MBE Advance Access published July 20, 2012

Loss of the DnaK-DnaJ-GrpE Chaperone System among the Aquificales

Tobias Warnecke*

Centre for Genomic Regulation (CRG) and UPF, Barcelona, Spain

***Corresponding author:** E-mail: Tobias.Warnecke@crg.es.

Associate editor: Jennifer Wernegreen

Abstract

The DnaK-DnaJ-GrpE (KJE) chaperone system functions at the fulcrum of protein homeostasis in bacteria. It interacts both with nascent polypeptides and with proteins that have become unfolded, either funneling its clients toward the native state or ushering misfolded proteins into degradation. In line with its key role in protein folding, KJE has been considered an essential building block for a minimal bacterial genome and common to all bacteria. In this study, I present a rigorous survey of 1,233 bacterial genomes, which reveals that the entire KJE system is uniquely absent from two members of the order Aquificales, *Desulfurobacterium thermolithotrophum*, and *Thermovibrio ammonificans*. The absence of KJE from these free-living bacteria is surprising, particularly in light of the finding that individual losses of *grpE* and *dnaJ* are restricted to obligate endosymbionts with highly reduced genomes, whereas *dnaK* has never been lost in isolation. Examining protein features diagnostic of DnaK substrates in *Escherichia coli*, radical changes in protein solubility emerge as a likely precondition for the loss of KJE. Both *D. thermolithotrophum* and *T. ammonificans* grow under strictly anaerobic conditions at temperatures in excess of 70°C, reminiscent of hyperthermophilic archaea, which—unlike their mesophilic cousins—also lack KJE. I suggest that high temperature promotes the evolution of high intrinsic protein solubility on a proteome-wide scale and thereby creates conditions under which KJE can be lost. However, the shift in solubility is common to all Aquificales and hence not sufficient to explain the restricted incidence of KJE loss.

Key words: Aquificales, chaperones, *Thermovibrio ammonificans*, *Desulfurobacterium thermolithotrophum*, DnaK, GrpE.

Introduction

Some genes form part of the very fabric of cellular organization as we understand it. Considered universal across a certain domain of life, they function at the core of basic cellular processes, such as replication, transcription, or translation, and constitute the minimal building blocks required to specify a self-replicating organism. Among the genes regarded as archetypal in bacteria is the DnaK-DnaJ-GrpE (KJE) chaperone system (Gil et al. 2004; Wong and Houry 2004). KJE is involved in both the folding of nascent proteins and the refolding of proteins that have become denatured. In either case, the basic mechanism of chaperone action is the same: DnaK transiently binds to exposed hydrophobic patches, thereby shields the polypeptide from unproductive interactors and allows stepwise exploration of the folding landscape (Hartl and Hayer-Hartl 2009). Substrate binding and release are ATP dependent and chiefly regulated by two cochaperones, DnaJ and GrpE (Burkholder et al. 1996; Brehmer et al. 2001; Hartl and Hayer-Hartl 2009). For some clients, repeated cycles of binding and release are sufficient to attain the native conformation; for others, KJE acts as a go-between, channeling folding intermediates to the downstream chaperonin complex GroEL/ES.

In *Escherichia coli*, ~700 mostly cytosolic proteins have recently been shown to interact habitually with DnaK under standard growth conditions (Calloni et al. 2012). By and large, these regular clients exhibit average physicochemical properties although they are enriched for proteins that have low intrinsic solubility, encode the topologically complex structural classification of proteins (SCOP) fold c.37, participate in hetero-oligomeric complexes, and/or have an elevated density of "DnaK binding sites," hydrophobic patches flanked by positively charged residues (Van Durme et al. 2009; Calloni et al. 2012). In addition, proteins that interact with DnaK tend to be expressed at below-average levels, consistent with selection promoting the evolution of chaperone-independent folding pathways at high levels of expression (Warnecke and Hurst 2010). Its regular interactions with a large and diverse array of substrates and its extensive interconnections with other chaperones and proteases place KJE at the fulcrum of protein homeostasis in bacteria (Calloni et al. 2012).

In this study, conducting a rigorous survey of 1,233 complete bacterial genomes, I report that the entire KJE system is absent from *Desulfurobacterium thermolithotrophum* and *Thermovibrio ammonificans*, two thermophilic bacteria that have been isolated independently from deep-sea hydrothermal vents in the Atlantic and Pacific, respectively.

Although the loss of genes from the "minimal bacterial gene set" (Gil et al. 2004) is not unprecedented, such cases have principally been reported in the context of highly reduced endosymbionts or obligate pathogens (McCutcheon

and Moran 2012). For example, alanyl-tRNA synthetase (*alaS*), ordinarily considered an indispensable part of bacterial physiology, has been lost from the tiny (246 kb) genome of Candidatus *S. muelleri* GWSS, a long-term endosymbiont of the glassy-winged sharpshooter (McCutcheon et al. 2009a). Similarly, some mollicutes, such as the goat pathogen *Mycoplasma capricolum*, lack *groEL/ES*, the only essential chaperone in *E. coli* (Wong and Houry 2004; Williams and Fares 2010). Although some of these losses—including that of *groEL/ES*—remain mysterious, others, especially those relating to biosynthetic capacities, can be explained by the bacterium's ability to acquire essential products from the host or a coendosymbiont (McCutcheon and Moran 2012). In other words, although a gene may be physically lost from the focal genome, its function or output remains available as part of a tightly integrated genetic system.

The absence of a chaperone module as pivotal as KJE from free-living bacteria is rather more curious. What factors have enabled the loss of this otherwise ubiquitous molecular machine? Have the folding requirements of nascent or unfolded proteins changed so dramatically as to make KJE functions dispensable? Or are the services normally provided by KJE still very much sought after but provided by a different mechanism? If so, has there been something akin to a like-for-like replacement, where a different chaperone plugs the gap left by the absence of KJE or did the loss go hand in hand with a more radical reorganization of protein folding pathways?

To address these questions and elucidate why KJE has been lost from the two genomes, I compare *D. thermolithotrophum* and *T. ammonificans* with seven other members of the order Aquificales, shedding light on the evolution of client proteins before and after the loss of KJE and highlighting unique features of their respective chaperone systems.

Materials and Methods

Establishing Presence and Absence of KJE Genes

The sequences of 1,233 bacterial genomes were obtained from the National Center for Biotechnology Information (NCBI) (ftp://ftp.ncbi.nih.gov/genomes/Bacteria/) in May 2011. The genome of Candidatus *Tremblaya princeps* was added at a later date. To establish whether *dnaK*, *dnaJ*, and *grpE* are present in a given genome, the following strategy was pursued: first, using *E. coli* (strain MG1655) versions of each gene as queries, BLASTP searches were conducted against every bacterial proteome as annotated in GenBank applying a comparatively lenient *E*-value threshold of 0.001. For all three genes, the majority of BLAST searches returned strong hits that unambiguously supported their presence in the target genome (supplementary table S1, Supplementary Material online). Whenever querying with *E. coli* genes did not return a significant hit, the query was repeated using a homolog from a more closely related genome that had itself previously been identified using the *E. coli* query. Further relaxing the threshold to an *E* value of 1, using PSI-BLAST (Altschul et al. 1997) or screening the protein domain database Pfam (Punta et al. 2012) did not yield additional candidate homologs.

