Title

Multiple Modes of Adaptation: Regulatory and Structural Evolution in a Small Heat Shock Protein Gene.

Journal

Molecular Biology and Evolution, 35(9)

ISSN

0737-4038

Authors

Tangwancharoen, Sumaetee
Moy, Gary W
Burton, Ronald S

Publication Date

2018-09-01

DOI

10.1093/molbev/msy138

Peer reviewed

Multiple modes of adaptation*:* regulatory and structural evolution in a small heat shock protein gene

Authors

Sumaetee Tangwancharoen, Gary W. Moy and Ronald S. Burton

Marine Biology Research Division
Scripps Institution of Oceanography
University of California, San Diego
9500 Gilman Drive
La Jolla, CA 92093-0202
U.S.A

Correspondence

Sumaetee Tangwancharoen (stangwan@ucsd.edu)

Abstract

Thermal tolerance is a key determinant of species distribution. Despite much study, the genetic basis of adaptive evolution of thermal tolerance, including the relative contributions of transcriptional regulation versus protein evolution, remains unclear. Populations of the intertidal copepod *Tigriopus californicus* are adapted to local thermal regimes across their broad geographic range. Upon thermal stress, adults from a heat tolerant southern population (SD) upregulate several heat shock proteins (HSPs) to higher levels than those from a less tolerant northern population (SC). Suppression of a specific HSP, HSPB1, significantly reduces *T. californicus* survival following acute heat stress. Sequencing of HSPB1 revealed population specific nucleotide substitutions in both promoter and coding regions of the gene. HSPB1 promoters from heat tolerant populations contain two canonical heat shock elements (HSEs), the binding sites for heat shock transcription factor (HSF), while less tolerant populations have mutations in these conserved motifs. Allele specific expression of HSPB1 in F1 hybrids between tolerant and less tolerant populations showed significantly biased expression favoring alleles from tolerant populations and supporting the adaptive divergence in these *cis*-regulatory variants. The functional impact of population-specific non-synonymous substitutions in HSPB1 coding sequences was tested by assessing the thermal stabilization properties of SD versus SC HSPB1 protein variants. Recombinant HSPB1 from the southern SD population showed greater capacity for protecting protein structure under elevated temperature. Our results indicate that both regulatory and protein coding sequence evolution within a single gene appear to contribute to thermal tolerance phenotypes and local adaptation among conspecific populations.

Introduction

Populations of species that range across heterogeneous habitats frequently show evolutionary adaptation to their local environments. Adaptive phenotypes can stem from genetic variation in coding sequences, gene regulatory sequences or both. The relative contributions to adaptation from structural and gene regulatory variation are the subject of considerable debate (Carroll 2005; Hoekstra and Coyne 2007; Wray 2007). Clear evidence exists for both structural (Place and Powers 1979; Alahiotis 1982; Perutz 1983; Wirgin, et al. 2011) and regulatory (Schulte, et al. 2000; Juneja, et al. 2016) variations that lead to adaptive physiological traits that correspond to organisms' local habitats.

In organisms ranging from bacteria to vertebrates, thermal stress induces the expression of Heat Shock Proteins (HSPs) that help mitigate cellular damage from misfolded proteins (Lindquist 1986; Parsell and Lindquist 1993; Feder and Hofmann 1999). Heat shock response is among the best-established models for studying gene regulatory mechanisms {e.g., *Drosophila* HSP70 gene expression (Perisic, et al. 1989; Fernandes, et al. 1995)}. Differential expression of HSPs has been linked to differences in thermal tolerance within (Schoville, et al. 2012; Gleason and Burton 2015) and among (Tomanek and Somero 1999) species and available evidence suggests that the evolution of thermal tolerance may be at least partially driven by changes in HSP gene expression (Sørensen, et al. 2003). The eukaryotic HSP gene regulatory system is well-studied. Heat shock response is largely transcriptionally regulated by heat shock transcription factor (HSF) (Wu 1995). HSF is known to bind to a specific DNA sequence motif called the heat shock element (HSE) (Amin, et al. 1988) upon thermal stress and mediate transcriptional response of HSPs (Pelham 1982). An HSE unit consists of three inverted tandem repeats of a 5 base pair motif with 3 conserved base pairs in the middle as 'nGAAn'. The 15 bp units with all consensus sequence among eukaryotic lineages is called the 'perfect' or 'canonical' HSE. Each 5bp motif binds a subunit of HSF which is trimeric when active (Fernandes, et al. 1994). Though the mechanism of HSF-HSP transcription regulation has been well characterized, to date few examples (Lerman and Feder 2001; Lerman, et al. 2003) exist that demonstrate a causal connection between point mutations in heat shock gene regulatory sequences and adaptation to different temperature regimes across a species range.

Although intraspecific HSP regulatory variation appears to contribute to the evolution of thermal tolerance (25,26), additional modes of adaptation, such as structural variation in the HSPs themselves, have not been widely considered. There is some evidence for correlations between small HSP genotypes and thermal environment (Frydenberg, et al. 2003; Healy, et al. 2010; Graham, et al. 2012). Although these correlations suggest a contribution of HSP structural variation to the evolution of thermal tolerance, meaningful comparison of the functional difference of HSP alleles, especially for the small HSP family, is lacking.

The copepod *Tigriopus californicus* inhabits high intertidal pools along the west coast of North America, spanning a broad latitudinal gradient from Alaska, USA (Dethier 1980) to Baja California, Mexico (Ganz and Burton 1995). Previous studies have shown that populations along the coast exhibit different tolerances to acute thermal stress, with southern populations being significantly more tolerant of high temperatures than northern populations (Willett 2010; Kelly, et al. 2013). Populations differ in HSP gene expression following exposure to heat stress; specifically, a heat tolerant San Diego (SD) population (32^{o}45 N 117^{o}15W) showed a greater degree of upregulation of HSPs than a less tolerant Santa Cruz (SC) population (36^{o}57 N 122^{o}03W) (Schoville, et al. 2012). Among upregulated genes following 1-hour acute heat stress in *T. californicus*, HSPB1 (Accession number: JW506233) showed > 100X increase in transcript abundance in the SD population while only 5X upregulation was observed in the SC population. In addition to differential expression, there is also significant non-synonymous variation in the protein coding region of HSPB1 gene, making this an attractive system to evaluate the functional consequences of both structural and regulatory variation in adaptive phenotypes.

One difficulty encountered in assessing the role of single HSP genes in adaptation is that in many organisms (including *T. californicus*), there are multiple families of HSP genes and even within families, each HSP gene can have multiple copies suggesting functional redundancy (de Jong, et al. 1998). However, a few studies had shown that small HSPS can have an essential and non-redundant functional role. A study in *Drosophila* showed that different levels of one small HSP transcripts can lead to differences in thermal tolerance in *Drosophila* larvae (Lockwood, et al. 2017). HSPB1 knocked out mice did not show compensation by other HSPs both at mRNA and protein levels and HSPB1 knocked out cell-line showed less viability after heat treatment (Huang, et al. 2007). Similarly, in *T. californicus*, when RNAi was used to knock down HSPB1

112 expression, mortality dramatically increased after heat stress (average 5-day survivorship
113 following stress was reduced by approximately 80%), indicating there was no direct back-up
114 capacity for its function (Barreto, Schoville, et al. 2015). These results suggest a critical role for
115 HSPB1 and its transcriptional regulation in survivorship following heat stress exposure and
116 possibly in driving local adaptation among *T. californicus* populations..

117 The current study examines the potential role of DNA sequence variation in both the proximal
118 promoter and protein coding regions of the small HSP gene, HSPB1, in generating population
119 differences in thermal tolerance in *T. californicus*. We hypothesize that the observed divergence
120 in HSPB1 promoter sequences account for differences in transcript abundances across
121 populations while coding sequence variation results in allelic differences in the thermal
122 protectant properties of HSPB1; combined, the experimental results present a compelling case
123 for the roles of both regulatory and structural gene evolution in molding adaptation to local
124 thermal regimes across a species range.

