fish Shellfish Immunol. 34, 712–716. doi:10.1016/j.fsi.2012.11.029

Zhang, L.J., Wu, Z.L., Wang, K.F., Liu, Q., Zhuang, H.M., Wu, G., 2015. Trade-off between thermal tolerance and insecticide resistance in *Plutella xylostella*. Ecol. Evol. 5, 515–530. doi:10.1002/ece3.1380

Zhao, C., Zhang, X., Li, F., Huan, P., Xiang, J., 2013. Functional analysis of the promoter of the heat shock cognate 70 gene of the Pacific white shrimp, *Litopenaeus vannamei*. Fish Shellfish Immunol. 34, 397–401. doi:10.1016/j.fsi.2012.09.022

Zhou, J., Wang, W.-N., He, W.-Y., Zheng, Y., Wang, L., Xin, Y., Liu, Y., Wang, A.-L., 2010. Expression of HSP60 and HSP70 in white shrimp, *Litopenaeus vannamei* in response to bacterial challenge. J. Invertebr. Pathol. 103, 170–178. doi:10.1016/j.jip.2009.12.006

Tables:

Table 1. Primers used for cytosolic-HSP70 molecular characterization.

Primers were designed by aligning the cytosolic-HSP70 available in the public databases, based on variable or conserved sequence regions, specific to non-organellar HSP70. Amplifications were obtained by combining different forward and reverse primers.

Primers names	Direction	Sequences (5' -> 3')
70Cg1	Forward	GCGAGWCDAARACCTTCAAYC
70Cg2	Forward	CAMGRCASSTGCAGTGGCATTG
70Chom	Forward	GAAACAGTAATTGGTATTGACCTGGG
70Cuniv	Forward	CTGYGAGAGVGCCAAGMGHAC
ATGcrab	Forward	ATGKCHAAGGGAGCWGCWGT
HS0	Forward	GTGTTGGWGTNTTCCAGCATGS
HS3	Forward	TKGGWGGBGARGAYTTYGAC
70Cg1	Reverse	GTACTGYTCRGCCTCRTTSAC
70Cg3	Reverse	CCTCAAGCCTGTTCTTGACTTC

70CHuniv	Reverse	TTARTCGACYTCCTCGAYKGTG
HS2	Reverse	CCCTCRAASARAGAGTCRATC

Table 2: Gene induction data of cytosolic hsp70 in response to heat shocks. Expression profiles were obtained by quantitative PCR experiments and stress inductions are presented as relative over-expression, compared to the control condition. Induction-fold values were reported from the text when possible or graphically deduced (NS: statistically not significant change). Heat shock temperature, time of exposure, time of recovery and tissue analysed, have been reported for each experiment. GenBank accession numbers were associated with sequence annotation (HSP70 or HSC70) and group classification. HSP70 forms considered as heat inducible in the literature appear in red. Sequences from Malacostraca gathered in HSP70 A and B (Table 2A) and sequences from chelicerates and hexapods belong to HSP70 A/B and C (Table 2B).