In the third step, to account for incomplete genome annotations, genomic DNA was scanned for regions of high-sequence homology to the queries (TBLASTN; *E*-value threshold: 0.001). For four genomes that—according to current annotations—lack at least one gene of the KJE triad, strong genomic hits were recovered. On the basis of the genomic fragments identified, I reconstructed open reading frames using a closely related homolog as a guide. These reading frames are uninterrupted by stop codons and closely resemble the guide homolog, strongly suggesting that the relevant gene is, in fact, present and failed to be spotted during initial annotation (see supplementary text, Supplementary Material online, for details and reconstructed proteins).

Where the pipeline up to this point did not uncover a viable homolog, raw sequencing data (supplementary table S2, Supplementary Material online) were examined to determine whether the apparent gene loss might have resulted from genome misassembly. Although this may seem a remote possibility, in one case (*Acinetobacter baumnannii* strain SDF), several sequencing reads exhibit high sequence similarity to the *dnaK* and *grpE* homologs from a closely related genome (supplementary table S3, Supplementary Material online). As elaborated further in the supplementary text, Supplementary Material online, the absence of *dnaK* and *grpE* in the SDF strain might therefore be an artifact of genome misassembly. In contrast, wherever a gene is inferred to be genuinely absent, raw read data provide little, if any, support for misassembly as the underlying cause (see supplementary text, Supplementary Material online, for details).

Phylogenetic Reconstruction

A set of orthologs common to the nine available Aquificales genomes was derived as follows: first, all nine annotated proteomes were blasted (BLASTP; *E*-value threshold: 0.001) against each other. Reciprocal best hits were retained and a set of orthologs present in all Aquificales defined as the intersection of all pairwise comparisons. Thereafter, putative orthologs that had less than 40% amino acid identity and differed by more than 10% in protein length were eliminated, resulting in a final set of 267 proteins, which were used for phylogenetic and downstream analyses (see supplementary fig. S1, Supplementary Material online, for the distribution of pairwise amino acid identities in this final set). Orthologs were aligned at the protein level using MUSCLE (version 3.8.31) (Edgar 2004) with an additional automated refinement step. Subsequently, alignments were back-translated to nucleotides and submitted to Gblocks (Castresana 2000) (minimum block size: five residues) to filter out poorly aligned and likely uninformative positions. PhyML (Guindon et al. 2010) was then used to reconstruct phylogenetic trees for each single-protein alignment and a concatenated alignment comprising all proteins. Alignment curation and tree reconstruction were repeated using protein alignments as input. This yields the same consensus topology (not shown). Rates of

nonsynonymous divergence were computed using the codeml algorithm in PAML (Yang 2007).

Properties of KJE Client Proteins

Protein solubility was predicted using CCSOL (Agostini et al. 2011), a support vector machine that was trained and tested using experimental data on protein solubility in *E. coli*. Analysis was confined to proteins predicted to localize to the cytoplasm by PSORTB (Yu et al. 2010), although results are robust when considering the entire proteome. DnaK binding sites across the same set of cytosolic proteins were predicted using LIMBO (Van Durme et al. 2009). Finally, the SUPERFAMILY database (Wilson et al. 2009) was screened to establish what fraction of genes in each genome encode the SCOP fold c.37.

Gene Content Comparison

To identify proteins that are present in *D. thermolithotrophum* and *T. ammonificans* but absent in all other Aquificales genomes and vice versa, gene content was compared based on cluster of orthologous groups (COG) annotations obtained from NCBI (http://www.ncbi.nlm.nih.gov/COG/). For a core set of chaperones and proteases, homologs were identified using the BLAST procedure described earlier. *Pyrococcus abyssi* proteins were used as queries to capture prefoldin homologs (see Results). The majority of queries recovered either no hit or a single hit in all nine genomes. Where more than one region of significant homology was detected, reciprocal best hits are consistent across all genomes, so that lineage-specific additions can be confidently identified.

Three members of the Clp family of proteases (ClpA, ClpB, and ClpC) exhibit high levels of sequence similarity. Consequently, each one is recovered when querying with a different family member. For these proteins, the domain structure of all hits with reference to the Pfam database was examined to confirm that the same set of proteases is present across all Aquificales genomes.

Codon Usage Analysis

To estimate translational adaptation of individual genes as a proxy for their expression level, the tRNA adaptation index (tAI) (dos Reis et al. 2004) was calculated. tAI measures the degree to which a gene uses codons that correspond to more abundant/efficiently decoded tRNAs. The genomic copy numbers of all tRNA species (knowledge of which is required to compute tAI) were determined using tRNAscan-SE (version 1.21) with a bacterial tRNA model (Lowe and Eddy 1997). Only coding sequences with a minimum length of 280 nucleotides were used in the final analysis, a cutoff that excludes very short genes (for which estimates will be less reliable) but retains all chaperones of interest. To facilitate cross-species comparison of translational adaptation, tAI values were normalized within each genome using a *Z* score, *Z*(tAI).

Results

Loss of the Entire KJE System from Two Desulfurobacteriaceae

I surveyed 1,233 complete bacterial genomes for the presence of *dnaK*, *dnaJ*, and *grpE*, which together function as a key chaperone module in bacteria (Hartl and Hayer-Hartl 2009). Taking into account errors in genome annotation and assembly (see Materials and Methods, supplementary text, Supplementary Material online), the majority of genomes contain well-defined homologs of each member of the KJE triad (supplementary table S1, Supplementary Material online).

The few cases where *dnaK*, *dnaJ*, or *grpE* have been lost on their own concern highly reduced endosymbiont genomes (table 1). First, there are no *grpE* homologs in the genomes of either Candidatus *Hodgkinia cicadicola* DSEM or Candidatus *Sulcia muelleri* SMDSEM, both obligate endosymbionts of the same cicada species, *Diceroprocta semicincta* (McCutcheon et al. 2009a). Second, *dnaJ* is absent from only a single genome, that of the psyllid endosymbiont Candidatus *Carsonella ruddii*. Finally, the genome of Candidatus *Tremblaya princeps*, an endosymbiont of mealybugs, encodes an N-terminally truncated version of *grpE*. Whether the conserved C-terminus (see supplementary text, Supplementary Material online) remains expressed and provides GrpE-like functionality or whether Tremblaya is able to use exogenous GrpE produced by the host or its own endosymbiont, Candidatus *Moranella endobia*, will have to be established experimentally. Despite their highly reduced gene content, all these genomes, without exception, encode *dnaK*.

