Pervasive convergent evolution and extreme phenotypes define chaperone requirements of protein homeostasis

Yasmine Draceni[a] and Sebastian Pechmann[a,1]

[a]Département de Biochimie, Université de Montréal, Montréal, QC H3T 1J4, Canada

Edited by Angela M. Gronenborn, University of Pittsburgh School of Medicine, Pittsburgh, PA, and approved August 8, 2019 (received for review March 19, 2019)

Maintaining protein homeostasis is an essential requirement for cell and organismal viability. An elaborate regulatory system within cells, the protein homeostasis network, safeguards that proteins are correctly folded and functional. At the heart of this regulatory system lies a class of specialized protein quality control enzymes called chaperones that are tasked with assisting proteins in their folding, avoiding aggregation and degradation. Failure and decline of protein homeostasis are directly associated with conditions of aging and aging-related neurodegeneration. However, it is not clear what tips the balance of protein homeostasis and leads to onset of aging and diseases. Here, using a comparative genomics approach we report general principles of maintaining protein homeostasis across the eukaryotic tree of life. Expanding a previous study of 16 eukaryotes to the quantitative analysis of 216 eukaryotic genomes, we find a strong correlation between the composition of eukaryotic chaperone networks and genome complexity that is distinct for different species kingdoms. Organisms with pronounced phenotypes clearly buck this trend. *Northobranchius furzeri*, the shortest-lived vertebrate and a widely used model for fragile protein homeostasis, is found to be chaperone limited while *Heterocephalus glaber* as the longest-lived rodent and thus an especially robust organism is characterized by above-average numbers of chaperones. Strikingly, the relative size of chaperone networks is found to generally correlate with longevity in Metazoa. Our results thus indicate that the balance in protein homeostasis may be a key variable in explaining organismal robustness.

protein homeostasis | chaperones | tree of life | genome evolution | aging

Keeping one's proteome properly folded and functional through varying conditions and stresses is of critical importance for cellular and organismal survival (1). Conversely, the decline of the cell's capacity to keep its proteins in their correct shape, i.e., to maintain protein homeostasis, is a central hallmark of aging and onset of aging-associated diseases (2). Cells maintain protein homeostasis through a complex regulatory network that integrates protein synthesis, folding, degradation, and trafficking pathways (3). The central players of the protein homeostasis network are a class of specialized protein quality control enzymes called chaperones (4, 5). Chaperones assist proteins in their folding; protect them from aberrant aggregation while promoting functional assembly; and, if needed, sequester and target them for degradation (4).

Eukaryotic genomes often contain more than ca. 50 to 300 different chaperone genes that classify into distinct families based on the structure and function of the encoded proteins. Hsp90-type chaperones are the main stress responders that stabilize and refold stress-denatured proteins (6, 7). They are usually encoded by only a few genes (8) that are strongly activated under unfavorable conditions (9). The family of Hsp70s comprises key enzymes involved in de novo folding, thus determining the fate of newly made proteins (10, 11). Hsp40 chaperones primarily act as cochaperones and nucleotide exchange factors to Hsp70s, providing specificity in guiding protein folding, (dis)assembly, and translocation (12, 13). Small heat-shock proteins (sHsp or Hsp20) stabilize partially denatured proteins and prevent their aggregation (14), while enzymes from the Hsp100 chaperone family are primarily involved in disaggregation and proteolysis of proteins (15). Hsp60s encode subunits of eukaryotic chaperonins, which are heteromeric protein complexes that provide a fully enclosed protective cavity for the folding of select proteins (16, 17). Together with their individual functional specializations, the ensemble of chaperones in the cell provides a powerful network to control the quality of proteins.

Chaperones generally bind to insoluble, sticky proteins that are at risk for aggregation (18). However, unlike in bacteria where each chaperone family is represented by only 1 or very few genes, chaperone genes in higher eukaryotes markedly increase in numbers. This is particularly true for chaperones of the Hsp70 and Hsp40 families that can be found with strongly increasing diversity in eukaryotes (19). The reasons for this expansion are far from understood. In specifically interacting with select pools of client proteins, chaperones are likely key regulators of cellular networks (20).

Accordingly, the comparison of genomes (21) and dynamically regulated proteomes (22) offers the opportunity to gain

Significance

Survival of all organisms depends on their ability to maintain protein homeostasis, i.e., to keep their proteins in folded and functional shapes. Failure of protein homeostasis is directly associated with aging-related protein misfolding diseases that include Alzheimer's and Parkinson's. However, our understanding of how cells maintain protein homeostasis remains incomplete. Here, by performing a large-scale comparative genomics study of protein homeostasis in 216 eukaryotes we find that the relative size of organisms' chaperone networks, the core components of cellular protein homeostasis, directly links to species longevity. Our work thus provides an elegant example of harnessing the power of evolution and comparative genomics to address fundamental open questions in biology with direct relevance to human diseases.

Author contributions: S.P. designed research; Y.D. and S.P. performed research; Y.D. and S.P. analyzed data; and Y.D. and S.P. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The computer codes used in this paper have been deposited in GitHub, https://github.com/pechmannlab/chapevo, and in Zenodo, https://zenodo.org/search?page=1&size=20&q=3387770.

[1]To whom correspondence may be addressed. Email: sebastian.pechmann@umontreal.ca.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1904611116/-/DCSupplemental.

fundamental insights into chaperone function (23, 24) as well as the balance between chaperone network and proteome (25). For instance, an analysis of 16 eukaryotes revealed a general correlation between the size of chaperone networks and the number of protein-coding genes (21). Malignant tumor growth in cancers is aberrantly balanced by dynamically adjusted chaperone expression profiles (26), and even viral infections impose specific requirements on the host cell chaperone network to sustain their replication (27). Multiple independent duplication and gene loss events in the Hsp90 family (8) may reflect a perpetual challenge to stay in balance with an equally evolving proteome. In turn, the expansion of proteomes (28) and encoded protein networks (29, 30) themselves is promoted by chaperones.

