

# Evolutionary changes in heat-inducible gene expression in lines of *Escherichia coli* adapted to high temperature

Michelle M. Riehle,<sup>1</sup> Albert F. Bennett,<sup>1</sup> Richard E. Lenski,<sup>2</sup> and Anthony D. Long<sup>1</sup>

<sup>1</sup>Department of Ecology and Evolutionary Biology, University of California at Irvine, Irvine, California 92697-2525; and <sup>2</sup>Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, Michigan 48824-4320

Submitted 26 March 2002; accepted in final form 31 March 2003

**Riehle, Michelle M., Albert F. Bennett, Richard E. Lenski, and Anthony D. Long.** Evolutionary changes in heat-inducible gene expression in lines of *Escherichia coli* adapted to high temperature. *Physiol Genomics* 14: 47–58, 2003. First published April 2, 2003; 10.1152/physiolgenomics.00034.2002.—The involvement of heat-inducible genes, including the heat-shock genes, in the acute response to temperature stress is well established. However, their importance in genetic adaptation to long-term temperature stress is less clear. Here we use high-density arrays to examine changes in expression for 35 heat-inducible genes in three independent lines of *Escherichia coli* that evolved at high temperature (41.5°C) for 2,000 generations. These lines exhibited significant changes in heat-inducible gene expression relative to their ancestor, including parallel changes in *fkpA*, *gapA*, and *hslT*. As a group, the heat-inducible genes were significantly more likely than noncandidate genes to have evolved changes in expression. Genes encoding molecular chaperones and ATP-dependent proteases, key components of the cytoplasmic stress response, exhibit relatively little expression change; whereas genes with periplasmic functions exhibit significant expression changes suggesting a key role for the extracytoplasmic stress response in the adaptation to high temperature. Following acclimation at 41.5°C, two of the three lines exhibited significantly improved survival at 50°C, indicating changes in inducible thermotolerance. Thus evolution at high temperature led to significant changes at the molecular level in heat-inducible gene expression and at the organismal level in inducible thermotolerance and fitness.

adaptation; functional genomics; heat-shock response; stress genes; thermotolerance; arabinose-utilization phenotype

THE INVOLVEMENT of heat-shock genes (more generally, stress-response genes) in the acute acclimation response to high-temperature stress was first reported in *Drosophila melanogaster* 40 years ago (38), and it has now been characterized in Bacteria, Archaea, and Eukarya (5). This response has been extensively studied (reviewed in Refs. 4, 20, 44), and it is evolutionarily conserved across domains (5, 31). Although it is well established that the heat-shock and other inducible genes play important roles in the phenotypic response

to acute temperature stress (21), it is unclear how they contribute to genetic adaptation at prolonged high temperature. Distinguishing between genotypic and phenotypic stress responses is especially important given the need to understand the range of biological responses to global temperature change. It is possible that the same genes that are most important in acute acclimation responses also change during long-term evolution under stressful conditions. Alternatively, evolutionary responses may involve additional genes or even an entirely different set of genes.

To rigorously distinguish between these hypotheses, one must specify a priori the set of candidate genes, as opposed to formulating and testing hypotheses after data has been collected. The 35 candidate genes considered here are classified as heat inducible; they include genes whose protein products are synthesized at higher rates when a growing culture is shifted from 37°C to 42°C (43, 44). These increases in rates of protein synthesis are controlled primarily at the transcriptional level by two transcription factors,  $\sigma^E$  and  $\sigma^H$  (33, 44, 45). The 35 candidate genes include 9 that encode molecular chaperones, 9 that are associated with ATP-dependent proteases and the degradation of damaged proteins, 3 involved in protein folding and degradation of abnormal proteins in the extracytoplasmic space, 3 that encode RNA polymerases, 1 related to amino acid metabolism, 1 involved with carbohydrate metabolism, 1 involved in lipopolysaccharide metabolism, 1 GTP-binding protein, 1 involved in tRNA modification, and several others with various protective functions (44). The protein products of 8 candidates are membrane associated or function in the extracytoplasmic space, whereas the remaining 27 are cytoplasmic (Table 1). Genes other than these 35 candidates could be involved in evolutionary adaptation to high temperature, and some others might even be regarded as candidates using other criteria. However, these candidate genes can be used to test whether heat-inducible genes are also involved in evolutionary adaptation to high temperature.

Our study examines evolutionary changes in the expression of these 35 candidate genes in three lines of *Escherichia coli* that evolved at 41.5°C for 2,000 generations. We test the specific hypothesis that these heat-inducible genes were targets of selection during the evolutionary adaptation to high temperature.

Article published online before print. See web site for date of publication (<http://physiolgenomics.physiology.org>).

Address for reprint requests and other correspondence: M. M. Riehle, Dept. of Ecology and Evolutionary Biology, Univ. of California at Irvine, Irvine, CA 92697-2525 (E-mail: [mriehle@uci.edu](mailto:mriehle@uci.edu)).

Table 1. *Heat inducible genes, proteins products, and putative cellular functions*

Gene	Protein Product	Putative Cell Function*
<i>clpA</i>	ClpA, ATP-dependent Clp protease, ATP-binding subunit	Directs protease to specific substrates and complexed with clpP it appears to function in the degradation of unfolded/abnormal proteins; cytoplasmic.
<i>clpB</i>	ClpB	Thought to be an ATPase subunit of an intracellular ATP dependent protease.
<i>clpP</i>	ClpP, ATP-dependent Clp protease proteolytic subunit	Major role in the degradation of misfolded proteins; cytoplasmic.
<i>clpX</i>	ClpX, ATP-dependent Clp protease ATP-binding subunit	Directs partner protease to specific substrates.
<i>dnaJ</i>	DnaJ, HSP40	With GrpE, stimulates the ATPase activity of DnaK; cytoplasmic.
<i>dnaK</i>	DnaK, HSP70	Molecular chaperone; mutants confer temperature sensitivity.
<i>fkpA</i>	FkpA, peptidyl-prolyl <i>cis-trans</i> isomerase	Folding catalyst, accelerates the folding of proteins; periplasmic.
<i>ftsJ</i>	FtsJ, ribosomal RNA large subunit methyltransferase	23S rRNA Um522 methyltransferase involved in cell division and growth.
<i>fxsA</i>	FxsA/HslW	Overexpression alleviates the exclusion of phage T7 in F plasmid bearing cells; inner membrane associated.
<i>gapA</i>	GapA, glyceraldehyde-3-phosphate dehydrogenase	Catalyzes 1st step in the 2nd phase of glycolysis; cytoplasmic.
<i>grpE</i>	GrpE, HSP24	With DnaJ, GrpE stimulates the ATPase activity of DnaK.
<i>hflB</i>	HflB, ATP-dependent protease	Essential inner membrane protease with function in the degradation of $\sigma^{32}$ ; integral membrane protein, inner membrane.
<i>hflX</i>	HflX/HslY	GTP binding protein.
<i>hslT</i>	HslT, inclusion body protein	Member of the small heat shock protein family.
<i>hslU</i>	HslU, ATP-dependent protease	ATPase subunit that interacts with HslV, chaperone subunit of a proteasome like degradation complex; cytoplasmic.
<i>hslV</i>	HslV, ATP-dependent protease	Probable subunit of a proteasome-like degradation complex; cytoplasmic.
<i>htpG</i>	HtpG, HSP90	Molecular chaperone with ATPase activity; cytoplasmic.
<i>htpX</i>	HtpX, probable protease	Overexpression leads to an increased degradation of abnormal proteins; integral membrane protein, inner membrane.
<i>htrA</i>	HtrA, membrane-associated serine endoprotease	Involved in protein degradation; essential for survival at 42°C; periplasmic.
<i>lon</i>	Lon, DNA binding, ATP-dependent protease	Degrades abnormal proteins in the presence of ATP; cytoplasmic.
<i>metA</i>	MetA, homoserine <i>O</i> -succinyltransferase	Biosynthesis of methionine; cytoplasmic.
<i>miaA</i>	MiaA, tRNA delta-2-isopentenylpyrophosphate (IPP) transferase	Specific tRNA undermodification may be a switch to higher mutation frequency when cells are subjected to environmental stress.
<i>mopA</i>	MopA, HSP60	Molecular chaperone for the assembly of enzyme complexes; prevents misfolding and promotes refolding and proper assembly of unfolded polypeptides generated under stress conditions; cytoplasmic.
<i>mopB</i>	MopB, HSP10	Molecular chaperone, binds to HSP60; cytoplasmic.
<i>ppiD</i>	PpiD, peptidyl-prolyl <i>cis-trans</i> isomerase	Folding catalyst, accelerates folding of outer membrane proteins.
<i>pspA</i>	PspA, regulatory proteins for phage shock protein operon	Phage shock protein operon may play a role in competition for survival under nutrient or energy limited conditions; inner membrane associated.
<i>rfaD</i>	RfaD, ADP-L-glycero-D-manno-heptose-6-epimerase	Heat inducible lipopolysaccharide, allows high temperature growth.
<i>rpoD</i>	RpoD, RNA polymerase, $\sigma^{70}$ subunit	Initiates the vast majority of exponential phase transcription.
<i>rpoE</i>	RpoE, RNA polymerase sigma E subunit ( $\sigma^E$ )	Extracytoplasmic stress regulon, role in high temperature and oxidative stress response.
<i>rpoH</i>	RpoH, RNA polymerase $\sigma^{32}$ subunit	Initiates transcription off heat shock promoters.
<i>ycjF</i>	YcjF/HslG	Uncharacterized protein in the <i>pspE-tyrR</i> intergenic region.
<i>yhgH</i>	YhgH/HslQ	Function uncharacterized.
<i>yrfG</i>	YrfG/HslP	Putative enzyme.
<i>yrfH</i>	YrfH, HSP15	An abundant heat shock protein that binds RNA and DNA.
<i>yrfI</i>	YrfI, HSP3	Molecular chaperone activated by disulfide bond formation; protects thermally unfolding proteins from aggregation; cytoplasmic.