In this context, it is all the more surprising to discover two free-living bacteria that lack the entire KJE system. Sporting genome sizes an order of magnitude larger than the aforementioned endosymbionts (table 1), *D. thermolithotrophum* and *T. ammonificans* have been isolated from the walls of deep-sea hydrothermal vents and grow at temperatures ~70°C (table 2). Why do these bacteria lack KJE, while it remains encoded in all other bacterial genomes sequenced to date?

A Single-Loss Event Can Explain the Absence of KJE in Desulfurobacteriaceae

To probe the evolutionary forces behind the loss of KJE, I first reconstructed phylogenetic relationships between *D. thermolithotrophum*, *T. ammonificans*, and seven other members of the order Aquificales for which whole genome sequences are available. The consensus tree based on 267 protein-coding orthologs (fig. 1, see Materials and Methods) supports the notion that *D. thermolitotrophum* and *T. ammonificans* constitute sister taxa (family Desulfurobacteriaceae), consistent with previous reports that relied on 16S rRNA, intergenic transcribed spacer sequence, and single proteins (Ferrera et al. 2007; Hügler et al. 2007). Although the two genomes have diverged considerably in sequence (fig. 1), gene content (supplementary table S4, Supplementary Material online),

Table 1. Organisms with Full or Partial Loss of KJE.

Taxon	Order	Key features	Genome size (kb)	*dnaK*	*dnaJ*	*grpE*
D. thermolithotrophum DSM 11699	Aquificales	Thermophilic Chemolithotrophic Anaerobic	1,541	✖	✖	✖
T. ammonificans HB-1	Aquificales	Thermophilic Chemolithotrophic Anaerobic	1,760[a]	✖	✖	✖
Candidatus *Hodgkinia cicadicola* DSEM	Rhizobiales	Obligate endosymbiont	144	✓	✓	✖
Candidatus *Sulcia muelleri* SMDSEM	Flavobacteriales	Obligate endosymbiont	277	✓	✓	✖
Candidatus *Carsonella ruddii*	Unclassified (class: gamma-proteobacteria)	Obligate endosymbiont	160	✓	✖	✓
Candidatus *Tremblaya princeps*[b]	Unclassified (class: beta-proteobacteria)	Obligate endosymbiont	139	✓	✓	Ψ?
Acinetobacter baumannii SDF[b]	Pseudomonadales	Hematophagic pathogen isolated from the gut of body lice	3,477[a]	?	✓	?

[a]Includes plasmids.
[b]See main text and supplementary text, Supplementary Material online, for details; Ψ = evidence for pseudogenization.

Table 2. Genomic and Ecological Characteristics of the Aquificales.

Taxon	Genome size (kb)	GC (%)	Optimal growth temperature (°C)	Optimal pH	Optimal salinity (% [w/v] NaCl)	Oxygen requirements	References
D. thermolithotrophum DSM 11699	1,541	34.9	70 (40–75)	6 (4.4–7.5)	3.5 (1.5–7.0)	Strictly anaerobic	L'Haridon et al. 1998
T. ammonificans HB-1	1,760[a]	52.1	75 (60–80)	5.5 (5–7)	2 (0.5–4.5)	Strictly anaerobic	Vetriani et al. 2004
Persephonella marina EX-H1	1,983[a]	37.1	73 (55–80)	6 (4.7–7.5)	2.5 (1–4.5)	Microaerophilic	Götz et al. 2002
Sulfurihydrogenibium sp. YO3AOP1	1,838	32.0	ND	ND	ND	ND	NA
Sulfurihydrogenibium azorense Az-Fu1	1,640	32.8	68 (50–73)	6 (5.5–7)	0.1 (0–0.25)	Microaerophilic	Aguiar et al. 2004
Hydrogenobaculum sp. YO4AAS1	1,559	34.8	58	4	ND	Microaerophilic	Ferrera et al. 2007
Aquifex aeolicus VF5	1,590[a]	43.3	85	6.5	3 (1–5)	Microaerophilic	Eder and Huber 2002
Thermocrinis albus DSM14484	1,500	46.9	85	7	0–0.7	Microaerophilic	Eder and Huber 2002
Hydrogenobacter thermophilus TK-6	1,743	44.0	70–75	7	ND	Microaerophilic	Kawasumi et al. 1984

NOTE.—ND, not determined.
[a]Includes plasmids.

and organization (supplementary fig. S2, Supplementary Material online), the most parsimonious scenario at present is therefore that the loss of KJE occurred in the common ancestor of the two bacteria, which branch in a basal position in the Aquificales (fig. 1).

Interestingly, there is no indication that the KJE system has evolved toward reduced physiological significance among the Aquificales in general. On the contrary, *dnaK* is highly conserved (supplementary table S5, Supplementary Material online), indicative of persistent functional constraint and/or high expression (Pal et al. 2006).

Aquificales Proteins Exhibit Greatly Enhanced Solubility

Has there perhaps been a systematic shift in folding requirements or unfolding probabilities of client proteins that might have rendered the services of DnaK dispensable? To dissect whether substrate-level changes have contributed to KJE dispensability, I examined three protein features recently found to be enriched in regular DnaK clients in *E. coli*: protein solubility, which is lower for habitual client proteins, the density of DnaK binding sites, and the incidence of SCOP fold c.37 (both higher for client proteins) (Calloni et al. 2012). In addition, a signature of enhanced thermostability might indicate a lower risk of thermally induced unfolding and hence a reduced workload for KJE. Based on this rationale, protein composition was examined with respect to a set of amino acids (IVYWREL) previously found enriched in proteins from thermophilic organisms (Zeldovich et al. 2007).

Contrasting both global proteomes and orthologous proteins, there is no indication that local sequence features might have facilitated KJE loss. Compared with *E. coli*, Aquificales proteins exhibit a greater average density of DnaK binding sites (fig. 2C, also see supplementary fig. S3, Supplementary Material online) and a greater fraction of proteins contain c.37