125

126 **Results and discussion**

127 *HSPB1 promotor sequencing*

128 Given the evidence suggesting HSF regulation of HSP genes from other model systems and
129 taking advantage of the existing draft genome sequence for the SD *T. californicus* population
130 (see https://i5k.nal.usda.gov/Tigriopus_californicus), we examined the 5' flanking region of
131 HSPB1 for population differences in potential gene regulatory sequences. The gene does not
132 have introns; therefore, only flanking promoter sequences were searched for HSEs using
133 JASPAR (Mathelier, et al. 2014). From the SD sequence, we identified 2 HSEs within the
134 promoter region 584bp upstream of the transcription start site of HSPB1. The upstream sequence
135 beyond this region is a repeat sequence. Both of these HSEs have the three inverted repeats of
136 the 'canonical HSE'.

137 We then sequenced the proximal promoter region of the HSPB1 gene from a set of 11 *T.*
138 *californicus* populations from distinct geographic regions spanning from Baja California to
139 Vancouver Island, Canada (Figure 1A). By sequencing multiple individuals from a population,
140 we identified two geographic regions with different promoter genotypes. Southern populations

from Southern California and Baja California all have two 'canonical HSEs' while northern populations have nucleotide substitutions within the conserved regions of both HSEs (Figure 1B, full promoter alignment Figure S1). The AB population from Los Angeles County, California, appears to be a transition between the two regions, with one intact 'canonical HSE' and one HSE with polymorphic site in the conserved GAA motif.

Though there are additional sequence polymorphisms in the HSPB1 promoter among populations, research in other systems suggests that the observed variation in the HSEs alone could result in different gene expression phenotypes, with promoter strength declining when nucleotide substitutions result in deviations from the canonical HSEs with conserved GAA sequences (Fernandes, et al. 1994, 1995; Dierick, et al. 2007). Unlike *Drosophila* (Lerman, et al. 2003), we found no evidence of transposon insertion in the promoter region between the transcription start sites and the HSEs in the populations sequenced.

RNA sequencing and allele specific expression

In order to determine if the observed mutations in the HSEs within the HSPB1 promoter actually result in differential transcription, we tested for differences in allele specific expression (ASE) in F1 hybrids between the SD and the SC populations. In hybrids, the two HSPB1 alleles are present in the same cellular environment including all *trans*-acting factors, so differences in allelic expression are isolated to the effects of different *cis*-regulatory elements (Tirosh, et al. 2009; McManus, et al. 2010). Given its canonical HSEs, we hypothesized that expression of the SD HSPB1 allele would be favored in SD/SC F1 hybrids.

Using RNA-seq analysis, we found only low levels of HSPB1 transcripts in both parental and F1 hybrid animals under control temperature (20^{o}C) conditions. In agreement with previous findings in parental populations SD and SC (Schoville, et al. 2012), HSPB1 was strongly upregulated after heat stress (Table 1). HSPB1 expression was significantly biased in hybrids under heat stress conditions, with the SD allele elevated in both biological replicates in both reciprocal crosses (i.e., all four tests). Notably, this strong expression bias was only observed under the heat stress conditions when the HSF/HSE mediated upregulation of expression is expected to occur (Table 2). The strong bias in allelic expression in F1 hybrids suggests that the causal

mutation is in the *cis*-regulatory elements of the HSF gene regulatory network, most likely the substitutions in the conserved motif of HSEs in the promoter region. The hybrids also showed slightly biased HSPB1 expression (toward the SD allele) under control condition, which may be due to either a low level of mapping bias between two alleles (Table S2) due to higher polymorphism in SC population (Pereira, et al. 2016), or it may reflect low levels of HSF/HSE mediated expression favoring the SD allele under the control conditions. However, any bias due to variation in mapping efficiency is relatively minor (>90% of hybrid reads under heat stress treatment mapped correctly) and could not account for the large expression bias between the alleles observed under heat stress condition.

Allele specific expression in additional interpopulation crosses

To further confirm the functional consequences of the substitutions in the HSEs, we examined levels of ASE in HSPB1 in hybrids between different population pairs, including pairs with the same HSE structure: BR and SD each have two canonical HSEs while PES and SC have substitutions in the conserved motif of both HSEs. We predicted that only the F1 hybrid between populations with different promoter structure (eg. BRxPES and SDxSC) would show ASE in HSPB1 after heat stress treatment; F1 hybrids between populations with similar promoters would not show allelic imbalance. To identify ASE, fragments of HSPB1 coding sequence were PCR amplified from both genomic DNA and cDNA followed by restriction digests to discriminate between the alleles. Following agarose gel electrophoresis, ASE was evaluated by comparing intensity of allele-specific fragments between F1 genomic DNA and cDNA templates (Figure 2A-D, Table 3). Band intensity of digested amplicons from genomic DNA template from F1 hybrids should reflect equal proportions of template expected in HSPB1 heterozygotes; thus, comparing cDNA template band intensity to genomic DNA template reveals any biased expression. To validate this approach, SDxSC genomic DNA and mRNA from the RNAseq experiment were used also analyzed by restriction digest.

No evidence for ASE was observed in F1 hybrids between SDxBR and SCxPES, confirming that no ASE occurs in hybrids between populations with the same structure of HSEs in the promoter of HSPB1 (for both heat tolerant and heat sensitive population pairs). In contrast, BRxPES showed significant ASE; under heat stress treatment, there was substantial bias toward the BR allele (containing two canonical HSEs in the promoter region) over the PES allele similar to

SDxSC (Figure 2A). The average band intensity of BR allele in F1 cDNA was approximately three times higher than the intensity of the genomic DNA template band (Figure 2E). ASE results from BRxPES hybrids further reinforces the SDxSC RNA-seq evidence for functional divergence in *cis*-regulatory elements, with enhanced expression of HSPB1 alleles from more heat tolerant populations.

Taken together, our data suggest that divergence in *cis*-elements of HSF-HSP gene regulatory network may underlie differential HSPB1 gene expression and ultimately contribute to differences in thermal tolerance among *T. californicus* populations. Biased expression of HSPB1 alleles in F1 hybrids suggests that observed single nucleotide polymorphisms (SNPs) in the *cis*-regulatory HSEs have significant functional consequences on regulation of gene expression. Furthermore, expression bias toward southern alleles (SD allele in F1 SDxSC hybrids and BR allele in BRxPES hybrids) suggests that alleles with canonical HSEs act as stronger promoters than northern alleles that have substitutions in the conserved HSE motifs. The geographic pattern of stronger promoters for HSPB1 in southern *T. californicus* populations suggests that natural selection is favoring either regulatory variation that enhances the heat shock response in the warmer portion of the species range while that selection is relaxed in the cooler portion of the species range. Given that the HSEs from southern populations match the canonical eukaryotic sequence for HSEs, our working hypothesis is that those populations have the ancestral promoter However, this inference is not directly supported by current knowledge of the phylogeographic history of *T. californicus,* which has yet to resolve the ancestral distribution of the species. Edmands (2001) found reduced population differentiation among populations north of San Francisco Bay which might be partly attributed to post-glacial range expansion. This could explain HSPB1 promoter similarity between populations from the northern range (PAC and FHL) and Central California populations (SC, PES and PL). However, the phylogeographic relationships among populations between Central California (including SC), Southern California (including SD) and Mexican populations remain unresolved (Edmands 2001; Peterson, et al. 2013) (Figure 1).

Structural variation

Although our results strongly implicate adaptive variation in *cis*-regulation, amino acid sequence divergence in both HSF (*trans*-regulatory element) and HSPB1 itself may also contribute to

differential thermal tolerance and local adaptation across populations. From existing transcriptome data (Schoville, et al. 2012), we identified a single HSF gene in *T. californicus* with 529 amino acids. HSF comparison between SD and SC revealed 10 amino acid substitutions between populations (Figure S3). Three additional populations' HSF sequences were obtained from unpublished RNA-seq data including BR, AB, and PES. The functional significance (if any) of these amino acid substitutions in HSF is unknown; however dN/dS analyses (estimated using PAML 4.7) (Yang 2007) found no significant evidence of positive selection at HSF (ω, dN, dS: 0.1251 , 0.0066, 0.0525) (Table S3A). Any potential contribution of population differences in HSF *trans*-acting regulatory elements was factored out of the ASE studies by the experimental design.