\*Information on putative cell function from Ecogene (10, 39) (<http://bmb.med.miami.edu/EcoGene/EcoWeb>) or Swissprot from the Swiss Institute of Bioinformatics (2) (<http://ca.expasy.org/sprot/>).

Evolved changes in gene expression were examined by comparing the high-temperature-adapted lines with their ancestor, during growth at the same high temperature (41.5°C) at which the lines had evolved. These data allow us to test whether evolved changes in heat-inducible gene expression are associated with increased fitness at a stressful high temperature and whether similar changes in gene expression occurred repeatedly in the independently derived lines. In a separate experiment, we also measured another aspect

of the heat-shock response [survival at lethal temperature (50°C) following acclimation to a high, but non-lethal, temperature] to determine whether inducible thermotolerance at the level of organismal performance had changed during evolutionary adaptation to high temperature.

The pattern of evolutionary change in the expression of heat-inducible genes may vary depending on the nature of the adaptive response. If the expression of an individual gene were itself specifically important for

high-temperature adaptation, then one would expect to see its repeated, independent evolution in several lines. If, instead, the overall organism-level integration of the expression of the heat-inducible genes as a group were more important for high-temperature adaptation, then many different genetic solutions to this challenge would be possible, leading to heterogeneous patterns of gene expression in the independently derived lines. For example, high-temperature adaptation may involve changes in the expression of different molecular chaperone-encoding genes in each line, rather than altered expression of the same chaperone in all of the evolved lines.

A previous study of these high-temperature-evolved *E. coli* lines showed some evolutionary parallelism at the level of gene duplication and deletion events; however, none of the regions of gene duplication and deletion contains genes that are known to be heat inducible (37). The experimental design and levels of replication used in the present study allow us to determine whether evolved changes in heat-inducible gene expression are parallel or heterogeneous across lines. Alternatively, if the heat-inducible genes are important only in the acute acclimation response to temperature stress, then we should observe no evolutionary changes whatsoever in their expression.

## MATERIALS AND METHODS

### Derivation of *E. coli* Lines

The bacteria used in this study were derived from *E. coli* B strain Bc251, later designated REL606, which is unable to use arabinose (Ara<sup>-</sup>) due to a revertible mutation (28, 29). This strain carries no plasmids or functional phages, and thus it is strictly clonal. The derivation of the lines used in this study involved two stages of experimental evolution under defined conditions (Fig. 1). First, starting with REL606, a number of populations were serially propagated for 2,000 generations at 37°C in Davis minimal medium with 25 µg/ml glucose (designated DM25), as previously described (29). Each day, populations were diluted 100-fold into fresh medium, requiring about 6.6 (=log<sub>2</sub>100) generations of binary fission to offset the serial dilution. Evolutionary adaptation to these conditions was evidenced by gains of ~30%, on average, in competitive fitness relative to the parent strain (29). A clone (designated REL1206) was isolated from one

of these 37°C-evolved populations, and another clone (REL1207) was obtained from it that differed only by a single mutation that restored the ability to grow on arabinose (Ara<sup>+</sup>) (9). The Ara marker is selectively neutral in the glucose medium at the temperatures used in our study, and it is used as a marker in competitive fitness assays (9). The Ara<sup>-</sup> clone served as the ancestor for three new populations that were propagated for 2,000 additional generations under the identical conditions, except these lines were maintained at 41.5°C. Single clones were then isolated from each line; they are individually designated 42-1, 42-2, and 42-3, and they are collectively referred to as the “42 group” (Fig. 1). These high-temperature lines experienced an average fitness gain of about 40% based on competitions at 41.5°C with the reciprocally marked 37°C-evolved ancestor (7). In fitness assays, both competitors were inoculated from freezer stocks and separately acclimated to the assay environment prior to mixing them; therefore, any phenotypic differences must result from underlying genetic changes that occurred during the experimental evolution.

### Measurement of Inducible Thermotolerance

An experiment was performed to measure differences in survival at 50°C between the ancestral and high-temperature-evolved lines, following acclimation to either 37°C or 41.5°C. First, the Ara<sup>+</sup> ancestor and the three evolved clones (42-1, 42-2, and 42-3) were separately inoculated from freezer stocks into 10 ml of Luria broth and grown overnight at 37°C. Each culture was then diluted 10,000-fold into DM25 and incubated at 37°C for 24 h. Next, equal volumes of an evolved clone and the ancestor were mixed, diluted 100-fold into fresh DM25, and grown together for 24 h under two acclimation conditions: 37.0 ± 1.0°C (in an air incubator) or 41.5 ± 0.5°C (in a water bath). All three evolved-ancestor mixtures were replicated sixfold at each acclimation temperature. Samples of each mixed culture were plated on tetrazolium-arabinose indicator agar (29), on which Ara<sup>-</sup> and Ara<sup>+</sup> cells produce red and white colonies, respectively, to determine the initial densities of the two types. The cultures were then transferred to 50°C (in another water bath), and after 4 h, a final sample was plated to determine the density of survivors of the evolved and ancestral types. The survival of each type was calculated as the natural log-transformed ratio of its final to initial density; the larger (less negative) this number, the greater was the survival at 50°C. By examining the survival of the ancestral and evolved types in mixture, any slight variation in temperature is experienced by both types.

### Bacterial Culture Conditions and Total RNA Extraction

Cultures were inoculated from freezer stocks into 10 ml Luria broth and incubated at 37 ± 1.0°C for 24 h; then diluted 100-fold into 10 ml DM500 and incubated for 24 h at 37°C; and finally diluted 100-fold into the same media and incubated for 24 h at 41.5 ± 0.5°C in a shaking water bath at 120 rpm. The next day, cultures were again diluted 100-fold into 10 ml DM500 and incubated at 41.5°C in a shaking water bath, and total RNA was isolated (RNAqueous, Ambion) from cells in mid log-phase growth (Fig. 2). Ten milliliters of culture was poured over ice in a sterile, RNase-free 50-ml centrifuge tube and spun down at 3,000 g for 3 min. The cell pellets were used for RNA extraction. Once isolated, RNA was treated with DNase (Promega) twice to ensure it was free of DNA contamination. Following DNase treatment, total RNA was resuspended in DEPC-treated water and used in cDNA synthesis reactions. All RNA preparations were performed with RNase-free supplies and DEPC-treated liq-

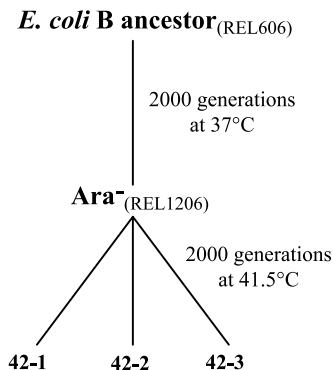


Fig. 1. Derivation of the three high-temperature-evolved lines of *E. coli* (42-1, 42-2, and 42-3) and their ancestor (Ara<sup>-</sup>).



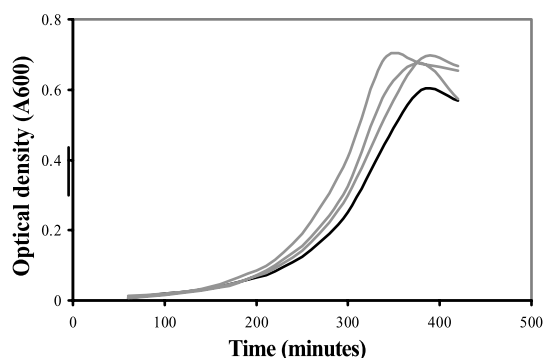


Fig. 2. Growth kinetics at 41.5°C. Gray curves are the three high-temperature-evolved lines; the black curve shows the ancestor. RNA was isolated from cultures in mid log-phase growth (optical density between 0.32 and 0.42, indicated by the bar along the y-axis).

uids. Total RNA was isolated from three independent log-phase cultures for each of the clones (3 evolved and 1 ancestral), for a total of 12 RNA preparations. RNA isolations were performed in blocks (Fig. 3), such that the four RNA samples hybridized to a given membrane were isolated on the same day from cultures that had grown in the same water bath. Data were analyzed by pairing ancestral and selected samples isolated in the same block.

#### Random Hexamer-primed cDNA Synthesis and Labeling

For cDNA synthesis, 20  $\mu$ g of total RNA and 37.5 ng of random hexamers were heated at 70°C for 3 min and then snap cooled on ice. cDNA synthesis was performed at 42°C for 3 h in a 60- $\mu$ l reaction containing the previously denatured RNA and random hexamer combination, reverse transcriptase buffer (Stratagene), 1 mM each of dATP, dTTP, and dGTP (New England Biolabs), 50  $\mu$ Ci [ $\alpha$ -<sup>32</sup>P]dCTP (New England Nuclear), 20 U RNasin (Promega), and 100 U SuperScript reverse transcriptase (GIBCO-BRL). The labeled cDNA was separated from unincorporated nucleotides using Probe Quant G-50 Micro columns (Amersham Pharmacia).