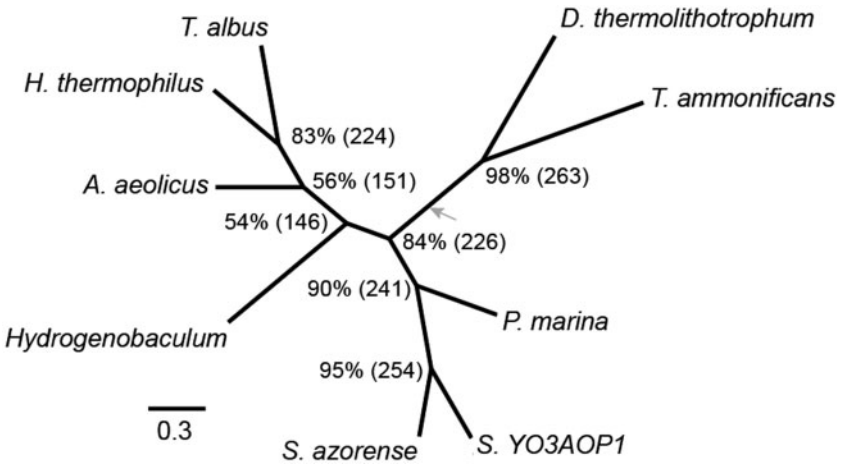

Fig. 1. Phylogeny of the Aquificales based on the nucleotide sequences of 267 protein-coding genes. Nodes are labeled according to the fraction (absolute number in parentheses) of individual gene trees that support the same trifurcation at that node. A subset of 127 genes with an ortholog in *Thermotoga maritima* was used to root the tree. A gray arrow indicates the position of the root. Bootstrap support for the concatenate was 100% for all nodes. The tree was constructed with PhyML (see Materials and Methods) and rendered with TreeDyn (Chevenet et al. 2006).

folds (supplementary fig. S4, Supplementary Material online). In contrast, there are striking differences in predicted protein solubility (fig. 2*A*), consistent with reduced reliance on KJE for folding. As expected, IVYWREL content is also higher in the thermophilic Aquificales than in mesophilic *E. coli* (fig. 2*B*). However, in neither case do the Desulfurobacteriaceae fall outside the range of the other Aquificales (fig. 2*A* and *B*). Thus, enhanced solubility may constitute an important precondition for KJE loss but does not by itself provide a complete explanation.

It should be noted at this point that the solubility values reported herein can only constitute an approximation of in vivo solubility. Although predictions are based exclusively on sequence features, the algorithm was trained on experimentally determined solubilities of *E. coli* proteins. Clearly, observed solubility reflects not only intrinsic features of the protein but also the conditions prevailing in the cytosol or in vitro, respectively. It is noteworthy, then, that thermophilic cytosols exhibit systematic differences in the concentration of ions and organic solutes, which not only affect osmoregulation but also solubility and protein thermostability (Martins et al. 1996; Santos and da Costa 2002). Putting such solutes into the equation will almost inevitably affect solubility estimates. Importantly, however, this might well enhance rather than diminish in vivo solubility differences between *E. coli* and the Aquificales, which would lend further credence to the hypothesis that solubility differences are implicated in KJE loss.

Is Temperature a Critical Factor in KJE Dispensability?

Changes in thermostability and solubility are consistent with adaptation to life at high temperatures where proteins encounter aggravated aggregation problems. Further consistent with a temperature-driven process is the fact that KJE has never been detected in hyperthermophilic archaea but is present in several of their mesophilic cousins. The patchy distribution of KJE genes among archaea may be due to differential gene loss, sporadic horizontal gene transfers, or a combination of the two processes (Gupta 1998; Gribaldo et al. 1999; Laksanalamai et al. 2004). Regardless of the evolutionary process at work, temperature-driven evolution, for example, of globally enhanced solubility, might determine the selective benefit of having KJE and hence lead to differential retention and/or acquisition.

Although temperature might have played a prominent role in facilitating the loss of KJE, there are no differences in optimal growth temperature (or pH or salinity) that single out the Desulfurobacteriaceae (table 2). The one systematic difference in evidence concerns oxygen requirements. In contrast to the remaining Aquificales examined, which live under microaerophilic conditions, *D. thermolithotrophum* and *T. ammonificans* have adopted a strictly anaerobic lifestyle (L'Haridon et al. 1998), which is reflected in their divergent gene content related to energy metabolism (supplementary table S4, Supplementary Material online). Evoking differences in respiratory mode to explain the loss of KJE might appear far fetched. However, some aspects of protein folding are linked to energy metabolism. Notably, the formation of disulfide bonds is coupled to the electron transport chain (Glockshuber 1999). Although disulfide bonds are rare in soluble proteins of aerobes due to the reductive environment of the cytosol, they seem to be a common feature of cytoplasmic proteins from hyperthermophilic bacteria and archaea (Mallick et al. 2002; Ladenstein and Ren 2008). I therefore tentatively suggest that systematic differences in the redox environment might have contributed to different folding requirements of proteins in Desulfurobacteriaceae.

Divergent Organization of Protein Homeostasis in Desulfurobacteriaceae

With substrate characteristics largely indistinguishable from other Aquificales and the contribution of energy metabolism uncertain, I hypothesized that differences in the