Species	Reference	Heat shock (expo. / recov.)	Tissue	HSP70 A sequences			HSP70 B sequences		
				Accession number (annotation)	Classification	Induction	Accession number (annotation)	Classification	Induction
Palaemonetes varians	Cottin et al. (2010)	+18°C (1 h / 2 h)	Abdomen	FJ875279 (hsc70 form 2)	HSP70 A1	NS	FJ875280 (hsp70 form 1)	HSP70 B2	15-fold
Rimicaris exoculata	Cottin et al. (2010)	+20°C (1 h / 2 h)	Abdomen	FJ356149 (hsc70 form 3)	HSP70 A1	NS	DQ534065 (hsp70 form 1)	HSP70 B2	373-fold
		"	"				FJ268954 (hsp70 form 2)	HSP70 B2	132-fold
Macrobrachium nipponense	Xiu et al. (2014)	+10°C (12 h / 48 h)	Gill	DQ660140 (hsc70-1)	HSP70 A1	2.2-fold			
		"	"	KC460343 (hsc70-2)	HSP70 A1	11.67-fold			
		"	Hepatopancreas	DQ660140 (hsc70-1)	"	down regulation			
		"	"	KC460343 (hsc70-2)	"	NS			
Litopenaeus vannamei	Wu et al. (2008)	+ 10°C (6 h / 1 h)	Gill	EF495128 (hsc70)	HSP70 A1	11-fold			
	Qian et al. (2012)	+10°C (1 h / none)	hepatopancreas	AY645906 (hsp70)	HSP70 A1	9.5-fold			
		"	"	EF495128 (hsc70)	"	3.7-fold			
		+10°C (6 h / none)	"	AY645906 (hsp70)	"	~3.5-fold			
		"	"	EF495128 (hsc70)	"	~3-fold			
Fenneropenaeus chinensis	Luan et al. (2010)	+10°C (2 h / none)	cephalothorax	AY748350 (hsc70)	HSP70 A1	~5-fold	FJ167398 (hsp70)	HSP70 B1	~100-fold
		+10°C (6 h / 6 h)	"	"	"	~5-fold			~4-fold
Scylla serrata	Fu et al. (2013)	+10°C (2 h / none)	Gill	JQ780845 (hsp70)	HSP70 A1	~2.5-fold			
		+15°C (2 h / none)	"	"	"	~4.8-fold			
Scylla paramamosain	Yang et al. 2013	+11°C (6 h : none)	haemocytes	EU754021 (hsp70)	HSP70 A1	23.3-fold			
Chaceon affinis	Mestre et al. (2015)	+ 14°C (1 h / 2 h)	muscle	KC907347 (hsp70 form 1)	HSP70 A3	NS	KC907348 (hsp70 form 2)	HSP70 B1b	NS
Euphausia crystallorophias	Cascella et al. (2015)	+ 6°C (3 h / none)	abdomen	KM067139 (hsp70 form A)	HSP70 A1	NS	KM067142 (hsp70 form E)	HSP70 B2	NS
		"	"	KM067140 (hsp70 form B)	HSP70 A1	~2.5-fold			
		"	"	KM067141 (hsp70 form C)	HSP70 A3	~2.5-fold			

Species	Reference	Heat shock (expo. / recov.)	Tissue	HSP70 A / B sequences			HSP70 C sequences		
				Accession number (annotation)	Classification	Induction	Accession number (annotation)	Classification	Induction
Panonychus citri	Yang et al. (2012)	+ 16°C (1 h / none)	whole	GQ495083 (hsp70-1)	HSP70 A / B	NS			
				GQ495084 (hsp70-2)	"	28.4-fold			
Ericerus pela (female)	Liu et al. (2014)	+8°C (1 h / none)	whole	KC161301 (hsc70)	HSP70 A / B	NS	KC161300 (hsp70)	HSP70 C	145.31-fold
		+14°C (1 h / none)	"	"	"	~3.5-fold	"	"	~55-fold
Ericerus pela (male)	Liu et al. (2014)	+8°C (1 h / none)	whole	KC161301 (hsc70)	HSP70 A / B	NS	KC161300 (hsp70)	HSP70 C	~1.9-fold
		+14°C (1 h / none)	"	"	"	down regulation	"	"	~1.2-fold

(A)

Tribolium castaneum	Altincicek et al. (2008)	+ 11°C (20 min / 16 h)	whole		XM_969349 (hsp68)	HSP70 C	~3-fold

(B)

Figure legends:

Figure 1. Phylogenetic relationships between the Malacostraca cytosolic-HSP70 proteins. Analyses were performed on 84 translated-sequences of 530 to 560 amino-acid length, including the C-terminal region, with four statistical methods. Phylogenetic tree was rooted using Endoplasmic Reticulum HSP70 (ER-HSP70) sequences. The Jones-Taylor-Thornton model was defined as the best-fitting, for our dataset. Node support values for Bayesian inference, Maximum Likelihood, Neighbour-Joining and Maximum Parsimony analyses are indicated between brackets. The phylogenetic tree was based on Bayesian inference, calculated with 3,000,000 generations until the standard deviation of split frequencies was below 0.01. Results were consolidated by Neighbour-joining and Maximum Parsimony analyses (1000 bootstrap replicates), only node values higher than 60 are presented. Sequences from this study were bold and highlighted with the symbol ●. Distinct cyt-HSP70 forms were numbered and boxed according to the corresponding groups, A (blue) and B (red). Heat stress induction-fold values are mapped, according to Table 2, into three categories: NS (statistically not significant changes), * (inferior to 9-fold induction) and ** (superior to 9-fold induction).