#### DNA High-density Array Hybridization

Panorama *E. coli* high-density gene arrays (Sigma Genosys) have all 4,290 *E. coli* K-12 open-reading frames (ORFs) arrayed in duplicate on 12  $\times$  24-cm nylon membranes. The membranes were soaked in 2 $\times$  SSPE for 10 min, then were prehybridized in 10 ml of hybridization solution (5 $\times$  SSPE, 2% SDS, 1 $\times$  Denhardt's, 0.1 mg/ml sheared salmon sperm DNA) for 1 h at 65°C. After separation from unincorporated nucleotides, the <sup>32</sup>P-labeled cDNA probe was boiled for 10 min and snap cooled on ice. This probe was added to 7.5 ml of hybridization solution and left to hybridize overnight at 65°C. After hybridization, each membrane was washed with 50 ml of 0.5 $\times$  SSPE containing 0.2% SDS at room temperature three times for 3 min each, followed by three washes in the same solution at 65°C for 20 min each. Each membrane was wrapped in plastic and exposed to a phosphor screen for 48 h, which was scanned at 88  $\mu$ m using a Molecular Dynamics PhosphorImager. Membranes were stripped as previously described (37). Note that the use of the K-12 genomic sequence in the construction of the high-density array precludes examining any *E. coli* B genes absent in K-12. However, published work (supplementary information in Ref. 37) hybridizing total genomic DNA to these arrays indicates that all 35 heat-inducible genes identified in K-12 are also present in the B strain that was the ancestor in the high-temperature selection experiment. Also, none of these heat-

inducible genes is present in the large duplications previously reported (37), although the methods used were unable to reliably detect small duplications (<2 ORFs). Because we are comparing the 37°C-evolved ancestor to the high-temperature-evolved lines, any genetic changes that occurred before the isolation of this ancestor are not relevant to the present study.

#### Data Acquisition

We used DNA Arrayvision 4.0 (Imaging Research, London, Ontario, Canada) software to grid the phosphor image, record the pixel density of each spot, and perform background subtractions. The average background was subtracted from each measurement on the same membrane, because background readings were fairly constant across a particular membrane. Background-subtracted values were used in the statistical analyses.

#### Experimental Design and Statistical Analyses

The experimental design for this study entailed three independent cDNA labelings and hybridizations for each of the four clones, the Ara<sup>-</sup> ancestor and three high-temperature-evolved derivatives (42-1, 42-2, and 42-3). Each preparation was hybridized to one of three membrane arrays, and an entire block was hybridized to the same membrane array (Fig. 3).

The expression level for each spot was normalized by scaling the background-subtracted expression level to the total count on the membrane. Thus a positive value indicates expression above background, whereas a value equal to zero indicates expression at (or slightly below) background. Because every ORF is arrayed in duplicate on each membrane, the two values were averaged for each gene. Differences between the natural log-transformed expression values of each evolved line and the ancestor were calculated, and these differences were then averaged across replicates. The standard deviation is then the standard deviation of the three differences associated with each replicate within an evolved/ancestral line comparison. Differences were tested statistically using the paired-data analysis option from Cyber-T (<http://www.igb.uci.edu/servers/cybert/>), a web-based program written in the R statistical programming language (<http://www.r-project.org>) for the paired *t*-test using a Bayesian estimate of the standard deviation to assign *P* values (6, 32). Ancestral and derived samples isolated on the same day were paired in the analysis. The paired *t*-test evaluates whether the mean difference between two samples (e.g., ancestor and 42-1) is significantly different from zero; it employs the mean difference between paired samples relative to the variance in these differences, whereas a standard *t*-test uses the variance within samples. The Bayesian estimate of the standard deviation is a weighted average of the observed standard deviation and a prior estimate of the standard deviation based on genes with similar expression

Probing Order	Membrane/Block		
	A	B	C
1	Ara <sup>-</sup>	42-3	42-2
2	42-1	42-2	Ara <sup>-</sup>
3	42-2	42-1	42-3
4	42-3	Ara <sup>-</sup>	42-1

Fig. 3. Experimental design for expression array studies of the Ara<sup>-</sup> ancestor and its three high-temperature-evolved derivatives. In each of three blocks (labeled A–C), the four clones were grown simultaneously and corresponding hybridizations used the same membrane. Statistical analyses of data within a block used a paired approach, in which evolved samples were paired with the ancestral sample from the same day (see MATERIALS AND METHODS).

levels. The Bayesian  $P$  values were obtained from Cyber-T for each paired  $t$ -statistic using 11 df [ $(n = 3) + (v_0 = 10) - 2$ ], a weighting of 10 for the Bayesian prior ( $v_0$ ), and a sliding window of 101 genes (the latter two being the default settings for Cyber-T). Because the paired analysis evaluates differences in expression values and not the raw values, we provided Cyber-T with an estimate of the absolute expression level by summing the natural log-transformed expression levels for the relevant lines, to obtain the Bayesian estimate of the standard deviation.

We chose to employ the paired analysis because the “block effect” associated with the day of RNA isolation was substantial; by performing paired analyses, one effectively eliminates this source of variation. The paired analysis compares only the samples that were isolated on the same day from cultures growing in the same water bath. This approach may be especially useful for studies at high temperature, where even small fluctuations have been demonstrated to have large effects on organismal performance in this system (8).

**Analysis of heat-inducible genes.** Genes that encode 43 heat-inducible proteins (including molecular chaperones, ATP-dependent proteases, and others) were previously identified (44). Of these 43 candidates, 35 gave expression readings above background for all 3 replicates in all 4 clones in this study, and these 35 are included in our analyses (the others are not considered further). A total of 1,964 genes produced readings above background for all replicates in all lines.

**Analysis of the heat-inducible genes as a group.** To analyze evolved expression changes in the heat-inducible genes as a group, a statistic  $\tau$  was calculated that summarizes the statistical support integrated over all the candidate genes;  $\tau$  equals the sum of the natural logarithms of the  $P$  values over all 35 genes over all three evolved lines (a total of 105 values). The significance of this statistic is assessed using a chi-square test (41) with 210 df ( $2 \times 105$ ). Alternatively, using all other above background gene from the array, one can obtain the empirical distribution of  $\tau$ . This is accomplished by randomly sampling groups of 35 genes (from those that gave readings above background) 10,000 times without replacement and calculating  $\tau$  for each sample. This analysis can also be carried out at the level of individual evolved lines by randomly sampling groups of 35 genes 10,000 times without replacement within a given line.

**Analysis of individual candidate genes over evolved lines.** To analyze expression changes in individual heat-inducible genes over evolved lines, the natural logarithms of the three  $P$  values for a given gene (one value from each derived-versus-ancestral comparison) were summed, and the significance was tested using a chi-square test with 6 df ( $2 \times 3$  lines). Notice that this test statistic could be significant for the group as a whole even if changes are in the opposite direction in independently derived lines, provided the  $P$  values are sufficiently strong for the individual lines.

**Analysis of individual candidate genes in independently derived lines.** To visualize patterns of gene expression, a custom program was written in the R language to create “Eisen diagrams” (Ref. 18; the R code for the program is available as Supplemental Material at the *Physiological Genomics* web site),<sup>1</sup> except that here the color scale is proportional to the paired  $t$ -statistics (i.e., statistical support for change) rather than to fold change. Previous work in *E. coli* and *D. melanogaster* (3, 23, 32) has shown that statistical

significance in replicated expression profiling experiments does not necessarily scale with fold change. Paired  $t$ -statistics were obtained from Cyber-T and tested with 11 df [i.e.,  $(n = 3) + (v_0 = 10) - 2$ ]. Genes with reduced expression in the derived lines relative to the ancestor are shown in shades of blue, and those with increased expression relative to the ancestor are shown in yellow (Fig. 4). The genes indicated in brighter shades have larger fold changes in expression, smaller variation in expression (suggesting tight regulation), or some combination thereof. That is, large absolute values of the  $t$ -statistic are expected when the paired differences between ancestral and evolved samples are large relative to the variance across replicate pairs. We defined those genes that had changes in the same direction (i.e., increased or decreased) in all three evolved lines, with each significant at  $P < 0.05$ , as exhibiting strong evolutionary parallelism, or replicability. Genes showing moderate levels of replicability were defined as those with changes in the same direction across all three lines, of which only two were significant ( $P < 0.05$ ).

