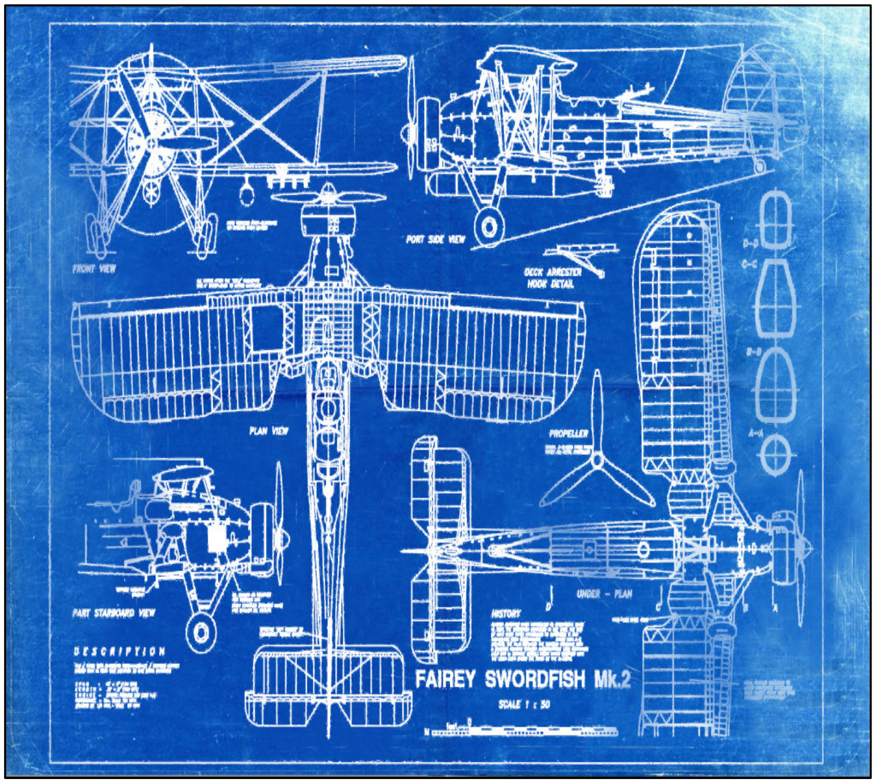
Evolution of resilience in protein interactomes across the tree of life

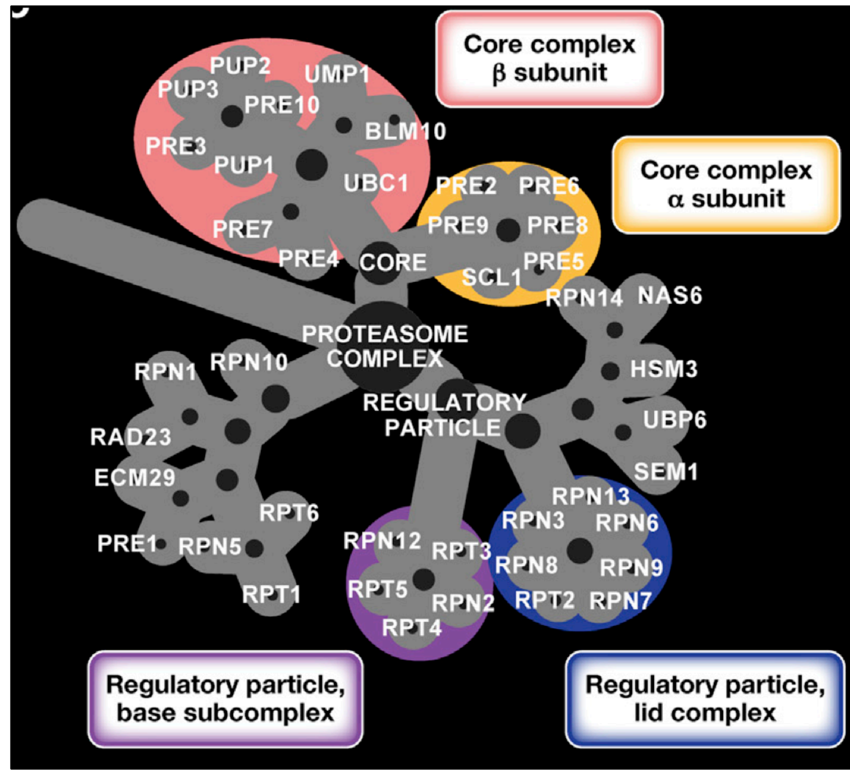
Marinka Zitnik, Rok Sosic,
Marcus Feldman, Jure Leskovec

Stanford University

Protein interaction network: Backbone of activity in a cell



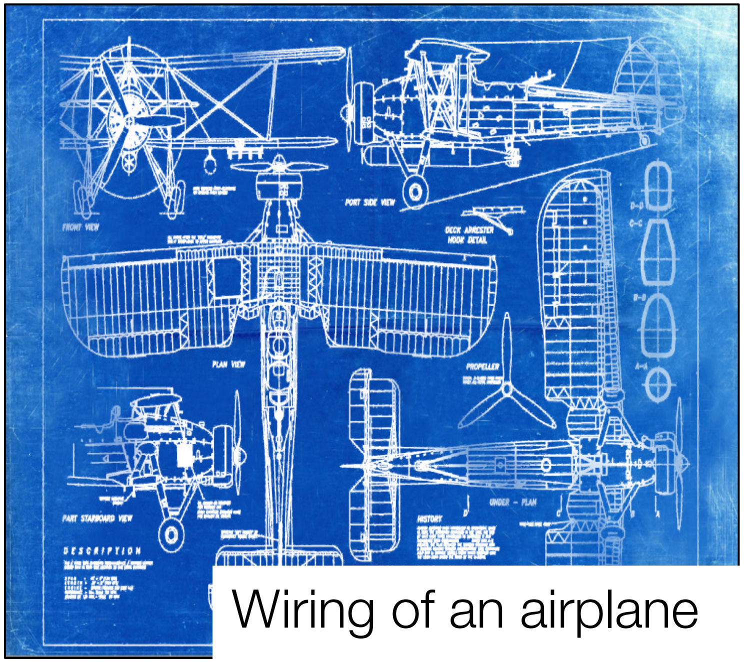
Physical interactions between an airplane's parts