To this end, the study of extreme phenotypes affords fascinating insights into the trade-offs of protein homeostasis underlying organismal fitness. At the protein level, thermophilic organisms counter the challenge of thermodynamic protein stability with reduced surface hydrophobicity to avoid aggregation yet an increase in buried sequence hydrophobicity to stabilize their structural core (31). Similarly, specifically evolved ice-binding proteins allow species to thrive at subzero temperatures (32). At the cellular and organismal levels, the *Dictyostelium discoideum* proteome with an extremely high number of aggregation-prone prion domains is balanced by an increase in Hsp100 disaggregase capacity (33). The African killifish *Northobranchius furzeri* as the shortest-lived vertebrate with a life expectancy of only 4 to 6 mo (34, 35) progresses through many of the hallmarks associated with human aging, including pronounced loss of protein homeostasis. In contrast, the naked mole rat *Heterocephalus glaber* as the longest-living rodent with a life expectancy of over 30 y (36) starkly differs from strongly related rodents such as *Mus musculus* that live for only around 3 y (36). The genetic origins of the comparably short and long lifespans, respectively, are far from understood. However, in all cases aging is characterized by the loss of protein homeostasis (37) and the accumulation of misfolded and aberrantly aggregated proteins (38), underlining the importance of better understanding how cells keep their proteomes in balance.

Here, we present a comparative genomics analysis of protein homeostasis across the eukaryotic tree of life. By expanding an initial review of 16 eukaryotic genomes (21) to the quantitative analysis of 216 eukaryotes, we report general principles of keeping the balance of successful protein homeostasis. General strong similarity in eukaryotic chaperone profiles suggests pervasive convergent evolution that maintains a universal composition of eukaryotic chaperone networks. Moreover, we found a systematic and much more pronounced link than previously reported between the size of chaperone networks and proteome complexity that is distinct for individual species kingdoms. Extreme phenotypes deviated from this trend in systematic fashion. The short-lived *N. furzeri* was found to be chaperone limited, while *H. glaber* as an especially robust organism was characterized by above-average numbers of chaperones. Strikingly, the relative size of chaperone networks was found to generally correlate with longevity in Metazoa. Our work thus indicates that the balance in protein homeostasis may be a key variable in explaining organismal robustness.

Results

To investigate principles of the evolution of protein homeostasis in eukaryotes, we collected a comprehensive high-quality dataset of annotated proteomes that we could analyze in the context of their phylogenetic relationships. Phylogenetic models across a diverse set of species are best computed based on a single marker (39), in eukaryotes usually the highly conserved 18S rRNA gene. To efficiently identify 18S rRNA in sequenced genomes, we developed a pipeline based on the mapping of synthetic rRNA reads derived from validated 18S sequences and the RNAmmer software (40). Notably, our pipeline led to a clear improvement over using the established tools BLAST or RNAmmer only (Fig. 1*A* and *SI Appendix*, Fig. S1*A*). Additional conservative data quality control (*Methods*; Fig. 1*B*; and *SI Appendix*, Fig. S1 *B* and *C* and Tables S1 and S2) resulted in a set of 216 eukaryotes spanning all 4 eukaryotic kingdoms of life, for which we computed a phylogenetic tree (Fig. 1*C*). The obtained phylogenetic model was validated by good consistency between independent runs and found to be in excellent agreement with taxonomy annotations (*Methods*; Fig. 1*C*; and *SI Appendix*, Fig. S1*D* and Tables S1 and S2).

Next, all occurrences of genes encoding chaperones of the Hsp20, Hsp40, Hsp60, Hsp70, Hsp90, and Hsp100 families were identified (*Methods*). Remarkably, patterns of the chaperone counts directly correlated with similarities and divisions in the previously computed phylogeny (Fig. 2). Increasing organism complexity, from unicellular fungi to multicellular animals to plants with their additional chloroplast genomes, was accompanied by an increase in diversity in all chaperone families (Fig. 2 and *SI Appendix*, Fig. S2). The shift to larger chaperone networks for more complex genomes was noticeable for all chaperone classes, but especially standing out were increased counts of Hsp20-, Hsp40-, and Hsp70-type chaperones (Fig. 2 and *SI Appendix*, Fig. S2). Within the different kingdoms, plants showed by far the largest variance (Fig. 2 and *SI Appendix*, Fig. S2), likely due to their prevalent polyploidy (41). While the overall relative composition of eukaryotic chaperone networks appeared strongly preserved, individual organisms stood out by well below- or above-average numbers in individual chaperone families (Fig. 2). Taken together, our analysis of chaperone networks across eukaryotes suggested fundamentally conserved similarities with pronounced exceptions.

We sought to more systematically evaluate these observed (dis)similarities in the composition of eukaryotic chaperone networks in the context of their evolutionary relations. Chaperone network similarity was first quantified by average pairwise correlation coefficients of the chaperone counts between each species and all others and compared to the average pairwise evolutionary distances obtained from the branch lengths of the phylogenetic tree (*Methods*). Focusing on the Fungi, Metazoa, and Plantae, we observed very strong correlations between the compositions of chaperone networks independent of their evolutionary distances (Fig. 3*A*). These results suggested that chaperone networks in eukaryotes are generally composed of very similar relative numbers of Hsp20-, Hsp40-, Hsp60-, Hsp70-, Hsp90-, and Hsp100-type chaperones. Because these different chaperones perform overlapping yet specialized functions, it was not

