Evolution of resilience in protein interactomes across the tree of life

Marinka Zitnik,[1] Rok Sosič,[1] Marcus W. Feldman,[2] Jure Leskovec[1,3,*]

[1]Department of Computer Science, Stanford University, Stanford, CA, USA
[2]Department of Biology, Stanford University, Stanford, CA, USA
[3]Chan Zuckerberg Biohub, San Francisco, CA, USA
[*]To whom correspondence should be addressed; E-mail: jure@cs.stanford.edu

Phenotype robustness to environmental fluctuations is a common biological phenomenon. Although most phenotypes are governed by multiple proteins that interact with each other, the basic principles of how such interactome networks respond to environmental unpredictability and change during evolution are largely unknown. Here we study interactomes of 1,840 species across the tree of life involving a total of 8,762,166 protein-protein interactions. Our study focuses on the resilience of interactomes to protein failures and finds that interactomes become more resilient during evolution, indicating that a species' position in the tree of life is predictive of how robust its interactome is to protein failures. In bacteria, we find that a more resilient interactome is in turn associated with the greater ability of the organism to survive in a more complex, variable and competitive environment. We find that at the protein family level, proteins exhibit a coordinated rewiring of interactions over time and that a resilient interactome arises through gradual change of the network topology. Our findings have implications for understanding molecular network structure both in the context of evolution and environment.

Significance statement: The interactome network of protein-protein interactions captures the structure of molecular machinery and gives rise to a bewildering degree of life complexity. The resilience to protein failures is a critical property of the interactome as the breakdown of proteins may lead to cell death or disease. By studying interactomes from 1,840 species across the tree of life, we find that evolution leads to more resilient interactomes, providing evidence for a longstanding hypothesis that interactomes evolve favoring robustness against protein failures. We find that a highly resilient interactome has an astonishingly beneficial impact on the organism to survive in complex, variable, and competitive habitats. Our findings reveal how interactomes change through evolution and how these changes impact their response to environmental unpredictability.

The enormous diversity of life shows a fundamental ability of organisms to adapt their phenotypes to changing environments (*1*). Most phenotypes are the result of an interplay of many molecular components that interact with each other and the environment (*2–5*). The study of life's diversity has a long history and extensive phylogenetic studies have demonstrated evolution at the DNA sequence level (*6–8*). While studies based on sequence data alone have unraveled evolution of genomes, mechanistic insights into how evolution shapes interactions between proteins in an organism remain elusive (*9, 10*).

DNA sequence information has been used to associate genes with their functions (*11*), determine properties of ancestral life (*12, 13*), and understand how the environment affects genomes (*14*). Despite these advances in the understanding of DNA sequence evolution, little is known about basic principles that govern the evolution of interactions between proteins. In particular, evolution of DNA and amino acid sequences could lead to pervasive rewiring of protein-protein interactions and create or destroy the ability of the interactions to perform their biological function.

The importance of protein-protein interactions has spurred experimental efforts to map all

interactions between proteins in a particular organism, its interactome, namely the complex network of protein-protein interactions in that organism. A large number of high-throughput experiments have reported high-quality interactomes in a number of organisms (*15–19*). Because interactomes underlie all living organisms, it is critical to understand how these networks change during evolution and elucidate key principles of their design.

Here, we use protein interactions measured by these large-scale interactome mapping experiments and study the evolutionary dynamics of the interactomes across the tree of life. Our protein interaction dataset contains a total of 8,762,166 physical interactions between 1,450,633 proteins from 1,840 species, encompassing all current protein interaction information at a cross-species scale (SI Appendix, Section 1 and Table S4). We group these interactions by species and represent each species with a separate interactome network, in which nodes indicate a species' proteins and edges indicate experimentally documented physical interactions, including protein-protein and regulatory interactions, metabolic pathway interactions, and kinase-substrate interactions measured in that species. We integrate into the dataset the evolutionary history of species provided by the tree of life constructed from small subunit ribosomal RNA gene sequence information (*12*) (SI Appendix, Section 2). Using network science, we study the network organization of each interactome; in particular, its resilience to protein failures, a critical factor determining the function of the interactome (*20–23*). We identify the relationship between the resilience of an interactome and evolutionary time and use this resilience to uncover relationships with natural environments in which organisms live. Although the interactomes are incomplete and biased toward much-studied proteins and model species (SI Appendix, Section 1 and Fig. S7), our analyses give results that are consistent across taxonomic groups, and that are not sensitive to network data quality and network size change (SI Appendix, Section 8 and Fig. S8), and indicate that our conclusions will still hold when more protein interaction data become available.