In addition to HSF amino acid sequence divergence, we examined structural variation of HSPB1 across populations of *T. californicus*. Small HSPs including HSPB1 are characterized by an α-crystallin domain towards the C-terminal end of the protein (de Jong, et al. 1998). There are two α-crystallin domains (pfam00011) in *T. californicus* HSPB1. There is substantial structural variation of HSPB1 between SD and SC populations: 17 amino acid substitutions and one indel occur within the 277 amino acids (Figure S2). The amino acid substitutions between SD and SC populations were found throughout the gene including inside the α-crystallin domains suggesting the potential for functional difference between two HSPB1 alleles. We aligned HSPB1 coding sequences of five populations and found a relatively elevated dN/dS (ω = 0.440, dN=0.0393, dS=0.0886) (Table S3B) compared to transcriptome-wide mean (ω = 0.120 between SD and SC populations)(Barreto, et al. 2011). We further identified polymorphisms (> 1% of mapped reads) within SD and SC populations from our transcriptomes for calculating Neutrality Index (NI) (McDonald and Kreitman 1991) and Direction of Selection (DoS) (Stoletzki and Eyre-Walker 2011) to look for signal of positive selection (NI < 1, and DoS > 0). Both indices (NI = 0.642 (Fisher exact test p = 0.47) and DoS = -0.04) do not suggest any significant evidence of positive or purifying selection in the HSPB1 gene between SD and SC populations.

However, given the high number of fixed amino acid substitutions between SD and SC HSPB1alleles, we used *in vitro* functional assays to directly test the potential adaptive significance of the extensive population differentiation in HSPB1 coding sequence. A previous study found that thermal tolerance of *E. coli* was increased when expressing a truncated

Tigriopus japonicus HSPB1 homolog with only one α-crystallin domain (Seo, et al. 2006). Based on this evidence and HSPB1's putative function as a member of the small heat shock protein family, we hypothesized that adaptive evolution would lead to enhanced thermal protectant properties in southern alleles where populations are more frequently exposed to high temperatures. Using an *in vitro* thermal protection assay (Gong, et al. 2009), we tested the function of SD and SC variants of HSPB1 proteins expressed in *E. coli*. Purified HSPB1 protein (Figure S3) from each population was found to reduce *in vitro* aggregation of a test protein (porcine citrate synthase, CS) held at a high temperature. Furthermore, the SD allele consistently out-performed the SC allele in reducing the measured protein aggregation observed in each temperature treatment and in all four HSPB1 concentrations tested (Figure 3); although each of these tests was not replicated, we view the consistent differences across all temperatures and all HSPB1 concentrations as appropriate validation of the functional differences between the SD and SC variants, especially because the relevant *in vivo* concentrations are unknown. We further analyzed the functional differences between SD and SC HSPB1 alleles using an enzyme activity assay (Hristozova, et al. 2016). Adding recombinant HSPB1 protected citrate synthase from enzymatic activity loss by high temperature, $F_{2,44} = 21.72688$, $p = <0.0001$ (Table 4). Again, the SD allele outperformed SC allele in retaining CS enzyme activity (Figure 4) with Tukey pairwise comparison test showed a significant difference between SD and SC allele ($p < 0.0001$) (Table 5). Results from both experiments confirmed our hypothesis that HSPB1 from more heat tolerant population has enhanced thermoprotectant properties.

Conclusions

In this study, we demonstrate that variation in HSPB1 expression and function among populations of *T. californicus* can, in part, be attributed to both *cis*-regulatory variation and coding sequence variation in the HSPB1 gene. ASE assays in F1 interpopulation hybrids confirmed the functional significance of SNPs in *cis*-regulatory elements between populations that differ in thermal tolerance phenotypes. Additionally, *in vitro* assays showed that HSPB1 produced by the heat tolerant SD was more potent at preventing protein aggregation and preserving enzyme activity at high temperature than HSPB1 from the less heat tolerant SC population. Both findings, in *cis*-regulatory sequences and protein function, are consistent with geographic differences in the thermal regimes experienced by different copepod populations.

Together with previous studies verifying the key role of HSPB1 in thermal response in *T. californicus* (Schoville, et al. 2012; Barreto, Schoville, et al. 2015), the present study demonstrates that selection can act on both protein structure and regulation of expression within a single gene, and that each mode of selection may contribute to local adaptation among populations.

Materials and methods

Copepod culturing and hybridization

Copepods were collected from high intertidal rock pools along the Pacific coast of North America (Figure 1, Table S1). Stock populations were maintained in 400 ml beakers filled with 250 ml of filtered (0.45 μm) seawater under constant 20ºC and 12 hour light:dark photoperiod. Copepods were fed ground *Spirulina* wafer fish food *ad libitum*. *T. californicus* F1 hybrids were produced following Barreto et. al. (Barreto, Pereira, et al. 2015). For the RNA sequencing experiment, each replicate of each reciprocal cross between SD and SC consisted of F1 hybrids that came from more than 30 successfully mated females.

Heat stress experiment and RNA sequencing

Both parental populations and reciprocal F1 hybrids between SD and SC were exposed to constant 20 ºC as a control treatment; experimental heat stress presumed to activate HSF involved increasing temperatures by 5 ºC per hour up to 35 ºC. After 60 minutes at 35 ºC, animals were sacrificed and RNA was immediately extracted with Tri Reagent (Sigma) following the manufacturer's protocol. Each sample contained 50 adult copepods of equal sex ratio and each treatment had two biological replicates. cDNA libraries were prepared using oligo-dT priming and Illumina's Truseq standard mRNA protocol. Libraries were sequenced (100-bp single-read) on the Illumina 4000 platform. Reads were aligned to a transcriptome reference (Barreto, Pereira, et al. 2015) that included paired orthologs from both populations. Parameters for mapping included a cutoff at 0.8 length fraction and 0.98 sequence similarity. Full analysis of ASE across the transcriptome will be presented elsewhere; here we focused only on the expression of HSPB1 alleles. Only reads that uniquely mapped to one of the two alleles

were counted. We performed binomial tests with 5% FDR to identify significantly biased HSPB1 allelic expression in hybrids.

HSPB1 promoter sequencing and HSE identification

To obtain genomic DNA for PCR amplification, individual copepods were put in 15 µl lysis buffer with Proteinase K (Willett and Burton 2001) then heated to 55 ºC for 90 minutes followed by 95 ºC for 15 minutes. For pooled extraction, 15-50 individuals from each population were used following DNeasy blood and tissue kit protocol (QIAGEN). We use the SD population draft genome to design primers to amplify a 361-bp product including 2 HSEs in the promoter region (forward primer: 5'-ACTAGTTGTCCGATACACAAACAAACTAT-3', reverse primer: 5'-GAAACAAAAGAGCCATGGTTTA-3'). We sequenced the promoter region from at least 10 individuals from each population. Sequences were aligned using Sequencher and Geneious.

Restriction digest

In addition to the RNA sequencing experiment involving the SD x SC hybrids, three further crosses were made to assess the role of promoter sequence in ASE. BRxPES cross is an independent test with different SNPs in the HSEs corresponding to SDxSC cross in the RNA-seq experiment. SDxBR and SCxPES (two southern and two northern populations, respectively) are crosses between populations with similar SNPs in the HSEs. Hybrids were raised in the same conditions as described above. For SCxPES and BRxPES crosses, F1 hybrids from a single female were used as a biological replicate. For SDxSC cross, we used pooled F1 hybrids from multiple females from independent crosses as a biological replicate. For each replicate, genomic DNA was obtained from F1 male using the same methods described above. cDNA was made from RNA extracted from F1 animals subjected to the same heat stress treatment as in the RNA sequencing experiment. Both genomic DNA and cDNA were used as template for PCR amplification of HSPB1 coding sequences. We used HSPB1 coding sequences of the four populations (unpublished data) to design PCR primers and identify population specific restriction cut sites using Webcutter 2.0 (Maarek, et al. 1997). The primers and the restriction enzymes used in the experiment are listed in Table S4. Restriction digest reactions were performed on 200 ng of PCR products following the manufacturer protocols. Each digested sample was run on gel electrophoresis up to 3 times to estimate the variability of the band intensity. Gel images were analyzed for band intensity using Image Lab software (Bio-Rad). We performed nested ANOVA

using the package nlme (Pinheiro, et al.) in R 3.4.2 (R core team, 2017) on percent band intensity of the largest digested band between genomic DNA and cDNA templates. We used the band intensity of the sample with genomic DNA template as a baseline for unbiased expression. Significant difference in band intensity in samples with cDNA template indicates biased ASE.

