

# Promoter Complexity and Tissue-Specific Expression of Stress Response Components in *Mytilus galloprovincialis*, a Sessile Marine Invertebrate Species

Chrysa Pantzartzi<sup>1</sup>, Elena Drosopoulou<sup>1</sup>, Minas Yiangou<sup>1</sup>, Ignat Drozdov<sup>2,3</sup>, Sophia Tsoka<sup>2</sup>, Christos A. Ouzounis<sup>2,4\*</sup>, Zacharias G. Scouras<sup>1\*</sup>

**1** Department of Genetics, Development & Molecular Biology, School of Biology, Faculty of Sciences, Aristotle University of Thessaloniki, Thessaloniki, Greece, **2** Centre for Bioinformatics, School of Physical Sciences & Engineering, King's College London, London, United Kingdom, **3** BHF Centre of Research Excellence, Cardiovascular Division, School of Medicine, James Black Centre, Denmark Hill Campus, King's College London, London, United Kingdom, **4** Computational Genomics Unit, Institute of Agrobiotechnology, Centre for Research & Technology Hellas, Thessaloniki, Greece

## Abstract

The mechanisms of stress tolerance in sessile animals, such as molluscs, can offer fundamental insights into the adaptation of organisms for a wide range of environmental challenges. One of the best studied processes at the molecular level relevant to stress tolerance is the heat shock response in the genus *Mytilus*. We focus on the upstream region of *Mytilus galloprovincialis* Hsp90 genes and their structural and functional associations, using comparative genomics and network inference. Sequence comparison of this region provides novel evidence that the transcription of Hsp90 is regulated via a dense region of transcription factor binding sites, also containing a region with similarity to the Gamera family of LINE-like repetitive sequences and a genus-specific element of unknown function. Furthermore, we infer a set of gene networks from tissue-specific expression data, and specifically extract an Hsp class-associated network, with 174 genes and 2,226 associations, exhibiting a complex pattern of expression across multiple tissue types. Our results (i) suggest that the heat shock response in the genus *Mytilus* is regulated by an unexpectedly complex upstream region, and (ii) provide new directions for the use of the heat shock process as a biosensor system for environmental monitoring.

**Citation:** Pantzartzi C, Drosopoulou E, Yiangou M, Drozdov I, Tsoka S, et al. (2010) Promoter Complexity and Tissue-Specific Expression of Stress Response Components in *Mytilus galloprovincialis*, a Sessile Marine Invertebrate Species. *PLoS Comput Biol* 6(7): e1000847. doi:10.1371/journal.pcbi.1000847

**Editor:** Philip E. Bourne, University of California San Diego, United States of America

**Received:** December 3, 2009; **Accepted:** June 2, 2010; **Published:** July 8, 2010

**Copyright:** © 2010 Pantzartzi et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** ID holds a pre-doctoral fellowship from the BHF Centre of Excellence, Cardiovascular Division, King's College London. Early stages of this work have been supported in part by the Network of Excellence ENFIN (contract number LSHG-CT-2005-518254) funded by the European Commission. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: christos.ouzounis@kcl.ac.uk (CAO); scouras@bio.auth.gr (ZGS)

## Introduction

The majority of molluscan species go through two principal developmental phases, a larval embryo (motile phase) followed by a clumping structure (sessile phase), when they are permanently attached to an underwater substrate. This lifecycle, common amongst marine invertebrates, poses challenges for adaptation and tolerance for a wide range of conditions at the littoral zone, including steep salinity or temperature gradients. Key model organisms for molluscan biology include species from the genus *Mytilus*, in particular *M. edulis*, *M. galloprovincialis* and *M. californianus*. Crucially, the latter species is a target organism for a genome sequencing project, whose results are eagerly expected by the community (<http://www.jgi.doe.gov/sequencing/why/3090.html>).

The *Mytilus* species group provides an ideal model both for fundamental questions of animal adaptation to stress response, as well as biotechnological applications, primarily as a pollution biosensor [1]. Its use extends into biomimetics [2], in particular protein-based medical adhesives [3], with potential applications in fields such as dentistry [4]. Moreover, its relatively complex developmental structure and higher taxonomic status as an

invertebrate, combined with the fact that it can suffer from mussel haemic neoplasia, renders this organism a potential model for human leukemia and an ideal biomarker for pollution-induced disease [5]. In this context, it is important to understand the mechanisms by which mussels tolerate and cope with environmental stress, given that their behavioral options are highly restricted, due to the sessile phase of their lifecycle.

In the past, comparisons between motility and sessility for higher organisms have been primarily confined to animals versus plants [6], with follow-up studies focusing on comparisons between large animals, e.g. humans, versus large plants, e.g. trees, and the trade-offs for the tree body plan [7]. Less attention has been paid to adaptations by sessile animals, in particular intertidal invertebrates (or, “plant” equivalents) [8–9], and the molecular mechanisms through which they achieve tolerance to stress. One exception is represented by heat shock response, a key factor for temperature adaptation that has been studied in this context to a certain extent [8], and specifically in *Mytilus* with regard to the Hsp70 [10] and Hsp90 [11] genes.

Transcriptional regulation can be achieved either by an extensive repertoire of paralogs and transcription factors (‘gene content strategy’) or a complex structure of promoters (‘gene

## Author Summary

Adaptation of sessile animals, such as molluscs, to stress is achieved by a number of molecular mechanisms, few of which are clearly understood. Insights from this research can provide clues about stress tolerance both for sessile and mobile organisms. The Mediterranean mussel, of the genus *Mytilus*, is a model organism for the study of stress at the molecular level, with sufficient gene structure and function data available. We have thus investigated a key stress response gene, Hsp90, and in particular its upstream region, using a combination of sequence and expression analysis approaches. We demonstrate that this region, responsible for the regulation of heat shock-associated gene expression, exhibits an unparalleled structural and functional complexity compared to other model organisms, as well as subtle gene expression patterns across multiple tissues. These results form the basis upon which the heat shock response can be used as a molecular biosensor for environmental monitoring in the future.

structure strategy'). Analysis of comprehensive datasets has clearly demonstrated that transcription factors (TFs) and transcription-associated proteins (TAPs) are not universally distributed but highly taxon-specific and that relative TF gene content increases with the taxonomic scale [12–13]. Such comparisons have been later extended by follow-up studies that analyzed TAP complements and their expansion rates in plants [14–15]. Thus, it is now known that one way by which plants, sessile organisms *par excellence*, achieve a finer degree of regulation is by the expansion of TF/TAP complements and a 'gene content strategy'. Yet, it is unclear whether similar trends are followed in sessile animals, since entire genome sequences for those are lacking so far, limiting the range of comparative genome-wide studies that can be performed.

As far as paralogs are concerned, recent studies that have focused on the heat shock response in plants, and in particular *Arabidopsis thaliana*, have revealed that the process involves up to 21 known TFs and four heat shock protein (Hsp) families (Hsp20/70/90/100) [16–17]. Despite a cursory resemblance to mammals, in *Drosophila* thermal sensing is achieved by a unique repertoire of genes [18], including thermostat systems not exclusively involving heat shock proteins [19]. In other words, and probably for different reasons, a gene content strategy might prevail in both model organisms for plants (*A. thaliana*) and motile invertebrates (*Drosophila*). Thus, it is worth examining what are the mechanisms through which stress response is regulated in sessile marine invertebrates in general, and the *Mytilus* genus in particular, and which strategy dominates gene expression.