Results

Modeling Resilience of Interactomes

Natural selection has influenced many features of living organisms, both at the level of individual genes (*24*) and at the level of whole organisms (*13*). To determine how natural selection influences the structure of interactomes, we study the resilience of interactomes to protein failures (*20, 22, 23*). Resilience is a critical property of an interactome as the breakdown of proteins can fundamentally affect the exchange of any biological information between proteins in a cell (Fig. 1a). Protein failure could occur through the removal of a protein (*e.g.*, by a nonsense mutation) or the disruption of a protein-protein interaction (*e.g.*, by environmental factors, such as availability of resources). The removal of even a small number of proteins can completely fragment the interactome and lead to cell death and disease (*4, 5*) (SI Appendix, Section S5.1 and Table S3). Failures of proteins and disruptions of interactions can thus affect the interactome to the extent that its connectivity can be completely lost and the interactome loses its biological function and increases the risk of disease (*5*).

We formally characterize the resilience of an interactome of a species by measuring how fragmented the interactome becomes when a fraction f of the proteins (nodes) are randomly removed from the interactome (Fig. 1a). The resulting isolated network clusters then determine the interactome fragmentation. A network cluster is defined as a connected component of the interactome in which any two nodes can reach each other by a path of edges. The smaller the network cluster, the fewer nodes can be reached from any given node in the cluster. To characterize how interactome fragments into clusters we use the Shannon's diversity index (Fig. 1b; SI Appendix, Fig. S1). If the fragmented interactome has one large cluster and only a few small broken-off clusters, then Shannon's diversity is low, providing evidence that the interactome has network structure that is resilient to protein failures (*20*) (SI Appendix, Fig. S2). In contrast, if the interactome breaks into many small clusters, it becomes fragmented, and its Shannon's di-

versity is high (SI Appendix, Fig. S2), indicating that the interactome is not resilient to protein failures.

To fully characterize the interactome resilience of a species we measure fragmentation of the species' interactome across all possible protein failure rates (SI Appendix, Fig. S3). Mathematically, we define resilience of each species' interactome using the Shannon's diversity index integrated over all possible protein failure rates f (SI Appendix, Section 5 and Fig. S3). Consider the interactome of the pathogenic bacterium *H. influenzae* and the interactome of humans, which have different resilience (Fig. 1c). In the *H. influenzae* interactome, on removing small fractions of all nodes many network clusters of varying sizes appear, producing a quickly increasing Shannon's diversity. In contrast, the human interactome fragments into a few small clusters and one large cluster whose size slowly decreases as small clusters break off, resulting in Shannon's diversity that increases linearly with the protein failure rate (Fig. 1c). Thus, unlike the fragmentation of the *H. influenzae* interactome, the human interactome stays together as a large cluster for very high values of protein failure rate, providing evidence for the topological stability of the interactome.

Resilience of Interactomes Throughout Evolution

We characterize systematically the resilience of interactomes for all species in the dataset (Fig. 1d and SI Appendix, Table S5). We find that species display varying degrees of interactome resilience to protein failures (Fig. 1e). At a global, cross-species scale, we find that a greater amount of genetic change is associated with a more resilient interactome structure (LOWESS fit; $R^2 = 0.36$; Fig. 1f), and this association remains strong even after statistical adjustments for the influence of many other variables (SI Appendix, Section 8). The more genetic change a species has undergone, the more resilient is its interactome. The evolutionary time of a species, which is defined by total branch length to the leaf taxon representing that species

in the tree of life (*12*), thus predicts resilience of the species' interactome, providing empirical evidence that interactome resilience is an evolvable property of organisms (*23*). It also suggests that the structure of present-day interactomes reflects their history or that interactomes must have a certain structure because that structure is well suited to the network's biological function. From an evolutionary standpoint, this suggests that evolutionary forces might shape protein interaction networks in such a way that their large-scale connectivity, *i.e.*, the network's biological function, remains largely unaffected by small protein failures as long the failures are random.

We also find that species from the same taxonomic domain have more similar interactome resilience than species from different domains ($p = 6 \cdot 10^{-11}$ for bacteria against eukaryotes; see SI Appendix, Fig. S10 for comparisons between other taxonomic groups). Furthermore, the degree of interactome resilience is significantly higher than expected by chance alone (SI Appendix, Fig. S9); that is, in a similar random network of identical size and degree distribution ($p = 5 \cdot 10^{-12}$), indicating that naturally occurring interactomes have higher resilience than their random counterparts. These findings are independent of genomic attributes of the species, such as genome size and the number of protein-coding genes, and are not direct effects of network size, the number of interactions in each species, broad-tailed degree distributions (*20*), or the presence of hubs in the interactome networks (SI Appendix, Fig. S8 and Table S1). Furthermore, these findings are consistent across a variety of assays that are used to measure the interactome (SI Appendix, Table S2).