Expression and purification of HSPB1

Full length SD and SC variants of HSPB1 were amplified from their respective cDNA with N-terminal 6xHis-tag, cloned into the pProEx Htb expression vector (Invitrogen) and transformed into *E. coli* BL21 (DE3) pLysE cells. Ten ml of an overnight culture in LB medium were added to 200ml and grown for 2 h at 37 ºC. Expression of the recombinant HSPB1 was induced by the addition of isopropyl-ß-D-1-thiogalactopyranoside (IPTG) to a final concentration of 1 mM. After an additional 2.5 h incubation at 37 ºC, cells were harvested by centrifugation at 3,000g for 20 minutes at 4 ºC and then lysed in lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole, 8 M urea, pH 7.4) The cell lysate was centrifuged 15,000g for 15 minutes to pellet the insoluble material and to collect the supernatant fraction of the cell lysate. The supernatant was then loaded onto a His60 Ni Superflow Resin column (Clontech). The column was washed in wash buffer (50 mM NaH_2PO_4, 300 mM NaCl, 40 mM imidazole, 8 M urea, pH 7.4) and then the His-tagged recombinant HSPB1 was eluted in wash buffer containing 300 mM imidazole. We did not remove His-Tag from the recombinant protein. The eluted HSPB1 fractions were then dialyzed against phosphate buffered saline (PBS) and protein concentration determined using the BCA Protein Assay Kit (Pierce).

Citrate Synthase (CS) aggregation and activity assays

Thermal aggregation experiments were performed as described in (Gong, et al. 2009). For each test, 10 µg CS from porcine heart (C3260, Sigma, 9.4 µg/µl) was incubated with either SD or SC HSPB1 in 1 ml PBS at 45 ºC and aggregation monitored by measuring turbidity at 320 nm in a spectrophotometer. Thermal inactivation of CS activity was done as described in (Hristozova, et al. 2016). The reaction was performed with 0.5 µg/ml CS (0.329 units/µg), 0.45 mM Acetyl-coA, 0.5mM oxaloacetate, and 0.1 mM Ellman's reagent (DTNB) in PBS and followed for 3 min at room temperature. CS activity was fitted into a linear mixed model with time and allele fixed effects and replicates (SD = 5, SC = 4, and no HSP = 6 with 56 total observations) as a random

effect using R package nlme (Pinheiro, et al.). We performed Tukey's pairwise comparison among two alleles and no HSP control using R package multcomp (Hothorn, et al. 2008).

Author contributions

S.T. and R.S.B. conceived and planned the project. S.T. conducted the RNA-seq experiment, HSPB1 promoter sequencing, and restriction digest experiment. G.W.M. conducted the cloning of HSPB1 and aggregation assay. S.T., G.W.M and R.S.B. wrote the manuscript.

Acknowledgment

This work was supported by US National Science Foundation grants (DEB1551466 and IOS1155030) to R.S.B. The authors thank Dr. Ricardo Pereira, Brian Hong, and Dr. Jon Shurin for providing some of the copepod populations. We also thank undergraduates Wenpei Li and Kiana Woodward for their help in this work. Dr. Tim Healy, Dr. Satomi Tsuboko-Ishii and Alice Harada provided helpful comments on the manuscript.

References

Alahiotis SN. 1982. Adaptation of *Drosophila* enzymes to temperature IV. Natural selection at the alcohol-dehydrogenase locus. Genetica 59:81-87.

Amin J, Ananthan J, Voellmy R. 1988. Key features of Heat-Shock Regulatory Elements. Molecular and Cellular Biology 8:3761-3769.

Barreto FS, Moy GW, Burton RS. 2011. Interpopulation patterns of divergence and selection across the transcriptome of the copepod *Tigriopus californicus*. Molecular Ecology 20:560-572.

Barreto FS, Pereira RJ, Burton RS. 2015. Hybrid Dysfunction and Physiological Compensation in Gene Expression. Molecular Biology and Evolution 32:613-622.

Barreto FS, Schoville SD, Burton RS. 2015. Reverse genetics in the tide pool: knock-down of target gene expression via RNA interference in the copepod *Tigriopus californicus*. Molecular Ecology Resources 15:868-879.

Carroll SB. 2005. Evolution at two levels: On genes and form. Plos Biology 3:1159-1166.

de Jong WW, Caspers GJ, Leunissen JAM. 1998. Genealogy of the alpha-crystallin - small heat-shock protein superfamily. International Journal of Biological Macromolecules 22:151-162.

409 Dethier MN. 1980. Tidepools as refuges: predation and the limits of the Harpacticoid copepod *Tigriopus*
410 *californicus* (Baker). Journal of Experimental Marine Biology and Ecology 42:99-111.
411 Dierick I, Irobi J, Janssens S, Theuns J, Lemmens R, Jacobs A, Corsmit E, Hersmus N, Van Den Bosch L,
412 Robberecht W, et al. 2007. Genetic Variant in the HSPB1 Promoter Region Impairs the HSP27 Stress
413 Response. Human Mutation 28:830-830.
414 Edmands S. 2001. Phylogeography of the intertidal copepod *Tigriopus californicus* reveals substantially
415 reduced population differentiation at northern latitudes. Molecular Ecology 10:1743-1750.
416 Feder ME, Hofmann GE. 1999. Heat-shock proteins, molecular chaperones, and the stress response:
417 Evolutionary and ecological physiology. Annual Review of Physiology 61:243-282.
418 Fernandes M, Xiao H, Lis JT. 1995. Binding of heat shock factor to and transcriptional activation of heat
419 shock genes in Drosophila. Nucleic Acids Research 23:4799-4804.
420 Fernandes M, Xiao H, Lis JT. 1994. Fine-structure analyses of the *Drosophila* and *Saccharomyces* heat-
421 shock factor - heat-shock element interactions. Nucleic Acids Research 22:167-173.
422 Frydenberg J, Hoffmann AA, Loeschcke V. 2003. DNA sequence variation and latitudinal associations in
423 hsp23, hsp26 and hsp27 from natural populations of *Drosophila melanogaster*. Molecular Ecology
424 12:2025-2032.
425 Ganz HH, Burton RS. 1995. Genetic differentiation and reproductive incompatibility among Baja
426 California populations of the copepod *Tigriopus californicus*. Marine Biology 123:821-827.
427 Gleason LU, Burton RS. 2015. RNA-seq reveals regional differences in transcriptome response to heat
428 stress in the marine snail *Chlorostoma funebralis*. Molecular Ecology 24:610-627.
429 Gong WN, Yue M, Xie BY, Wan FH, Guo JY. 2009. Inhibition of Citrate Synthase Thermal Aggregation In
430 Vitro by Recombinant Small Heat Shock Proteins. Journal of Microbiology and Biotechnology 19:1628-
431 1634.
432 Graham AM, Merrill JD, McGaugh SE, Noor MAF. 2012. Geographic Selection in the Small Heat Shock
433 Gene Complex Differentiating Populations of *Drosophila pseudoobscura*. Journal of Heredity 103:400-
434 407.
435 Healy TM, Tymchuk WE, Osborne EJ, Schulte PM. 2010. Heat shock response of killifish (*Fundulus*
436 *heteroclitus*): candidate gene and heterologous microarray approaches. Physiological Genomics 41:171-
437 184.
438 Hoekstra HE, Coyne JA. 2007. The locus of evolution: Evo devo and the genetics of adaptation. Evolution
439 61:995-1016.
440 Hothorn T, Bretz F, Westfall P. 2008. Simultaneous inference in general parametric models. Biometrical
441 Journal 50:346-363.
442 Hristozova N, Tompa P, Kovacs D. 2016. A Novel Method for Assessing the Chaperone Activity of
443 Proteins. Plos One 11.
444 Huang L, Min JN, Masters S, Mivechi NF, Moskophidis D. 2007. Insights into function and regulation of
445 small heat shock protein 25 (HSPB1) in a mouse model with targeted gene disruption. Genesis 45:487-
446 501.
447 Juneja P, Quinn A, Jiggins FM. 2016. Latitudinal clines in gene expression and *cis*-regulatory element
448 variation in *Drosophila melanogaster*. BMC Genomics 17.
449 Kelly MW, Grosberg RK, Sanford E. 2013. Trade-Offs, Geography, and Limits to Thermal Adaptation in a
450 Tide Pool Copepod. American Naturalist 181:846-854.
451 Lerman DN, Feder ME. 2001. Laboratory selection at different temperatures modifies heat-shock
452 transcription factor (HSF) activation in *Drosophila melanogaster*. Journal of Experimental Biology
453 204:315-323.
454 Lerman DN, Michalak P, Helin AB, Bettencourt BR, Feder ME. 2003. Modification of heat-shock gene
455 expression in *Drosophila melanogaster* populations via transposable elements. Molecular Biology and
456 Evolution 20:135-144.