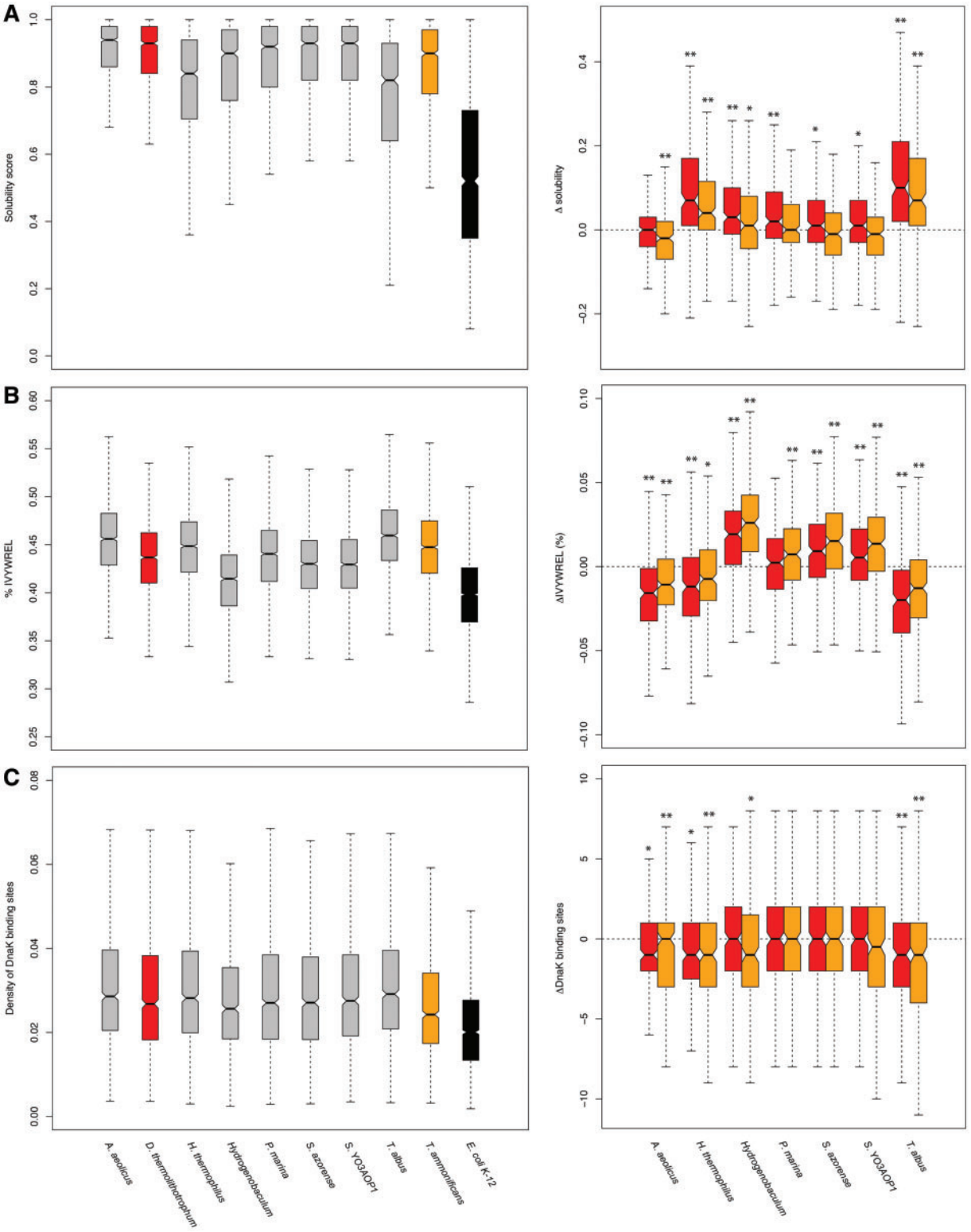

Fig. 2. Protein features associated with KJE dependency. The left-hand panels illustrate, for the global cytosolic proteomes of nine Aquificales genomes and *E. coli*, (*A*) protein solubility, (*B*) the density of DnaK binding sites per amino acid, and (*C*) the fraction of all amino acids in the protein that belong to a class of residues characteristic of thermostable proteins (IVYWREL). All Aquificales distributions are significantly different from the *E. coli* distribution ($P \ll 0.0001$, Wilcoxon test). The right-hand panels consider how these properties change between orthologs of a specific Aquificales taxon and either *D. thermolithotrophum* (red) or *T. ammonificans* (orange). Values greater than zero indicate larger values in *D. thermolithotrophum* or *T. ammonificans*, respectively. Distributions with means significantly different from zero based on a one-sample t-test are marked with $^{**}P < 0.001$ or $^{*}P < 0.05$.

Table 3. Presence/Absence of Major Chaperones and Proteases across Nine Aquificales Genomes.

	Absent in . . . genomes	Single homolog in . . . (genomes)	Additional homolog in . . .
GroEL		9	
GroES		8	*P. marina*
HtpX		7	*D. thermolithotrophum* *T. ammonificans*
HtpG	9		
ClpX		9	
ClpP		9	
ClpA/ClpB/ClpC[a]		9	
HslU		9	
HslV		9	
IbpA	9		
IbpB	9		
Prefoldin α subunit	7	*A. aeolicus* *T. ammonificans*	
Prefoldin β subunit	9		
Trigger factor		9	
HslO	4	*A. aeolicus* *D. thermolithotrophum* *H. thermophilus* *T. albus* *T. ammonificans*	

[a]See Materials and Methods.

organization of protein homeostasis might separate the Desulfurobacteriaceae from other (hyper)thermophiles.

One possibility here is that KJE has been rendered dispensable through quantitative changes in the expression of familiar chaperones such as GroEL/ES. This is plausible as long as there is sufficient functional redundancy between chaperone systems. Three pieces of evidence suggest that—at least to some extent—this is the case for KJE: 1) *dnaK* deletions in *E. coli* are viable under standard growth conditions, but the deletion of both *dnaK* and the trigger factor gene *tig* results in synthetic lethality above 30°C (Teter et al. 1999), 2) *groEL* overexpression can rescue *dnaK/tig* double-deletion strains (Vorderwülbecke et al. 2004), and 3) several chaperones and proteases are drastically upregulated in *dnaK/dnaJ* deletion strains of *E. coli*, including *IbpB* (~170-fold), *IbpA* (~48-fold), *ClpB* (~28-fold), *HslU/V* (~11-fold), and *GroEL/ES* (~8-fold) (Calloni et al. 2012). By implication, the constitutive upregulation of certain nodes in the protein homeostasis network may play some part in buffering KJE loss, although, notably, Hsp20-class genes (*ibpA* and *ibpB*) are absent from all Aquificales genomes (table 3). A second possibility concerns more qualitative changes, i.e., the integration of novel components into protein production, folding, and degradation circuits.

To gauge the contribution of both quantitative and qualitative changes to protein homeostasis networks, gene inventories were compiled for all members of the Aquificales (see Materials and Methods). Based on these inventories, it is evident that many key genes are present in single copy across all genomes. However, there are a number of exceptions (table 3 and supplementary table S4, Supplementary Material online) and two cases in particular warrant closer inspection.

First, *D. thermolithotrophum* and *T. ammonificans* uniquely encode not one but two genes with significant sequence similarity to the *E. coli* heat shock protein HtpX. Anchored to the inner membrane, HtpX cleaves the cytoplasmic fraction of membrane-bound proteins and has therefore been suggested to be involved in eliminating misfolded proteins from the membrane, although the specific properties of its substrates remain poorly characterized (Akiyama 2009). In both taxa, one of the homologs exhibits well-defined transmembrane domains and neatly conforms to the *E. coli* template. The other, which I dub HtpX-B, lacks such domains and is predicted to localize to the cytosol (supplementary fig. S5, Supplementary Material online). Second, the genome of *T. ammonificans* encodes an alpha subunit of prefoldin, which is noteworthy not only because it highlights archaea as a potential source of novel parts for bacterial folding circuits but also, more importantly, because prefoldin can be regarded as a (ATP independent) functional analog of DnaK, binding to unfolded or partially folded substrates and delivering some of them to downstream chaperonins (Leroux et al. 1999). If KJE loss were indeed facilitated by the presence of a like-for-like stand-in, prefoldin would be an obvious candidate for that role.