Relationship between Interactome Resilience and Ecology

We next ask if there is a relationship between species' interactome resilience and aspects of species' ecology (SI Appendix, Section 4). We examine the relationship between interactome resilience and the fraction of regulatory genes and find that bacteria with more resilient in-

teractomes have significantly more regulatory genes in their genomes ($R^2 = 0.32$; Fig. 2a). Bacteria with highly resilient interactomes can also survive in more diverse and competitive environments, as revealed by exceptionally strong effects between the resilience and the level of co-habitation and the environmental scope (Fig. 2b-c). Furthermore, using a categorization of bacteria into five groups based on their natural environments (NCBI classification for bacterial lifestyle (terrestrial, multiple, host cell, aquatic, specialized) ordered by the complexity of each category (*25*)), we find that terrestrial bacteria living in the most complex ecological habitats have the most resilient interactomes ($p = 7 \cdot 10^{-3}$; Mann-Whitney U test), and that host-associated bacteria have the least resilient interactomes ($p = 4 \cdot 10^{-6}$; Fig. 2d). Our analysis further reveals that interactome resilience is indicative of oxygen dependence; the most resilient interactomes are those of aerobic bacteria ($p = 8 \cdot 10^{-4}$), followed by facultative and then anaerobic bacteria, which do not require oxygen for growth (Fig. 2d).

These relationships suggest that molecular mechanisms that render a species' interactome more resilient might also allow it to cope better with environmental challenges. In the network context, high interactome resilience suggests that proteins can interact with each other even in the face of high protein failure rate. High interactome resilience indicates that a species has a robust interactome, in which many mutations represent protein failures that are neutral in a given environment, have no phenotypic effect on the network's function and are thus invisible to natural selection (*23*). However, neutral mutations may not remain neutral indefinitely, and a once neutral mutation may have phenotypic effects in a changed environment and be important for evolutionary innovation (*22*). Although a large number of mutations in a resilient interactome might not change the network's primary function, they might alter other network features, which can drive future adaptations as the environment of the species changes (*26*). Changes that are neutral concerning one aspect of the network's function could lead to innovation in other aspects, suggesting that a resilient interactome can harbor a vast reservoir of neutral mutations.

Structural Changes of Protein Network Neighborhoods

A resilient interactome may arise through changes in the network structure of individual proteins over time (Fig. 3a). To investigate such changes in local protein neighborhoods, we decompose species interactomes into local protein networks, using a 2-hop subnetwork centered around each protein in a given species as a local representation of a protein's direct and nearby interactions in the species' interactome (SI Appendix, Section 6.1 and Fig. S4). We obtain 81,673 protein network neighborhoods and then use orthologous relationships between proteins to group them into 2,224 protein families, with an average of 38 protein neighborhoods originating from 12 species in each family (SI Appendix, Section 3). Each family represents a group of orthologous proteins that share a common ancestral protein (Fig. 3b).

By examining protein families, we find that the number of isolated network clusters in protein network neighborhoods and the effective size of the neighborhoods (Fig. 3b; SI Appendix, Section 6.2) both decrease with evolutionary time ($p = 3 \cdot 10^{-8}$ and $p = 0.03$, respectively; Fig. 3c-d), indicating that protein neighborhoods become more connected during evolution. These structural changes in the neighborhoods suggest a molecular network model of evolution (Fig. 3b): For orthologous proteins in two species, as the evolutionary distance between the species increases, the proteins' local network neighborhoods become increasingly different and the neighborhood becomes more interconnected in the species that has undergone more genetic change.

Network Rewiring of Protein-Protein Interactions

To study evolutionary mechanisms of structural changes in the interactomes, we investigate network motifs. We first identify orthologous protein pairs from evolutionarily close species (SI Appendix, Section 3), resulting in 2,485,564 protein pairs, which we then use to calculate interaction rewiring rates (IRR) for selected network motifs (Fig. 4a). We calculate the number

of times each motif appears in each protein neighborhood and derive the IRR by comparing the motif occurrences between the interactomes of the older and the younger species of each protein pair (SI Appendix, Section 7.1). We find strong statistical evidence that network motifs rewire during evolution ($p < 10^{-33}$ for all network motifs; Fig. 4b), suggesting that rewiring of interactions is an important mechanism for the evolution of interactomes. For example, proteins in evolutionarily older species on average participate in a factor of 0.861 fewer protein-protein interactions compared to proteins in evolutionarily younger species ($\mathrm{IRR} = -0.215$; Fig. 4b). This significant negative correlation between each protein's number of interactions and the protein's evolutionary age confirms earlier studies in the *S. cerevisiae* (*27*). We also find that square motifs of interactions become more common in protein neighborhoods during evolution ($\mathrm{IRR} = 0.016$; Fig. 4b). A range of biological evidence (*18, 28, 29*) supports this positive rate of change in the number of square motifs: From a structural perspective (*29*), protein-protein interactions often require complementary interfaces; hence two proteins with similar interfaces share many of their neighbors. However, they might not interact directly with each other, which manifests in the interactome as a square motif of interactions (see SI Appendix, Fig. S6 for an illustration of interaction interfaces recognizing the binding sites in proteins). Evolutionary arguments following gene duplication (*18*) reach the same conclusion; proteins with multiple shared interaction partners are likely to share even more partners and thereby produce new square motifs of interactions. To test the predictive power of our motif-based model of structural network changes, we estimate the size of the whole human interactome by extrapolating the *S. cerevisiae* interactome, using interaction rewiring rates from Fig. 4b (SI Appendix, Section 7.3). Assuming one splice isoform per gene, we predict the number of interactions in humans to be $\sim$160,000. This prediction is in surprisingly good agreement with three previous estimates of the size of the human interactome, which range from $150{,}000$ to $370{,}000$ interactions (*15, 17, 30*) and have proved crucial to establishing the complexity of the human interactome (*19*).