457 Lindquist S. 1986. The Heat-Shock Response. Annual Review of Biochemistry 55:1151-1191.
458 Lockwood BL, Julick CR, Montooth KL. 2017. Maternal loading of a small heat shock protein increases
459 embryo thermal tolerance in *Drosophila melanogaster*. Journal of Experimental Biology 220:4492-4501.
460 Maarek YS, Jacovi M, Shtalhaim M, Ur S, Zernik D, BenShaul IZ. 1997. WebCutter: A system for dynamic
461 and tailorable site mapping. Computer Networks and Isdn Systems 29:1269-1279.
462 Mathelier A, Zhao X, Zhang AW, Parcy F, Worsley-Hunt R, Arenillas DJ, Buchman S, Chen C-y, Chou A,
463 Ienasescu H, et al. 2014. JASPAR 2014: an extensively expanded and updated open-access database of
464 transcription factor binding profiles. Nucleic Acids Research 42:D142-D147.
465 McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the ADH locus in *Drosophila*. Nature
466 351:652-654.
467 McManus CJ, Coolon JD, Duff MO, Eipper-Mains J, Graveley BR, Wittkopp PJ. 2010. Regulatory
468 divergence in *Drosophila* revealed by mRNA-seq. Genome Research 20:816-825.
469 Parsell DA, Lindquist S. 1993. The function of Heat-Shock Proteins in stress tolerance - degredation and
470 reactivation of damaged proteisn. Annual Review of Genetics 27:437-496.
471 Pelham HRB. 1982. A regulatory upstream promoter element in the *Drosophila* HSP 70 Heat-Shock gene.
472 Cell 30:517-528.
473 Pereira RJ, Barreto FS, Pierce NT, Carneiro M, Burton RS. 2016. Transcriptome-wide patterns of
474 divergence during allopatric evolution. Molecular Ecology 25:1478-1493.
475 Perisic O, Xiao H, Lis JT. 1989. Stable binding of *Drosophila* heat-shock factor to head-to-head and tail-to-
476 tail repeats of a conserved 5-bp recognition unit. Cell 59:797-806.
477 Perutz MF. 1983. Species adaptation in a protein molecule. Molecular Biology and Evolution 1:1-28.
478 Peterson DL, Kubow KB, Connolly MJ, Kaplan LR, Wetkowski MM, Leong W, Phillips BC, Edmands S.
479 2013. Reproductive and phylogenetic divergence of tidepool copepod populations across a narrow
480 geographical boundary in Baja California. Journal of Biogeography 40:1664-1675.
481 Pinheiro J, Bates D, DebRoy S, Sarkar D. R Core Team.(2017). nlme: Linear and Nonlinear Mixed Effects
482 Models [Software].
483 Place AR, Powers DA. 1979. Genetic variation and relative catalytic efficiencies: Lactate dehydrogenase B
484 allozymes of *Fundulus heteroclitus*. Proceedings of the National Academy of Sciences of the United
485 States of America 76:2354-2358.
486 Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression
487 analysis of digital gene expression data. Bioinformatics 26:139-140.
488 Schoville SD, Barreto FS, Moy GW, Wolff A, Burton RS. 2012. Investigating the molecular basis of local
489 adaptation to thermal stress: population differences in gene expression across the transcriptome of the
490 copepod *Tigriopus californicus*. BMC Evolutionary Biology 12.
491 Schulte PM, Glemet HC, Fiebig AA, Powers DA. 2000. Adaptive variation in lactate dehydrogenase-B
492 gene expression: Role of a stress-responsive regulatory element. Proceedings of the National Academy
493 of Sciences of the United States of America 97:6597-6602.
494 Seo JS, Lee YM, Park HG, Lee JS. 2006. The intertidal copepod *Tigriopus japonicus* small heat shock
495 protein 20 gene (Hsp20) enhances thermotolerance of transformed *Escherichia coli*. Biochemical and
496 Biophysical Research Communications 340:901-908.
497 Stoletzki N, Eyre-Walker A. 2011. Estimation of the Neutrality Index. Molecular Biology and Evolution
498 28:63-70.
499 Sørensen JG, Kristensen TN, Loeschcke V. 2003. The evolutionary and ecological role of heat shock
500 proteins. Ecology Letters 6:1025-1037.
501 Tirosh I, Reikhav S, Levy AA, Barkai N. 2009. A Yeast Hybrid Provides Insight into the Evolution of Gene
502 Expression Regulation. Science 324:659-662.

Tomanek L, Somero GN. 1999. Evolutionary and acclimation-induced variation in the heat-shock responses of congeneric marine snails (genus *Tegula*) from different thermal habitats: Implications for limits of thermotolerance and biogeography. Journal of Experimental Biology 202:2925-2936.
Willett CS. 2010. Potential fitness trade-offs for thermal tolerance in the intertidal copepod Tigriopus californicus. Evolution 64:2521-2534.
Willett CS, Burton RS. 2001. Viability of cytochrome C genotypes depends on cytoplasmic backgrounds in *Tigriopus californicus*. Evolution 55:1592-1599.
Wirgin I, Roy NK, Loftus M, Chambers RC, Franks DG, Hahn ME. 2011. Mechanistic Basis of Resistance to PCBs in Atlantic Tomcod from the Hudson River. Science 331:1322-1325.
Wray GA. 2007. The evolutionary significance of *cis*-regulatory mutations. Nature Reviews Genetics 8:206-216.
Wu C. 1995. Heat shock transcription factors: Structure and regulation. Annual Review of Cell and Developmental Biology 11:441-469.
Yang ZH. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. Molecular Biology and Evolution 24:1586-1591.

Table 1: Differential expression pattern of HSPB1 across genotypes. Significant values were obtained from Likelihood Ratio Test in edgeR package (Robinson, et al. 2010).

Crosses	Heat stress expression (average $\log_2$ CPM±SE)	$\log_2$ Fold Change (heat stress / control)	p value
SDfxSCm	11.05±0.83	6.81	4.70E-45
SCfxSDm	11.68±0.30	6.77	7.96E-37
SD	12.14±0.41	7.57	5.24E-65
SC	10.49±0.30	5.88	1.54E-40

Table 2: F1 allele specific expression of HSPB1(raw mapped reads to each parental reference) of two independent reciprocal crosses and significant values (5% FDR p values) from binomial tests of equal expression between the two alleles in F1 hybrids

Cross	Control			Heat Stress		
	SD counts	SC counts	5% FDR p value	SD counts	SC counts	5% FDR p value
SCfxSDm #1	25	13	0.28	16365	6252	1.87E-321
SCfxSDm #2	152	128	0.27	29701	12960	1.61E-321
SDfxSCm #1	119	60	1.37E-3	13452	4976	1.80E-321
SDfxSCm #2	141	123	0.43	23157	7906	1.31E-321

Table 3: Significant values from nested ANOVA model for the test for ASE by restriction digest

Cross	#family	#observations	F value	p value
SCxPES	7	34	0.00703	0.9338
BRxPES	6	31	145.79364	<0.001
SDxBR	4	16	0.06362	0.8055

Table 4: Significant values from linear mixed model for the CS activity assay.

Source of variation	df	F value	p value
Intercept	1	33.21778	<0.0001
Time	1	243.71921	<0.0001
Allele	2	21.72688	<0.0001
Time x Allele	2	2.61886	0.0842

Table 5: Significant values from Tukey pairwise comparison among SD and SC allele treatments and no HSP control on the CS activity assay.