Examining codon usage patterns as a surrogate for expression level, it is evident that the prefoldin subunit exhibits the highest level of translational adaptation of any of the major chaperones in *T. ammonificans* followed by *htpX* and *htpX-B* (fig. 3). *D. thermolithotrophum* on the other hand lacks the

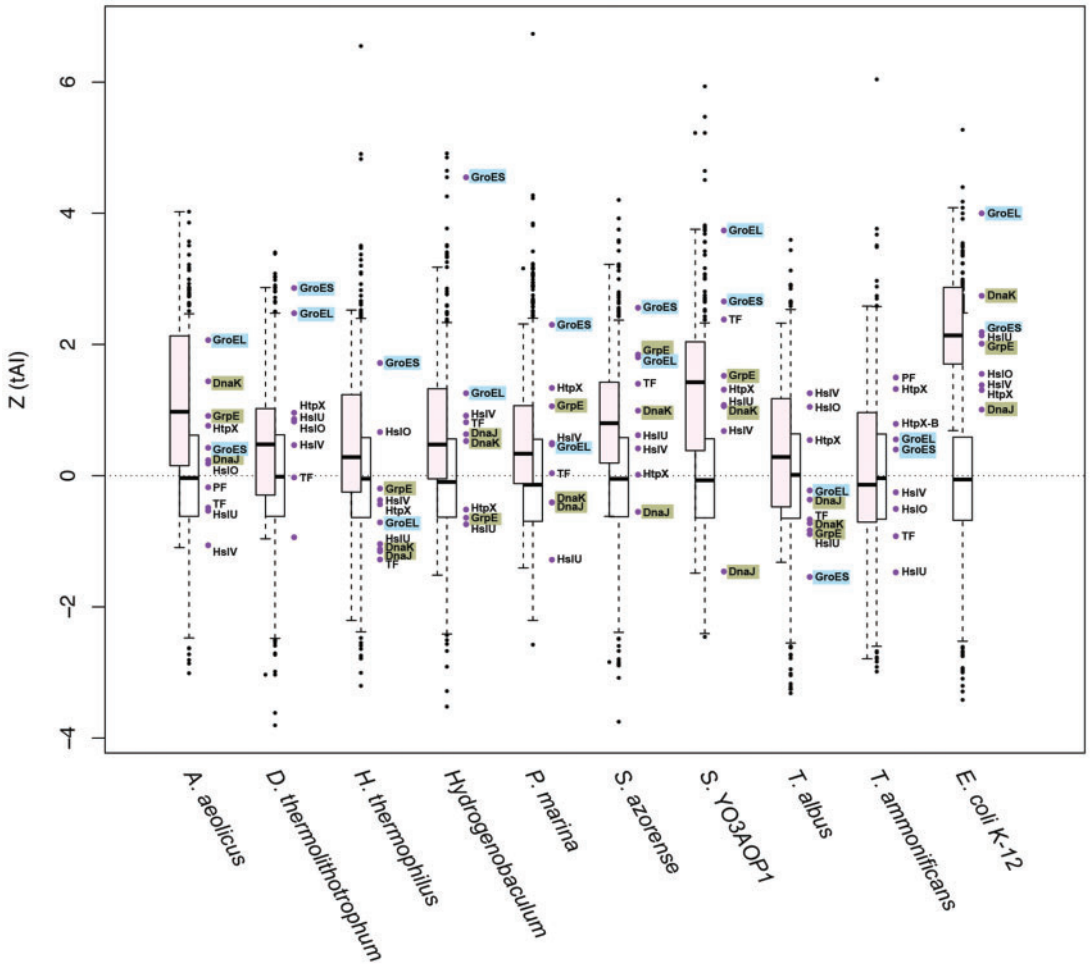

Fig. 3. Translational adaptation of Aquificales genes. White boxplots represent the distribution of normalized tAI values in each genome, where larger Z(tAI) values correspond to greater translational adaptation. Pink boxplots represent ribosomal proteins. The tAI values of a select group of chaperones are highlighted as purple dots. TF, trigger factor; PF, prefoldin.

prefoldin subunit and translational adaptation of *htpX-B* seems limited, whereas *groEL* and *groES* exhibit codon usage patterns consistent with high expression. Thus, although the substrate characteristics of potential KJE clients in *D. thermolithotrophum* and *T. ammonificans* are similar (fig. 2), the same cannot be said for their respective systems of protein homeostasis, as far as one can glean from gene set analysis and codon usage patterns.

Assuming my initial appraisal is correct and KJE loss occurred in the common ancestor of *T. ammonificans* and *D. thermolithotrophum*, the evidence presented earlier would suggest that substantial remodeling of chaperone systems happened after, and is likely unrelated to, KJE loss, obscuring which factors were critical in enabling KJE loss. If, on the other hand, we were instead dealing with parallel losses, one might be led to conclude that *T. ammonificans* and *D. thermolithotrophum* have followed rather different pathways to life without KJE. Without greater phylogenetic resolution, such a convergent loss scenario cannot be ruled out with confidence.

Discussion

The absence, from free-living bacteria, of a major chaperone system such as KJE, which occupies a central role in our understanding of how protein folding is organized in bacteria, is puzzling.

Although *dnaK* deletions can be viable under permissive conditions, they are lethal under heat shock (Teter et al. 1999; Vorderwülbecke et al. 2004). In addition, KJE is involved in protein secretion (Wild et al. 1992) and essential for the synthesis of certain polypeptides, such as flagellar proteins (Shi et al. 1992), that are likely essential in the wild. This strongly implies that the loss of KJE would compromise fitness in the medium term, in keeping with the observation that *dnaK* in particular is consistently highly conserved and—the case reported here apart—universal to bacterial life.

In search of the factors enabling KJE loss, there is a dramatic shift in predicted protein solubility in the Aquificales compared with mesophilic *E. coli*. I suggest that high ambient temperature (perhaps in conjunction with a strictly anaerobic lifestyle and changes in solute content) has prompted this global shift toward greater protein solubility, laying the foundations for KJE-independent folding. This hypothesis is consistent with the absence of KJE from hyperthermophilic archaea (Laksanalamai et al. 2004). However, solubility ranges are comparable across the Aquificales yet KJE has only been lost from the Desulfurobacteriaceae, suggesting

that temperature-driven changes in solubility only provide a partial explanation. What, then, are the additional changes that enabled KJE to be lost in the Desulfurobacteriaceae? First, further changes at the substrate level might have played a role. For example, there is a slight reduction in the number of sites targeted by DnaK—and hence likely refractory to efficient folding—in the Desulfurobacteriaceae (fig. 2*C*, right panel) compared with most other Aquificales. However, there is insufficient temporal resolution to distinguish whether these subtle changes facilitated the loss of KJE or reflect subsequent selection against aggregation-prone regions normally monitored by DnaK. Along with substrate-level evolution, qualitative and/or quantitative changes to the protein homeostasis setup might also have been a contributing factor. Although I have focused on other chaperones known to interact with KJE in *E. coli*, it is important to point out that proteases might play an important role in this regard. Notably, it has been suggested that *Mycoplasma* species that have lost *groEL/ES* might have undergone a shift toward greater reliance on protein degradation vis-à-vis efficient folding (Wong and Houry 2004). Although this remains speculative, the discovery of a unique cytosolic variant of *htpX* in Desulfurobacteriaceae (*htpX-B*) might point in the same direction, especially for *T. ammonificans*, where codon usage patterns suggest that *htpX-B* is expressed at high levels. However, there were no obvious systematic differences in tAI across a wider range of proteases (not shown), which might have signaled a switch to a degradation-heavy strategy in the Desulfurobacteriaceae. In addition, note that Mycoplasma proteins also show greatly enhanced solubility when compared with *E. coli* (supplementary fig. S6, Supplementary Material online), so a similar mechanism of preconditioning chaperone loss might be at work.