We focus on the Hsp90 family as a case study for stress response in sessile animals and examine the structure and function of the Hsp90 upstream region in *M. galloprovincialis*. Previously, two distinct Hsp90 genes with the same genomic organization have been isolated from *M. galloprovincialis* [11], herein called Mghsp90 genes. Detailed sequence analysis revealed that the two genes contain nine exons and exhibit great similarities in both the 5' non-coding and the coding regions but differ in their 3' non-coding regions, as well as in three introns, due to the presence of repeated sequences [11]. The 5' non-coding region of both genes contains a non-translated exon and multiple binding sites for various transcription factors, highly suggestive of potential interactions of these factors with the Hsp90 promoter and subtle patterns of gene regulation [11].

A comparative analysis of Hsp90 gene content across all taxa with available sequence data has clearly shown that invertebrate genomes contain a relatively small number of Hsp90 genes (3–4 genes), compared to those of vertebrates (>5 genes) [20]. Thus, it appears that the *Mytilus* genome might contain a relatively small number of TFs (e.g. heat shock factors or HSFs – no such factors can be detected in the *Mytilus californianus* EST collection, not shown) and/or Hsp90 genes, raising the question how the expression of Hsp90 and other heat shock genes is regulated in sessile invertebrates.

In the present work, we perform a detailed analysis of the Mghsp90 upstream region in terms of structure and expression, and reveal the presence of previously undetected sequence elements of unknown function. Based on tissue-specific expression data, we also delineate the potential associations of Mghsp90 with another 174 genes that are involved in a complex pattern of expression across tissues. These two discoveries are discussed within the context of existing knowledge and are expected to contribute towards a deeper understanding of the heat shock response in sessile organisms.

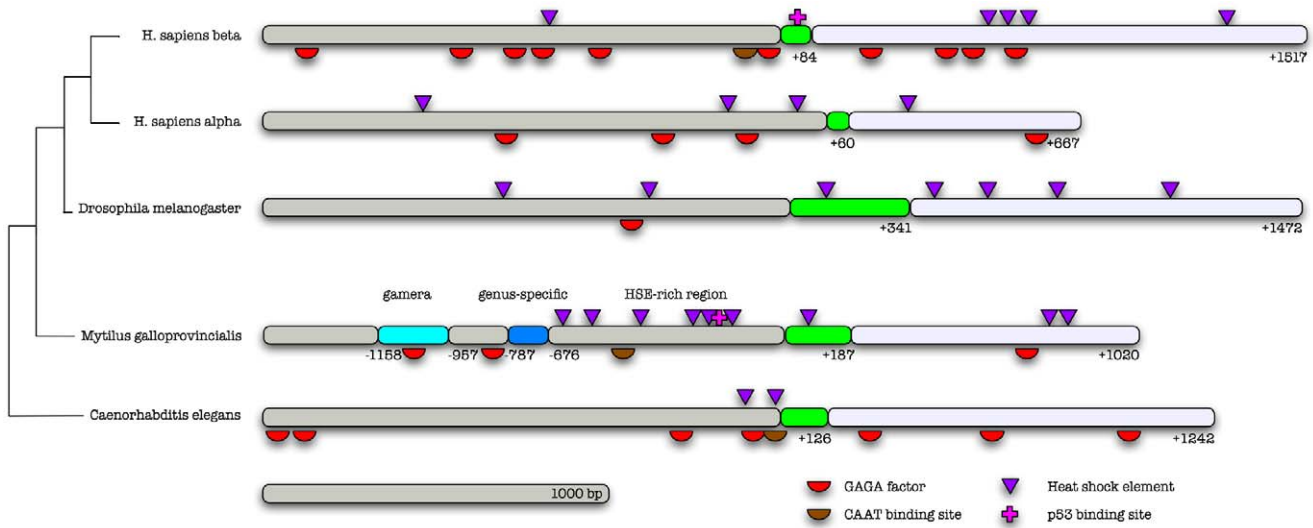
## Results/Discussion

### Comparative analysis of the Mghsp90 upstream region reveals an unexpected complexity

The comparison of the 5' upstream region of Mghsp90 genes to their homologs in two model organisms for which there is extensive genomic evidence and humans reveals an increase of complexity in TF binding sites including heat shock elements (HSEs, binding sites for HSFs – see **Methods**). The *M. galloprovincialis* Hsp90 region exhibits a peculiar degree of unexpected complexity with regard to its phylogenetic context, not only in terms of quantity of predicted elements but also in fine structure of the promoter (**Figure 1**). The *Mytilus* region contains more regulatory sites than the *D. melanogaster* region (namely, 14 sites *vs.* 8), a total count similar to that of the human Hsp90 beta gene (17 sites – **Figure 1**). Moreover, it is host to two newly identified elements (Gamera and a genus-specific sequence), both of unknown function (represented by blue bars, **Figure 1**), followed by a HSE-rich region with a CAAT binding site and a putative p53 binding site (see also below, and **Figure 1 in Protocol S1**).

### Detection of a LINE-like repetitive sequence in the Mghsp90 upstream region

Curiously, upstream of the first exon (1158 nucleotides) of Mghsp90, there exists a 201-base pair (bp) sequence element with a putative GAGA factor binding site (**Figure 1**), 69% identical over 181 nucleotides to the medaka fish *Oryzias curvinotus* LINE-like repetitive sequence Gamera [21]. The similarity extends over positions 1907–2085 of the *O. curvinotus* 4493-bp sequence entry (Genbank accession number AB081572, GI:19570857) and more specifically over the 'open reading frame' b (defined at positions 1353–3052) [21]. Thus, this region of approximately 200 nucleotides is only a fraction of the putative ORF b and, to our knowledge, it is the first time this segment is reported outside the *Oryzias* genus and its closest relatives [21] (**Figure 2**). Moreover, multiple copies of this region can also be identified in the genome of the blood fluke *Schistosoma mansoni* [22] (**Figure 2, Figure 2.1 in Protocol S1**). Fragments of this sequence are also present in (i) the Expressed Sequence Tag (EST) database, more specifically in the neural transcriptome and thus genome of the gastropod *Aplysia californica* [23], the termite *Hodotermopsis sjostedti* [24], the African cichlid fish *Oreochromis niloticus* (Lee et al., unpublished, GI:



**Figure 1. Comparative analysis of the *M. galloprovincialis* sequence 5' region of Hsp90 genes and three animal species, including human.** Dark grey bars correspond to the 5' untranslated region, green bars correspond to the first exon, and light grey bars to the first intron. Two regions found only in the *M. galloprovincialis* Hsp90 genes are marked (in blue). Numbering is relative to the transcription initiation site (first exon). Transcription factor binding sites and heat shock elements in different colors (bottom right) and scale (bottom left) are clearly shown. Relationships of the five genes are shown with a relative dendrogram (not drawn to scale). doi:10.1371/journal.pcbi.1000847.g001

253867024), the mollusc *Lymnaea stagnalis* [25] and the sea anemone *Nematostella vectensis* [26] (in that order of sequence similarity – **Figure 2.2 in Protocol S1**); (ii) the unfinished high-throughput genomic sequence database (**Figure 2.3 in Protocol S1**), in the genome of sea urchin *Strongylocentrotus purpuratus* [27] and (iii) the Whole-Genome-Shotgun Sequence database (**Figure 2.4 in Protocol S1**) in the genome of the hemichordate *Saccoglossus kowalevskii* (unpublished).

The functional significance of this element is not clear, yet given that the region can be identified in at least ten – highly unrelated and primarily aquatic – species, the presence of a transposable element of a highly mobile nature (or its evolutionary relic) is indicated (**Figure 2**). In *M. galloprovincialis*, it has also been shown that mobile elements reside within introns of the Hsp70 genes [10], however there is no detectable sequence similarity between those elements and the Mghsp90 Gamera-like sequence presented here.