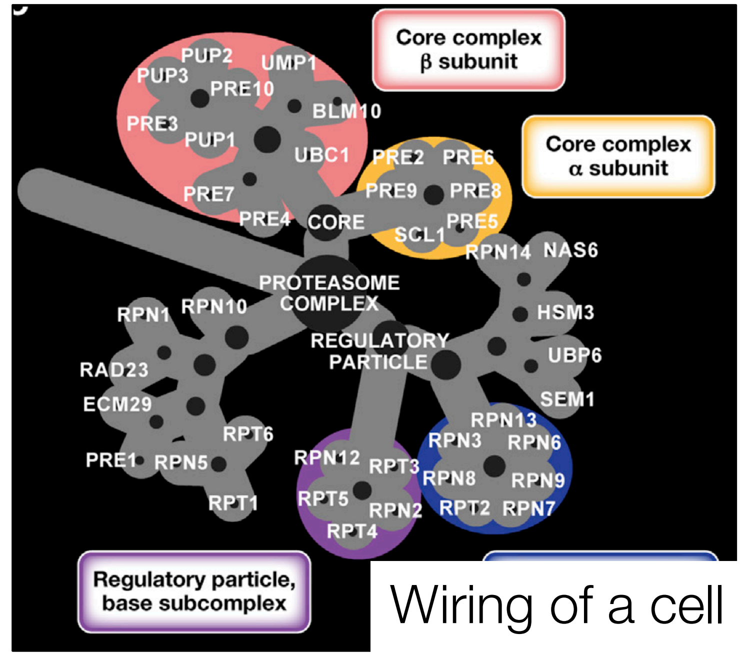
Carvunis & Ideker, Cell'14

Physical interactions between a cell's molecular components

How do protein networks evolve?



Wiring of an airplane



Wiring of a cell

Carvunis & Ideker, Cell'14

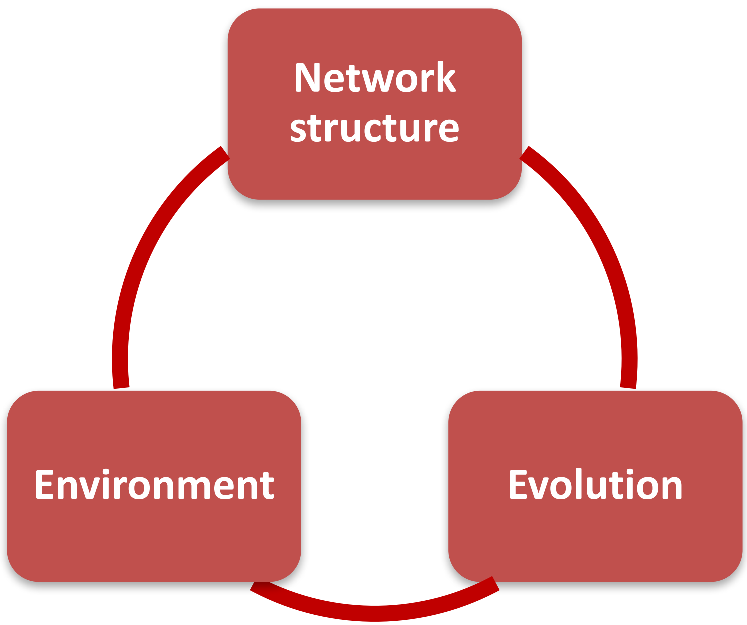
But we do not know how networks change with evolution!

- *Whether or not natural selection shapes the evolution of protein-protein interaction networks remains unclear* [Nature'15, '16, '17]
- *Whether network rewiring is a consequence of sequence divergence or a driver of evolution remains an open question* [Science'17]

Today's Talk

1) How **protein-protein interaction networks** change with evolution?

2) How **network changes** affect phenotypes and species' survival in **natural habitats**?



Why is modeling network evolution hard?

Massive time span and rare data samples

- Species separated by millions of years of evolution

Messy, incomplete network data

- Lack of high-coverage protein interaction data, e.g.,
  - humans: 20 thousand genes → need to test ~200 million protein pairs for interaction
  - <30% of human protein pairs tested in last 20 years [Rolland et al., Cell'14]

Many possible confounders

- Genome size, number of protein-coding genes, etc.
- Network size, degree distributions, presence of hub nodes, etc.
- Investigative biases towards model organisms

Our Approach

1. Build a dataset by integrating evolutionary, interactome, and ecological data

2. Use dataset to study evolution of interactomes:
   - How protein-protein interaction networks change with evolution?
   - How changes in these networks affect phenotypes and species' survival in natural environments?

Key Element: Evolutionary Dataset

Objective: Capture all documented protein-protein interactions across all species

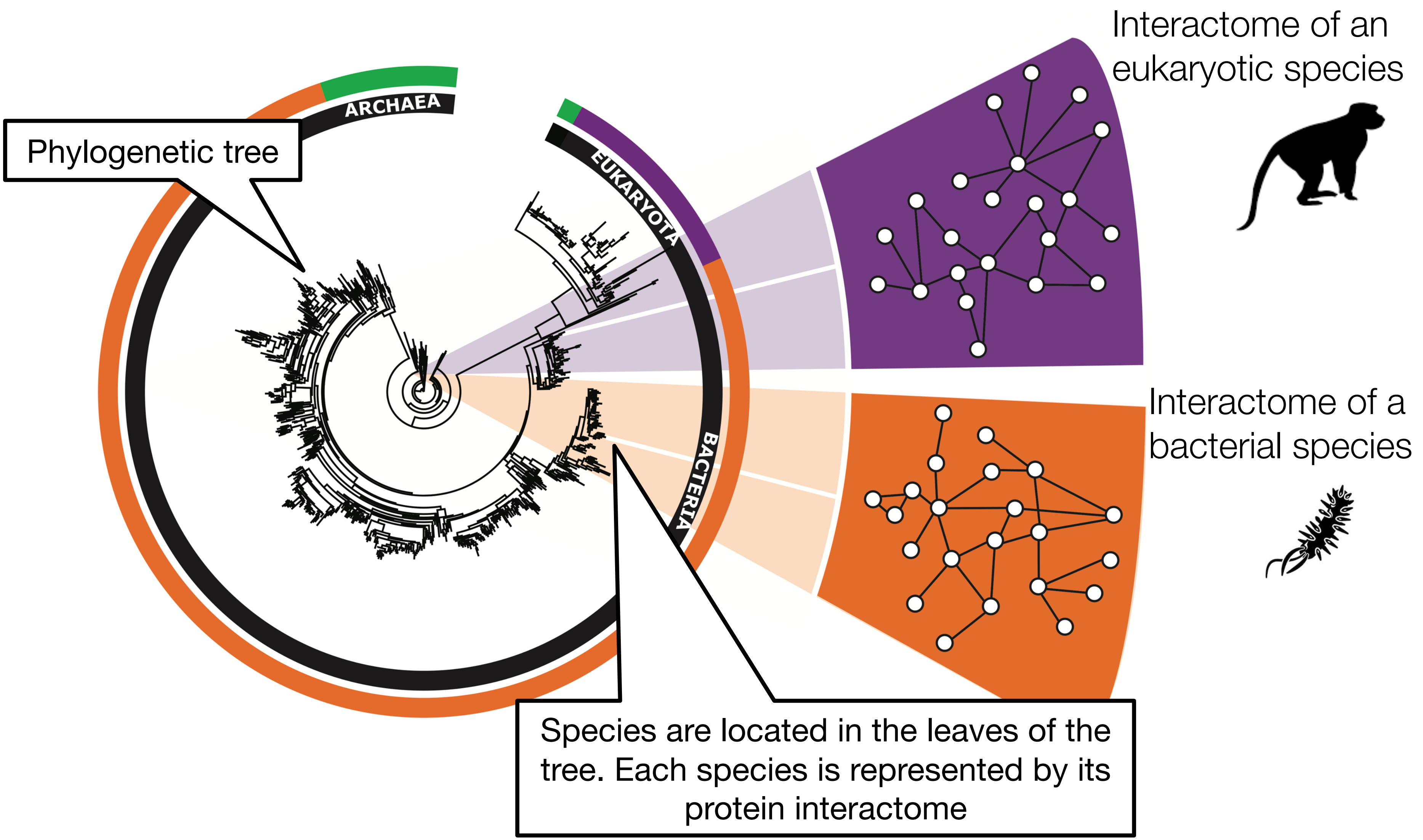
We build a unique dataset:

- 1,840 networks: 1,539 bacteria, 111 archaea, 190 eukarya
- 1,450,633 nodes: Species' proteins
- 8,762,166 edges: Physical protein-protein interactions (PPIs)

- Protein interactome: Species represented by their PPI networks
- Tree of life: Evolutionary history of species
- Ecology: Complexity of habitats in which species live

>300X larger dataset than previous studies

Tree of Networks



Modeling Tasks

- **Data:** Tree of networks
- **Two main tasks:**
  1) Characterize resilience of interactomes to network failures
  2) Identify network and evolutionary mechanisms of resilience

Why resilience?

- **Resilience** to network failures is critical:
  - Breakdown of proteins affects the exchange of biological information in the cell [Huttlin et al., Nature'17]
  - Failures lead to cell death and disease [Chen et al., Nat. Genet.'18]

How to characterize resilience to network failures?

Define **interactome resilience** measure:

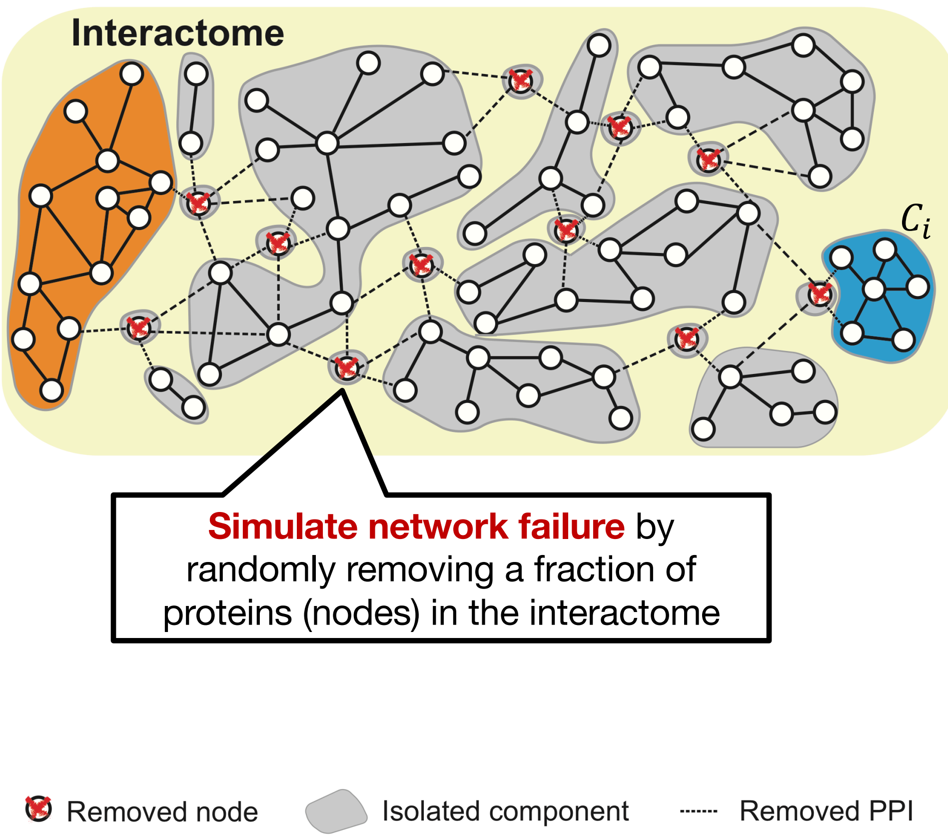
- Information-theoretic formulation
- Shannon diversity theory [Sheldon'69]

Resilience measure has three key elements:

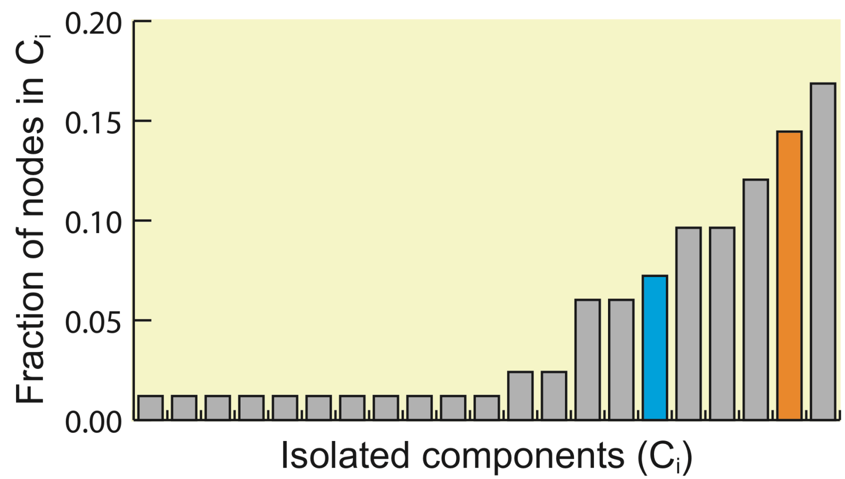
1. Simulate network failure at a particular rate
2. Measure how fragmented the interactome becomes
3. Repeat 1-2 across all possible failure rates