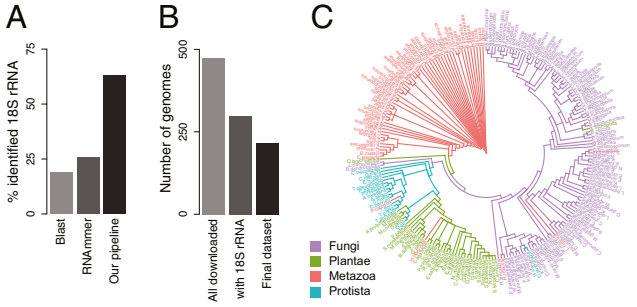

Fig. 1. Phylogenetic classification of eukaryotes. (*A*) Identification of 18S rRNA sequences. The developed pipeline improves on the efficient identification of 18S rRNA sequences from sequenced genomes (*Methods* and *SI Appendix*). (*B*) Conservative data quality control yielded a set of 216 genomes (*Methods* and *SI Appendix*, Table S2). (*C*) Eukaryotic tree of life colored by species kingdom. Strong clustering of species within their kingdoms supports a good phylogenetic model.

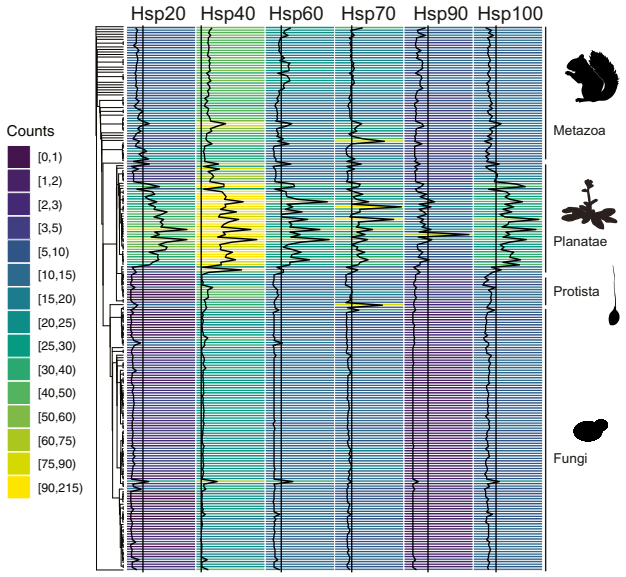

Fig. 2. Evolution of chaperone networks. The counts of Hsp20-, Hsp40-, Hsp60-, Hsp70-, Hsp90-, and Hsp100-type chaperone genes from 216 eukaryotes are visualized as a heatmap along the previously computed phylogeny. Individual counts (black curves) relative to the median (black lines) for each chaperone family further highlight patterns of chaperone network expansion that link to species kingdom annotation.

surprising to find all of them proportionally present. Moreover, the observation that the generally very high similarity in chaperone profiles across eukaryotes was independent of evolutionary distance (Fig. 3*A*) indicated a fundamental selective pressure on chaperone network composition, likely through convergent evolution.

Notable exceptions stood out. The fungus with the lowest average correlation coefficient was *Rhizophagus irregularis*, a symbiotic fungus used in soil agriculture and from a different taxonomy division than the rest of the fungi in our data. The plants with the lowest average correlation coefficient were *Rosa chinesis* and *Capsicum annuum*, 2 highly domesticated plants (Fig. 3*A*). Of the more diverse animal cluster, the most dissimilar chaperone profiles were found in the Pacific oyster *Crassostrea gigas* and the sea urchin *Strongylocentrotus purpuratus* even though they exhibited only short average evolutionary distances (Fig. 3*A*).

To better understand constraints on chaperone network composition, we decided to analyze in depth the chaperone networks of these most dissimilar Metazoa *C. gigas* and *S. purpuratus* together with *N. furzeri* and *H. glaber*, 2 model organisms of known extreme phenotypes that directly link to especially fragile and robust protein homeostasis, respectively. Interestingly, *C. gigas* was characterized by an extraordinarily high count of over 80 Hsp70 chaperones (Fig. 3 *B* and *C*) yet had one of the lowest counts of Hsp100s of all animals analyzed (*SI Appendix*, Fig. S3). *S. purpuratus* had a high number of Hsp70s (Fig. 3 *B* and *C*) as well as one of the highest counts of Hsp100s (*SI Appendix*, Fig. S3). Notably, the naked mole rat *H. glaber* displayed one of the highest counts of Hsp40 chaperones and ranked among the highest in counts of both Hsp70s and Hsp90s (Fig. 3 *B* and *C*). In contrast, the African killifish *N. furzeri* stood out for one of the lowest counts of Hsp90s (Fig. 3 *B* and *C*). Taken together, the most dissimilar examples or cases that represent known extreme phenotypes were also characterized by extreme counts of specific chaperone types.

Extending this quantification of chaperone network similarity, we performed hierarchical clustering of the Metazoa chaperone profiles based on cosine distance that is more sensitive to deviations in counts of individual chaperone types. Importantly, this analysis reinforced the general ultraconservation of chaperone network composition, evident by prevalent very low cosine distances (Fig. 3*D*). A small cluster of species including the sexually reproducing nematode *Caenorhabditis remanei* and the rat parasite *Strongyloides ratti* could now additionally be identified as particularly Hsp20 rich (Fig. 3*D*). *H. glaber* clustered with the black fruit bat *Pteropus alecto* based on high Hsp90 counts while *N. furzeri* appeared to have a chaperone network with standard composition (Fig. 3*D*). The continued sequencing of additional species will likely clarify these distinct adaptations that currently stand out as exceptions from a generic eukaryotic chaperone profile.