Hypothesis	p value
SD = no HSP	< 0.0001
SC = no HSP	0.775
SD = SC	< 0.0001

Figure 1

HSEs in the promoter region of HSPB1 gene across populations of *T. californicus*

A map showing populations of *T. californicus* along the Pacific coast of North America (A) with corresponding HSEs sequences in the HSPB1 promoter. Red dots indicate the populations that we sequenced individually. Orange dots indicates the populations that we performed pooled DNA extraction. (B). The numbers mark the positions upstream from transcription start site for the SD population. Full promotor alignment is shown in the supplementary figure S1. Red characters indicate the consensus GAA motif of HSEs. Blue characters indicate SNPs or polymorphism in the conserved part of HSEs that deviate from the consensus sequence. Green characters indicate SNPs outside of the conserved motif.

Figure 2

HSPB1 Allele Specific Expression in restriction digest experiment Pictures form gel-electrophoresis showing restriction cut bands in each F1 hybrid and parent pairs (A-C). The boxes encircle the largest restriction cut bands in both genomic DNA and cDNA template used for evaluating ASE. Box plots showing percent band intensity between cut bands between genomic DNA and cDNA templates (D).

Figure 3

Effects of HSPB1 on thermal aggregation of CS.

SD or SC HSPB1 at 6.25, 12.5, 25 and 50 µg/ml were incubated at 45ºC with CS (10 µg/ml). Insoluble CS aggregates formed over time were detected by light scattering at OD320. ADD PLOTS.

Figure 4

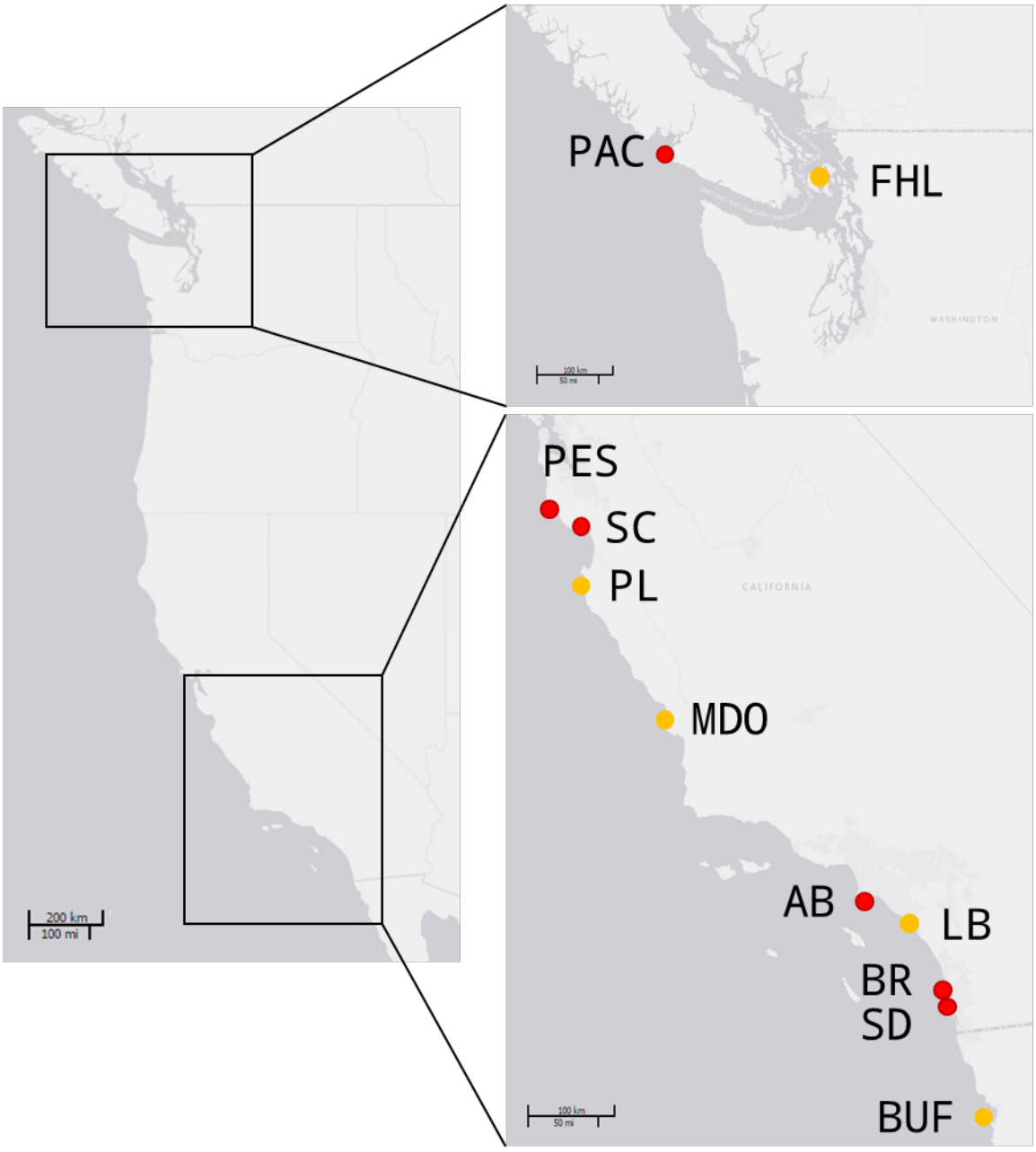
Effects of HSPB1 on thermal inactivation of CS activity.

SD or SC HSPB1 (172 µg/ml for each) were incubated with 0.5 µg/ml CS (0.329 units/µg) CS at 44°C. The deactivation of CS is shown as the % relative remaining activity. ERROR BARS??