Discussion

Most evolutionary analyses of biological networks have focused on model organisms with high-coverage protein interaction data, such as *S. cerevisiae*, *M. musculus*, and humans. Interactomes mapped by unbiased tests of all possible pairwise combinations of proteins on the same platform remain scarce, an important limitation of the present study. In addition, experimentally documented protein interactions are currently subject to a high number of false positives and negatives. As more protein interaction data are collected, and more genomes become available, the generalizability of our findings can be further evaluated. However, our results are strikingly consistent across both different subsets of protein interaction data (SI Appendix, Table S2) and different phylogenetic lineages (SI Appendix, Fig. S10), and are not explained by a multitude of possible genomic and network confounders (SI Appendix, Section 8, Fig. S8, and Table S1), thus providing confidence that our key findings cannot be attributed to biases in the datasets.

Our study presents a new paradigm for evolutionary studies by demonstrating that interactomes reveal the fundamental design principles of molecular networks. We examine the interactomes of 1,840 species across the tree of life. Our findings highlight molecular evolution as an important predictor of structural network change and show that the evolutionary time of a species predicts resilience of the species' interactome to protein failures. They offer quantitative evidence for the biological proposition that an organism that has undergone more genetic change has a more resilient interactome, which, in turn, is associated with the greater ability of the organism to survive in a more complex, variable and competitive environment. Our findings can also help clarify the mechanisms of how interactomes change during evolution, why currently observed network structures exist and how they may change in the future, and facilitate the extrapolation of functional information from experimentally characterized proteins to their orthologous proteins in poorly studied organisms.