The Loss of Individual KJE Genes

The survey conducted here also highlights that losses of individual KJE genes are limited to obligate endosymbionts. Despite a strong trend toward reductive genome evolution in these bacteria, the loss of KJE components is surprising. There is no evidence that KJE has assumed a minor physiological role in these organisms, which would render the loss of *grpE* in particular somewhat more palatable. On the contrary, quantitative proteomic data from Candidatus *Hodgkinia cicadicola* and Candidatus *Sulcia muelleri* SMDSEM revealed that DnaK ranks alongside GroEL/ES as one of the most highly expressed proteins in the cell (McCutcheon et al. 2009a, b); a situation that closely resembles the status quo in other endosymbionts such as Buchnera, where a major fraction of total protein production is dedicated to GroEL/ES (Wernegreen and Moran 1999). Typically, this has been interpreted as an evolutionary response to long-term small effective population size. As selection can no longer efficiently purge slightly deleterious mutations, these mutations accumulate over time in a ratchet-like fashion (Moran 1996; van Ham et al. 2003). In addition, with each mutation that destabilizes protein structure, the workload for chaperones increases. Assuming that a similar regime governs chaperone evolution in Hodgkinia and Sulcia, it is difficult to see how GrpE, a critical regulator of DnaK activity, would have become dispensable. Its disappearance becomes all the more curious when considering that experimental data from *E. coli* strongly suggest that KJE acts as a finely tuned system. *GrpE* deletions are unconditionally lethal (Ang and Georgopoulos 1989). Conversely, overexpression of *grpE* and *dnaK* causes abnormal morphology and growth defects (Blum et al. 1992; Sugimoto et al. 2008). In short, upsetting the balance between DnaK, DnaJ, and GrpE can have dramatic fitness effects.

What is intriguing in this regard is that Candidatus *Hodgkinia cicadicola* and Candidatus *Sulica muelleri* SMDSEM, the two taxa that have lost *grpE*, live inside the same cicada host, whereas Sulcia species associated with different hosts have retained a recognizable *grpE* homolog (McCutcheon et al. 2009a). Whether the shared host environment can somehow account for the coincident losses by providing GrpE homologs that interact with DnaK or whether DnaK has evolved to operate independently of its normal regulators are interesting questions to be addressed experimentally in the future.

Supplementary Material

Supplementary text, figures S1–S6, and tables S1–S5 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

The author thanks Romain Derelle for a curated set of mitochondrial GrpE sequences, John McCutcheon and Masahira Hattori for providing raw sequencing data, Joost van Durme as well as Federico Agostini and Gian Tartaglia for calculating LIMBO and protein solubility scores, respectively, Oxana Arkhipova and Tobias Maier for valuable discussions, Ulrich Hartl for sharing data prior to publication, and three anonymous referees for comments on the manuscript. This work was supported by a European Molecular Biology Organisation long-term fellowship and an International Interdisciplinary Postdoctoral Programme fellowship, co-funded by Marie Curie Actions.

References

Agostini F, Vendruscolo M, Tartaglia GG. Sequence-based prediction of protein solubility. *J Mol Biol.*, Advance Access published December 9, 2011, doi:10.1016/j.jmb.2011.12.005.

Aguiar P, Beveridge TJ, Reysenbach AL. 2004. Sulfurihydrogenibium azorense, sp. nov., a thermophilic hydrogen-oxidizing microaerophile from terrestrial hot springs in the Azores. *Int J Syst Evol Microbiol.* 54:33–39.

Akiyama Y. 2009. Quality control of cytoplasmic membrane proteins in *Escherichia coli*. *J Biochem.* 146:449–454.

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25:3389–3402.

Ang D, Georgopoulos C. 1989. The heat-shock-regulated grpE gene of *Escherichia coli* is required for bacterial growth at all temperatures but is dispensable in certain mutant backgrounds. *J Bacteriol.* 171: 2748–2755.

Blum P, Ory J, Bauernfeind J, Krska J. 1992. Physiological consequences of DnaK and DnaJ overproduction in *Escherichia coli*. *J Bacteriol.* 174: 7436–7444.

Brehmer D, Rudiger S, Gassler CS, Klostermeier D, Packschies L, Reinstein J, Mayer MP, Bukau B. 2001. Tuning of chaperone activity of Hsp70 proteins by modulation of nucleotide exchange. *Nat Struct Biol.* 8: 427–432.

Burkholder WF, Zhao X, Zhu X, Hendrickson WA, Gragerov A, Gottesman ME. 1996. Mutations in the C-terminal fragment of DnaK affecting peptide binding. *Proc Natl Acad Sci USA.* 93: 10632–10637.

Calloni G, Chen T, Schermann SM, Chang H-C, Genevaux P, Agostini F, Tartaglia GG, Hayer-Hartl M, Hartl FU. 2012. DnaK functions as a central hub in the E. coli chaperone network. *Cell Rep.* 1: 251–264.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol.* 17: 540–552.

Chevenet F, Brun C, Bañuls AL, Jacq B, Christen R. 2006. TreeDyn: towards dynamic graphics and annotations for analyses of trees. *BMC Bioinformatics* 7:439.

dos Reis M, Savva R, Wernisch L. 2004. Solving the riddle of codon usage preferences: a test for translational selection. *Nucleic Acids Res.* 32: 5036–5044.

Eder W, Huber R. 2002. New isolates and physiological properties of the Aquificales and description of Thermocrinis albus sp. nov. *Extremophiles* 6:309–318.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32:1792–1797.

Ferrera I, Longhorn S, Banta AB, Liu Y, Preston D, Reysenbach AL. 2007. Diversity of 16S rRNA gene, ITS region and aclB gene of the Aquificales. *Extremophiles* 11:57–64.

Gil R, Silva FJ, Pereto J, Moya A. 2004. Determination of the core of a minimal bacterial gene set. *Microbiol Mol Biol Rev.* 68: 518–537.

Glockshuber R. 1999. Protein folding.. Where do the electrons go? *Nature* 401:30–31.

Götz D, Banta A, Beveridge TJ, Rushdi AI, Simoneit BR, Reysenbach AL. 2002. Persephonella marina gen. nov., sp. nov. and Persephonella guaymasensis sp. nov., two novel, thermophilic, hydrogen-oxidizing microaerophiles from deep-sea hydrothermal vents. *Int J Syst Evol Microbiol.* 52:1349–1359.

Gribaldo S, Lumia V, Creti R, Conway de Macario E, Sanangelantoni A, Cammarano P. 1999. Discontinuous occurrence of the hsp70 (dnaK) gene among Archaea and sequence features of HSP70 suggest a novel outlook on phylogenies inferred from this protein. *J Bacteriol.* 181:434–443.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol.* 59:307–321.