**A conserved genus-specific sequence of unknown function**

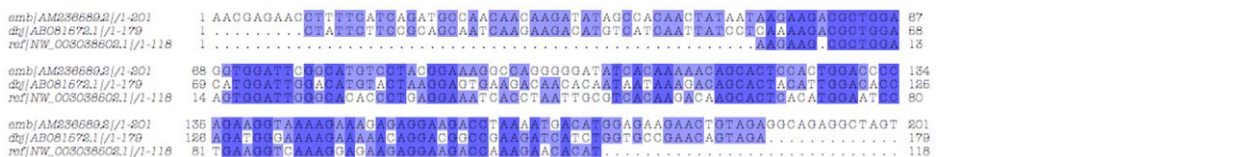
Another feature of the Mghsp90 upstream region is a genus-specific sequence, approximately 100-bp long, located 787 positions before the first exon of Mghsp90 genes (**Figure 1**). This region is much more phylogenetically restricted than the Gamera element, found only in the genus *Mytilus*, namely the *M. galloprovincialis* mytilin B precursor gene [28–29] – (accession

number: AF177540.1, positions 777–815 antisense strand, non-coding region), a lysozyme gene (AF334662.1, positions 1016–1050 sense strand, second intron) [30] and a cDNA (AM878017.1) both from *M. edulis*, and a cDNA sequence from *M. californianus* (GE753693.1) (**Figure 3**). This genus-specific sequence does not contain any transcription factor binding sites (**Figure 1**), thus its functional significance is not known at present.

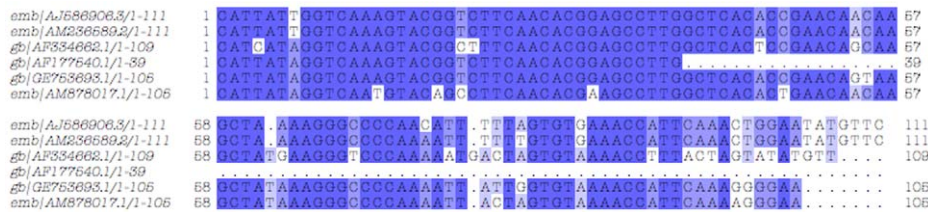
It is worth noting that similarly to the mytilin B gene, another antimicrobial peptide gene, the *M. galloprovincialis* defensin 2 (MGD2) gene, contains a 160-bp long element with similarities to the *M. edulis* lysozyme gene (fourth intron), two glycosidase gene introns (endo-1,4-beta-D-glucanase – AJ308548.1, 2<sup>nd</sup> intron; endo-1,4-mannanase – AJ271365.2, 5<sup>th</sup> intron), the 3'-UTR of the *M. galloprovincialis* Hsp70-1 gene, all being similar to an ISSR sequence (AJ938114), indicating the presence of a transposable element [31]. The above mentioned genes all have catabolic roles and might indeed be connected to defense mechanisms, broadly associated with stress. Further study is required in order to understand the role of these genus-specific sequences in the molecular physiology of the above mentioned loci.

**The putative p53 binding site and the molluscan neoplasia connection**

A putative binding site for p53 is located between two HSEs in the 5' regulatory region of the Mghsp90 genes [11] (**Figure 1**),



**Figure 2. Alignment of the LINE-like upstream region in Mghsp90.** Upstream sequence is identical in both genes, Mghsp90-1 is shown only (AM236589.2), and similar to the LINE-like Gamera element in *O. curvinotus* (AB081572.1) and *S. mansoni* (NW\_003038502.1). Identities in all sequences are represented as dark blue boxes and in two sequences in light blue boxes. Alignment visualized by JalView [91]. Sequence positions are relative to the alignment and do not correspond to those of the database entries. doi:10.1371/journal.pcbi.1000847.g002



**Figure 3. Alignment of the genus-specific upstream region in Mghsp90.** Mghsp90-1/2 genes (AM236589.2 / AJ586906.3 respectively) are similar to the *M. edulis* lysozyme gene (AF334662.1), the *M. galloprovincialis* mytilin B precursor gene (AF177540.1) and two cDNAs from *M. californianus* and *M. edulis* (GE753693.1 / AM878017.1, respectively) Identities in all sequences are represented as dark blue boxes, other identity blocks in light blue boxes. Alignment visualization and sequence positions as in **Figure 2**.

doi:10.1371/journal.pcbi.1000847.g003

being identical to the consensus binding site of human p53 to retinoblastoma susceptibility gene [32]. This binding site is evidently absent from other species, including *C. elegans* and *D. melanogaster*, but present in the human Hsp90 beta gene [33] (**Figure 1**). The p53 proteins from two *Mytilus* species exhibit very high similarity to their human homologs, and especially in the DNA binding domain, the transcriptional activation domain (TAD) and the nuclear localization signal. In addition, residues mutated in various human cancers are also conserved in the *Mytilus* p53 proteins [34]. It should be noted that p53 is phylogenetically restricted to animals while the molluscan versions (Decapodiformes, Bivalvia and *Haliotis* sp.) exhibit a very high similarity to the vertebrate sequences (not shown). The prediction of the p53 binding site in *Mytilus* is based on the known association of p53 with the upstream region of the human Hsp90 beta gene [33], the conservation of the *Mytilus* p53 genes [34] and the observation that an identical site is present in human Hsp90 (an Mghsp90 homolog) [11].

In order to further establish the validity of the predicted p53 binding site in a phylogenetic context, we have searched the non-redundant nucleotide database with the Mghsp90 genes as queries (see **Methods**). We subsequently identified 215 homologous target regions, with the closest sequence-similar entries carefully selected to exclude cDNA clones or partial coding sequences, across a wide taxonomic spectrum (**Figure 1 in Protocol S1**). These sequences were scanned for putative p53 binding sites (732 matches in total, see **Methods**), conditioned on the p53 phylogenetic distribution mentioned above; in other words, sites found in organisms known to encode for p53 were considered as positive cases (727 in total), while the remainder were treated as negative cases (5 in total). Despite well-understood limitations, e.g. the under-representation of certain species in terms of comparable Hsp90 sequence data and the over-representation of others in terms of redundant sequences, it is evident that p53-containing species exhibit a high number of predicted p53 binding sites (primarily chordates), while other organisms (such as fungi or plants), present a sporadic pattern of false positive hits, as expected. The exception in this otherwise consistent picture is the molluscs (Bivalvia and *Haliotis* sp.), having a small number of predicted p53 binding sites (**Figure 1 in Protocol S1**). The shortage of sequence information for molluscs, coupled with a possibly non-canonical sequence motif, leaves the question open for the unambiguous detection and experimental confirmation of the elusive molluscan p53 binding site.

The presence of a putative p53 binding site in the promoter region of the *Mytilus* Hsp90 genes raises questions about the possible implication of Hsp90 proteins in molluscan leukemia. Very recent studies on the association of p53 with heat shock response [35], the differential expression of p53 in mussel haemic neoplasia [5], and the impact of pollutants on p53 expression [36]

underline the potential involvement of p53 in both heat shock response and neoplasia and its irregular similarity to vertebrate homologs [37], as well as its potential use as a marker for environmental monitoring [34]. In other species, namely soft-shell clams, certain results also indicate that environmentally induced alterations in p53 might contribute to leukemia [38–39].

Indeed, expression studies have established that Hsp genes and a p53-like gene are abundant in *M. galloprovincialis* [40], especially in pollutant exposed mussels [41], now searchable through the Mytibase resource [42]. Moreover, there is evidence from proteomics studies that Hsp proteins are expressed in stress conditions and can potentially be used as pollution biomarkers [43–44] or temperature biosensor [45].

