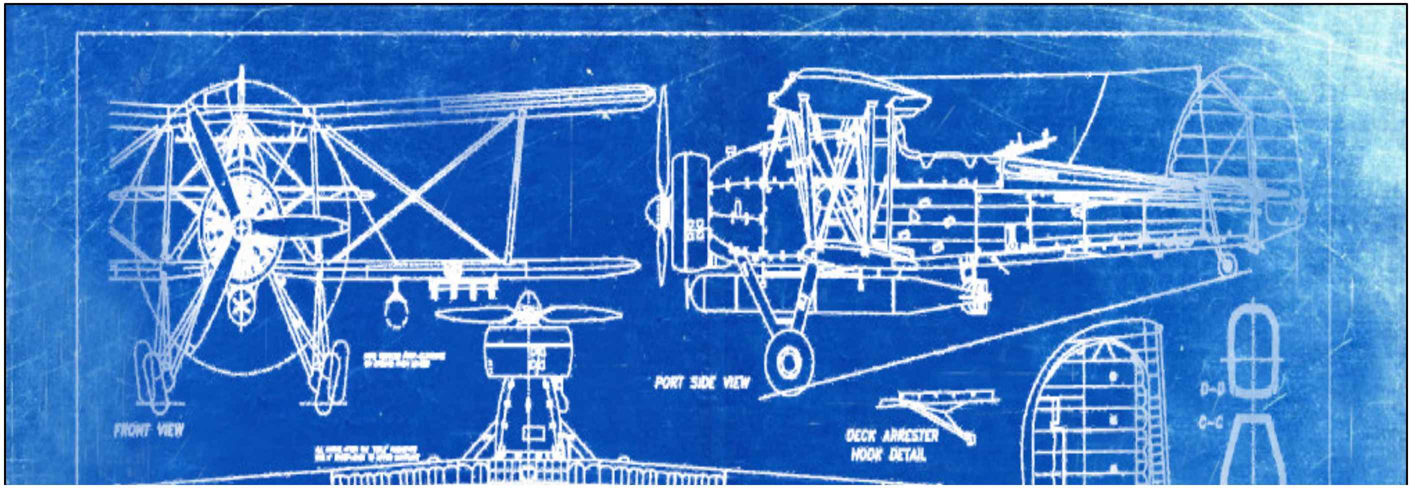
Evolution of resilience in protein interactomes across the tree of life