To test whether the expansion of individual chaperone families linked to distinct expansions of coevolved proteomes, we evaluated putative relationships between chaperone counts and sequence hydrophobicity, aggregation propensity, and intrinsic disorder in proteomes (*Methods*). On average, fungi and animals showed higher aggregation and disorder scores than plants (*SI Appendix*, Fig. S3*B*). This difference may reflect the contribution of chloroplasts that maintain the character of their archaeal and bacterial origins to plant genomes (42). The strongest individual signal was obtained for an increase in protein disorder in animal proteomes that directly correlated with increasing numbers of Hsp40s and Hsp70s (Fig. 3*E* and *SI Appendix*, Fig. S3*C*). Albeit weaker in explained variance, an increase in protein aggregation propensity correlated with increasing numbers of Hsp20 and Hsp100 chaperones in both animals and plants (*SI Appendix*, Fig. S3*C*). Thus, even at the level of chaperone family size an increase in chaperone counts could be directly linked to an increase in organism and proteome complexity (43).

Because chaperones preferentially bind to aggregation-prone proteins, we next mapped the aggregation propensity of the proteomes onto the inferred phylogeny. This highlighted many clusters of similar aggregation profiles across closely related species (Fig. 4*A*). To directly evaluate how proteome complexity and aggregation burden are balanced with sufficient protein homeostasis capacity, we compared the proteome aggregation scores, computed as the sum of the individual predicted protein aggregation propensities and thus reflecting the size of a proteome weighted by its propensity to aggregate, with the sizes of chaperone networks. Remarkably, we found strong correlations between proteome aggregation score and chaperone network size that followed distinct trends for individual species kingdoms (Fig. 4*B*). Of note, almost equivalent results could be obtained for the comparison of chaperone network to just the proteome sizes (*SI Appendix*, Fig. S4*A* and Table S3).

To further understand this correlation for the more diverse cluster of Metazoa, we first validated that the proteomes of *H. glaber*, *N. furzeri*, *C. gigas*, and *S. purpuratus* did not exhibit particularly different aggregation scores (Fig. 4*C*). We observed a general and very strong linear correlation between the size of the core chaperone network defined as Hsp40, Hsp70, and Hsp90 and the proteome aggregation score (Fig. 4*D* and *SI Appendix*, Fig. S4*B*) that was much more pronounced than that previously reported for a subset of 16 eukaryotes (21). Organisms that clearly diverted from this trend included the ones previously identified. The naked mole rat *H. glaber* was found far to the right, suggesting an excess of chaperones compared to its proteome aggregation burden (Fig. 4*D*). This observation matched very well the remarkable robustness and longevity of *H. glaber*. *C. gigas* could also be found far to the right of the general trend (Fig. 4*D*). In contrast, the African killifish *N. furzeri* placed on the left side of the general correlation (Fig. 4*D*), reflective of its known fragile protein homeostasis. Similarly, *M. musculus* as one of the shortest-lived mammals equally placed clearly to the left side of this correlation, as did the nematodes *Caenorhabditis*

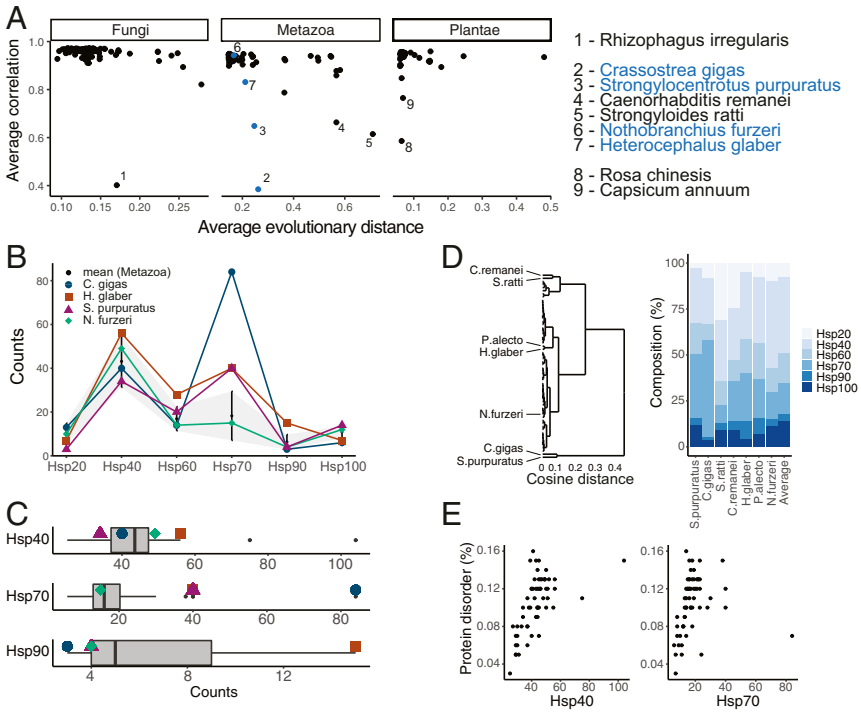

Fig. 3. Diversity and convergent evolution in chaperone networks. (*A*) Similarity of the composition of chaperone networks within phylogenetic clusters as a function of evolutionary distance. Average evolutionary distance denotes the average of pairwise evolutionary distances computed from the branch lengths. Average correlation coefficients represent the average of pairwise correlation coefficients between the compositions of chaperone networks of a query species and of the other species. (*B*) Composition of the chaperone networks of 4 exemplary species that are characterized by pronounced known extreme phenotypes, namely the short-lived vertebrate *N. furzeri*, the long-lived rodent *H. glaber*, the pacific oyster *C. gigas*, and the sea urchin *S. purpuratus*. The distributions of the chaperone counts in all Metazoa (black dots) ± SD (gray area) are shown as reference. (*C*) Distribution of the numbers of Hsp40, Hsp70, and Hsp90 chaperones in animal genomes. (*D*) Hierarchical clustering of chaperone networks in Metazoa based on their relative composition and cosine distance. Prevalent extremely short branch lengths suggest strong conservation of the relative composition of chaperone networks, likely through convergent evolution. The distinct compositions of the very few species with lower chaperone network similarity are highlighted. (*E*) Expansion of Hsp40 and Hsp70 chaperone families correlates with an increase in protein disorder in Metazoa proteomes.

elegans and *C. remanei* (Fig. 4*D*). *S. purpuratus* fell just left of the general trend despite high life expectancy, suggesting that its expanded chaperone network may compensate other putative constraints (Fig. 4*D*). Importantly, the same observations could be made when considering full chaperone networks (*SI Appendix*, Fig. S4*A*).