## RESULTS

### 41.5°C Is Stressful

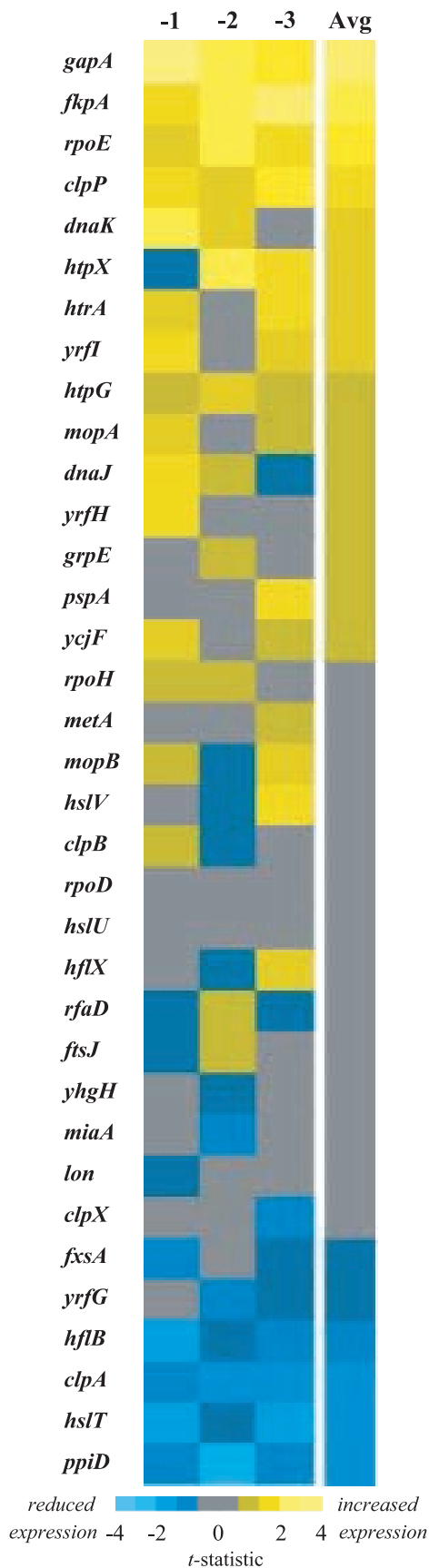
Despite being  $<5^\circ\text{C}$  above the ancestral temperature of  $37^\circ\text{C}$ ,  $41.5^\circ\text{C}$  is stressful to the ancestral strain in terms of reducing both growth rate and yield. The growth rate of the ancestor is about 10% lower at  $41.5^\circ\text{C}$  than at  $37^\circ\text{C}$ , and its biovolume yield (i.e., total cell volume at stationary phase) is 56% lower at  $41.5^\circ\text{C}$  than at  $37^\circ\text{C}$  (unpublished data, A. F. Bennett and A. J. Cullum). The three lines that evolved at  $41.5^\circ\text{C}$  are 42% more fit, on average, than the ancestor when competing at that temperature (7).

### Evolutionary Change In Heat-Inducible Gene Expression

Table 2 and Fig. 4 summarize the evolutionary changes in expression patterns of the heat-inducible candidate genes. These changes include both increases and decreases in expression levels of individual genes. The set of Ara<sup>-</sup>-derived lines (42-1, 42-2, and 42-3) exhibits a significant difference from the ancestor in the overall pattern of expression of the 35 candidate genes ( $\tau = -129.76$ ). Based on a random sampling of 10,000 groups of 35 genes each, the probability of obtaining a  $\tau$ -statistic as or more extreme than the value observed is only 0.007. Based on chi-square with 210 df ( $2 \times 3$  lines  $\times$  35 genes per line), this difference in expression is also significant ( $P < 0.01$ ). At the level of individual evolved lines, the group of 35 candidate genes shows changes in expression compared with the ancestor (42-1,  $\tau = -44.74$ ,  $P = 0.18$ ; 42-2,  $\tau = -42.33$ ,  $P < 0.05$ ; 42-3,  $\tau = -42.69$ ,  $P < 0.03$ ). Despite the lowest  $\tau$ -statistic, the change is not significant in 42-1, because of a greater number of statistically significant differences among noncandidate genes; however, if one tests the  $\tau$  statistic for 42-1 ( $-44.74$ ) against  $\tau$  distributions for 42-2 and 42-1, the result is highly significant ( $P < 0.02$ ).

At the level of individual candidate genes, 7 of 35 exhibit significant differences in expression between

<sup>1</sup>The Supplementary Material for this article (R code for clustering diagrams) is available online at <http://physiolgenomics.physiology.org/cgi/content/full/00034.2002/DC1>.



the group of high-temperature-evolved lines and their ancestor, as assessed by chi-square tests with 6 df ( $2 \times 3$  lines). The genes with significant changes in expression include three involved with the stress response in the periplasmic space: *rpoE*, an RNA polymerase gene; and *fkpA* and *ppiD*, both peptidyl-prolyl *cis-trans* isomerases that promote protein folding. Of the remaining 4, *clpP* is the proteolytic subunit of ATP-dependent proteases; *hstT* encodes a heat-shock protein that binds heat-denatured proteins and prevents their aggregation; *gapA* encodes the protein that converts glyceraldehyde 3-phosphate into 1,3-bisphosphoglycerate during a pivotal step in glycolysis; and *htpX* is a probable protease. Of these seven genes, six exhibit parallel changes in the evolved lines, with four having increased (*rpoE*, *fkpA*, *clpP*, *gapA*) and two (*ppiD*, *hslT*) decreased gene expression in all three lines. The remaining gene, *htpX*, shows heterogeneous expression changes across lines.

We now move from analyzing genes with changed expression in the set of three evolved lines, taken as a group, to considering the pattern of changes in individual lines. Of the 35 candidate genes, 27 show no significant differences in expression between any individual line and the ancestor, 5 show significant changes in one line, 2 show changes in two lines, and 1 gene exhibits expression differences from the ancestor in all three of the independently derived lines (all with  $P < 0.05$ ). The probability of a particular gene showing significant changes in all three independently derived lines by chance alone is  $0.05^3$ ; the probability that one or more genes of 35 would exhibit this pattern by chance alone is  $1 - (1 - 0.05^3)^{35} \cong 0.004$ . Of the 35 candidate genes, *gapA* had increased expression in all three derived lines ( $P < 0.05$ ), indicating a high degree of evolutionary replicability; whereas *fkpA* and *hslT* showed moderate replicability (i.e., all three lines with expression changes in the same direction, with two significant at  $P < 0.05$ ). None of the genes show significant changes in opposite directions in different evolved lines, but two (*dnaK*, *htpX*) show significant changes in expression in one direction in one line and nonsignificant changes in the opposite direction in another line (Table 2). The average number of expression changes within an evolved line is 4.

It has been customary in analyzing data from expression profiling experiments to use an arbitrary threshold of twofold to identify genes of biological importance in the context of the experiment. However, when adequate levels of experimental replication and statistical tests are employed, many genes showing

Fig. 4. Evolutionary changes in expression of the 35 heat-inducible candidate genes. The three high-temperature-evolved lines are indicated along the top, with their average, and gene names are shown on the left. The color of each cell shows the change in gene expression relative to the ancestral level; yellow indicates increased expression in the evolved line, whereas blue is reduced expression. The intensity of color is based on a *t*-statistic, which quantifies the magnitude of the change relative to the variation among the replicates. Values above 2.2 and below -2.2 are significant at  $P < 0.05$  with 11 df.