Figure 2. Arthropod cytosolic-HSP70s classification using aggregative methodology. Analysis was performed on 426 sequences (327 to 2010 nucleotides in length). Sequences of each major subphylum are represented in different colours: Crustacea (blue), Chelicerata (pink), Hexapoda (green) and Myriapoda (cyan). Resulting sequence order can be read as a trigonometric circle, beginning in the middle of the right part of the rose (see also supplementary Table S1). ER-HSP70 and HSP90 coding sequences were included in the

analysis, to confirm HSP family relationships. Groups were annotated according to the phylogenetic analysis results. Ambiguous sequences of A and B groups are annotated A / B.

Figure 3. 3-Dimensional representation of cytosolic-HSP70 A-2, from *Procambarus clarkii* (KU613184), with associated sequence. 3-D model was constructed with Phyre2 and visualized with the UCSF chimera program. Colours are attributed for the different domains: NBD (blue), Linker (yellow), SBD (green) and CTD (magenta). Discriminating patterns are highlighted in red and indicated with stars.

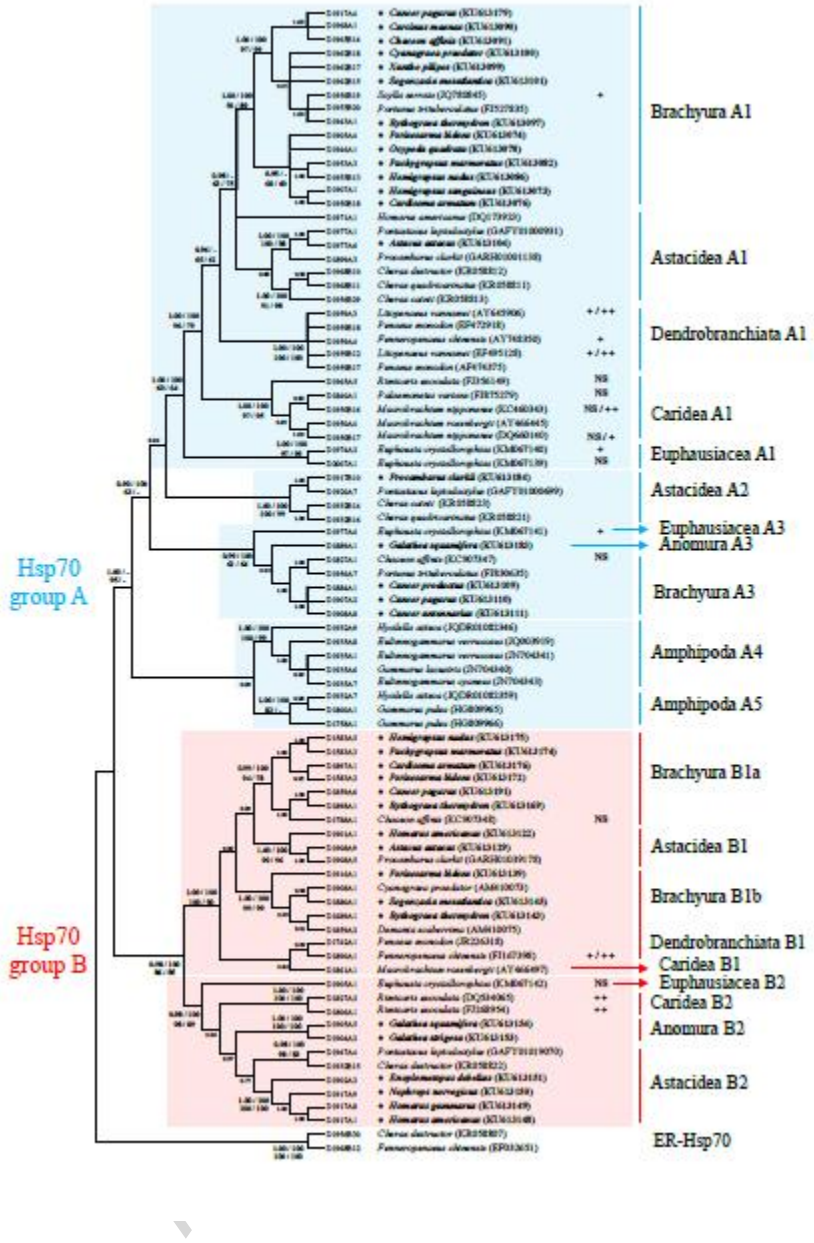


Figure 1

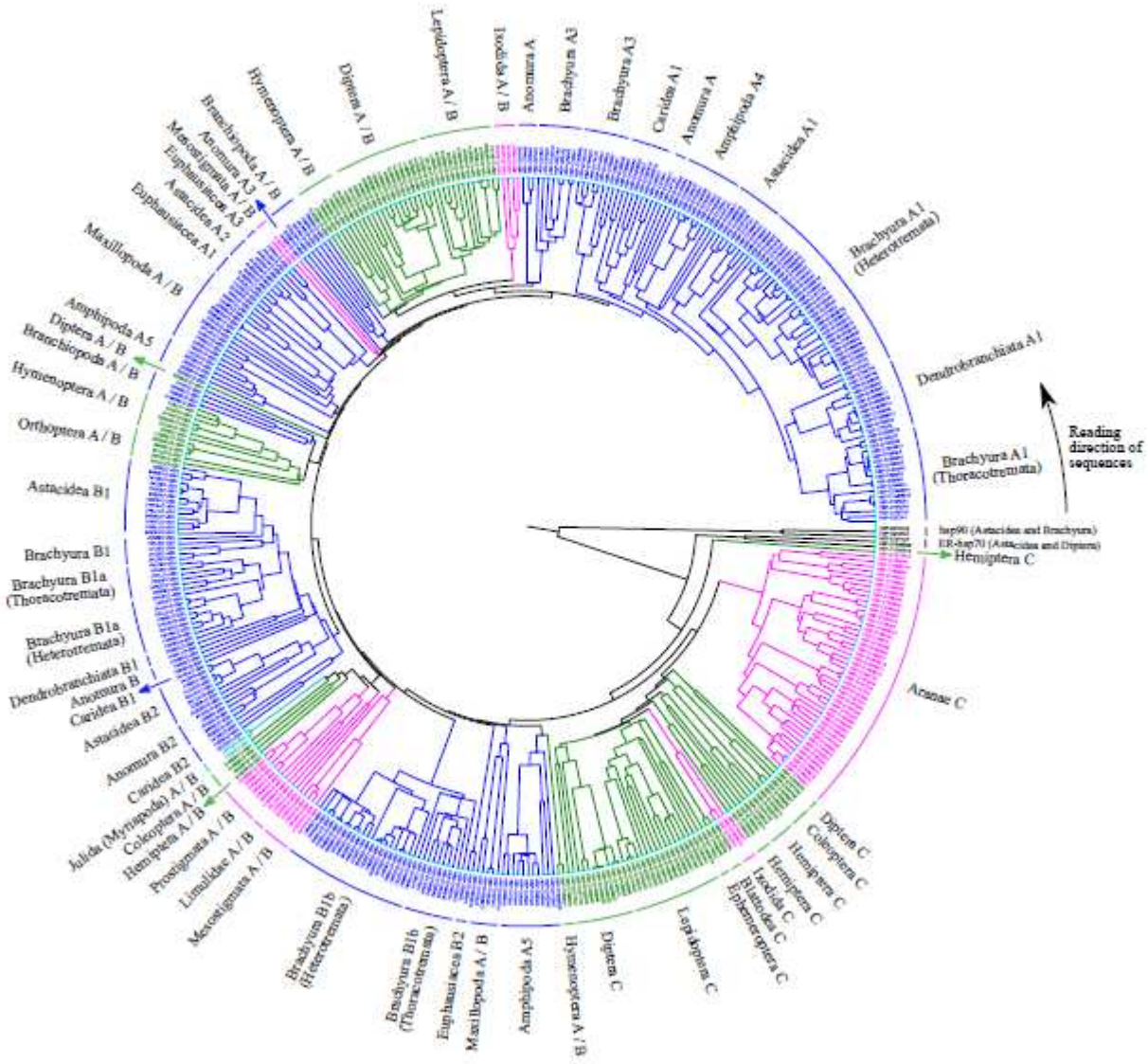
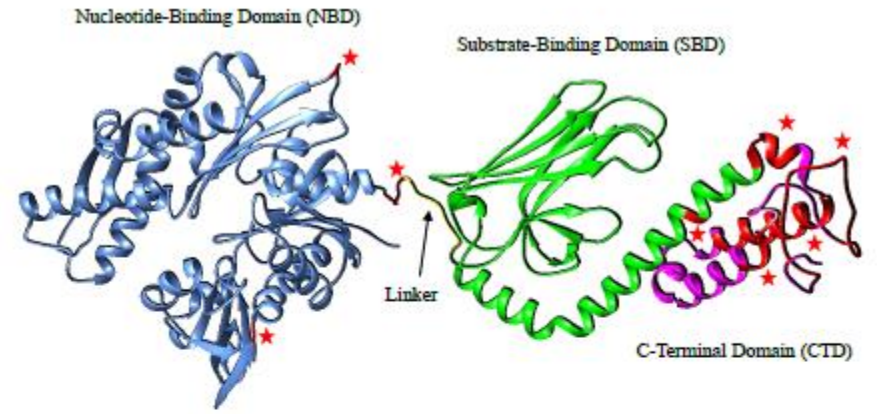