Finally, to test whether deviation from the correlation between the size of chaperone networks and proteome complexity was generally indicative of organismal robustness, we compared the deviation of individual species from this general trend, quantified by the residuals of the robust regression fit, to species longevity (44). Strikingly, we could observe a clear anticorrelation between relative chaperone network size and species maximum age (Fig. 4*E*). Specifically, organisms with disproportionally small chaperone networks have in general a much lower life expectancy than organisms with large chaperone networks relative to their proteomes (Fig. 4*D* and *SI Appendix*, Table S4). The main exceptions from this trend were found to be some heavy animals including *Loxodonta africana*, the African bush elephant, and the alligator species *Alligator mississippiensis* and *Alligator sinensis*. Another clear exception was *Homo sapiens* who live much longer than suggested by their predicted robustness (Fig. 4*E* and *SI Appendix*, Table S4), likely the result of modern medicine and high-tech societies. Of note, the same anticorrelation, albeit weaker, could be observed when normalizing longevity by body weight (*SI Appendix*, Fig. S4*C*) (45). Taken together, our results thus suggest a fundamental link between the balance in protein homeostasis and organismal robustness.

Discussion

We performed a comparative genomics analysis of protein homeostasis in 216 eukaryotes. Our results suggest strong convergent evolution to maintain the overall composition of eukaryotic chaperone networks; a systematic and pronounced link between the size of chaperone networks and organism complexity that is distinct for individual species kingdoms; and, remarkably, a direct correlation between relative chaperone network size and species longevity.

Organisms that divert from the general chaperone profile are often accompanied by specific phenotypic constraints that match fascinating facets of species-specific adaptations. The sea urchin *S. purpuratus* is characterized by a remarkable lifespan of over a century as well as high fecundity (46, 47). With a high degree of polymorphisms in its population (47), *S. purpuratus* contained one of the largest counts of Hsp100 chaperones. The proteolytic activity of Hsp100s may assist the complex and sophisticated innate immune system (48) in defending against a plethora of stressors, as well as the biosynthesis and assembly of its pronounced spikes. Another maritime organism living in the highly stressful intertidal zone, the Pacific oyster *C. gigas* has evolved unusually high numbers of Hsp70 chaperones to adapt to harsh living conditions and the specific biosynthesis requirements of making the hard oyster shells (49).

The observation that the deviation from the general trend between chaperone network size and proteome complexity directly correlates with longevity in Metazoa is striking. In the cases of *N. furzeri* and *H. glaber* this links primarily to opposing numbers of Hsp90 proteins. The short-lived *N. furzeri* has

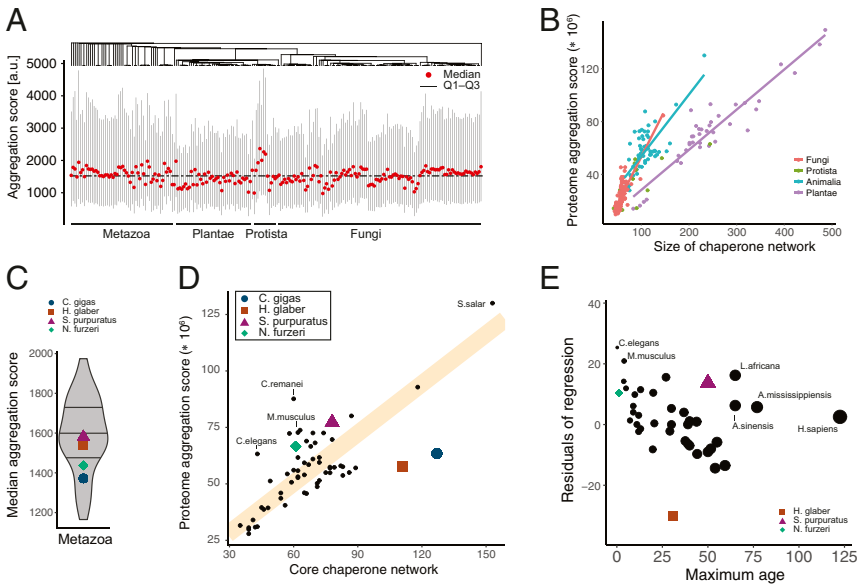

Fig. 4. Balanced protein homeostasis and organismal robustness. (*A*) Distribution of predicted protein aggregation propensities in proteomes across the previously computed phylogeny. Shown are median, 25% (Q1), and 75% (Q3) percentiles of the distributions of per-protein predicted aggregation scores for each species. (*B*) Link between proteome aggregation score, which was computed as size of a proteome weighted by its predicted propensity to aggregate, and size of chaperone networks. Regression lines highlight distinct trends for individual species kingdoms. Similarly strong trends can be observed between the proteome and chaperone network sizes, respectively (*SI Appendix*, Fig. S4 and Table S4). (*C*) Distribution of median aggregation scores across the animal kingdom. (*D*) Relationship between the sizes of the core chaperone networks composed of Hsp40, Hsp70, and Hsp90 chaperones and the proteome aggregation score. The regression fit (orange line) separates organisms that are predicted to be more fragile (above line) vs. more robust (below line). (*E*) Diversion from balanced protein homeostasis correlates with longevity. The residuals of the regression fit (*D*) serve as a proxy for organism fragility or robustness and are plotted against data on species maximum age.

one of the lowest and the long-lived *H. glaber* one of the highest numbers of Hsp90 genes among the animals analyzed. The naked mole rat *H. glaber* is often compared to *M. musculus* for their stark contrast in longevity (50), which is equally supported by our analyses (Fig. 4*D* and *SI Appendix*, Fig. S4*C*). Another short-lived model organism for aging research, the nematode *C. elegans*, is similarly found to be chaperone limited (Fig. 4*D*). The individual living conditions and phenotypes of different organisms are as complex as the protein homeostasis network that will ultimately have to be understood at a much more detailed level.