Table 2. Expression values for 35 candidate genes

"b" Number	Gene Name	Line 42-1				Line 42-2				Line 42-3				Across Lines	
		Average	Bayesian SD	Bayesian P value	Fold change	Average	Bayesian SD	Bayesian P value	Fold change	Average	Bayesian SD	Bayesian P value	Fold change	Overall P value	Mean fold change
b0882	<i>clpA</i>	-0.1921	0.3001	0.291	-1.21	-0.4205	0.3763	0.079	-1.52	-0.3222	0.3553	0.145	-1.38	0.076	-1.37
b2592	<i>clpB</i>	0.1846	0.3271	0.349	1.20	-0.1816	0.4099	0.459	-1.20	0.0449	0.4469	0.865	1.05	0.683	1.02
b0437	<i>clpP</i>	0.3501	0.3880	0.146	1.42	0.4151	0.5124	0.188	1.51	0.6814	0.4885	<b>0.034</b>	1.98	<b>0.030</b>	1.62
b0438	<i>clpX</i>	0.0167	0.2800	0.920	1.02	-0.0568	0.3501	0.784	-1.06	-0.2166	0.3680	0.330	-1.24	0.825	-1.09
b0015	<i>dnaJ</i>	0.2678	0.2389	<b>0.078</b>	1.31	0.1738	0.3392	0.394	1.19	-0.1265	0.3426	0.536	-1.13	0.223	1.11
b0014	<i>dnaK</i>	0.5133	0.3111	<b>0.016</b>	1.67	0.2822	0.3691	0.212	1.33	-0.0582	0.5652	0.862	-1.06	0.068	1.28
b3347	<i>fkpA</i>	0.2651	0.2511	0.095	1.30	0.4161	0.2595	<b>0.018</b>	1.52	0.7318	0.4216	<b>0.012</b>	2.08	<b>0.001</b>	1.60
b3179	<i>ftsJ</i>	-0.2806	0.6436	0.466	-1.32	0.2202	0.5634	0.512	1.25	-0.1110	0.6156	0.761	-1.12	0.756	-1.06
b4140	<i>fxsA</i>	-0.2322	0.3780	0.310	-1.26	-0.0308	0.3970	0.896	-1.03	-0.1569	0.4206	0.531	-1.17	0.700	-1.15
b1779	<i>gapA</i>	0.6175	0.2748	<b>0.003</b>	1.85	0.5823	0.3640	<b>0.018</b>	1.79	0.4878	0.3434	<b>0.032</b>	1.63	<b>&lt;0.001</b>	1.76
b4143	<i>groEL/mopA</i>	0.2061	0.2988	0.257	1.23	0.0471	0.3045	0.794	1.05	0.1313	0.3634	0.544	1.14	0.624	1.14
b4142	<i>groES/mopB</i>	0.1686	0.4085	0.490	1.18	-0.2034	0.4297	0.430	-1.23	0.2229	0.3794	0.331	1.25	0.502	1.06
b2614	<i>grpE</i>	0.0876	0.3325	0.657	1.09	0.2174	0.3794	0.342	1.24	0.0494	0.2963	0.778	1.05	0.746	1.13
b3178	<i>hflB</i>	-0.2841	0.2268	0.053	-1.33	-0.1300	0.2504	0.388	-1.14	-0.1997	0.3263	0.312	-1.22	0.120	-1.23
b4173	<i>hflX</i>	-0.0606	0.3267	0.754	-1.06	-0.1928	0.4425	0.466	-1.21	0.2170	0.3636	0.323	1.24	0.630	-1.01
b3687	<i>hslT</i>	-0.5303	0.4050	<b>0.044</b>	-1.70	-0.1720	0.4041	0.476	-1.19	-0.5989	0.4511	<b>0.042</b>	-1.82	<b>0.029</b>	-1.54
b3931	<i>hslU</i>	-0.0528	0.2835	0.753	-1.05	-0.0248	0.3134	0.893	-1.03	0.0124	0.3171	0.947	1.01	0.989	-1.02
b3932	<i>hslV</i>	-0.0292	0.4316	0.909	-1.03	-0.3847	0.7371	0.385	-1.47	0.4518	0.4097	0.083	1.57	0.313	1.01
b0473	<i>htpG</i>	0.2025	0.3967	0.395	1.22	0.2508	0.3789	0.276	1.29	0.1366	0.4206	0.585	1.15	0.481	1.22
b1829	<i>htpX</i>	-0.1346	0.2970	0.449	-1.14	0.4899	0.3177	<b>0.022</b>	1.63	0.3042	0.2857	0.092	1.36	<b>0.029</b>	1.25
b0161	<i>htrA</i>	0.1442	0.2399	0.320	1.16	0.0556	0.2744	0.732	1.06	0.3934	0.3806	0.101	1.48	0.278	1.22
b0439	<i>lon</i>	-0.0704	0.2290	0.605	-1.07	-0.0572	0.2954	0.744	-1.06	0.0748	0.2755	0.647	1.08	0.872	-1.02
b4013	<i>metA</i>	0.0037	0.2734	0.982	1.00	-0.0297	0.3166	0.874	-1.03	0.1469	0.3580	0.492	1.16	0.943	1.04
b4171	<i>miaA</i>	-0.0374	0.2740	0.817	-1.04	-0.2240	0.3058	0.231	-1.25	0.0298	0.3141	0.872	1.03	0.729	-1.08
b0441	<i>ppiD</i>	-0.1334	0.2268	0.330	-1.14	-0.8719	0.5166	<b>0.014</b>	-2.39	-0.2604	0.3065	0.169	-1.30	<b>0.026</b>	-1.52
b1304	<i>pspA</i>	-0.1223	0.8212	0.801	-1.13	0.1232	0.8866	0.814	1.13	0.5736	0.6038	0.128	1.77	0.549	1.21
b3628	<i>rfaD</i>	-0.1236	0.4081	0.610	-1.13	0.2275	0.4611	0.411	1.26	-0.1874	0.5186	0.544	-1.21	0.679	-1.03
b3067	<i>rpoD</i>	0.0620	0.3005	0.728	1.06	-0.0640	0.3132	0.730	-1.07	0.0605	0.3695	0.782	1.06	0.941	1.02
b2573	<i>rpoE</i>	0.2009	0.2552	0.200	1.22	0.4075	0.2555	<b>0.018</b>	1.50	0.3273	0.2836	0.071	1.39	<b>0.011</b>	1.37
b3461	<i>rpoH</i>	0.1541	0.4883	0.596	1.17	0.2433	0.5838	0.485	1.28	0.0475	0.6135	0.896	1.05	0.845	1.16
b1322	<i>ycjF</i>	0.2267	0.3910	0.337	1.25	0.0052	0.4731	0.985	1.01	0.1391	0.4233	0.581	1.15	0.771	1.13
b3413	<i>yhgH</i>	0.1010	0.3938	0.666	1.11	-0.2892	0.6593	0.463	-1.34	-0.0287	0.4690	0.917	-1.03	0.866	-1.08
b3399	<i>yrfG</i>	-0.0987	0.5250	0.751	-1.10	-0.5489	0.7690	0.242	-1.73	-0.1791	0.4278	0.483	-1.20	0.561	-1.32
b3400	<i>yrfH</i>	0.3470	0.3501	0.114	1.41	0.0131	0.4004	0.956	1.01	0.0507	0.4590	0.852	1.05	0.576	1.15
b3401	<i>yrfI</i>	0.3870	0.3567	0.087	1.47	0.0502	0.4192	0.839	1.05	0.2604	0.4325	0.319	1.30	0.276	1.26

P values <0.05 are shown in bold. "Average" indicates mean of 3 replicate differences in log-transformed expression level between the evolved and ancestral line. "Overall P value" ("Across Lines") indicates combined P values ( $\Sigma \ln P$ ) significance assessed with  $\chi^2$ , df = 6.

significant changes in expression fall below this arbitrary threshold. In fact, all seven of the heat-inducible genes that show significant differences between the ancestor and the group of high-temperature-evolved lines differ, on average, by less than twofold (Table 2). The largest proportional change between the ancestor and the mean of the evolved lines was +1.76-fold for *gapA*. The expression of *rpoE* increased on average by only +1.37-fold in the evolved lines, but this change was nonetheless significant because its magnitude was highly consistent across the three lines. Only 2 of the 12 significant changes in individual lines show greater than twofold changes (*ppiD* in 42-2, *fkpA* in 42-1). Moreover, some noncandidate genes (see next section) that changed more than twofold are not significant ( $P > 0.05$ ), whereas others that changed less than twofold are significant. Thus there is no general correspondence between this arbitrary fold-change criterion and statistical significance.

### Whole Genome Results

We focused our attention on candidate genes for the biological reasons explained in the Introduction. We also sought to apply rigorous statistical analyses, while avoiding becoming mired in the numerous signals,

including false positives, that one expects when thousands of dependent variables (the number of ORFs on the arrays) are simultaneously subjected to statistical tests (12). Nonetheless, the whole-genome expression patterns are briefly summarized here, to contrast candidate and noncandidate genes. In the Ara<sup>-</sup> lines, there were 1,929 noncandidate genes that showed expression levels above background level in all three evolved lines and the ancestor. Of these, 137 (7%) differed significantly ( $P < 0.05$ ; based on chi-square with 6 df) between the ancestor and the average evolved line. Although this proportion is slightly above the 5% expected by chance alone, it is significantly less than the 7 of 35 (20%) among the candidate genes ( $P = 0.0037$ , one-tailed chi-square test). Furthermore, candidate genes are more likely to exhibit either moderate or high levels of replicability across independent lines, as defined previously (3/35 candidate genes vs. 36/1,929 noncandidates:  $P = 0.005$  by chi-square test). In summary, although there were significant changes in expression of some noncandidate genes, the 35 candidate genes identified from previous work on short-term, acclimation responses to high-temperature stress (44) were significantly more likely to have altered expression levels during the long-term, evolu-

tionary response to the high-temperature selection regime.

### Evolutionary Changes in Inducible Thermotolerance

Following acclimation at 41.5°C for 24 h, all three high-temperature-evolved lines survived a 4-h exposure to 50°C better than when they were acclimated instead to 37°C ( $P < 0.05$ , Fig. 5). This difference confirms the persistence of inducible thermotolerance in the evolved lines. Moreover, two of the three lines (42-1 and 42-2) survived significantly better than did the ancestor at 50°C after acclimation to 41.5°C (both  $P < 0.01$ ), as did the three lines considered as a group ( $P < 0.05$ ), indicating that the degree of thermotolerance had evolved.

### DISCUSSION

Three lines of *E. coli* evolved for 2,000 generations at 41.5°C, a stressful but nonlethal temperature (Fig. 1). Compared with the ancestor, these lines show significant and extensive changes in competitive fitness at 41.5°C (7), growth rate at 41.5°C (Fig. 2), heat-inducible gene expression (Table 2, Fig. 4), and inducible thermotolerance (Fig. 5). Because of the fact that all strains are removed from the freezer and acclimated in parallel to the same experimental conditions, we know that significant differences in phenotype must be caused by underlying genetic changes, even though the precise mutations are unknown. These results support the hypothesis that increased fitness at high temperature is associated with altered patterns of heat-inducible gene expression compared with the ancestor. Many of the evolved changes in heat-inducible expression were parallel, such that the three lines either all increased or all decreased their expression levels. Of the seven candidate genes whose expression changed significantly in the evolved lines as a group, six show

parallel changes: four have increased (*rpoE*, *fkpA*, *clpP*, *gapA*) and two (*ppiD*, *hslT*) have decreased gene expression in all three lines; the seventh such gene (*htpX*) has diverged in its expression across the evolved lines (Table 2). None of the 35 candidate genes show significant changes in opposite directions in different lines, but two genes (*dnaK*, *htpX*) show significant expression changes in one direction in one line and nonsignificant changes in the opposite direction in another line. Therefore, in addition to parallel evolutionary changes in expression of heat-inducible genes, there appear to be various combinations of expression changes that can increase fitness at high temperature. Three replicate evolved lines are inadequate to fully resolve the parallel vs. divergent pattern of changes; many more replicate lines would be useful for addressing this issue. Also, not all of the heat-inducible genes participated in evolutionary adaptation to high temperature. Therefore, there are substantial differences in gene expression between bacteria phenotypically acclimated to high temperature (i.e., ancestors grown at 41.5°C) and those that have adapted genetically as well as acclimated phenotypically to high temperature (i.e., evolved lines selected and grown at 41.5°C).