#### Differential gene expression analysis consistent with tissue specificity

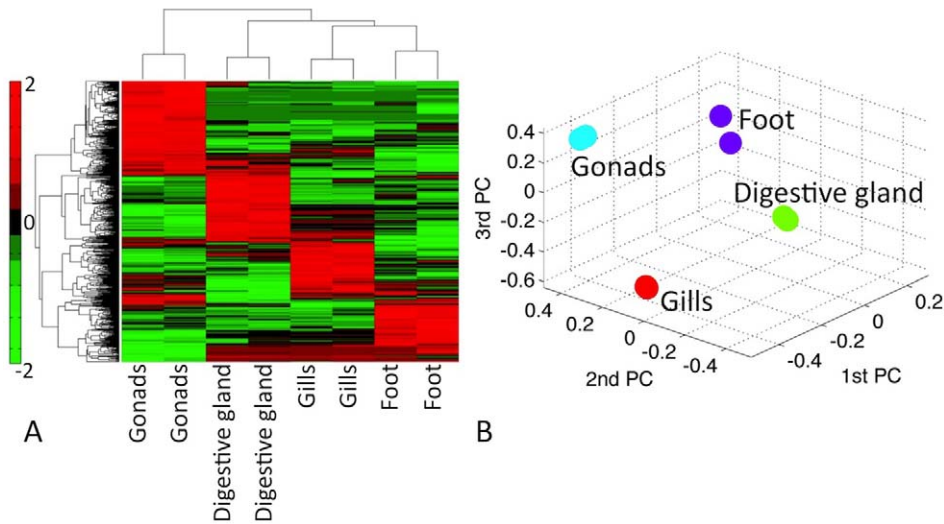
In order to investigate co-expression patterns for Mghsp90 genes, we have extracted tissue-specific gene expression data available in Mytibase, encompassing 3840 cDNA sequences [42]. Following normalization (see **Methods**), we detected 547 genes (14% of total, in the ‘original’ network) that are differentially expressed across all four tissue types under investigation (namely gills, gonads, foot and digestive gland – **Figure 4**).

A two-way clustering across genes and tissues confirms that the four tissue types can be accurately detected (**Figure 4A**). This step also suggests that the 547 differentially expressed genes can be clustered into four distinct classes corresponding to the four tissues, with relatively low overlap (**Figure 4A**). A Principal Component Analysis of the original network further confirms the inter-replicate reproducibility and tissue specificity, indicating the high quality and consistency of the initial gene expression data (**Figure 4B**).

#### Co-expression analysis through network clustering across tissues

To infer gene associations via co-expression profiles, PCCs (see **Methods**) were computed for all possible pair-wise gene permutations of the original network. High PCC values correspond to a large similarity in expression profiles across four tissue types. Only those gene pairs with PCC>0.90 were further considered. This step yielded a global co-expression network, defined as the ‘inferred’ network, containing 3692 nodes and 57697 edges (**Figure 5**). The inferred network represents 96% of all cDNA clones in the original network. Such high coverage may be explained by the limited number of experimental replicates provided in the dataset.

To ensure that only significant associations are considered, MCL clustering (see **Methods**) was performed to produce a ‘clustered’ network with 1719 nodes and 43286 associations (**Table 1**). The clustered network represents a subset of the

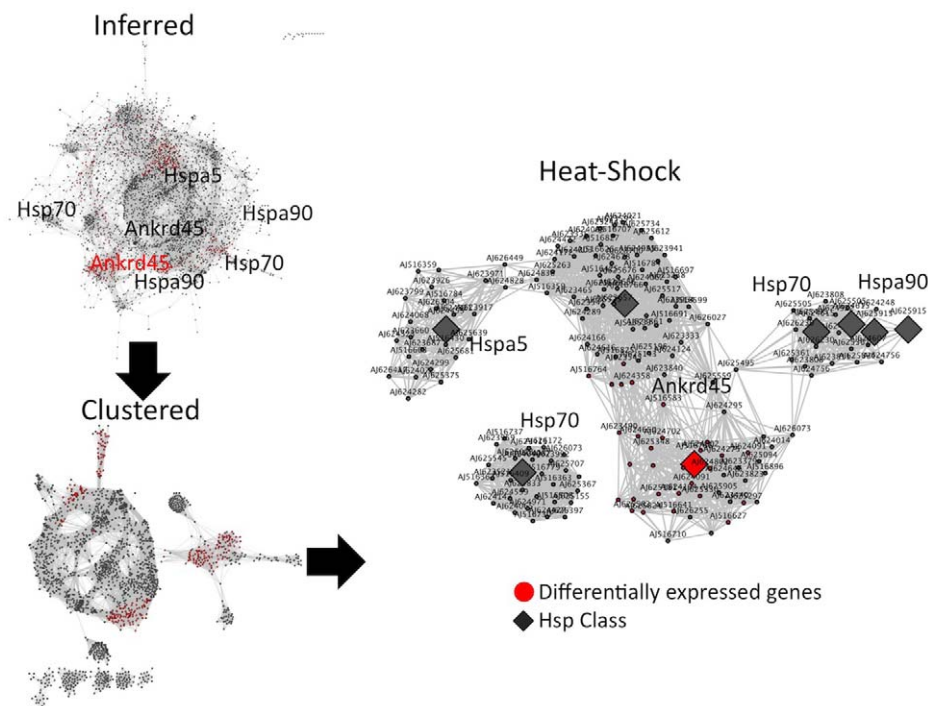


**Figure 4. Differential gene expression patterns of tissue samples from *M. galloprovincialis* under normal conditions.** (A) A two-way clustering identifies 547 differentially expressed genes across two replicates over four normal tissue types. Low-high gene expression values are represented by a green-red color scale; (B) Principal Component Analysis (PCA) of the data further confirms the inter-replicate reproducibility and tissue specificity. In this representation the 1<sup>st</sup>/2<sup>nd</sup>/3<sup>rd</sup> principal components (PC) are mapped onto x-, y-, and z- axis respectively. doi:10.1371/journal.pcbi.1000847.g004

inferred network enriched with the most highly connected genes with the strongest co-expression values (Figure 5). Interestingly, of the 547 differentially expressed genes obtained initially, 271 (~50%) are present in the clustered network, signifying a sufficient coverage of differential expression. This enriched network thus maintains 75% (43286/57697) of network edges, from which more reliable associations can then be extracted.

#### Associations of the Hsp class of genes

To delineate the involvement of the wider Hsp class of genes in normal *M. galloprovincialis* tissue, 8 cDNA sequences corresponding to 4 distinct *Mytilus* Hsp genes, labeled as Hspa5 (Grp78 homolog), Hsp70, Hspa90 (Mghsp90), and Ankrd45 (similar to heat shock 70 KD protein C precursor) were identified in the clustered network (Figure 5). The “Ankrd45”-like sequence (e.g.



**Figure 5. Composite cross-tissue networks of *M. galloprovincialis* co-expression and Hsp gene class associations.** These networks are annotated as ‘differentially expressed genes’ (red) and ‘Hsp gene class’ (grey) (see text and Methods). The gene names for the 4 unique Hsp class members are shown, corresponding to 8 individual cDNA clones. doi:10.1371/journal.pcbi.1000847.g005

**Table 1.** Statistical information for co-expression networks.

Network	Nodes	Edges	Connectivity (avg)	Connectivity (max)	Degree (avg)	Out degree (max)	In degree (max)
Inferred	3692	57697	31.3	120	17.2	118	115
Clustered	1719	43286	50.4	119	25.9	116	113
Hsp	174	2226	25.6	56	13.3	47	44

doi:10.1371/journal.pcbi.1000847.t001

XP\_290882.1) warrants description: its N-terminal part contains ankyrin repeats most similar to the ankyrin repeat domain of the human p53 binding protein [46], while its C-terminal part is similar to Grp78, a homolog of Hsp70 (**Figure 3 in Protocol S1**). Structural evidence indicates that the ankyrin repeats of p53 binding proteins (53BP2) bind to the L2 loop of p53 [47], implicating a configuration of ankyrin repeats such as the one found in Ankrd45, in a potentially mediated p53-Hsp70 domain interaction.