A shortcoming of our analysis lies in the inference from the number of chaperone genes to their protein-folding capacity in the cell, which equally strongly depends on their expression profiles. Nonetheless, from an engineering and systems perspective there are fundamental differences between having few loci that are highly expressed vs. multiple gene copies, even if both result in similar final protein levels. As organisms age and mutations accumulate, the biosynthesis of chaperones themselves may be negatively affected. By distributing the genetic information across several loci the system becomes more redundant and thus robust (51). Equally importantly, diversity in chaperones likely yields a gain of specificity and fidelity in regulation, thus affording more control over adaptive responses to stress that is critical for long-term survival.

Finally, protein homeostasis is achieved through a truly complex regulatory network and protein folding under chaperone assistance comprises only 1 central aspect of it (52). Many parts of this regulatory system are tightly controlled and coupled, and even RNA molecules are chaperoned (53). An alternative means to remove misfolded and aggregated proteins is provided by powerful degradation systems whose malfunctioning is equally implicated in organismal aging and disease (54). The continued sequencing of genomes will without doubt enable researchers to substantially expand evolutionary and comparative genomics efforts to learn about protein homeostasis as well its link to aging and aging-related diseases.

Methods

Code Availability. Computer code to reproduce all analyses is available at https://github.com/pechmannlab/chapevo and archived at doi:10.5281/zenodo.3387770.

Data Sources. All eukaryotic genomes from NCBI RefSeq that also had corresponding validated proteomes available in the Uniprot database were downloaded, yielding genomic sequences for 472 eukaryotes. From RefSeq, both annotated RNA sequences and whole-genome FASTA files were obtained. Draft genome assemblies of *H. glaber* (36) and *N. furzeri* (34, 35) were downloaded additionally. The corresponding protein sequences for each organism were retrieved from the Uniprot database using the nonredundant FASTA files without isoforms and in the case of *N. furzeri* from the species-specific database (http://nfingb.leibniz-fli.de/) and filtered for 1 isoform per gene based on gene IDs. Data on species longevity were obtained from the AnAge database (44).

Identification of Chaperone Networks. Counts of genes encoding Hsp20-, Hsp40-, Hsp60-, Hsp70-, Hsp90-, and Hsp100-type chaperones were identified with expert-curated profile hidden Markov models (HMMs) from the HSPIR database (55). The hmmsearch module of the Hmmer software (56) was used to search proteomes; hits with e-value $< 10^{-6}$ were considered.

Phylogenetic Tree of Eukaryotes. The phylogenetic relationship between eukaryotic genomes was computed based on 18S rRNA sequences. The 18S rRNA sequences were first identified by aligning synthetic 50-nt reads from the validated 18S rRNA sequences of *C. elegans*, *Drosophila melanogaster*, *H. sapiens*, *M. musculus*, and *Saccharomyces cerevisiae* with HISAT2 (57) to the target genome, followed by RNAmmer (40) analysis of the candidate locus with the highest coverage (*SI Appendix*). The 18S rRNA sequences that contained "N" characters or introduced more than 15% gaps across pairwise alignments were removed (*SI Appendix*, Table S2) before constructing a high-quality sequence alignment with the structure-based SSU-align program (58). Phylogenetic trees were constructed with PhyloBayes (59) using the CAT/GTR model and 3 independent MCMC chains of length 10,000. The

first 1,000 trees were cast off as burn-in. For our final dataset of 216 eukaryotic genomes the largest discrepancy between 3 independent chains was maxdiff = 0.27, thus indicating a good phylogenetic model (59). Individual evolutionary distances were computed from the branch lengths with the ETE3 toolkit (60). Species taxonomy information was retrieved from the NCBI taxonomy database with the R package myTAI (61). We assigned species with missing kingdom annotations to the cluster of "Protists."

Data Analysis. The propensity of proteins to aggregate was predicted from the protein amino acid sequences with the Tango software (62). Tango predicts per-residue aggregation scores that were summed up to yield protein aggregation scores. Finally, proteome aggregation scores were computed as the sum of the individual protein aggregation scores, thus reflecting the size of the proteome weighted by its aggregation propensity. Protein sequence hydrophobicity was computed with the Kyte–Doolittle hydrophobicity scale (63), and protein intrinsic disorder was predicted from individual protein sequences with IUPred2 (64). Robust regression fits were performed by quantile regression with the R package "quantreg."

ACKNOWLEDGMENTS. We are grateful to Dr. E. Noutahi and Dr. S. Laurin-Lemay for assistance with the PhyloBayes calculations. This work was supported by an Natural Sciences and Engineering Research Council of Canada discovery grant and the Université de Montréal, as well as by access to supercomputers managed by Calcul Québec and Compute Canada. S.P. holds the Canada Research Chair in Computational Systems Biology.