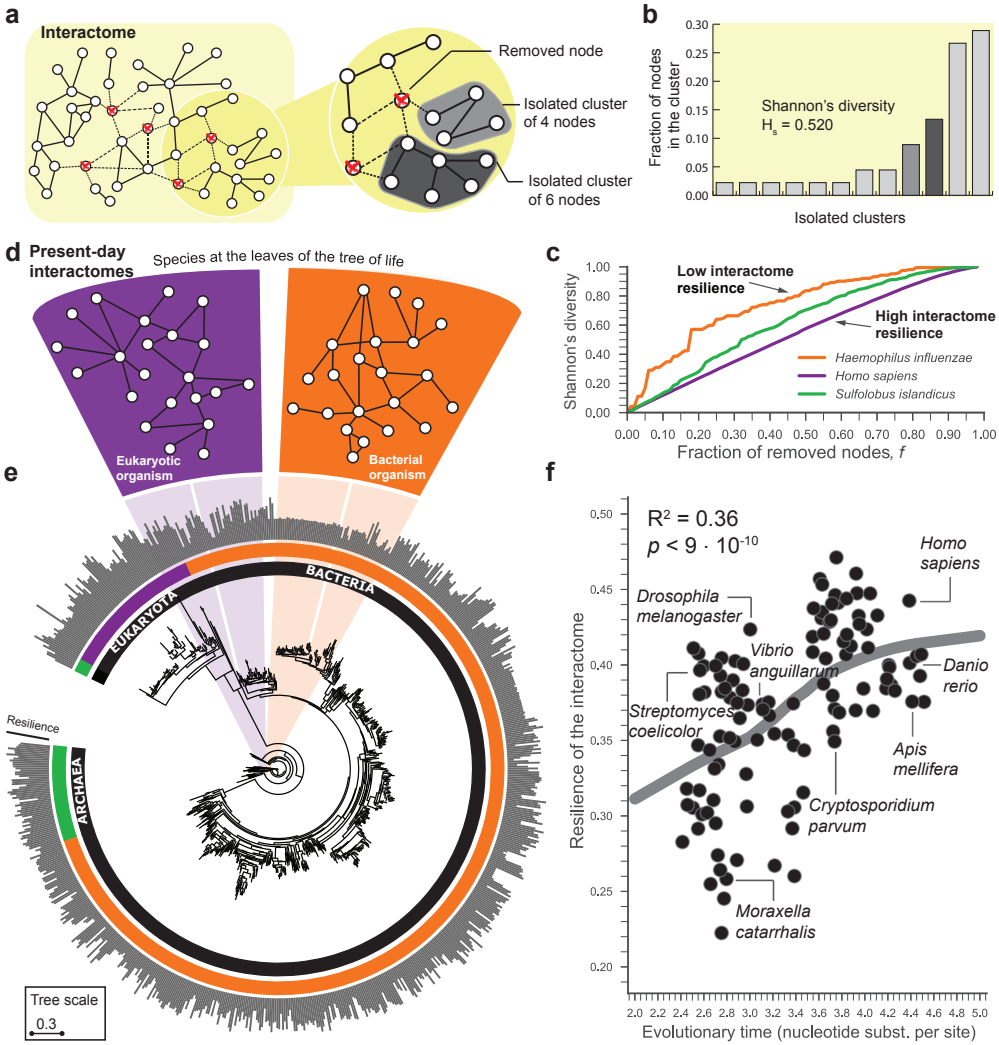

Figure 1: Protein interaction data of 1,840 species consisting of 8,762,166 interactions by 1,450,633 proteins reveal the resilience of interactomes across vast evolutionary distances. (**a**) The interactome of an organism consists of all physical interactions between proteins in the organism. When a certain fraction ($f = 5/45$ in this example) of the proteins are removed from the interactome, the interactome fragments into a number of isolated network clusters. (**b**) Shannon's diversity index (SI Appendix, Section 5) measures how the interactome fragments into clusters at a given failure rate f. (**c**) The resilience of the interactome integrates Shannon's diversity index across all possible failure rates f (SI Appendix, Section 5). Resilience value 1 indicates the most resilient interactome, and resilience value 0 indicates a complete loss of the connectivity of the interactome (SI Appendix, Fig. S3). *H. sapiens* (*H. influenzae*) has the most (least) resilient interactome (their resilience is 0.461 and 0.267, respectively) among the three selected organisms. (**d**) A small neighborhood of the interactome in a eukaryotic and a bacterial species. As ancestral species have gone extinct, older interactomes have been lost, and only present-day interactomes are available for analysis. (**e**) Phylogenetic tree showing 1,539 bacteria, 111 archaea, and 190 eukarya (*12*). Evolutionary time of a species is defined as the total branch length (nucleotide substitutions per site) from the root to the corresponding leaf in the tree (SI Appendix, Section 2). The outside circle of bars shows the interactome resilience of every species. Current protein-protein interaction data might be prone to notable selection and investigative biases (SI Appendix, Section 1); the subsequent plot shows the interactome resilience for 171 species with at least 1,000 publications in the NCBI Pubmed (SI Appendix, Fig. S7). (**f**) Across all species, evolutionary time of a species predicts resilience of the species' interactome to protein failures (LOWESS fit; $R^2 = 0.36$); more genetic change implies a more resilient interactome. Three species with the largest evolutionary time (far right on the x-axis) have on average 20.4% more resilient interactome than the three species with the shortest evolutionary time (far left on the x-axis).