Simulate failure and measure fragmentation of the interactome

Upon network failure, interactome fragments into isolated components. **Entropy of component sizes!**



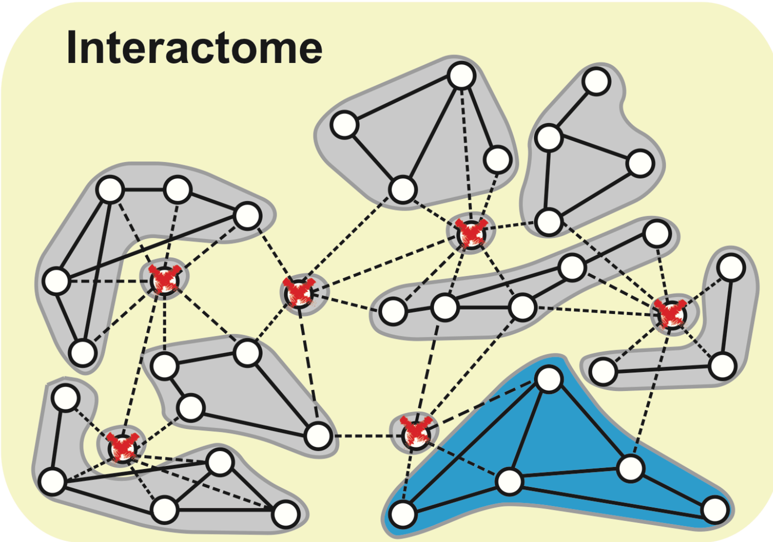
Simulate network failure by randomly removing a fraction of proteins (nodes) in the interactome