Marinka Zitnik, Rok Sosic,
Marcus W. Feldman, Jure Leskovec

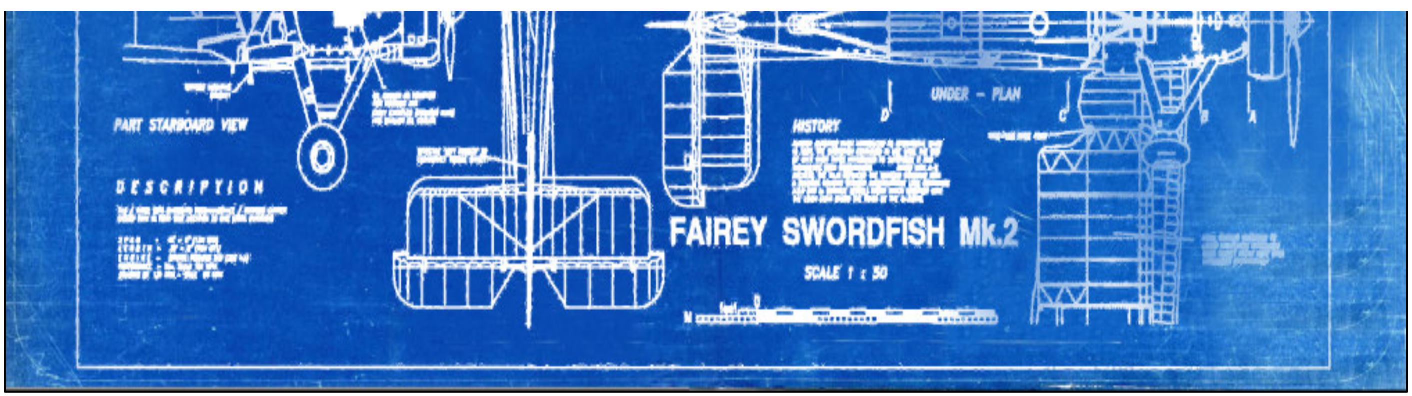
Stanford | ENGINEERING
Computer Science

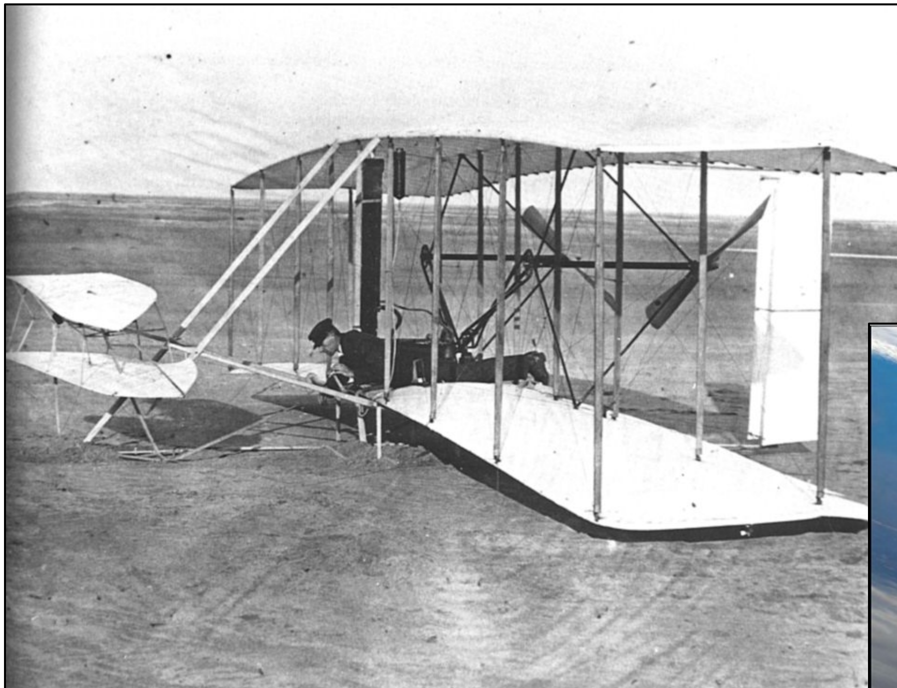
Stanford | Biology

CHAN ZUCKERBERG BIOHUB



Many components, parts that need to work together for the airplane to function properly





The Wright Flyer, 1903

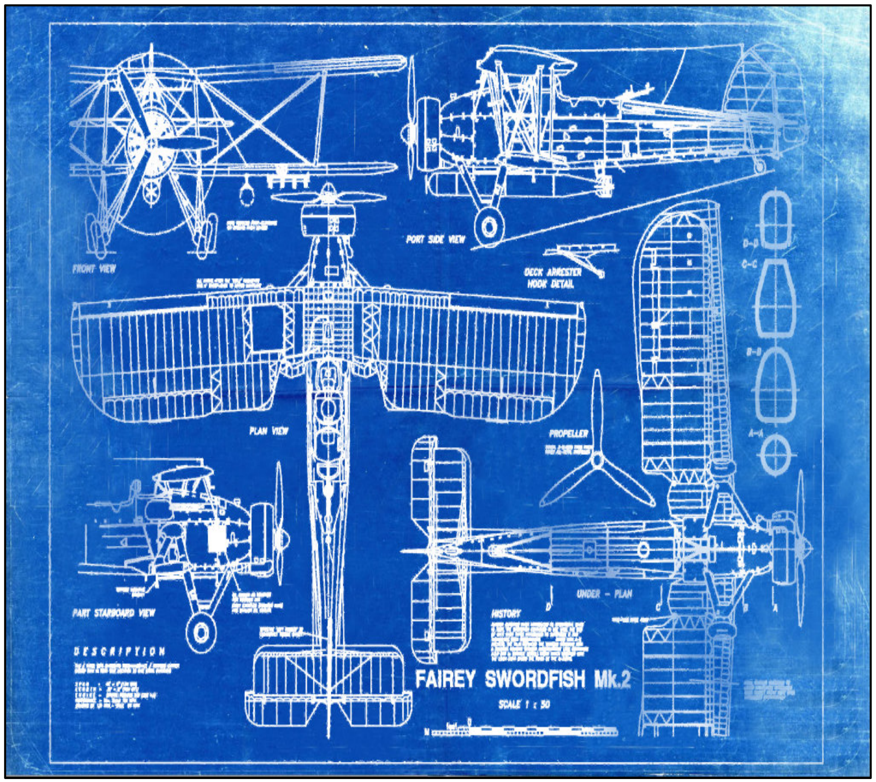
Boeing, 2010

Advancements in engineering have tremendously improved airplanes since the 19th century