In fact, since the initial discovery that the Hsp70 promoter is regulated by p53 [48], there is mounting evidence that these two proteins are involved in various processes, including oral dysplasia [49], endometrial carcinomas [50], gastric cancers [51], ischemia [52] and wound healing [53]. These interactions have been reviewed elsewhere [54–55]. Similarly, it has been demonstrated that p53 requires the activity of Hsp90s [56] and the structural [57] and biochemical [58] basis of this interaction has been deciphered. In fact, it appears that p53, Hsp70 and Hsp90 are involved in a complex interplay during carcinogenesis [59].

To examine Hsp-related associations in greater detail, the nearest-neighbor members of 8 Hsp cDNA clones were selected, defined as the Hsp network (**Figure 5**). This network contained 174 genes and 2226 associations, accounting for 4.5% of genes in the original network (**Tables 1 and 2 in Protocol S1** – node labels refer to MyArray1.0 identifiers, see **Methods**). The Hsp network contains clones with similarity to perlucin (a biomineralization-associated protein) [60] and the *M. edulis* polyphenolic adhesive protein [61], among others (**Table 1 in Protocol S1**); it is curious that in this set, there is also a clone highly similar to the *M. edulis* gene for endo-1,4-mannanase, discussed above.

Remarkably, 30/547 (5.5%) of differentially expressed genes are found to be co-expressed with the Ankrd45 clone. This suggests that members of the Hsp class are involved in complex transcription patterns across multiple tissue types rather than a single one. Indeed, the closest co-expression neighbors of Mghsp90 are two cDNAs for calreticulin – a calcium-binding chaperone (AJ624756/AJ625361) known to be associated with Hsp proteins [62] (**Figure 6**). Given the high-quality, yet limited data, the gene expression analysis outlined here strongly indicates that the known Hsp-associated genes in *Mytilus* are involved in intricate ways with each other, are possibly controlled by a small number of TFs over a number of tissues and conditions. It is thus possible that a mechanism for heat response might involve a ‘gene structure’ strategy, with few genes involved in a multitude of gene expression pathways.

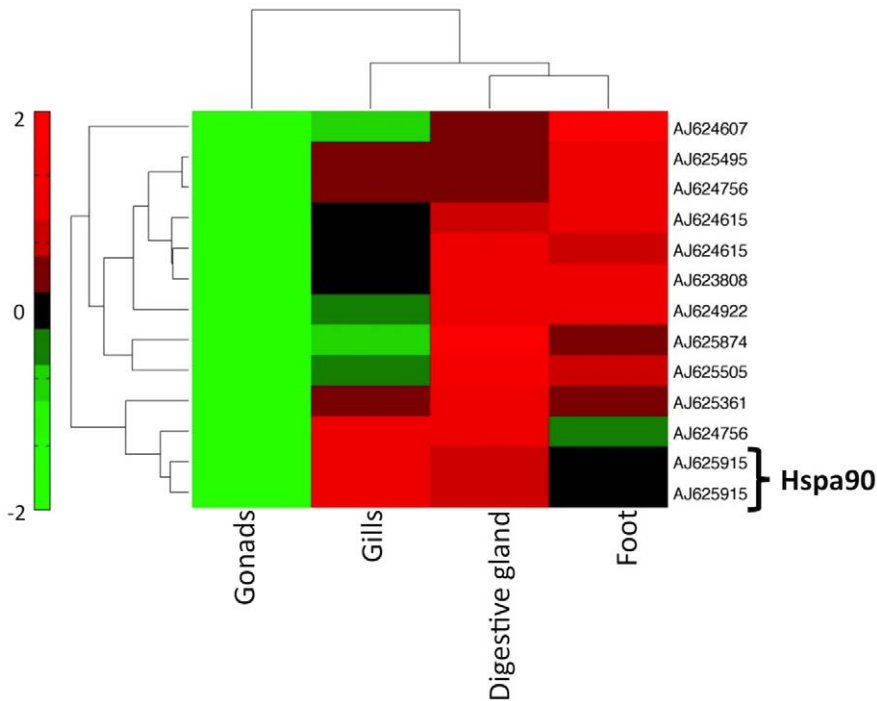
## Conclusion and future perspectives

In this study, we have dissected computationally the upstream region of the Mghsp90 genes to investigate its structure and function. The structural complexity of this region strongly suggests

that the transcription of Hsp90 stress response is tightly regulated via a dense region of heat shock elements and other regions of varying phylogenetic dispersion (**Figure 1**). Compared to other model organisms, such as *C. elegans* and *D. melanogaster*, this regulation appears to be achieved through a ‘gene structure’ strategy, i.e. a complex gene structure. In addition, expression analysis of the heat shock response indicates that a handful of key molecules belonging to the heat-shock class, exhibit a differential tissue-specific expression profile, possibly in gills and the digestive gland, while at the same time maintaining a multitude of associations through a complex co-expression network (**Figure 5**). Our results are consistent with current knowledge about chaperone function both within molecular [63–64] and ecological contexts [65–66], and demonstrate the efficacy of both comparative genomics and systems biology for the elucidation of complex relationships between genotype, environment and phenotype.

The nature of sessile animals, with the *Mytilus* genus as a model organism, can shed light into their metabolic capabilities [67], stress responses [68] and resilience of evolutionary extinction [69]. The stress response for sessile animals is of particular interest, especially in cases where different ecological niches can be compared for close relatives, e.g. different growth potential in varying hydrostatic pressure or temperature [70]. Heat shock proteins in particular are used as indicators of thermal stress [68]; for instance, in the case of marine snails (*Tegula* genus), the time course and magnitude of the heat shock response was measured in a field study by monitoring the synthesis of heat shock proteins [71]. In another field study on the Oregon coast, *M. californianus* and its predator *Pisaster ochraceus* were examined for production of the Hsp70 heat shock proteins; it was found that while mussels (a sessile species) have an increased production of Hsp70, its starfish predators (a mobile species) do not, potentially exhibiting decreased heat shock adaptation compared to their prey [72]. Sessile marine invertebrates have been studied in the context of rising sea temperatures, including *M. edulis* [73] and *Rhopaloeides odorabile*, a common Great Barrier Reef sponge [74].

In the future, the thermal ecology of stress response can potentially inform policy decisions for environmental management in the context of climate change [75] – including the analysis of biogeographical range shifts [76], particularly important for sessile animals, the understanding of complex prey-predator interactions e.g. the above mentioned pair of *P. ochraceus* and *M. californianus* [77], and instigate a more integrated approach that will eventually include both weather records and niche-level measurements [78]. Currently, more established approaches for the use of *Mytilus* relate to its use as a biosensor system for the environmental monitoring of coastal water pollution [79], heavy metals or organic pollutants [80] – including manufactured substances such as fiberglass [81]. In conclusion, this work forms a basis upon which the stress response in *Mytilus* will be better understood at the molecular level.