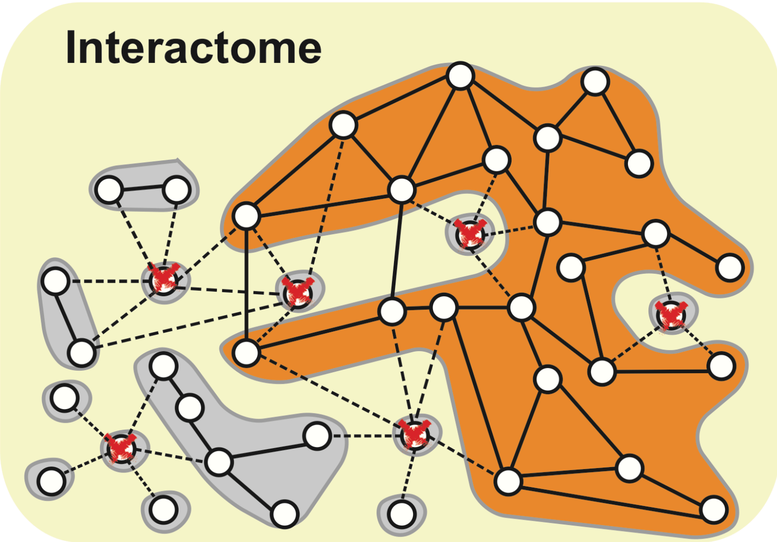


$$H = -\sum_{C_i} \frac{C_i}{N} \log \frac{C_i}{N}$$

C_i/N is fraction of all nodes N in isolated component C_i

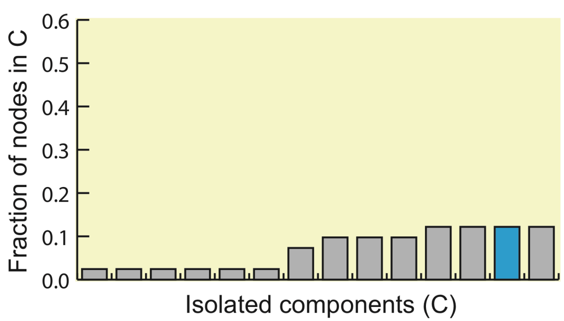
Fragmentation: Example





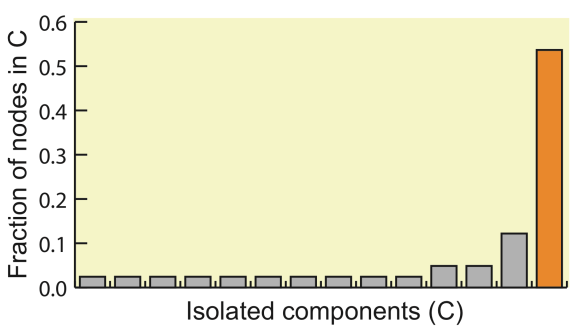
High entropy

Many small isolated components, all of approximately the same size

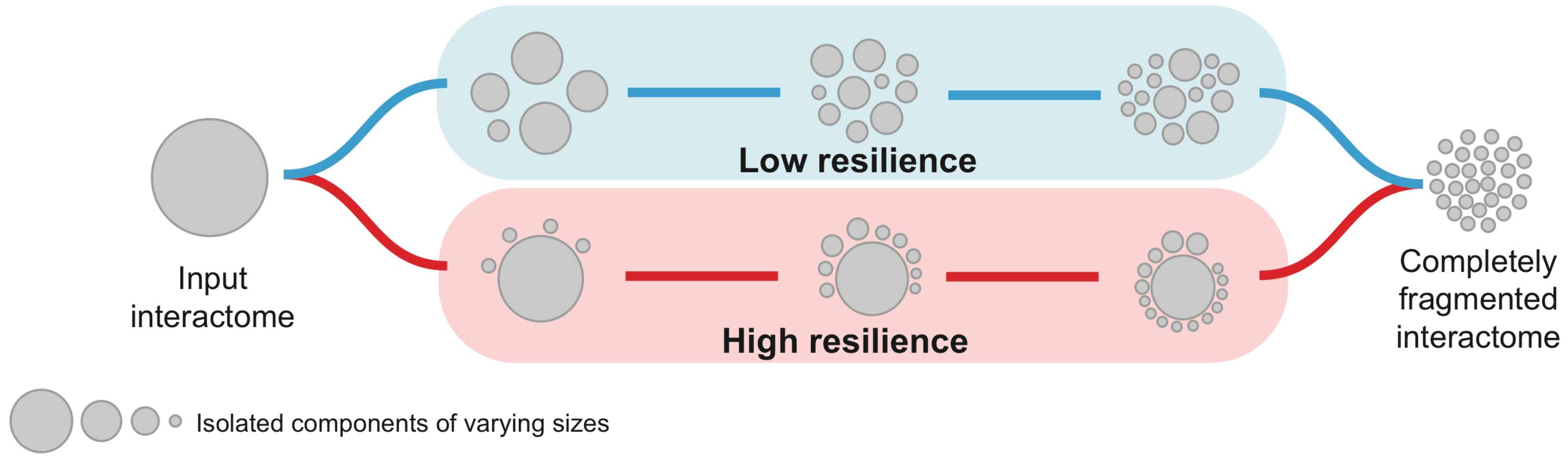


Low entropy

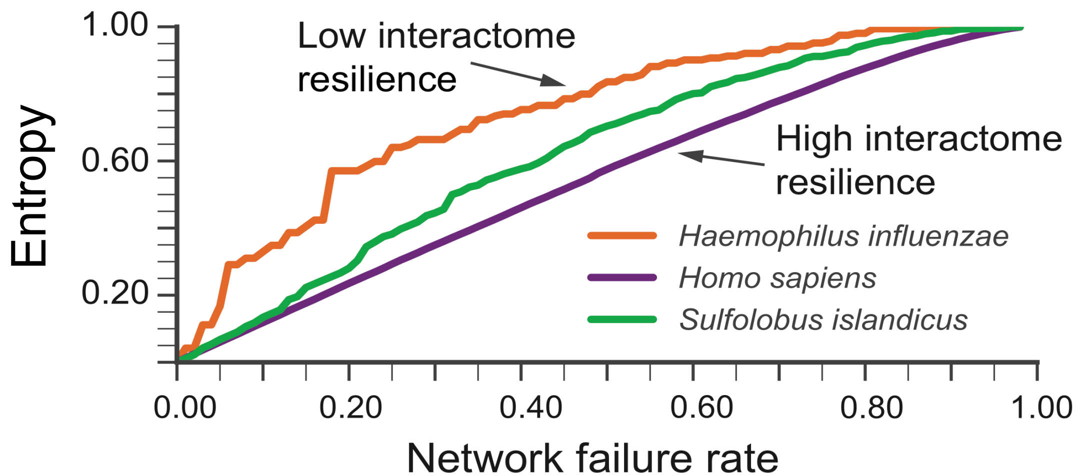
Large isolated component, only a few small broken-off components



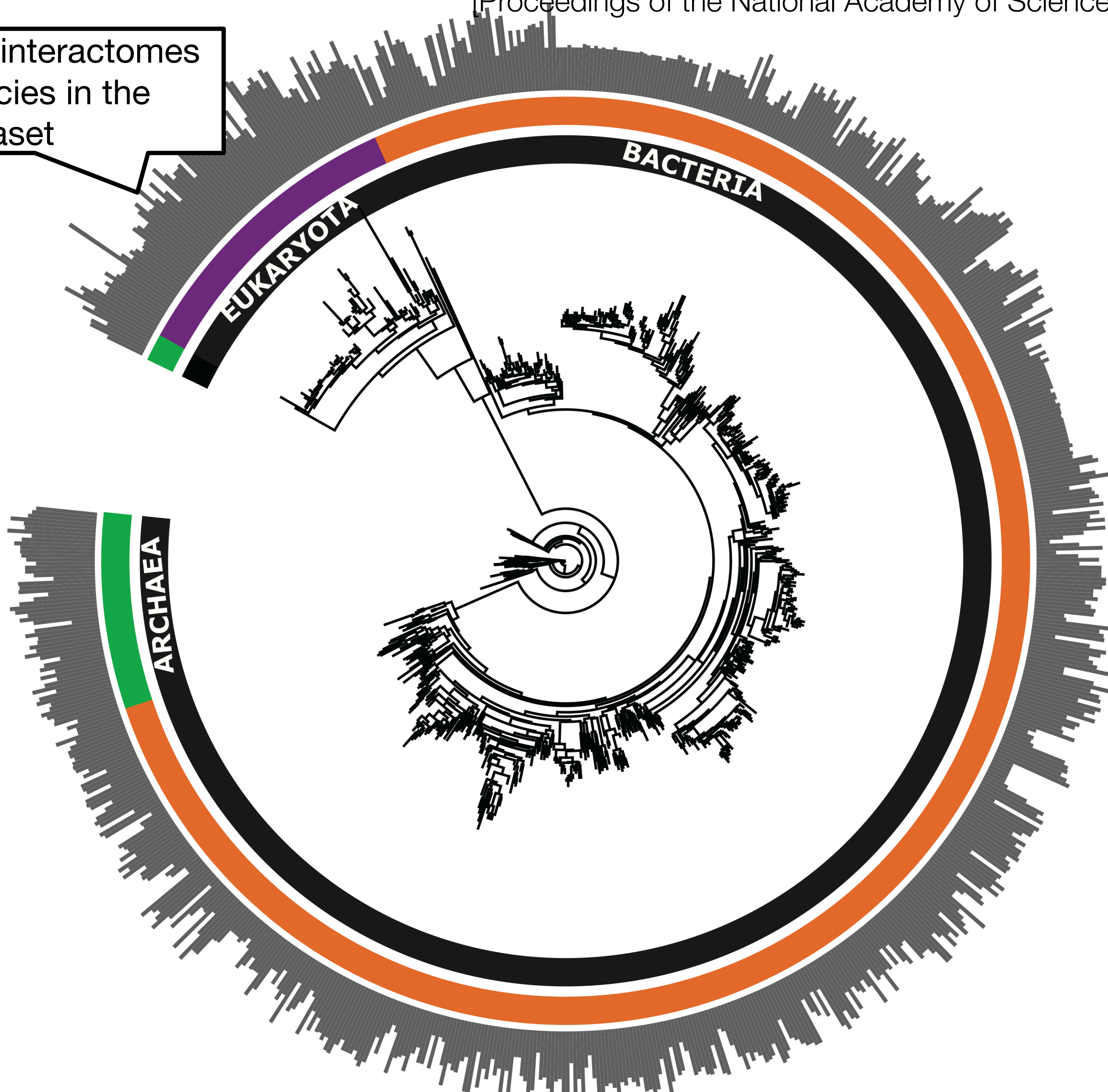
Resilience: Fragmentation integrated across all possible failure