571

572 Figure 1

573 A

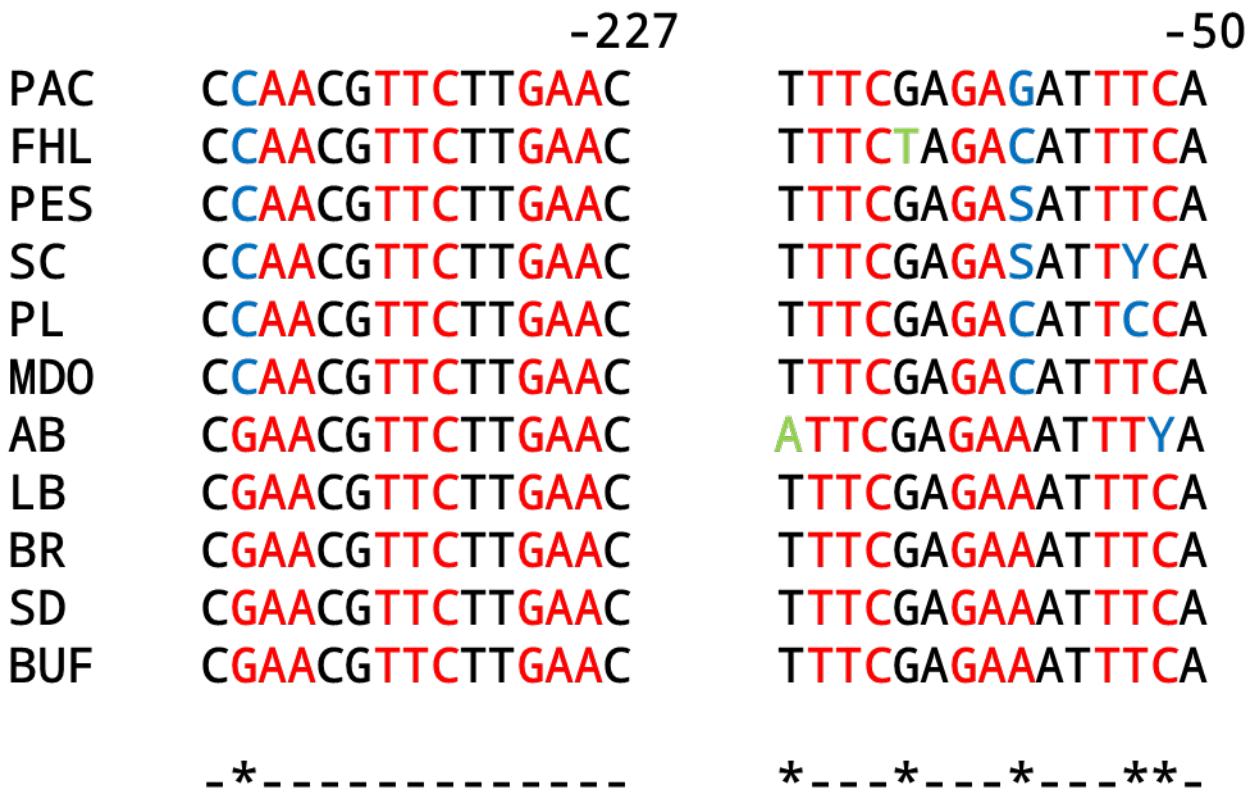


574

575

576

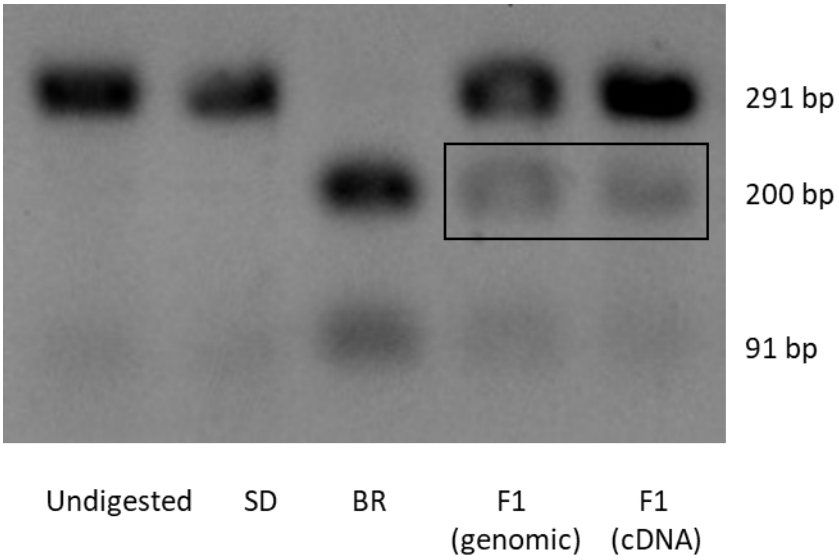
B


-227
-50
PAC CCAACGTTCTTGAAC TTTCGAGAGATTCA
FHL CCAACGTTCTTGAAC TTTCTAGACATTCA
PES CCAACGTTCTTGAAC TTTCGAGASATTCA
SC CCAACGTTCTTGAAC TTTCGAGASATTYCA
PL CCAACGTTCTTGAAC TTTCGAGACATTCCA
MDO CCAACGTTCTTGAAC TTTCGAGACATTTCA
AB CGAACGTTCTTGAAC ATTCGAGAAATTTYA
LB CGAACGTTCTTGAAC TTTCGAGAAATTTCA
BR CGAACGTTCTTGAAC TTTCGAGAAATTTCA
SD CGAACGTTCTTGAAC TTTCGAGAAATTTCA
BUF CGAACGTTCTTGAAC TTTCGAGAAATTTCA
-*----------- *---*---*---**-

588

589 Figure 2

590 A SDxBR



591

592

593

594

595

596

597

598

599

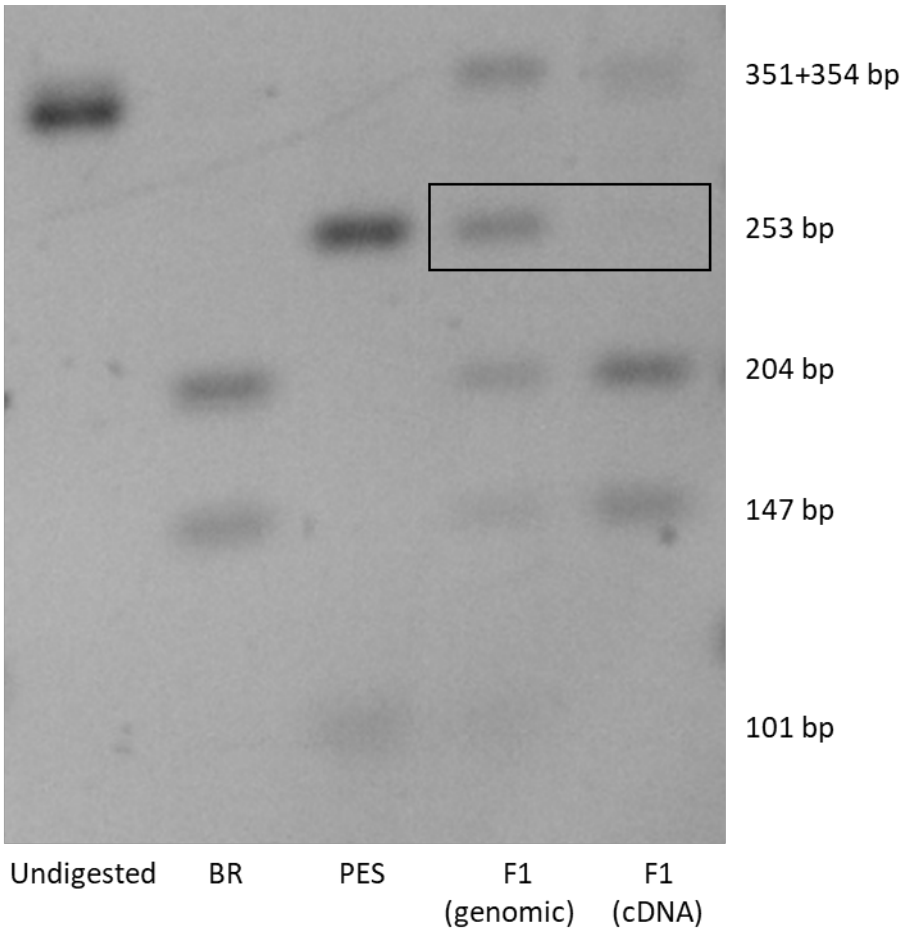
600

601

602

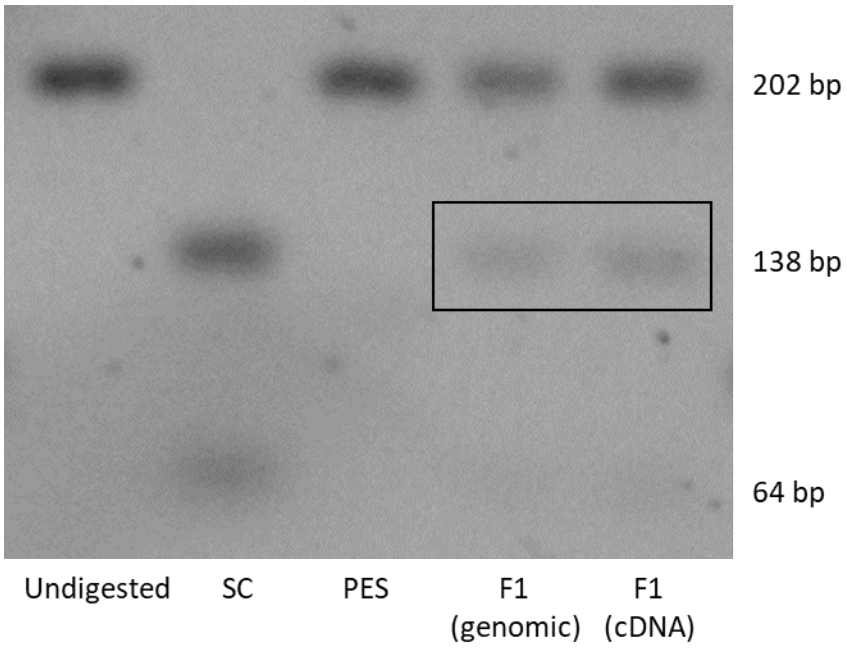
603

B BRxPES


351+354 bp
253 bp
204 bp
147 bp
101 bp
Undigested
BR
PES
F1 (genomic)
F1 (cDNA)