### Comparison of Candidate and Noncandidate Genes

Our decision to focus on the heat-inducible genes as candidates for evolutionary adaptation was based on two considerations. First, because the heat-inducible genes are involved in physiological acclimation, they are functionally meaningful candidates for involvement in genetic adaptation as well. Second, with more than 4,000 genes in the *E. coli* genome, it is difficult to make sense of all the expression changes simultaneously. The decision to focus on an a priori set of candidate genes is supported by our finding that a significantly higher proportion of these genes underwent evolutionary changes in expression than did all other noncandidate genes as a group. The fact that 20% of the heat-inducible genes exhibited an evolved response across evolved lines confirms overlap between the genes involved in physiological acclimation and genetic adaptation. The expression changes observed in the noncandidate genes will require additional study to determine how many and which ones are biologically meaningful as the 7% statistically significant noncandidate genes is only slightly above the 5% proportion expected by chance alone.

### Evolved Changes in Expression of Several Heat-inducible Genes

Genes that show similar changes in expression across the evolved lines provide evidence for the repeatability of evolutionary adaptation at the level of gene expression. These repeatable changes in expression complement and extend previous evidence for parallel gene duplications in some of these same evolved lines (37). However, none of the heat-inducible genes examined in this study are present in the previously identified regions of gene duplication.

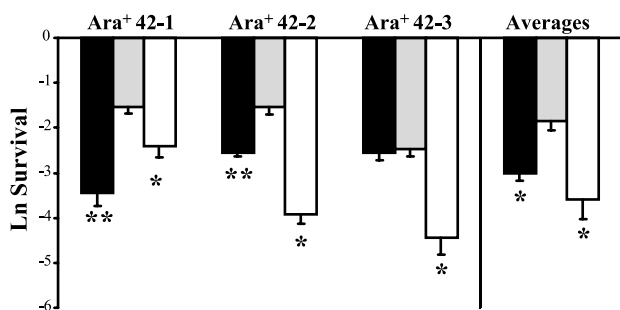


Fig. 5. Evolved changes in inducible thermotolerance. Gray bars and white bars show the survival of the three high-temperature-evolved lines at 50°C following acclimation for 24 h to 41.5°C and 37°C, respectively. Black bars show the survival of the paired ancestral controls at 50°C after acclimation to 41.5°C. Data are natural log-transformed ratios of final to initial counts during 4 h at 50°C. Higher (less negative) numbers indicate better survival. Error bars are standard errors based on 6 assays for each line under each condition. The evolved lines show significant heat-inducible thermotolerance (white bars vs. gray bars:  $*P < 0.05$  for all three lines and the group mean). Survival is also greater in the evolved lines than the ancestral controls following acclimation at 41.5°C (black vs. gray bars:  $**P < 0.01$  for two of three lines,  $*P < 0.05$  for the group mean).



The *gapA* gene exhibited the most pronounced evolutionary parallelism of the 35 candidate genes, with all three lines showing increased expression that was significant for each line individually as well as the set of three lines taken together (all  $p < 0.05$ ). This gene encodes GapA, the most efficient of three *E. coli* isoenzymes that phosphorylate glyceraldehyde 3-phosphate to produce 1,3-biphosphoglycerate in a crucial step during glycolysis (40). The *gapA* gene is transcribed from one of four promoters, depending on environmental conditions, thereby apparently ensuring survival under varying and poor conditions. The second promoter is recognized by the heat-shock RNA polymerase,  $\sigma^H$ , such that transcription of *gapA* is maintained under temperature stress (11).

Two other genes, *fkpA* and *hslT*, exhibit evolutionary parallelism in expression that is less pronounced than for *gapA* but nonetheless is compelling. In *fkpA* and *hslT*, there were significant changes in the high-temperature-evolved lines as a group and significant expression changes ( $P < 0.05$ ) in two of the three individual lines, and the third line showed changed expression in the same direction. The expression of *fkpA* increased, whereas that of *hslT* declined, in the evolved lines (Fig. 4). FkpA, the protein product of *fkpA*, is one of three soluble peptidyl-prolyl *cis-trans* isomerases (PPIases) in the periplasm of *E. coli* (13). In general, PPIases accelerate protein folding. FkpA has been shown to interact with early protein-folding intermediates, thus preventing their aggregation, and to play a role in the reactivation of inactive proteins by binding to unfolded polypeptides and releasing others to a synthesis pathway. Null mutations in *fkpA* have been shown to stimulate transcription of another gene, *htrA* (*degP*), that encodes a periplasmic protease, evidently because the absence of FkpA compromises extracytoplasmic protein folding and thereby causes an increased need for extracytoplasmic proteases (34). In a previous study of the response to an acute temperature stress (sudden shift from 37–50°C) in *E. coli* strain K-12 (36), neither *gapA* nor *fkpA* exhibited altered expression levels, possibly highlighting the difference between the short-term acclimation response and the long-term evolutionary response, although differences in strains and other culture conditions might also be involved particularly the difference in responses to 41.5°C and 50°C.

The expression of *hslT* was reduced in the lines that evolved at 41.5°C (Fig. 4). HslT is a small heat-shock protein that binds heat-denatured proteins and holds them in a nonaggregating state (33). Although overexpression of HslT does increase resistance to heat and other stresses, null mutants are no more susceptible to high temperature than are wild-type cells, indicating that HslT is nonessential at high temperature (24). As was previously suggested (24), other chaperones may compensate for the lost HslT function. In any case, our finding of lower *hslT* expression after long-term evolutionary adaptation to high temperature stands in contrast to the large increase in *hslT* expression following acute temperature stress (36).

Five other genes, *clpP*, *dnaK*, *htpX*, *ppiD*, and *rpoE*, show significant changes in expression in only one of the three evolved line, although in four of them (*clpP*, *htpX*, *ppiD*, and *rpoE*) other lines had sufficiently similar responses that the overall change in the evolved lines as a group was also significant (Fig. 4). The expression level of *clpP* increased in the evolved lines; it encodes the proteolytic subunit of two ATP-dependent proteases that are important in the degradation of misfolded proteins. The expression of *dnaK* also tended to increase in the evolved lines; it encodes the prokaryotic homolog of the HSP70 molecular chaperone gene found in eukaryotes. The expression of *htpX* increased in two lines and declined in a third; it encodes a putative protease, and mutants with *htpX* disrupted can grow at all temperatures and show no obvious phenotypic effects (26). Expression of *ppiD* tended to decline in the evolved lines; it encodes a PPIase and is the only member of the  $\sigma^H$  regulon that has been shown to date to participate in folding noncytoplasmic proteins. Null mutations in *ppiD* reduce the folding of outer-membrane proteins and cause induction of the periplasmic stress response (15). The expression of *rpoE* increased in the high-temperature-evolved lines. It encodes an RNA polymerase,  $\sigma^E$ , that controls transcription of genes involved in the extracytoplasmic stress response, which leads to proper protein folding in the periplasmic space as well as inner and outer membranes (35). In an earlier study of the acute response of *E. coli* to high temperature, *rpoE* was up-regulated (36), but the extent of this phenotypic response was much greater (5- to 15-fold) than the evolutionary response during long-term propagation at constant high temperature. Similar differences between acute and evolved responses exist for *clpP*, *dnaK*, and *htpX* expression. Such differences emphasize the need to distinguish between phenotypic acclimation and genetic adaptation in understanding changes in gene-expression profiles.

#### *Absence of Evolutionary Change In Expression of Other Candidates*

Although the 8 candidate genes discussed above underwent evolutionary changes in expression in one or more of the evolved lines, the other 27 did not. For example, in contrast to the parallel acclimation and evolutionary stress responses of *rpoE* expression, *rpoD* exhibits an acclimation response but showed no evolutionary change in its level of expression. The *rpoD* encodes the RNA polymerase,  $\sigma^{70}$ , responsible for the transcription of a majority of genes expressed during exponential growth (21). The constancy of *rpoD* expression during evolutionary adaptation to high temperature presumably ensures the appropriate transcription of many genes that are generally important for bacterial growth. It should also be noted that changes in *rpoD* expression during other growth phases, or under other culture conditions, cannot be excluded. Our expression data were obtained for cells in mid log-phase growth, at the same temperature and in the same

medium in which the experimental evolution took place, except that a higher glucose concentration was used for this array study to provide a higher yield of RNA.