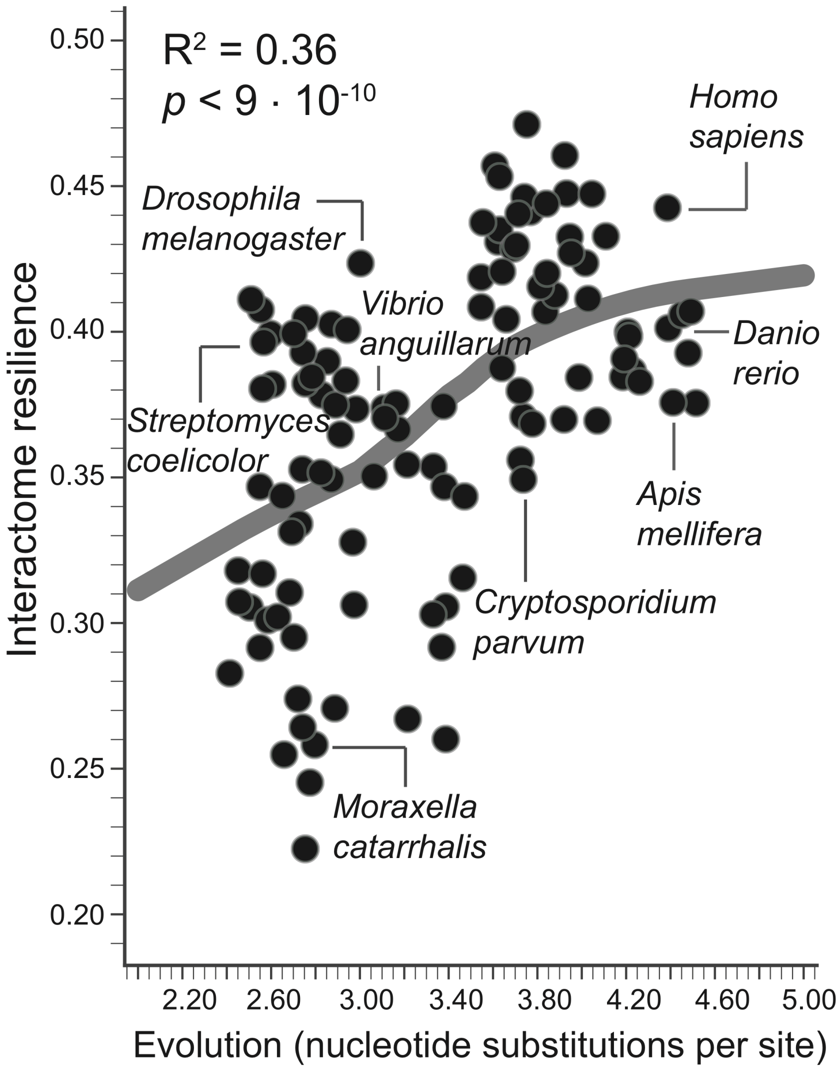
E.g., resilience for three species:



Resilience of interactomes for all species in the dataset



Evolution leads to resilience



Protein interactomes become **more resilient with evolution**

More genetic change a species has undergone, **more resilient** is its interactome

Protein interactomes become **more resilient to network failures** over time

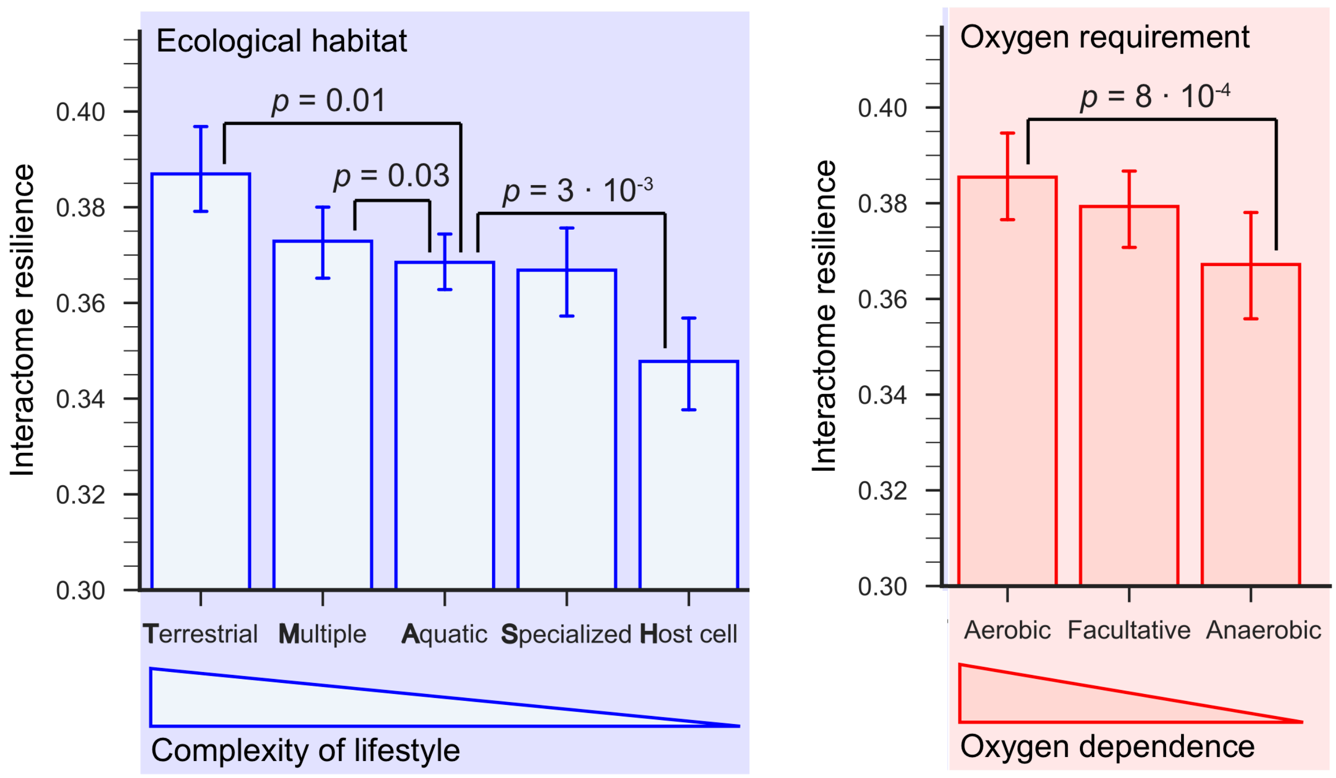
[Proceedings of the National Academy of Sciences (PNAS)'19]