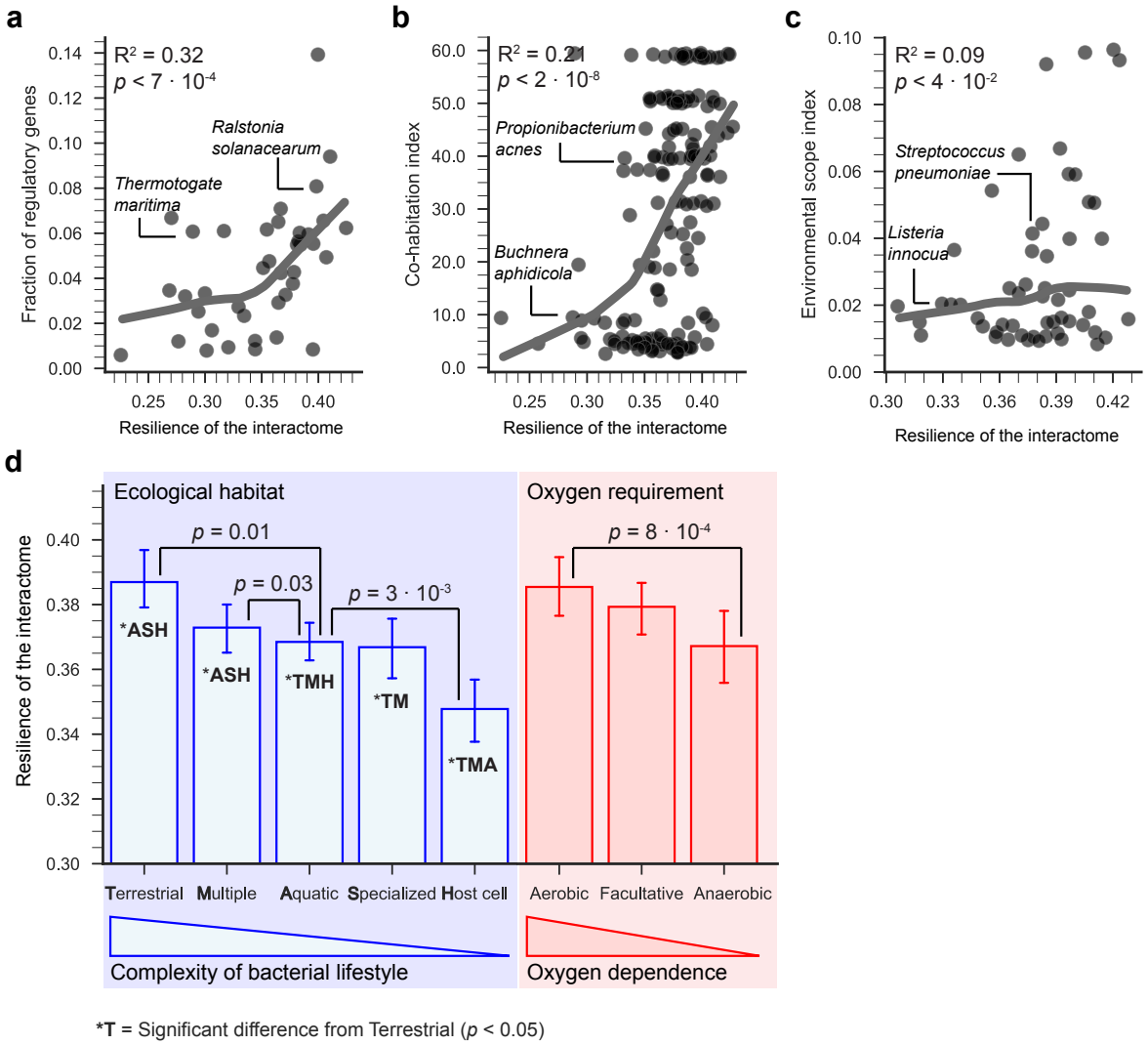

Figure 2: Bacteria with more resilient interactomes survive in more complex, variable and competitive environments. We use ecological information for 287 bacterial species (*25*) to examine the relationship between species' interactome resilience and their ecology (SI Appendix, Section 4). (**a**) Interactome resilience positively correlates with the fraction of regulatory genes in bacteria, an established indicator of environmental variability of species' habitats (*25*) ($R^2 = 0.32$). (**b** and **c**) For environmental viability of a species, we use a co-habitation index that records how many organisms populate each environment in which the species is viable (*i.e.*, the level of competition in each viable environment), and an environmental scope index that records a fraction of the environments in which the species is viable (*i.e.*, species' environmental diversity) (*25*). The resilience of the interactome positively correlates with the level of co-habitation encountered by bacteria ($R^2 = 0.21$), and bacteria with resilient interactomes tend to thrive in highly diverse environments ($R^2 = 0.09$). (**d**) Terrestrial bacteria have the most resilient interactomes ($p = 7 \cdot 10^{-3}$), and host-associated bacteria have the least resilient interactomes ($p = 4 \cdot 10^{-6}$). In bacteria, interactome resilience is indicative of oxygen dependence. Aerobic bacteria have the most resilient interactomes ($p = 8 \cdot 10^{-4}$), followed by facultative and the anaerobic bacteria. Error bars indicate 95% bootstrap confidence interval; p values denote the significance of the difference of the means according to a Mann-Whitney U test.