Figure 2



>D1917B10_KU613184
MPKAVGIDLGTTFSCVGVFQHGKVEIIANDQGNRTTPSYVAFTDAERLIGDAAKNQVAMNPNNTVFDAKRLIGRKFNDASVQADMK
HWPFTVISDGGKPKIQIEYKGETKSFYPEEISSMVLIKMKETAEAYLGSTVKDAVITVPAYFNDSQRQATKDAGTISGMNVLRIIN
EPTAAAIAYGLDKKVGGERNVFIFDLGGGTFDVSILPIEDGIFEVKSTAGDTHLGGEDFDNRMVTHFIQEFKRKYKKDMSDNKRAV
RRLRTACERAKRTLSSSTQASIEIDSLYEGVDFYTSITRARFEELCADLFRGALDPVEKSLRDAKMDKSQINEIVMVGGSTRIPKI
QKLLQDFFNGKELNKSINPDEAVAYGAAVQAAILCGDKSEAVQDLLLLDVAPLSLGIETAGGVMTALIKRNTTIPTKQTQTFTTYS
DNQPGVLIQVYEGERAMTKDNNLLGKFELTGIPPAPRGVPQIEVTFDIDANGILNVSAADKSTGKENKITITNDKGRLSKEEIEKM
VQDAEKYKADDENQRERIAAKNSLESYCFNMKSTVEDEKFKDKVSSTDRSKILDACNEAIKWLDSNQSAEKDEFEHKQKEVEQICN
PIITKMYASSGGAPGAPGGAPEGASTGGGPTIEEVD

Figure 3

List of abbreviations

HSP70: Heat shock protein 70 kDa

HSC70: Cognate heat shock protein 70 kDa

Cyt-HSP70: cytosolic heat shock protein 70 kDa

ER-HSP70: endoplasmic reticulum heat shock protein 70 kDa

HSP90: Heat shock protein 90 kDa

NCBI: National Center of Biotechnology Information

NEF: Nucleotide exchange factor

gDNA: genomic DNA

cDNA: complementary DNA

PCR: Polymerase Chain Reaction

ML: Maximum Likelihood

MP: Maximum Parsimony

NJ: Neighbour-Joining

NBD: Nucleotide-Binding Domain

SBD: Substrate-Binding Domain

CTD: C-Terminal Domain