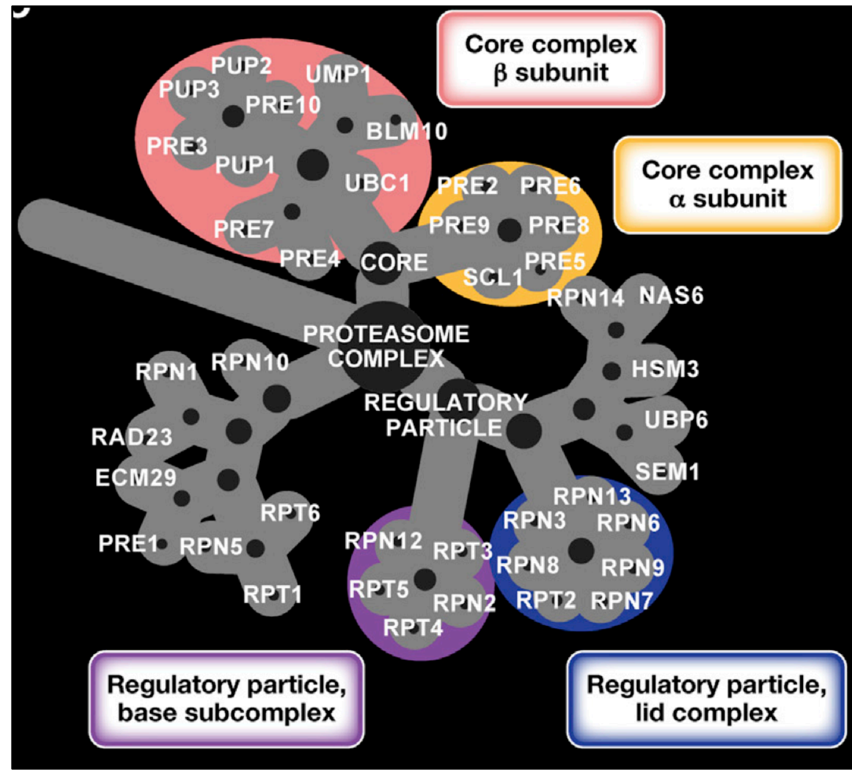
Proceedings of the National Academy of Sciences, 116, 10:4426-4433, 2019