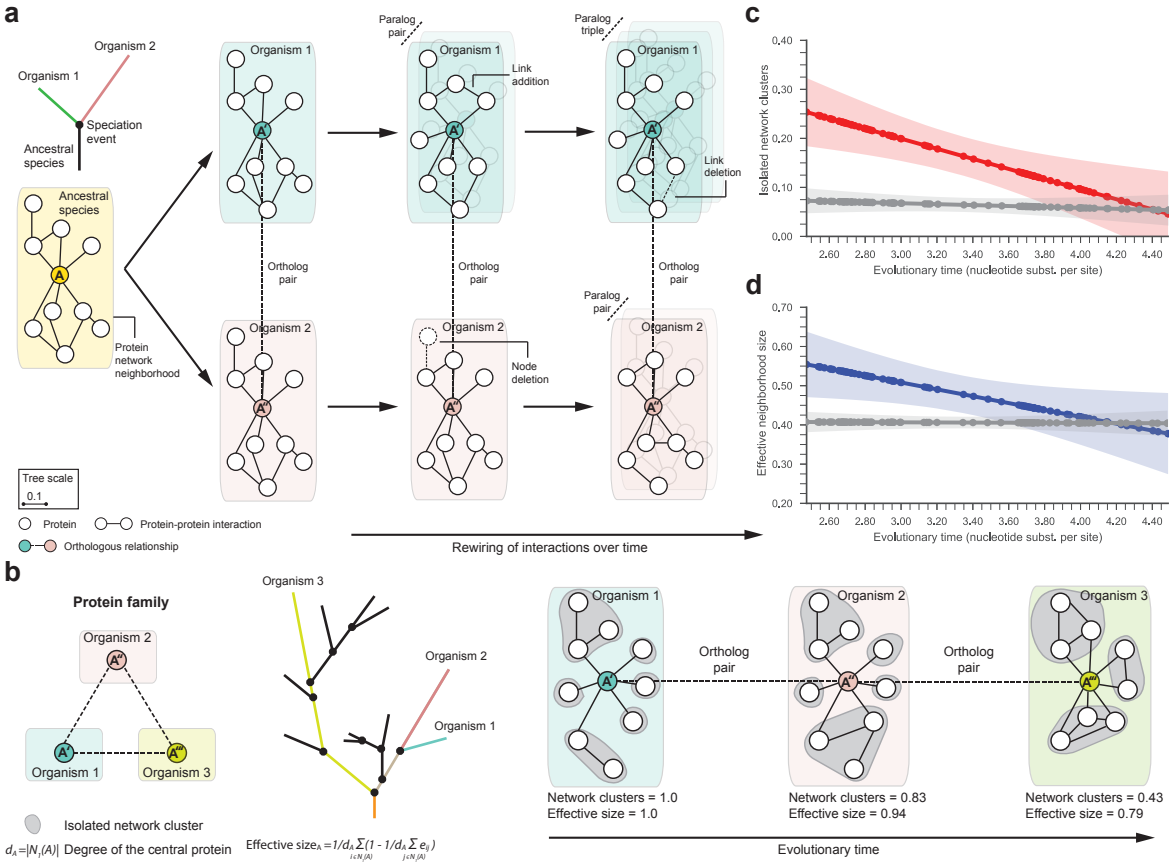

Figure 3: Evolutionary forces may mitigate network structural changes of local protein neighborhoods. (**a**) A hypothetical phylogenetic tree illustrates a speciation event that gives rise to two lineages according to the speciation-divergence model (*11*) and leads to present-day organisms '1' (green) and '2' (pink). In this example, a single ancestral protein A that was present in the ancestral species gives rise to proteins A' and A'' upon speciation; A' and A'' form an orthologous protein pair. As the two newly arising species diverge and protein sequences evolve, protein network neighborhoods (SI Appendix, Fig. S4) in their interactomes can rewire independently over time. Shown are also in-paralogs, proteins which arise through gene duplication events in species '1' and '2' after speciation. (**b**) A hypothetical protein family with three protein members (A', A'', A'''), each from a different organism. In the phylogenetic tree, organism '1' is located at the tip of the lineage with the shortest evolutionary time, whereas organism '3' is in the lineage with the longest evolutionary time. We represent the protein family by a sequence of orthologous proteins ordered by the evolutionary time of proteins' originating species (SI Appendix, Section 3). We then characterize the network neighborhood of each protein in the family by calculating two network metrics (SI Appendix, Fig. S5). Isolated clusters are given by the degree-adjusted number of connected components in the neighborhood that arise when the central protein is removed from the interactome (gray) (SI Appendix, Section 6). The neighborhood size down-weighted by the redundancy of local interactions gives the effective size of the neighborhood (SI Appendix, Section 6). (**c** and **d**) The number of isolated network clusters and the effective size of protein neighborhoods both decrease with evolutionary time ($p = 3 \cdot 10^{-8}$ and $p = 0.03$, respectively; Spearman's ρ rank correlation), suggesting that local interaction neighborhoods rewire via a coordinated evolutionary mechanism. Lines in **c** and **d** show the LOWESS fit of median-aggregated network metric values for 81,673 proteins from 2,224 protein families; color bands indicate 95% confidence band for the LOWESS fit; gray lines show random expectation.