Gupta RS. 1998. Protein phylogenies and signature sequences: a reappraisal of evolutionary relationships among archaebacteria, eubacteria, and eukaryotes. *Microbiol Mol Biol Rev.* 62:1435–1491.

Hartl FU, Hayer-Hartl M. 2009. Converging concepts of protein folding in vitro and in vivo. *Nat Struct Mol Biol.* 16:574–581.

Hügler M, Huber H, Molyneaux SJ, Vetriani C, Sievert SM. 2007. Autotrophic CO2 fixation via the reductive tricarboxylic acid cycle in different lineages within the phylum Aquificae: evidence for two ways of citrate cleavage. *Environ Microbiol.* 9:81–92.

Kawasumi T, Igarashi Y, Kodama T, Minoda Y. 1984. Hydrogenobacter thermophilus gen. nov.,sp. nov., an extremely thermophilic, aerobic, hydrogen-oxidizing bacterium. *Int J Syst Bacteriol.* 34:5–10.

Ladenstein R, Ren B. 2008. Reconsideration of an early dogma, saying "there is no evidence for disulfide bonds in proteins from archaea." *Extremophiles* 12:29–38.

Laksanalamai P, Whitehead TA, Robb FT. 2004. Minimal protein-folding systems in hyperthermophilic archaea. *Nat Rev Microbiol.* 2:315–324.

Leroux MR, Fändrich M, Klunker D, Siegers K, Lupas AN, Brown JR, Schiebel E, Dobson CM, Hartl FU. 1999. MtGimC, a novel archaeal chaperone related to the eukaryotic chaperonin cofactor GimC/prefoldin. *EMBO J.* 18:6730–6743.

L'Haridon S, Cilia V, Messner P, Ragúenès G, Gambacorta A, Sleytr UB, Prieur D, Jeanthon C. 1998. *Desulfurobacterium thermolithotrophum* gen. nov., sp. nov., a novel autotrophic, sulphur-reducing bacterium isolated from a deep-sea hydrothermal vent. *Int J Syst Bacteriol.* 48(Pt 3): 701–711.

Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25:955–964.

Mallick P, Boutz DR, Eisenberg D, Yeates TO. 2002. Genomic evidence that the intracellular proteins of archaeal microbes contain disulfide bonds. *Proc Natl Acad Sci USA.* 99:9679–9684.

Martins LO, Carreto LS, da Costa MS, Santos H. 1996. New compatible solutes related to Di-myo-inositol-phosphate in members of the order Thermotogales. *J Bacteriol.* 178:5644–5651.

McCutcheon JP, McDonald BR, Moran NA. 2009a. Convergent evolution of metabolic roles in bacterial co-symbionts of insects. *Proc Natl Acad Sci USA.* 106:15394–15399.

McCutcheon JP, McDonald BR, Moran NA. 2009b. Origin of an alternative genetic code in the extremely small and GC-rich genome of a bacterial symbiont. *PLoS Genet.* 5:e1000565.

McCutcheon JP, Moran NA. 2012. Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol.* 10:13–26.

Moran NA. 1996. Accelerated evolution and Muller's rachet in endosymbiotic bacteria. *Proc Natl Acad Sci USA.* 93:2873–2878.

Pal C, Papp B, Lercher MJ. 2006. An integrated view of protein evolution. *Nat Rev Genet.* 7:337–348.

Punta M, Coggill PC, Eberhardt RY, et al. (16 co-authors). 2012. The Pfam protein families database. *Nucleic Acids Res.* 40: D290–D301.

Santos H, da Costa MS. 2002. Compatible solutes of organisms that live in hot saline environments. *Environ Microbiol.* 4: 501–509.

Shi W, Zhou Y, Wild J, Adler J, Gross CA. 1992. DnaK, DnaJ, and GrpE are required for flagellum synthesis in *Escherichia coli*. *J Bacteriol.* 174: 6256–6263.

Sugimoto S, Saruwatari K, Higashi C, Sonomoto K. 2008. The proper ratio of GrpE to DnaK is important for protein quality control by the DnaK-DnaJ-GrpE chaperone system and for cell division. *Microbiology* 154:1876–1885.

Teter SA, Houry WA, Ang D, Tradler T, Rockabrand D, Fischer G, Blum P, Georgopoulos C, Hartl FU. 1999. Polypeptide flux through bacterial Hsp70: DnaK cooperates with trigger factor in chaperoning nascent chains. *Cell* 97:755–765.

Van Durme J, Maurer-Stroh S, Gallardo R, Wilkinson H, Rousseau F, Schymkowitz J. 2009. Accurate prediction of DnaK-peptide binding

via homology modelling and experimental data. *PLoS Comput Biol.* 5: e1000475.

van Ham RC, Kamerbeek J, Palacios C, et al. (16 co-authors). 2003. Reductive genome evolution in Buchnera aphidicola. *Proc Natl Acad Sci U S A.* 100:581–586.

Vetriani C, Speck MD, Ellor SV, Lutz RA, Starovoytov V. 2004. *Thermovibrio ammonificans* sp. nov., a thermophilic, chemolithotrophic, nitrate-ammonifying bacterium from deep-sea hydrothermal vents. *Int J Syst Evol Microbiol.* 54:175–181.

Vorderwülbecke S, Kramer G, Merz F, Kurz TA, Rauch T, Zachmann-Brand B, Bukau B, Deuerling E. 2004. Low temperature or GroEL/ES overproduction permits growth of *Escherichia coli* cells lacking trigger factor and DnaK. *FEBS Lett.* 559:181–187.

Warnecke T, Hurst LD. 2010. GroEL dependency affects codon usage-support for a critical role of misfolding in gene evolution. *Mol Syst Biol.* 6:340.

Wernegreen JJ, Moran NA. 1999. Evidence for genetic drift in endosymbionts (Buchnera): analyses of protein-coding genes. *Mol Biol Evol.* 16:83–97.

Wild J, Altman E, Yura T, Gross CA. 1992. DnaK and DnaJ heat shock proteins participate in protein export in *Escherichia coli*. *Genes Dev.* 6:1165–1172.

Williams TA, Fares MA. 2010. The effect of chaperonin buffering on protein evolution. *Genome Biol Evol.* 2:609–619.

Wilson D, Pethica R, Zhou Y, Talbot C, Vogel C, Madera M, Chothia C, Gough J. 2009. SUPERFAMILY—sophisticated comparative genomics, data mining, visualization and phylogeny. *Nucleic Acids Res.* 37:D380–386.

Wong P, Houry WA. 2004. Chaperone networks in bacteria: analysis of protein homeostasis in minimal cells. *J Struct Biol.* 146:79–89.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 24:1586–1591.

Yu NY, Wagner JR, Laird MR, et al. (11 co-authors). 2010. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* 26:1608–1615.

Zeldovich KB, Berezovsky IN, Shakhnovich EI. 2007. Protein and DNA sequence determinants of thermophilic adaptation. *PLoS Comput Biol.* 3:e5.