1. F. Chiti, C. M. Dobson, Protein misfolding, functional amyloid, and human disease. *Annu. Rev. Biochem.* **75**, 333–366 (2006).
2. S. Kaushik, A. M. Cuervo, Proteostasis and aging. *Nat. Med.* **21**, 1406–1415 (2015).
3. E. T. Powers, R. I. Morimoto, A. Dillin, J. W. Kelly, W. E. Balch, Biological and chemical approaches to diseases of proteostasis deficiency. *Annu. Rev. Biochem.* **78**, 959–991 (2009).
4. Y. E. Kim, M. S. Hipp, A. Bracher, M. Hayer-Hartl, F. U. Hartl, Molecular chaperone functions in protein folding and proteostasis. *Annu. Rev. Biochem.* **82**, 323–355 (2013).
5. M. Brehme *et al.*, A chaperome subnetwork safeguards proteostasis in aging and neurodegenerative disease. *Cell Rep.* **9**, 1135–1150 (2014).
6. A. J. McClellan *et al.*, Diverse cellular functions of the Hsp90 molecular chaperone uncovered using systems approaches. *Cell* **131**, 121–135 (2007).
7. M. Taipale, D. F. Jarosz, S. Lindquist, HSP90 at the hub of protein homeostasis: Emerging mechanistic insights. *Nat. Rev. Mol. Cell Biol.* **11**, 515–528 (2010).
8. B. Chen, D. Zhong, A. Monteiro, Comparative genomics and evolution of the HSP90 family of genes across all kingdoms of organisms. *BMC Genomics* **7**, 156 (2006).
9. V. Albanese, A. Y. Yam, J. Baughman, C. Parnot, J. Frydman, Systems analyses reveal two chaperone networks with distinct functions in eukaryotic cells. *Cell* **124**, 75–88 (2006).
10. S. Pechmann, F. Willmund, J. Frydman, The ribosome as a hub for protein quality control. *Mol. Cell* **49**, 411–421 (2013).
11. F. Willmund *et al.*, The cotranslational function of ribosome-associated Hsp70 in eukaryotic protein homeostasis. *Cell* **152**, 196–209 (2013).
12. P. Walsh, D. Bursać, Y. C. Law, D. Cyr, T. Lithgow, The J-protein family: Modulating protein assembly, disassembly and translocation. *EMBO Rep.* **5**, 567–571 (2004).
13. N. B. Nillegoda *et al.*, Evolution of an intricate J-protein network driving protein disaggregation in eukaryotes. *eLife* **6**, e24560 (2017).
14. Y. Sun, T. MacRae, Small heat shock proteins: Molecular structure and chaperone function. *Cell. Mol. Life Sci.* **62**, 2460–2476 (2005).
15. S. M. Doyle, S. Wickner, Hsp104 and ClpB: Protein disaggregating machines. *Trends Biochem. Sci.* **34**, 40–48 (2009).
16. A. Leitner *et al.*, The molecular architecture of the eukaryotic chaperonin TRiC/CCT. *Structure* **20**, 814–825 (2012).
17. A. Y. Yam *et al.*, Defining the TRiC/CCT interactome links chaperonin function to stabilization of newly made proteins with complex topologies. *Nat. Struct. Mol. Biol.* **15**, 1255–1262 (2008).
18. G. G. Tartaglia, C. M. Dobson, F. U. Hartl, M. Vendruscolo, Physicochemical determinants of chaperone requirements. *J. Mol. Biol.* **400**, 579–588 (2010).
19. X. B. Qiu, Y. M. Shao, S. Miao, L. Wang, The diversity of the Dnaj/Hsp40 family, the crucial partners for Hsp70 chaperones. *Cell. Mol. Life Sci.* **63**, 2560–2570 (2006).
20. C. Sőti, C. Pal, B. Papp, P. Csermely, Molecular chaperones as regulatory elements of cellular networks. *Curr. Opin. Cell Biol.* **17**, 210–215 (2005).
21. E. T. Powers, W. E. Balch, Diversity in the origins of proteostasis networks–a driver for protein function in evolution. *Nat. Rev. Mol. Cell Biol.* **14**, 237–248 (2013).
22. A. J. Sala, L. C. Bott, R. I. Morimoto, Shaping proteostasis at the cellular, tissue, and organismal level. *J. Cell Biol.* **216**, 1231–1241 (2017).
23. D. Bogumil, D. Alvarez-Ponce, G. Landan, J. O. McInerney, T. Dagan, Integration of two ancestral chaperone systems into one: The evolution of eukaryotic molecular chaperones in light of eukaryogenesis. *Mol. Biol. Evol.* **31**, 410–418 (2014).
24. M. O. Press *et al.*, Genome-scale co-evolutionary inference identifies functions and clients of bacterial Hsp90. *PLoS Genet.* **9**, e1003631 (2013).
25. D. M. Roth, W. E. Balch, Modeling general proteostasis: Proteome balance in health and disease. *Curr. Opin. Cell Biol.* **23**, 126–134 (2011).
26. E. A. Hadizadeh, A. Sverchkova, J. Saez-Rodriguez, A. Schuppert, M. Brehme, A systematic atlas of chaperome deregulation topologies across the human cancer landscape. *PLoS Comput. Biol.* **14**, e1005890 (2018).
27. R. Geller, S. Pechmann, A. Acevedo, R. Andino, J. Frydman, Hsp90 shapes protein and RNA evolution to balance trade-offs between protein stability and aggregation. *Nat. Commun.* **9**, 1781 (2018).
28. D. Caetano-Anollés, K. M. Kim, J. E. Mittenthal, G. Caetano-Anollés, Proteome evolution and the metabolic origins of translation and cellular life. *J. Mol. Evol.* **72**, 14–33 (2011).
29. S. Pechmann, J. Frydman, Interplay between chaperones and protein disorder promotes the evolution of protein networks. *PLoS Comput. Biol.* **10**, e1003674 (2014).