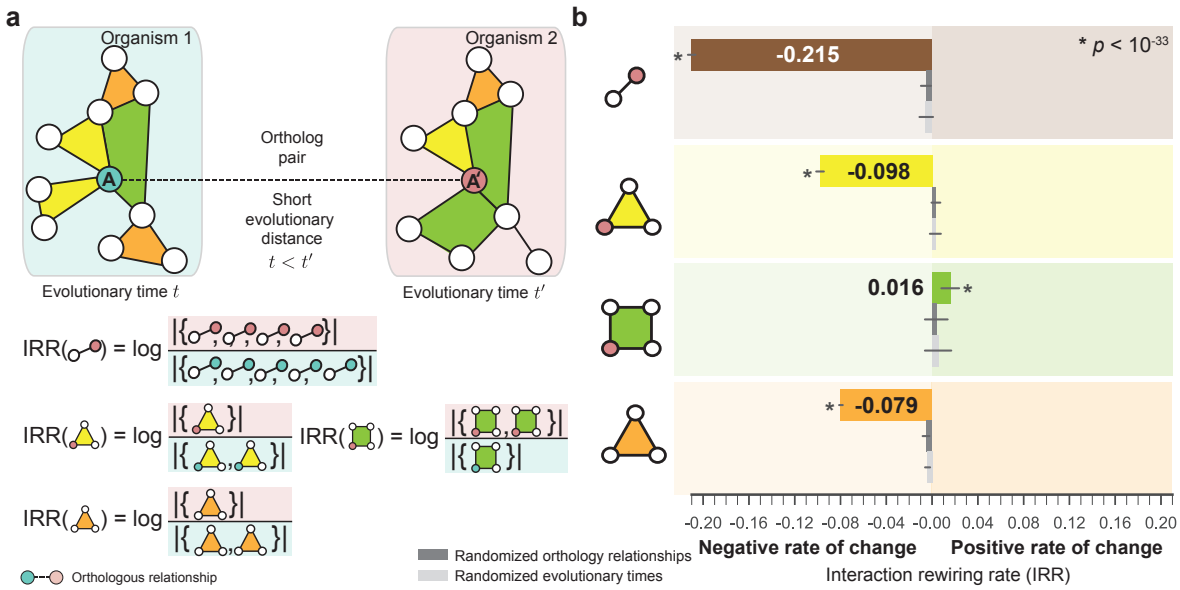

Figure 4: The rewiring rate of interactions in local protein neighborhoods varies with the topology of network motifs. (**a**) Interaction rewiring rate IRR measures the fold change between the probability of observing a particular network motif in the network neighborhood of protein A' relative to the probability of observing the same motif in the neighborhood of an evolutionarily younger orthologous protein A. A positive (negative) rate of change indicates the motif becomes more (less) common over time (SI Appendix, Section 7). Shown are the rewiring rates for interactions (*i.e.*, edges; the number of interactors of A' versus A), triangle motifs touching the orthologous protein (yellow), square motifs touching the orthologous protein (green), and triangle motifs in the protein network neighborhood (orange). (**b**) Square motifs become more common in protein neighborhoods during evolution ($p < 10^{-33}$), which is supported by a range of biological evidence (*18, 28, 29*). However, triangle motifs become less common over time ($p < 10^{-33}$ for both types of triangle motifs). Gray bars indicate random expectation (SI Appendix, Section 7), either for random orthology relationships (dark gray) or for random evolutionary times (light gray); error bars indicate 95% bootstrap confidence interval; p values denote the significance of the difference of IRR distributions using a two-sample Kolmogorov-Smirnov test.

Code and data availability

Software implementation of statistical methodology together with documentation and examples of usage will be made publicly available through snap.stanford.edu website. We will share all data used in the paper, including the processed interactomes, with the community with this publication.

Author contributions

M.Z. performed the statistical analysis. M.Z., R.S., M.W.F., and J.L. jointly analyzed the results and wrote the paper.

Author information

The authors declare no conflict of interest. Correspondence and requests for materials should be addressed to J.L. (jure@cs.stanford.edu).

References

1. C. J. Castelle, J. F. Banfield, *Cell* **172**, 1181 (2018).
2. A.-L. Barabási, Z. N. Oltvai, *Nature Reviews Genetics* **5**, 101 (2004).
3. J. X. Hu, C. E. Thomas, S. Brunak, *Nature Reviews Genetics* **17**, 615 (2016).
4. E. L. Huttlin, *et al.*, *Nature* **545**, 505 (2017).
5. S. Chen, *et al.*, *Nature Genetics* **50**, 1032 (2018).
6. R. J. Britten, *Science* **231**, 1393 (1986).
7. Z. Yang, B. Rannala, *Nature Reviews Genetics* **13**, 303 (2012).
8. S. Marsit, *et al.*, *Nature Reviews Genetics* **18**, 581 (2017).
9. R. A. Studer, *et al.*, *Science* **354**, 229 (2016).
10. T. R. Sorrells, L. N. Booth, B. B. Tuch, A. D. Johnson, *Nature* **523**, 361 (2015).
11. R. L. Tatusov, E. V. Koonin, D. J. Lipman, *Science* **278**, 631 (1997).
12. L. A. Hug, *et al.*, *Nature Microbiology* **1**, 16048 (2016).
13. D. H. Parks, *et al.*, *Nature Microbiology* **2**, 1533 (2017).
14. S. Mukherjee, *et al.*, *Nature Biotechnology* **35**, 676 (2017).
15. J.-F. Rual, *et al.*, *Nature* **437**, 1173 (2005).

16. A. Roguev, *et al.*, *Science* **322**, 405 (2008).

17. K. Venkatesan, *et al.*, *Nature Methods* **6**, 83 (2009).

18. Arabidopsis Interactome Mapping Consortium, *et al.*, *Science* **333**, 601 (2011).

19. T. Rolland, *et al.*, *Cell* **159**, 1212 (2014).

20. R. Albert, H. Jeong, A.-L. Barabási, *Nature* **406**, 378 (2000).

21. A. Wagner, *Science Signaling* **2003**, pe41 (2003).

22. A. Wagner, *FEBS Letters* **579**, 1772 (2005).

23. A. Wagner, *Robustness and Evolvability in Living Systems* (Princeton University Press, 2013).

24. T. V. Vo, *et al.*, *Cell* **164**, 310 (2016).

25. S. Freilich, *et al.*, *Genome Biology* **10**, R61 (2009).

26. A. Barve, A. Wagner, *Nature* **500**, 203 (2013).

27. H. B. Fraser, *et al.*, *Science* **296**, 750 (2002).

28. I. Ispolatov, A. Yuryev, I. Mazo, S. Maslov, *Nucleic Acids Research* **33**, 3629 (2005).

29. O. Keskin, N. Tuncbag, A. Gursoy, *Chemical Reviews* **116**, 4884 (2016).

30. G. T. Hart, A. K. Ramani, E. M. Marcotte, *Genome Biology* **7**, 120 (2006).