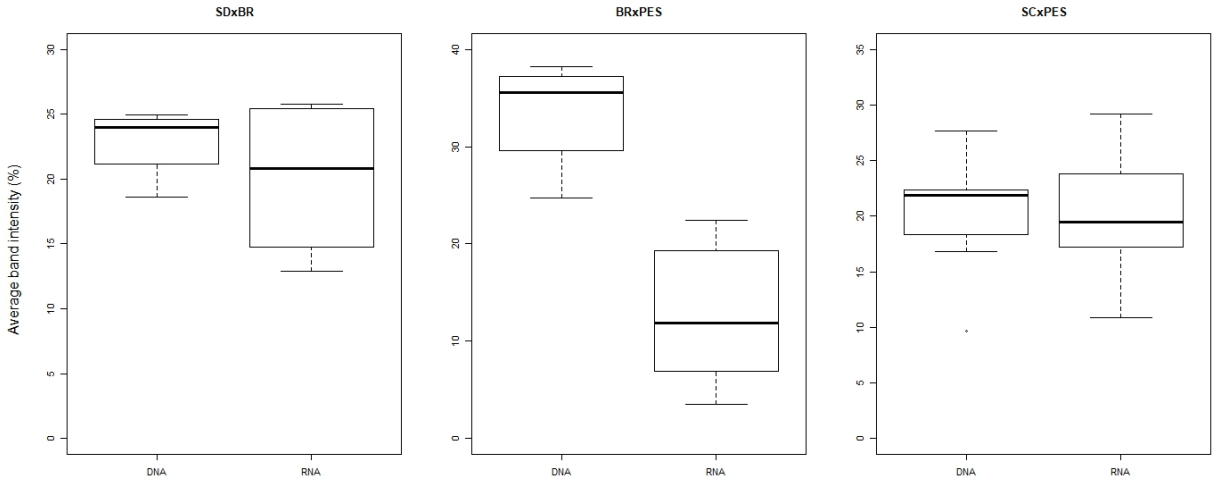
615

616 C SCxPES



617

618 D



619

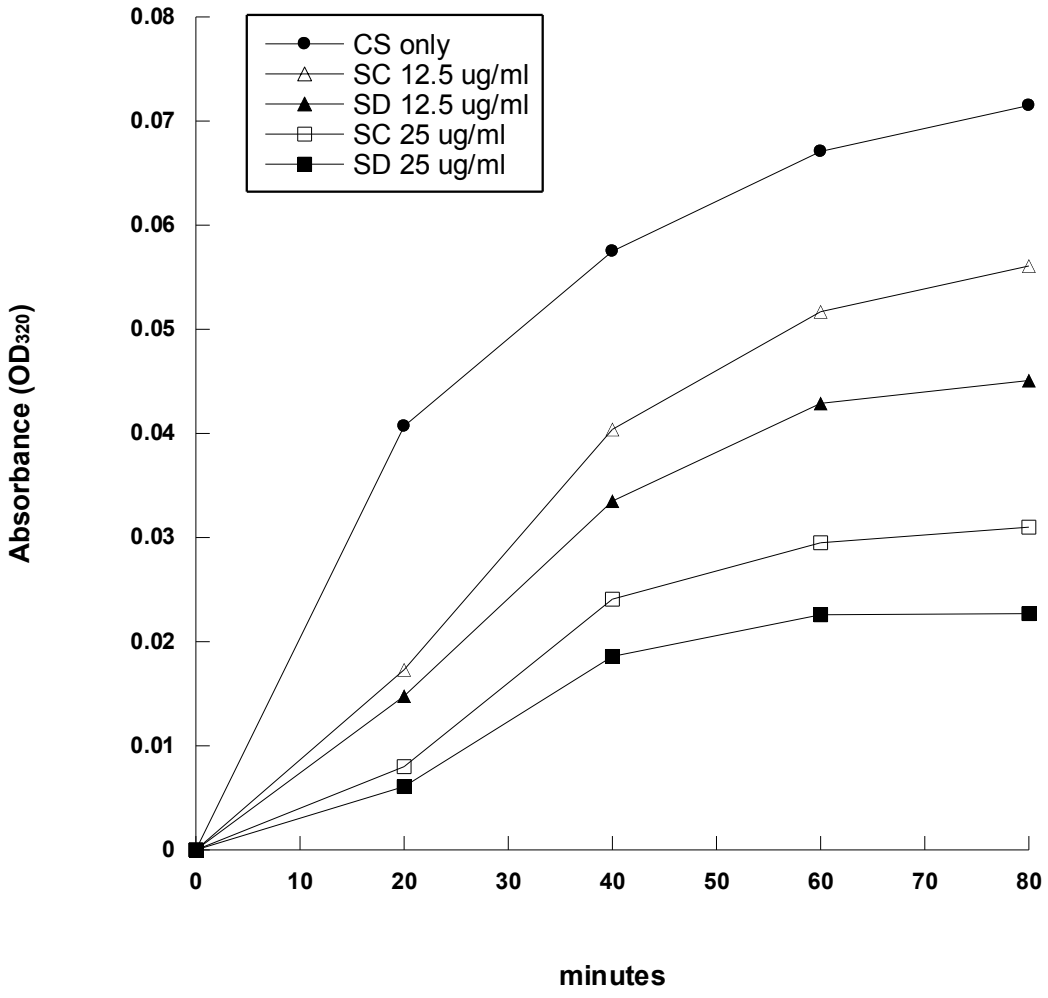
620

621

622

623

624 Figure 3



625

626

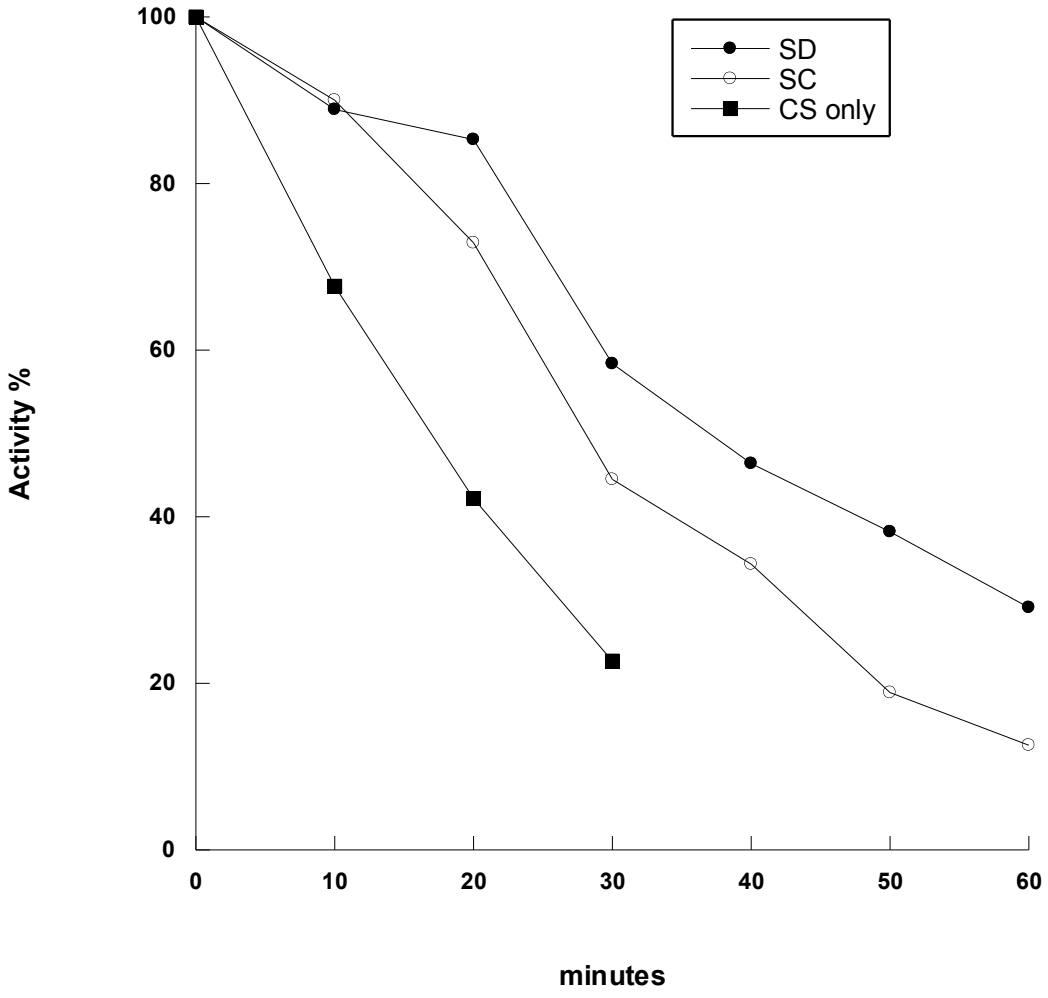
627

628

629

630

631 Figure 4



632

633

634

635