Findings are not due to data biases

- Consistent results across taxonomic groups
- Robust to network data quality and network size
- Consistent findings across biological assays
- Findings are not due to confounding:
  - Genomic attributes, e.g., genome size, protein-coding genes
  - Network properties, e.g., hub nodes, broad-tailed degree distributions, number of interactions in each species
  - Bias toward much-studied proteins and model species

Key findings will still hold when more protein interaction data become available

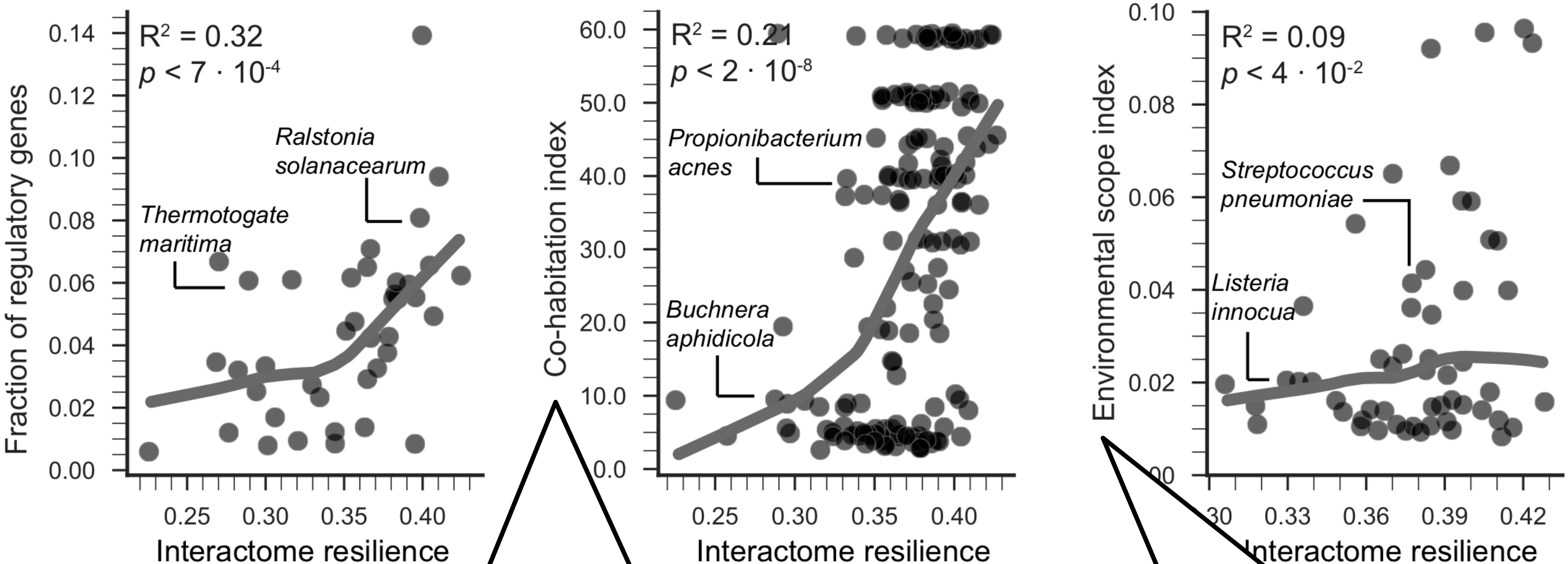
Resilience is beneficial



Organisms with **more resilient interactome** can survive in **more complex, diverse**, and **competitive habitats**

E.g., Terrestrial habitat + Oxygen → Highly resilient interactome

Resilience is beneficial

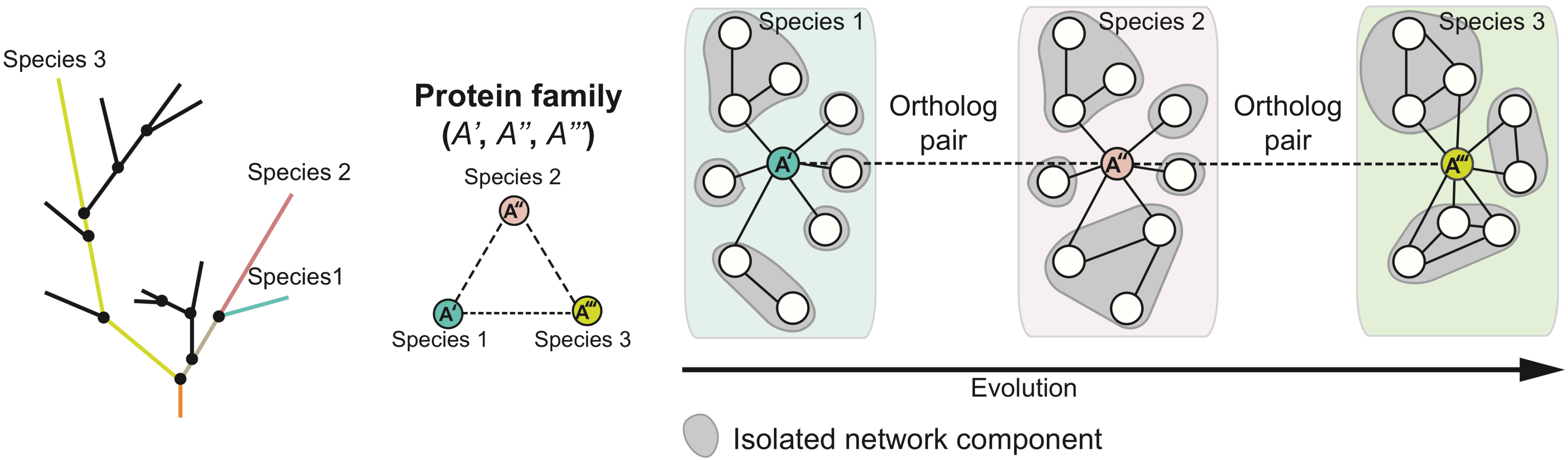


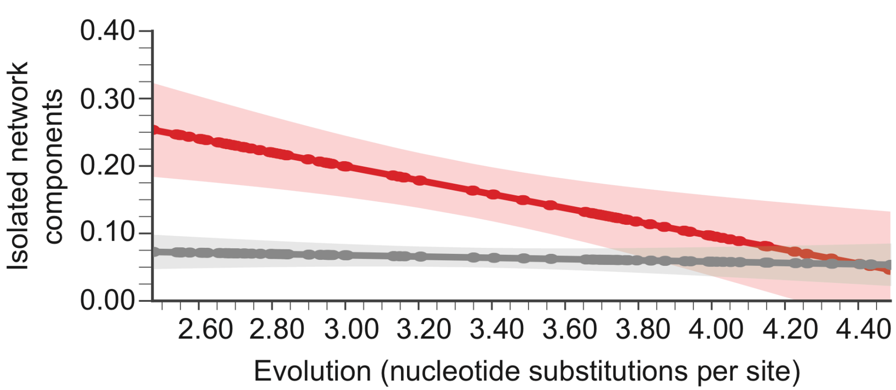
Level of competition in environments in which species live