30. T. Korcsmáros, I. A. Kovács, M. S. Szalay, P. Csermely, Molecular chaperones: The modular evolution of cellular networks. *J. Biosci.* **32**, 441–446 (2007).
31. A. M. Thangakani, S. Kumar, D. Velmurugan, M. S. M. Gromiha, How do thermophilic proteins resist aggregation?*Proteins* **80**, 1003–1015 (2012).
32. P. L. Davies, Ice-binding proteins: A remarkable diversity of structures for stopping and starting ice growth. *Trends Biochem. Sci.* **39**, 548–555 (2014).
33. L. Malinovska, S. Palm, K. Gibson, J. M. Verbavatz, S. Alberti, *Dictyostelium discoideum* has a highly Q/N-rich proteome and shows an unusual resilience to protein aggregation. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E2620–E2629 (2015).
34. D. R. Valenzano *et al.*, The African turquoise killifish genome provides insights into evolution and genetic architecture of lifespan. *Cell* **163**, 1539–1554 (2015).
35. K. Reichwald *et al.*, Insights into sex chromosome evolution and aging from the genome of a short-lived fish. *Cell* **163**, 1527–1538 (2015).
36. M. Keane *et al.*, The naked mole rat genome resource: Facilitating analyses of cancer and longevity-related adaptations. *Bioinformatics* **30**, 3558–3560 (2014).
37. K. Schneider, A. Bertolotti, Surviving protein quality control catastrophes - from cells to organisms. *J. Cell Sci.* **128**, 3861–3869 (2015).
38. D. C. David, Aging and the aggregating proteome. *Front. Genet.* **3**, 247 (2012).
39. A. Patwardhan, S. Ray, A. Roy, Molecular markers in phylogenetic studies - a review. *J. Phylogenet. Evol. Biol.* **2**, 131 (2014).
40. K. Lagesen *et al.*, RNAmmer: Consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* **35**, 3100–3108 (2007).
41. S. P. Otto, The evolutionary consequences of polyploidy. *Cell* **131**, 452–462 (2007).
42. I. Yruela, B. Contreras-Moreira, Protein disorder in plants: A view from the chloroplast. *BMC Plant Biol.* **12**, 165 (2012).
43. E. Schad, P. Tompa, H. Hegyi, The relationship between proteome size, structural disorder, and organism complexity. *Genome Biol.* **12**, R120 (2011).
44. R. Tacutu *et al.*, Human ageing genomic resources: New and updated databases. *Nucleic Acids Res.* **46**, D1083–D1090 (2018).
45. J. R. Speakman, Correlations between physiology and lifespan - two widely ignored problems with comparative studies. *Aging Cell* **4**, 167–175 (2005).
46. Consortium SUGS, The genome of the sea urchin Strongylocentrotus purpuratus. *Science* **314**, 941–952 (2006).
47. R. A. Cameron, M. Samanta, A. Yuan, D. He, E. Davidson, SpBase: The sea urchin genome database and web site. *Nucleic Acids Res.* **37**, D750–D754 (2009).
48. L. C. Smith, Innate immune complexity in the purple sea urchin: Diversity of the sp185/333 system. *Front. Immunol.* **3**, 70 (2012).
49. G. Zhang *et al.*, The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* **490**, 49–54 (2012).
50. K. A. Rodriguez *et al.*, Determinants of rodent longevity in the chaperone-protein degradation network. *Cell Stress Chaperones* **21**, 453–466 (2016).
51. J. Stelling, U. Sauer, Z. Szallasi, F. J. Doyle, J. Doyle, Robustness of cellular functions. *Cell* **118**, 675–685 (2004).
52. C. L. Klaips, G. G. Jayaraj, F. U. Hartl, Pathways of cellular proteostasis in aging and disease. *J. Cell Biol.* **217**, 51–63 (2018).
53. M. Rudan, D. Schneider, T. Warnecke, A. Krisko, RNA chaperones buffer deleterious mutations in E. coli. *eLife* **4**, 1–16 (2015).
54. I. Dikic, Proteasomal and autophagic degradation systems. *Annu. Rev. Biochem.* **86**, 31.3–31.32 (2017).
55. R. R. Kumar *et al.*, HSPIR: A manually annotated heat shock protein information resource. *Bioinformatics* **28**, 2853–2855 (2012).
56. S. R. Eddy, Accelerated profile HMM searches. *PLoS Comput. Biol.* **7**, e1002195 (2006).
57. D. Kim, B. Langmead, S. L. Salzberg, HISAT: A fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360 (2015).
58. E. P. Nawrocki, D. L. Kolbe, S. R. Eddy, Infernal 1.0: Inference of RNA alignments. *Bioinformatics* **25**, 1335–1337 (2009).
59. N. Lartillot, N. Rodrigue, D. Stubbs, J. Richer, PhyloBayes MPI: Phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. *Syst. Biol.* **62**, 611–615 (2013).
60. J. Huerta-Cepas, F. Serra, P. Bork, ETE3: Reconstruction, analysis, and visualization of phylogenomic data. *Mol. Biol. Evol.* **33**, 1635–1638 (2016).
61. H. G. Drost, A. Gabel, J. Liu, M. Quint, I. Grosse, myTAI: Evolutionary transcriptomics with R. *Bioinformatics* **34**, 1589–1590 (2018).
62. A. M. Fernandez-Escamilla, F. Rousseau, J. Schymkowitz, L. Serrano, Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat. Biotechnol.* **22**, 1302–1306 (2004).
63. J. Kyte, R. F. Doolittle, A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132 (1982).
64. B. Mészáros, G. Erdös, Z. Dosztányi, IUPred2A: Context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res.* **46**, W329–W337 (2018).