**Figure 6. Hierarchical clustering of 10 unique probes corresponding to genes co-expressed with Hspa90.** Expression of Hspa90 appears to be elevated in gills and digestive gland and down-regulated in gonads. The change in expression is not statistically significant ( $p > 0.05$ ). doi:10.1371/journal.pcbi.1000847.g006

## Methods

### Sequence comparison

The *M. galloprovincialis* Hsp90 sequence was analyzed as previously [11]. The Hsp90 upstream regions from three other representative animal species, namely *Caenorhabditis elegans*, *Drosophila melanogaster* and *Homo sapiens*, were analyzed in a similar fashion. Previously published data concerning Hsp90 genes from these species were also taken into consideration for annotation purposes [82–88]. Sequence database searches were performed by BLAST (for nucleotide sequences, blastn, version 2.2.22) [89]. Sequence alignments were computed using ClustalW [90] and visualized by JalView [91].

### Promoter analysis

Regulatory elements, in the 5' non-coding regions of the Mghsp90 genes were identified with Alibaba2 [92], P-MATCH [93] and the Transcription Element Search System (TESS) [94]. An extensive comparative analysis for p53 binding sites was performed using the Matrix Search analysis tool of the TRED database [95], scanning query sequences against the p53-specific sequence matrix (cut-off score 2) from the JASPAR collection [96].

### Expression analysis

Data were obtained from the Gene Expression Omnibus (GEO) database, representing one spotted cDNA array (Accession number: GSE2176) for gene expression in normal mussels (*M. galloprovincialis*), using the MyArray 1.0 platform targeting 1712 clones with a total of 3840 cDNA sequences [42]. This dataset encompasses the total RNA isolated from gills ( $n = 2$ ), gonads ( $n = 2$ ), foot ( $n = 2$ ), and digestive gland ( $n = 2$ ) [42]. Data normalization was performed by taking the binary logarithm ( $\log_2$ ) of normalized intensities (defined as test signal/reference

signal). Normalized data ('original' network) was subsequently subjected to statistical validation.

### Network modeling

Gene associations were identified by computing the Pearson Correlation Coefficient (PCC) for all gene pairs in the raw network. Gene pairs with positive correlations indicated by a  $PCC > 0.90$  were considered to be co-expressed. Co-expression patterns were represented as networks where each node corresponds to a unique gene and each edge represents a co-expression association. The final network ('inferred' network) was clustered using the Markov Clustering Algorithm (MCL) in order to both filter noisy associations and identify biologically meaningful clusters ('clustered' network), as previously described [97]. The inflation parameter for MCL was set to 3.0. Only clusters with  $> 10$  genes were further analyzed for biologically meaningful associations.

### Statistical analysis

Differential expression analysis was performed by applying Analysis of Variance (ANOVA) to all genes across four distinct tissue types. Only those genes with the overall p-value below 0.05 were considered as differentially expressed. Two-way unsupervised hierarchical clustering of differentially expressed gene signals was performed using Euclidean distance as a similarity measure. Principal component analysis (PCA) was also performed to confirm the validity of the analysis for the four tissue-specific datasets. All statistical analyses were performed with MATLAB (The MathWorks, Natick, MA – www.mathworks.com).

### Supporting Information

**Protocol S1** 18 Supplement files plus an index file: 3 Supplementary figures, 2 Supplementary tables - referenced in

text as Protocol S1; index provided with an explanation of the directory contents.

Found at: doi:10.1371/journal.pcbi.1000847.s001 (5.18 MB ZIP)

## Acknowledgments

We wish to thank Prof. Ben J. Blencowe (University of Toronto) for critical reading of the manuscript and valuable suggestions for future research.