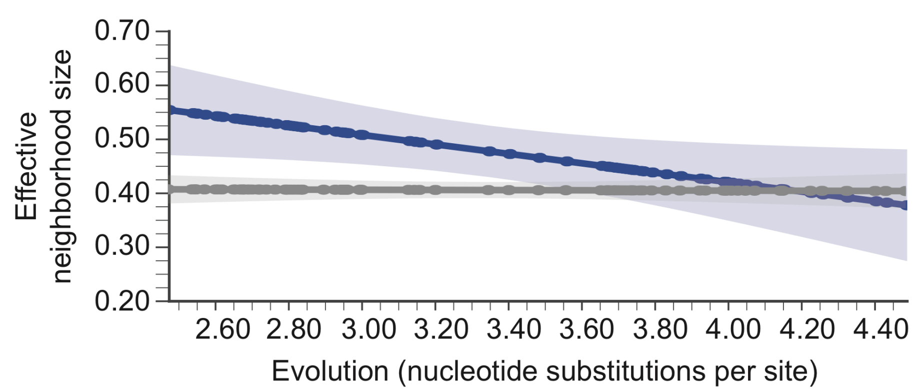
Diversity of environments in which species live

Organisms with **more resilient interactome** survive in **more complex, diverse**, and **competitive habitats**

Resilience arises through gradual change of network topology

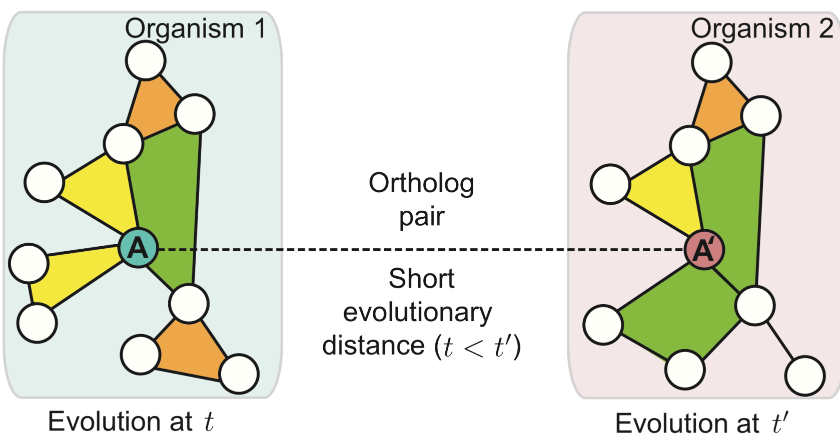




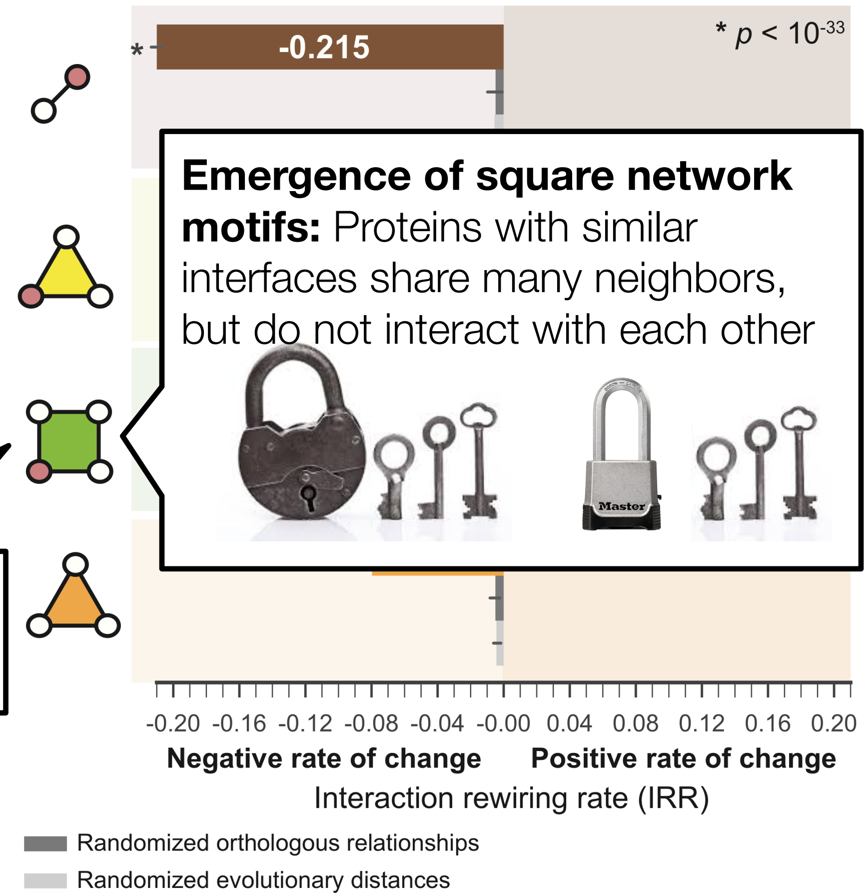


Network Mechanism of Resilience

Rewiring of protein-protein interactions in local protein neighborhoods



Square network motifs become more common with evolution



Protein Protein-protein interaction

Orthologous relationship

Key New Insights

Resilient interactome: Proteins able to interact in the face of network failures:

- Failures/changes are neutral in the current environment
- Neutral changes do not remain neutral indefinitely
- Crucial for survival in a changed environment

Resilient interactome is a **reservoir** that drives future evolution

Implications for ecology, network biology, design of robust systems