#### *Expression Changes Across Genes of Similar Function*

Many of the heat-inducible candidate genes encode either molecular chaperones or ATP-dependent proteases (Table 1). In fact, the nine genes in each of these functional categories together make up over half of the heat-inducible candidates examined in this study. The chaperone-encoding genes (*dnaK*, *dnaJ*, *grpE*, *hslT*, *htpG*, *mopA*, *mopB*, *yrfI*, and *yrfH*), as a group, had significantly different expression profiles in evolved line 42-1 compared with the ancestor ( $\tau = -17.85$ ,  $P < 0.0076$ ,  $df = 9 \text{ genes} \times 2 = 18$ ), whereas neither of the other evolved lines underwent significant changes in molecular-chaperone gene expression (line 42-2,  $P = 0.75$ ; line 42-3,  $P = 0.63$ ). This difference is particularly interesting because line 42-1 has significantly higher fitness at high temperature than do the other evolved lines; moreover, 42-1 is the only one of the high-temperature lines to have extended the upper limit of its thermal niche, and it is also the only one to have lost fitness at lower temperatures (8). This pattern suggests that expression of the chaperone-encoding genes may be responsible for these extreme responses, although this should be viewed at present as a hypothesis for further study. As a group, the nine heat-inducible genes encoding ATP-dependent proteases (*clpA*, *clpB*, *clpP*, *clpX*, *hflB*, *hslU*, *hslV*, *htpX*, and *lon*) show no significant changes in expression for any of the evolved lines relative to the ancestor (line 42-1,  $P = 0.47$ ; line 42-2,  $P = 0.20$ ; line 42-3,  $P = 0.10$ ). Although the genes that produce the molecular chaperones and ATP-dependent proteases represent a large fraction of the acute response to high-temperature stress, neither group appears to be the main player in the long-term evolutionary response to this same stress.

Both *ppiD* and *fkpA* encode peptidyl-prolyl *cis-trans* isomerases that accelerate outer-membrane and periplasmic protein folding (1; Table 1). The evolved lines as a group had significantly higher *fkpA* expression and significantly lower *ppiD* expression (Table 2), suggesting that evolution at 41.5°C favored a shift in the relative abundance of these functionally similar proteins. The promoters for these two protein-folding catalysts are recognized by different RNA polymerases: *fkpA* by  $\sigma^E$ , and *ppiD* by both  $\sigma^H$  and the Cpx two-component system (14, 15, 34). The increased expression of *rpoE*, which encodes  $\sigma^E$ , may have contributed to the increased expression of *fkpA*.

The heat-inducible candidates encode proteins that function in the cytoplasm, membrane, or periplasm (Table 1). It appears that the evolved response was focused on the expression of those genes encoding periplasmic functions. Three of the four genes with periplasmic functions (*fkpA*, *ppiD*, *rpoE*, but not *htrA*) evolved significant changes in expression in the high-temperature lines as a group, in contrast with only 4 of

the 31 cytoplasmic and membrane-associated genes (*clpP*, *gapA*, *hslT*, *htpX*). This difference is supported by a chi-square test ( $P < 0.005$ ), and it suggests that genes involved in the less studied extracytoplasmic stress response may play larger roles in long-term adaptation to high temperature than those genes in the well-characterized cytoplasmic stress response.

The evolved lines showed no overall trend toward higher or lower expression of the 35 heat-inducible candidate genes as a group relative to the ancestor. Heat-inducible genes with increased, decreased, and unchanged expression relative to the ancestor were all encountered, as discussed earlier. A potential problem with full-length ORF arrays is cross-hybridization of a given labeled cDNA probe with several ORFs on the array. This cross-hybridization may cause ORFs with high sequence similarity to show correlated changes in expression (36). However, an analysis of the 35 candidate genes studied here indicates that none has high DNA homology with other *E. coli* genes.

#### *Evolved Changes in Gene Expression in Other Systems*

Only two previously published studies have used DNA arrays to quantify changes in gene expression during experimental evolution, and neither one involved adaptation to thermal stress (12, 19). One examined three lines of yeast that had evolved in glucose-limited cultures (19). That work was the first in which arrays were used to study evolved responses, in contrast to acute changes in expression produced under different culture conditions. This yeast study found parallel expression changes across lines in a number of genes involved in efficient glucose utilization. No formal statistical tests were used in this earlier study, which relied on proportional changes in expression level. The design in our study used threefold replication of each evolved and ancestral clone, which provided information on both the magnitude of any expression change and the experimental error inherent in measuring that change. Given this replication, formal statistical tests were used to identify genes for which expression patterns had evolved. This approach allowed us to identify not only genes undergoing parallel changes across independently evolved lines, but also genes whose expression levels had diverged in replicate evolved lines. For example, the expression of *htpX* differed between lines 42-1 and 42-2 ( $P < 0.005$ ).

The other published study using genomic arrays to investigate evolved changes in gene expression employed *E. coli* strains closely related to those in this study (12). In particular, this earlier study compared the expression profiles of two lines that evolved at 37°C for 20,000 generations and their ancestor. The ancestor of the 37°C-evolved lines is strain REL606, from which the ancestor of the high-temperature lines was also derived following 2,000 generations at 37°C (Fig. 1). As with our study, this earlier study used replication and formal statistics to identify genes that underwent significant changes in expression. But unlike our study,

the 37°C selection regime did not suggest any obvious pool of candidate genes. The authors of this earlier study therefore considered all of the genes in the genome, then narrowed their focus to those genes that exhibited significant changes in both evolved lines. Strikingly, all 59 genes identified by this criterion showed parallel changes, in which both lines independently evolved either increased or decreased expression (12). Only one of the 59 genes, *rpoD*, that evolved such parallel changes in expression at 37°C was a heat-inducible candidate in this study, indicating the distinctive nature of these selection regimes that were identical in all respects except temperature (and duration). Moreover, the expression of *rpoD* declined significantly at 37°C (12) but was constant during evolution at 41.5°C (Table 2), providing further evidence that this small temperature difference and temperature approached the upper thermal niche substantially alter selection on patterns of gene expression.

### Inducible Thermotolerance

In addition to examining changes in the expression of heat-inducible genes in the high-temperature-evolved *E. coli* lines, we also measured inducible thermotolerance, which is perhaps the most widely studied aspect of the stress response. Survival during exposure to a lethal temperature following acclimation to a high, but nonlethal, temperature was significantly greater in the evolved lines than in their ancestors. Our study is similar in this respect to a study of *D. melanogaster* that found evolutionary changes in inducible thermotolerance in laboratory lines under high-temperature selection (25).

Comparative studies have identified differences in the induction temperature and duration of the stress response in organisms from various habitats (22, 42). Many of the studies have focused on threshold temperatures of heat-shock induction, and some have suggested that these induction temperatures are genetically hardwired (16, 17, 27). In *Drosophila*, investigations of stress-response regulation have identified changes in heat-shock activation temperature during laboratory evolution, as well as naturally occurring P-element insertions that are responsible for decreases in heat-shock protein expression (30, 46). These studies have contributed to an understanding of stress genes and proteins as well as their regulation and evolution. However, most such studies have focused on single genes, proteins, or regulatory elements, whereas expression arrays have allowed us to identify the relative importance of the extracytoplasmic stress response along with the cytoplasmic stress response in the adaptation to high temperature.

### Conclusions

This study extends previous work on experimental lines of *E. coli* that evolved at high temperature by simultaneously measuring changes in expression levels of 35 heat-inducible candidate genes. We found some evolutionary changes in common across all rep-

licate lines, whereas other changes were seen in only some of the lines. By examining this entire suite of genes, we expose the complicated evolution of their expression, which may be overlooked when examining the response of only one gene or protein at a time. Although we focused on changes in mRNA concentrations, and hence transcription, of heat-inducible genes during evolutionary adaptation to high temperature, further studies on protein synthesis and abundance would be complementary. An important feature of our study is that we focused on candidate genes identified by previous research; this focus facilitated data analysis, hypothesis testing, and interpretation. Finally, our study complements previous research on the acute acclimation response to thermal stress by examining the much slower, but also more persistent, evolutionary response. The fact that we found a significantly higher proportion of evolved changes in the expression of these candidate genes, relative to the background pool of noncandidate genes, demonstrates important overlap between these short- and long-term responses. By contrast, the absence of an overall evolutionary increase in the expression of these heat-inducible genes highlights important differences between phenotypic acclimation and genetic adaptation and the fact that the evolutionary response could be in the opposite direction as the acclimation response.

We thank W. Hatfield for helpful suggestions; L. Zhang for programming assistance; and P. McDonald and M. Bennett for laboratory assistance.

### DISCLOSURES

Our work was funded by National Science Foundation (NSF) Grant IBN 9905980 (to A. F. Bennett and R. E. Lenski) and National Institutes of Health Grant GM-58564 (to A. D. Long) and by an NSF Predoctoral Fellowship, an NSF Doctoral Dissertation Improvement Grant, funds from the Chao Family Grant for Functional Genomics, and a Graduate Fellowship in Bioinformatics and Bioengineering (to M. M. Riehle).