## References

- Shugart LR, McCarthy JF, Halbrook RS (1992) Biological markers of environmental and ecological contamination: an overview. *Risk Anal* 12: 353–360.
- Silverman HG, Roberto FF (2007) Understanding marine mussel adhesion. *Mar Biotechnol* (NY) 9: 661–681.
- Strausberg RL, Link RP (1990) Protein-based medical adhesives. *Trends Biotechnol* 8: 53–57.
- Holten-Andersen N, Waite JH (2008) Mussel-designed protective coatings for compliant substrates. *J Dent Res* 87: 701–709.
- Muttray AF, Schulte PM, Baldwin SA (2008) Invertebrate p53-like mRNA isoforms are differentially expressed in mussel haemic neoplasia. *Marine Env Res* 66: 412–421.
- Bradshaw AD (1972) Some of the evolutionary consequences of being a plant. *Evol Biol* 5: 25–47.
- Petit RJ, Hampe A (2006) Some evolutionary consequences of being a tree. *Annu Rev Ecol Evol Syst* 37: 187–214.
- Huey RB, Carlson M, Crozier L, Frazier M, Hamilton H, et al. (2002) Plants versus animals: do they deal with stress in different ways? *Integ Comp Biol* 42: 415–423.
- Borges RM (2005) Do plants and animals differ in phenotypic plasticity? *J Biosci* 30: 41–50.
- Kourtidis A, Drosopoulou E, Pantartzis CN, Chintiroglou CC, Scouras ZG (2006) Three new satellite sequences and a mobile element found inside HSP70 introns of the Mediterranean mussel (*Mytilus galloprovincialis*). *Genome* 49: 1451–1458.
- Pantartzis CN, Kourtidis A, Drosopoulou E, Yianguo M, Scouras ZG (2009) Isolation and characterization of two cytoplasmic hsp90s from *Mytilus galloprovincialis* (Mollusca: Bivalvia) that contain a complex promoter with a p53 binding site. *Gene* 431: 47–54.
- Coulson RM, Enright AJ, Ouzounis CA (2001) Transcription-associated protein families are primarily taxon-specific. *Bioinformatics* 17: 95–97.
- Coulson RM, Ouzounis CA (2003) The phylogenetic diversity of eukaryotic transcription. *Nucleic Acids Res* 31: 653–660.
- Shiu SH, Shih MC, Li WH (2005) Transcription factor families have much higher expansion rates in plants than in animals. *Plant Physiol* 139: 18–26.
- Richardt S, Lang D, Reski R, Frank W, Rensing SA (2007) PlanTAPDB, a phylogeny-based resource of plant transcription-associated proteins. *Plant Physiol* 143: 1452–1466.
- Swindell WR, Huebner M, Weber AP (2007) Plastic and adaptive gene expression patterns associated with temperature stress in *Arabidopsis thaliana*. *Heredity* 99: 143–150.
- Swindell WR, Huebner M, Weber AP (2007) Transcriptional profiling of *Arabidopsis* heat shock proteins and transcription factors reveals extensive overlap between heat and non-heat stress response pathways. *BMC Genomics* 8: 125.
- Dillon ME, Wang G, Garrity PA, Huey RB (2009) Thermal preference in *Drosophila*. *J Thermal Biol* 34: 109–119.
- Hamada FN, Rosenczweig M, Kang K, Pulver SR, Ghezzi A, et al. (2008) An internal thermal sensor controlling temperature preference in *Drosophila*. *Nature* 454: 217–220.
- Chen B, Zhong D, Monteiro A (2006) Comparative genomics and evolution of the HSP90 family of genes across all kingdoms of organisms. *BMC Genomics* 7: 156.
- Koga A, Hori H, Ishikawa Y (2002) Gamera, a family of LINE-like repetitive sequences widely distributed in medaka and related fishes. *Heredity* 89: 446–452.
- Berriman M, Haas BJ, LoVerde PT, Wilson RA, Dillon GP, et al. (2009) The genome of the blood fluke *Schistosoma mansoni*. *Nature* 460: 352–358.
- Moroz LL, Edwards JR, Puthanveetil SV, Kohn AB, Ha T, et al. (2006) Neuronal transcriptome of *Aplysia*: neuronal compartments and circuitry. *Cell* 127: 1453–1467.
- Yuki M, Moriya S, Inoue T, Kudo T (2008) Transcriptome analysis of the digestive organs of *Hodotermopsis sjostedti*, a lower termite that hosts mutualistic microorganisms in its hindgut. *Zool Sci* 25: 401–406.
- Feng ZP, Zhang Z, van Kesteren RE, Straub VA, van Nierop P, et al. (2009) Transcriptome analysis of the central nervous system of the mollusc *Lymnaea stagnalis*. *BMC Genomics* 10: 451.
- Putnam NH, Srivastava M, Hellsten U, Dirks B, Chapman J, et al. (2007) Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science* 317: 86–94.
- Sodergren E, Weinstock GM, Davidson EH, Cameron RA, Gibbs RA, et al. (2006) The genome of the sea urchin *Strongylocentrotus purpuratus*. *Science* 314: 941–952.
- Mitta G, Hubert F, Dyrzynda EA, Boudry P, Roch P (2000) Mytilin B and MGD2, two antimicrobial peptides of marine mussels: gene structure and expression analysis. *Dev Comp Immunol* 24: 381–393.
- Parisi MG, Li H, Toubiana M, Parrinello N, Cammarata M, et al. (2009) Polymorphism of mytilin B mRNA is not translated into mature peptide. *Mol Immunol* 46: 384–392.
- Bachali S, Jager M, Hassanin A, Schoentgen F, Jolles P, et al. (2002) Phylogenetic analysis of invertebrate lysozymes and the evolution of lysozyme function. *J Mol Evol* 54: 652–664.
- Boon E, Faure MF, Bierne N (2009) The flow of antimicrobial peptide genes through a genetic barrier between *Mytilus edulis* and *M. galloprovincialis*. *J Mol Evol* 68: 461–474.
- Shiio Y, Yamamoto T, Yamaguchi N (1992) Negative regulation of Rb expression by the p53 gene product. *Proc Natl Acad Sci USA* 89: 5206–5210.
- Zhang Y, Wang JS, Chen LL, Cheng XK, Heng FY, et al. (2004) Repression of hsp90beta gene by p53 in UV irradiation-induced apoptosis of Jurkat cells. *J Biol Chem* 279: 42545–42551.
- Muttray AF, Cox RL, St-Jean S, van Poppelen P, Reinisch CL, et al. (2005) Identification and phylogenetic comparison of p53 in two distinct mussel species (*Mytilus*). *Comp Biochem Physiol C* 140: 237–250.
- Murray-Zmijewski F, Slee EA, Lu X (2008) A complex barcode underlies the heterogeneous response of p53 to stress. *Nat Rev Mol Cell Biol* 9: 702–712.
- Banni M, Negri A, Rebelo M, Rapallo F, Boussetta H, et al. (2009) Expression analysis of the molluscan p53 protein family mRNA in mussels (*Mytilus* spp.) exposed to organic contaminants. *Comp Biochem Physiol C* 149: 414–418.
- Stifanic M, Micic M, Ramsak A, Blaskovic S, Ruso A, et al. (2009) p63 in *Mytilus galloprovincialis* and p53 family members in the phylum Mollusca. *Comp Biochem Physiol B Biochem Mol Biol* 154: 264–273.
- Barker CM, Calvert RJ, Walker CW, Reinisch CL (1997) Detection of mutant p53 in clam leukemia cells. *Exp Cell Res* 232: 240–245.
- Walker CW, Bottger SA (2008) A naturally occurring cancer with molecular connectivity to human diseases. *Cell Cycle* 7: 2286–2289.
- Venier P, Pallavicini A, De Nardi B, Lanfranchi G (2003) Towards a catalogue of genes transcribed in multiple tissues of *Mytilus galloprovincialis*. *Gene* 314: 29–40.
- Dondero F, Piacentini L, Marsano F, Rebelo M, Vergani L, et al. (2006) Gene transcription profiling in pollutant exposed mussels (*Mytilus* spp.) using a new low-density oligonucleotide microarray. *Gene* 376: 24–36.
- Venier P, De Pitta C, Bernante F, Varotto L, De Nardi B, et al. (2009) MytiBase: a knowledgebase of mussel (*M. galloprovincialis*) transcribed sequences. *BMC Genomics* 10: 72.
- Hamer B, Hamer DP, Müller WEG, Batel R (2004) Stress-70 proteins in marine mussel *Mytilus galloprovincialis* as biomarkers of environmental pollution: a field study. *Environ Int* 30: 873–882.
- Hamer B, Pavicic-Hamer D, Müller WEG, Zahn RK, Batel R (2005) Detection of stress-70 proteins of mussel *Mytilus galloprovincialis* using 2-D gel electrophoresis: a proteomics approach. *Fresenius Env Bull* 14: 605–611.
- Anestis A, Lazou A, Portner HO, Michaelidis B (2007) Behavioral, metabolic, and molecular stress responses of marine bivalve *Mytilus galloprovincialis* during long-term acclimation at increasing ambient temperature. *Am J Physiol Regul Integr Comp Physiol* 293: R911–921.
- FitzGerald JE, Grenon M, Lowndes NF (2009) 53BP1: function and mechanisms of focal recruitment. *Biochem Soc Trans* 37: 897–904.
- Gorina S, Pavletich NP (1996) Structure of the p53 tumor suppressor bound to the ankyrin and SH3 domains of 53BP2. *Science* 274: 1001–1005.
- Agoff SN, Hou J, Linzer DI, Wu B (1993) Regulation of the human hsp70 promoter by p53. *Science* 259: 84–87.
- Kaur J, Srivastava A, Ralhan R (1996) p53-HSP70 complexes in oral dysplasia and cancer: potential prognostic implications. *Eur J Cancer B Oral Oncol* 32B: 45–49.
- Nanbu K, Konishi I, Komatsu T, Mandai M, Yamamoto S, et al. (1996) Expression of heat shock proteins HSP70 and HSP90 in endometrial carcinomas. Correlation with clinicopathology, sex steroid receptor status, and p53 protein expression. *Cancer* 77: 330–338.
- Machara Y, Oki E, Abe T, Tokunaga E, Shibahara K, et al. (2000) Overexpression of the heat shock protein HSP70 family and p53 protein and prognosis for patients with gastric cancer. *Oncology* 58: 144–151.

## Author Contributions

Conceived and designed the experiments: CP ID ST CAO ZGS. Performed the experiments: CP ID CAO. Analyzed the data: CP ED MY ID ST CAO. Contributed reagents/materials/analysis tools: CP ED MY ST CAO ZGS. Wrote the paper: CP ED MY ID CAO ZGS. Coordinated study: CAO ZGS.



52. Laubriet A, Fantini E, Assem M, Cordelet C, Teyssier JR, et al. (2001) Changes in HSP70 and P53 expression are related to the pattern of electromechanical alterations in rat cardiomyocytes during simulated ischemia. *Mol Cell Biochem* 220: 77–86.
53. Barrow RE, Dasu MR (2005) Oxidative and heat stress gene changes in hypertrophic scar fibroblasts stimulated with interleukin-1beta. *J Surg Res* 126: 59–65.
54. Zyllicz M, King FW, Wawrzynow A (2001) Hsp70 interactions with the p53 tumour suppressor protein. *EMBO J* 20: 4634–4638.
55. Walerych D, Olszewski MB, Gutkowska M, Helwak A, Zyllicz M, et al. (2009) Hsp70 molecular chaperones are required to support p53 tumor suppressor activity under stress conditions. *Oncogene*: Sep 14.
56. Blagosklonny MV, Toretsky J, Bohlen S, Neckers L (1996) Mutant conformation of p53 translated in vitro or in vivo requires functional HSP90. *Proc Natl Acad Sci USA* 93: 8379–8383.
57. Rudiger S, Freund SM, Veprintsev DB, Fersht AR (2002) CRINEPT-TROSY NMR reveals p53 core domain bound in an unfolded form to the chaperone Hsp90. *Proc Natl Acad Sci USA* 99: 11085–11090.
58. Walerych D, Kudla G, Gutkowska M, Wawrzynow B, Muller L, et al. (2004) Hsp90 chaperones wild-type p53 tumor suppressor protein. *J Biol Chem* 279: 48836–48845.
59. Muller P, Hrstka R, Coomber D, Lane DP, Vojtesek B (2008) Chaperone-dependent stabilization and degradation of p53 mutants. *Oncogene* 27: 3371–3383.
60. Weiss IM, Kaufmann S, Mann K, Fritz M (2000) Purification and characterization of perlucin and perlustrin, two new proteins from the shell of the mollusc *Haliotis laevis*. *Biochem Biophys Res Commun* 267: 17–21.
61. Filipula DR, Lee SM, Link RP, Strausberg SL, Strausberg RL (1990) Structural and functional repetition in a marine mussel adhesive protein. *Biotechnol Prog* 6: 171–177.
62. Diz AP, Dudley E, MacDonald BW, Pina B, Kenchington EL, et al. (2009) Genetic variation underlying protein expression in eggs of the marine mussel *Mytilus edulis*. *Mol Cell Proteomics* 8: 132–144.
63. Zhao R, Davey M, Hsu YC, Kaplanek P, Tong A, et al. (2005) Navigating the chaperone network: an integrative map of physical and genetic interactions mediated by the hsp90 chaperone. *Cell* 120: 715–727.
64. Gong Y, Kakihara Y, Krogan N, Greenblatt J, Emili A, et al. (2009) An atlas of chaperone-protein interactions in *Saccharomyces cerevisiae*: implications to protein folding pathways in the cell. *Mol Syst Biol* 5: 275.
65. Hofmann GE, Buckley BA, Place SP, Zippay ML (2002) Molecular chaperones in ectothermic marine animals: biochemical function and gene expression. *Integ Comp Biol* 42: 808–814.
66. Hofmann GE (2005) Patterns of Hsp gene expression in ectothermic marine organisms on small to large biogeographic scales. *Integ Comp Biol* 45: 247–255.
67. Cavalier-Smith T (1992) Origins of secondary metabolism. *CIBA Found Symp* 171: 64–80.
68. Dahlhoff EP (2004) Biochemical indicators of stress and metabolism: applications for marine ecological studies. *Ann Rev Physiol* 66: 183–207.
69. Roy K, Hunt G, Jablonski D (2009) Phylogenetic conservatism of extinctions in marine bivalves. *Science* 325: 733–737.
70. Mestre NC, Thatje S, Tyler PA (2009) The ocean is not deep enough: pressure tolerances during early ontogeny of the blue mussel *Mytilus edulis*. *Proc Biol Sci* 276: 717–726.
71. Tomanek L, Somero GN (2000) Time course and magnitude of synthesis of heat-shock proteins in congeneric marine snails (Genus *tegula*) from different tidal heights. *Physiol Biochem Zool* 73: 249–256.
72. Petes LE, Mouchka ME, Milston-Clements RH, Momoda TS, Menge BA (2008) Effects of environmental stress on intertidal mussels and their sea star predators. *Oecologia* 156: 671–680.
73. Jones SJ, Mieszkowska N, Wethey DS (2009) Linking thermal tolerances and biogeography: *Mytilus edulis* (L.) at its southern limit on the east coast of the United States. *Biol Bull* 217: 73–85.
74. Whalan S, Ettinger-Epstein P, de Nys R (2008) The effect of temperature on larval pre-settlement duration and metamorphosis for the sponge, *Rhopaloeides odorabile*. *Coral Reefs* 27: 783–786.
75. Clarke A (2003) Costs and consequences of evolutionary temperature adaptation. *Trends Ecol Evol* 18: 573–581.
76. Tomanek L (2008) The importance of physiological limits in determining biogeographical range shifts due to global climate change: the heat-shock response. *Physiol Biochem Zool* 81: 709–717.
77. Broitman BR, Szathmary PL, Mislán KAS, Blanchette CA, Helmuth B (2009) Predator-prey interactions under climate change: the importance of habitat vs body temperature. *Oikos* 118: 219–224.
78. Helmuth B (2009) From cells to coastlines: how can we use physiology to forecast the impacts of climate change? *J Exp Biol* 212: 753–760.
79. Venier P, De Pitta C, Pallavicini A, Marsano F, Varotto L, et al. (2006) Development of mussel mRNA profiling: Can gene expression trends reveal coastal water pollution? *Mutat Res* 602: 121–134.
80. Kaloyianni M, Dailianis S, Chrisikopoulou E, Zannou A, Koutsogiannaki S, et al. (2009) Oxidative effects of inorganic and organic contaminants on haemolymph of mussels. *Comp Biochem Physiol C* 149: 631–639.
81. Galimany E, Ramon M, Delgado M (2009) First evidence of fiberglass ingestion by a marine invertebrate (*Mytilus galloprovincialis* L.) in a N.W. Mediterranean estuary. *Mar Pollut Bull* 58: 1334–1338.
82. Blackman RK, Meselson M (1986) Interspecific nucleotide sequence comparisons used to identify regulatory and structural features of the *Drosophila* hsp82 gene. *J Mol Biol* 188: 499–515.
83. Hickey E, Brandon SE, Smale G, Lloyd D, Weber LA (1989) Sequence and regulation of a gene encoding a human 89-kilodalton heat shock protein. *Mol Cell Biol* 9: 2615–2626.
84. Inoue T, Takamura K, Yamae H, Ise N, Kawakami M, et al. (2003) *Caenorhabditis elegans* DAF-21 (HSP90) is characteristically and predominantly expressed in germline cells: spatial and temporal analysis. *Dev Growth Differ* 45: 369–376.
85. Rebbe NF, Ware J, Bertina RM, Modrich P, Stafford DW (1987) Nucleotide sequence of a cDNA for a member of the human 90-kDa heat-shock protein family. *Gene* 53: 235–245.
86. Rebbe NF, Hickman WS, Ley TJ, Stafford DW, Hickman S (1989) Nucleotide sequence and regulation of a human 90-kDa heat shock protein gene. *J Biol Chem* 264: 15006–15011.
87. Shen Y, Liu J, Wang X, Cheng X, Wang Y, et al. (1997) Essential role of the first intron in the transcription of hsp90beta gene. *FEBS Lett* 413: 92–98.
88. Zhang SL, Yu J, Cheng XK, Ding L, Heng FY, et al. (1999) Regulation of human hsp90alpha gene expression. *FEBS Lett* 444: 130–135.
89. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215: 403–410.
90. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23: 2947–2948.
91. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ (2009) Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25: 1189–1191.
92. Grabe N (2002) AliBaba2: context specific identification of transcription factor binding sites. *In Silico Biol* 2: S1–15.
93. Chekmenev DS, Haid C, Kel AE (2005) P-Match: transcription factor binding site search by combining patterns and weight matrices. *Nucleic Acids Res* 33: W432–437.
94. Schug J (2008) Using TESS to predict transcription factor binding sites in DNA sequence. *Curr Protoc Bioinformatics* Chapter 2, pp Unit 2.6.
95. Jiang C, Xuan Z, Zhao F, Zhang MQ (2007) TRED: a transcriptional regulatory element database, new entries and other development. *Nucleic Acids Res* 35: D137–140.
96. Vlieghe D, Sandelin A, De Bleser PJ, Vleminx K, Wasserman WW, et al. (2006) A new generation of JASPAR, the open-access repository for transcription factor binding site profiles. *Nucleic Acids Res* 34: D95–97.
97. Enright AJ, Van Dongen S, Ouzounis CA (2002) An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res* 30: 1575–1584.