### REFERENCES

1. **Agashe VR and Hartl FU.** Roles of molecular chaperones in cytoplasmic protein folding. *Semin Cell Dev Biol* 11: 15–25, 2000.
2. **Appel RD, Barioch A, and Hochstrassed DF.** A new generation of information retrieval tools for biologists: the example of the ExPASy WWW server. *Trends Biochem Sci* 19: 258–260, 1994.
3. **Arfin SM, Long AD, Ito ET, Toller L, Riehle MM, Paegle ES, and Hatfield GW.** Global gene expression profiling in *Escherichia coli* K12: the effects of integration host factor. *J Biol Chem* 275: 29672–29684, 2000.
4. **Arsene F, Tomoyasu T, and Bakau B.** The heat shock response of *Escherichia coli*. *Int J Food Microbiol* 55: 3–9, 2000.
5. **Ashburner M.** The effects of heat shock and other stresses on gene activity: an introduction. In: *Heat Shock: From Bacteria to Man*, edited by Schlesinger MJ, Ashburner M, and Tissieres A. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, 1982, p. 1–9.
6. **Baldi P and Long AD.** A Bayesian framework for the analysis of microarray data: regularized *t*-test and statistical inferences of gene changes. *Bioinformatics* 17: 509–519, 2001.
7. **Bennett AF and Lenski RE.** Evolutionary adaptation to temperature. V. Adaptive mechanisms and correlated responses in experimental lines of *Escherichia coli*. *Evolution* 50: 493–503, 1996.



8. **Bennett AF and Lenski RE.** Evolutionary adaptation to temperature. II. Thermal niches of experimental lines of *Escherichia coli*. *Evolution* 47: 1–12, 1993.
9. **Bennett AF, Lenski RE, and Mittler JE.** Evolutionary adaptation to temperature. I. Fitness responses of *Escherichia coli* to changes in its thermal environment. *Evolution* 46: 16–30, 1992.
10. **Berlyn MKB.** Linkage map of *Escherichia coli* K-12, edition 10: The traditional map. *Microbiol Mol Biol Rev* 62: 814–984, 1998.
11. **Carpentier B and Branlant C.** The *Escherichia coli* *gapA* gene is transcribed by the vegetative RNA polymerase holoenzyme E-sigma-70 and by the heat shock RNA polymerase E-sigma-32. *J Bacteriol* 176: 830–839, 1994.
12. **Cooper TF, Rozen DE, and Lenski RE.** Parallel changes in gene expression after 20,000 generations in *Escherichia coli*. *Proc Natl Acad Sci USA* 100: 1072–1077, 2003.
13. **Danese PN and Silhavy TJ.** Targeting and assembly of periplasmic and outer-membrane proteins in *Escherichia coli*. *Annu Rev Genet* 32: 59–94, 1998.
14. **Dartigalongue C, Missiakas D, and Raina S.** Characterization of the *Escherichia coli*  $\sigma^E$  regulon. *J Biol Chem* 276: 20866–20875, 2001.
15. **Dartigalongue C and Raina S.** A new heat-shock gene *ppiD*, encodes a peptidyl-prolyl isomerase required for folding of outer membrane proteins in *Escherichia coli*. *EMBO J* 17: 3968–3980, 1998.
16. **Dietz TJ.** Acclimation of the threshold induction temperatures for 70-kDa and 90-kDa heat shock proteins in the fish, *Gillichthys mirabilis*. *J Exp Biol* 188: 383–388, 1994.
17. **Dietz TJ and Somero GN.** The threshold induction temperature of the 90kDa heat shock protein is subject to acclimatization in eurythermal goby fishes (genus *Gillichthys*). *Proc Natl Acad Sci USA* 89: 3889–3893, 1992.
18. **Eisen MB, Spellman PT, Brown PO, and Botstein D.** Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* 95: 14863–14868, 1998.
19. **Ferea TL, Botstein D, Brown PO, and Rosenzweig RF.** Systematic changes in gene expression patterns following adaptive evolution in yeast. *Proc Natl Acad Sci USA* 96: 9721–9726, 1999.
20. **Gross CA.** Function and regulation of heat shock proteins. In: *Escherichia coli and Salmonella Cellular and Molecular Biology*, edited by Neidhardt FC, Curtiss R, Ingraham JL, Lin ECC, Low KB, Magasanik B, Reznikoff WS, Riley M, Schaechter M, and Umberger HE. Washington, DC: Am. Soc. Microbiol., 1996, p. 1382–1400.
21. **Hiratsu K, Amemura M, Nasimoto H, Shinagawa H, and Makino K.** The *rpoE* gene of *Escherichia coli*, which encodes  $\sigma^E$ , is essential for bacterial growth at high temperature. *J Bacteriol* 177: 2918–2922, 1995.
22. **Hofmann GE and Somero GN.** Interspecific variation in thermal denaturation of proteins in the congeneric mussels *Mytilus trossulus* and *Mytilus galloprovincialis*: evidence from the heat-shock response. *Mar Biol (Berl)* 126: 65–75, 1996.
23. **Jin W, Riley RM, Wolfinger RD, White KP, Passador-Gorgel G, and Gibson G.** The contributions of sex, genotype and age to transcriptional variance in *Drosophila melanogaster*. *Nat Genet* 29: 389–395, 2001.
24. **Kitigawa M, Matsumura Y, and Tsuchido T.** Small heat shock proteins, IbpA and IbpB, are involved in resistances to heat and superoxide stresses in *Escherichia coli*. *FEMS Microbiol Lett* 184: 165–171, 2000.
25. **Koban M, Graham G, and Prosser CL.** Induction of heat-shock protein synthesis in teleost hepatocytes: effects of acclimation. *Physiol Zool* 60: 290–296, 1987.
26. **Kornitzer D, Teff D, Altuvia S, and Oppenheim AB.** Isolation, characterization, and sequence of an *Escherichia coli* heat shock gene, *htpX*. *J Bacteriol* 173: 2944–2953, 1991.
27. **Lansing E, Justensen J, and Loeschke V.** Variation in the expression of Hsp70, the major heat-shock protein, and thermotolerance in larval and adult selection lines of *Drosophila melanogaster*. *J Therm Biol* 25: 443–450, 2000.
28. **Lederberg S.** Genetics of host-controlled restriction and modification of deoxyribonucleic acid in *Escherichia coli*. *J Bacteriol* 91: 1029–1036, 1966.
29. **Lenski RE, Rose MR, Simpson SC, and Tadler SC.** Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2000 generations. *Am Nat* 138: 1315–1341, 1991.
30. **Lerman DN and Feder ME.** Laboratory selection at different temperatures modifies heat-shock transcription factor (HSF) activation in *Drosophila melanogaster*. *J Exp Biol* 204: 315–323, 2001.
31. **Lindquist S.** The heat-shock response. *Annu Rev Biochem* 55: 1151–1191, 1986.
32. **Long AD, Mangalam HJ, Chan BYP, Toller L, Hatfield GW, and Baldi P.** Improved statistical inference from DNA microarray data using analysis of variance and a Bayesian statistical framework. *J Biol Chem* 276: 19937–19944, 2001.
33. **Lund PA.** Microbial molecular chaperones. *Adv Microb Physiol* 44: 93–140, 2001.
34. **Ramm K and Plückerthun A.** High enzymatic activity and chaperone function are mechanistically related features of the dimeric *E. coli* peptidyl-prolyl-isomerase *FkpA*. *J Biol Chem* 275: 17106–17113, 2000.
35. **Riavio TL and Silhavy TJ.** Sensing and responding to envelope stress. In: *Bacterial Stress Responses*, edited by Storz G and Hengge-Aronis R. Washington DC: Am. Soc. Microbiol., 2000, p. 19–31.
36. **Richmond CS, Glasner JD, Mau R, Jin H, and Blattner FR.** Genome-wide expression profiling in *Escherichia coli* K-12. *Nucleic Acids Res* 27: 3821–3835, 1999.
37. **Riehle MM, Bennett AF, and Long AD.** Genetic architecture of thermal adaptation in *Escherichia coli*. *Proc Natl Acad Sci USA* 98: 525–530, 2001.
38. **Ritossa FM.** A new puffing pattern induced by a temperature shock and DNP in *Drosophila*. *Experientia* 18: 571–573, 1962.
39. **Rudd KE.** EcoGene: a genome sequence database for *Escherichia coli* K-12. *Nucleic Acids Res* 28: 60–64, 2000.
40. **Seta FD, Bowschi-Muller S, Vignais G, and Branlant C.** Characterization of *Escherichia coli* strains with *gapA* and *gapB* genes deleted. *J Bacteriol* 179: 5218–5221, 1997.
41. **Sokal RR and Rohlf FJ.** *Biometry*. New York: Freeman, 1995.
42. **Tomanek L and Somero GN.** 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, 2000.
43. **Yamamori T, Ito K, Nakamura Y, and Yura T.** Transient regulation of protein synthesis in *Escherichia coli* upon shift-up of growth temperature. *J Bacteriol* 134: 1138–1140, 1978.
44. **Yura T, Kanemori M, and Morita MT.** The heat shock response: regulation and function. In: *Bacterial Stress Responses*, edited by Storz G and Hengge-Aronis R. Washington DC: Am. Soc. Microbiol., 2000, p. 3–18.
45. **Yura T, Nagai H, and Mori H.** Regulation of the heat-shock response in bacteria. *Annu Rev Microbiol* 47: 321–350, 1993.
46. **Zatsepina OG, Velikodvorskaia VV, Molodtsov VB, Garbuz D, Lerman DN, Bettencourt BR, Feder ME, and Evgenev MB.** A *Drosophila melanogaster* strain from sub-equatorial Africa has exceptional thermotolerance but decreased Hsp70 expression. *J Exp Biol* 204: 1869–1881, 2001.