Protein interaction network: Backbone of activity in a cell

Carvunis & Ideker, Cell'14



Physical interactions between an airplane's parts

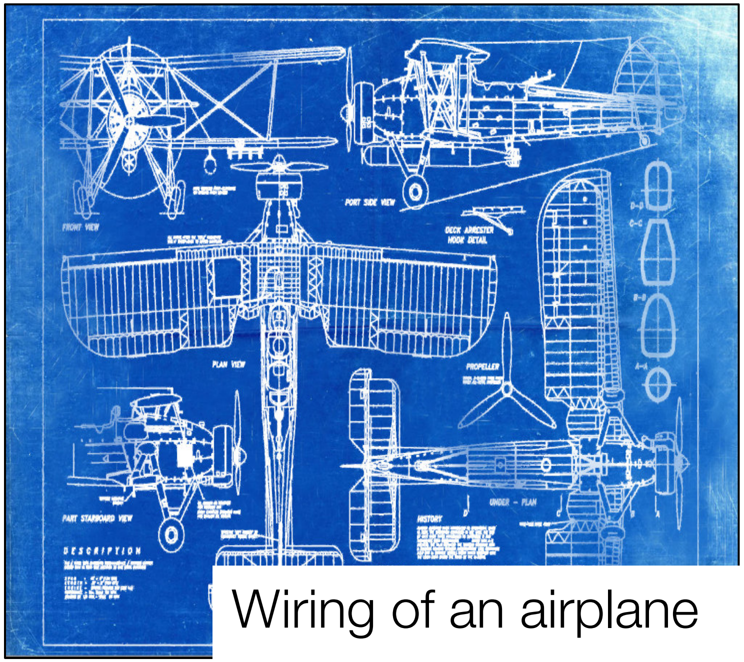


Physical interactions between a cell's molecular components

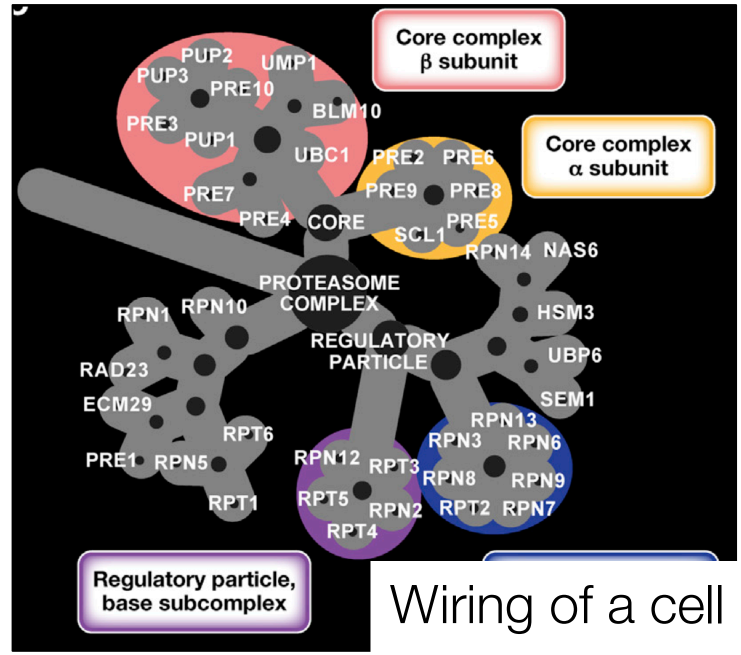
Proceedings of the National Academy of Sciences, 116, 10:4426-4433, 2019

How do protein networks evolve?

Carvunis & Ideker, Cell'14



Wiring of an airplane



Wiring of a cell

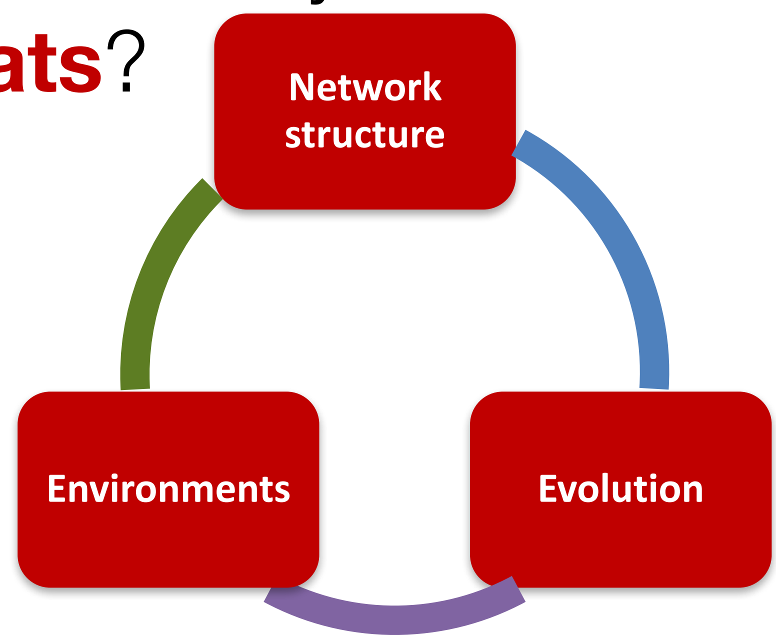
But we do not know how networks change with evolution!

- *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 organism's ability to survive 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 and combining data:
   - Species-specific protein-interaction networks
   - Phylogenetic species information
   - Ecological data on natural habitats in which species live

2. Use dataset to study evolution of protein networks:
   - How protein interaction networks change with evolution?
   - How network changes affect species' survival?

Proceedings of the National Academy of Sciences, 116, 10:4426-4433, 2019

Key Element: Evolutionary Dataset

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

We build a unique dataset:

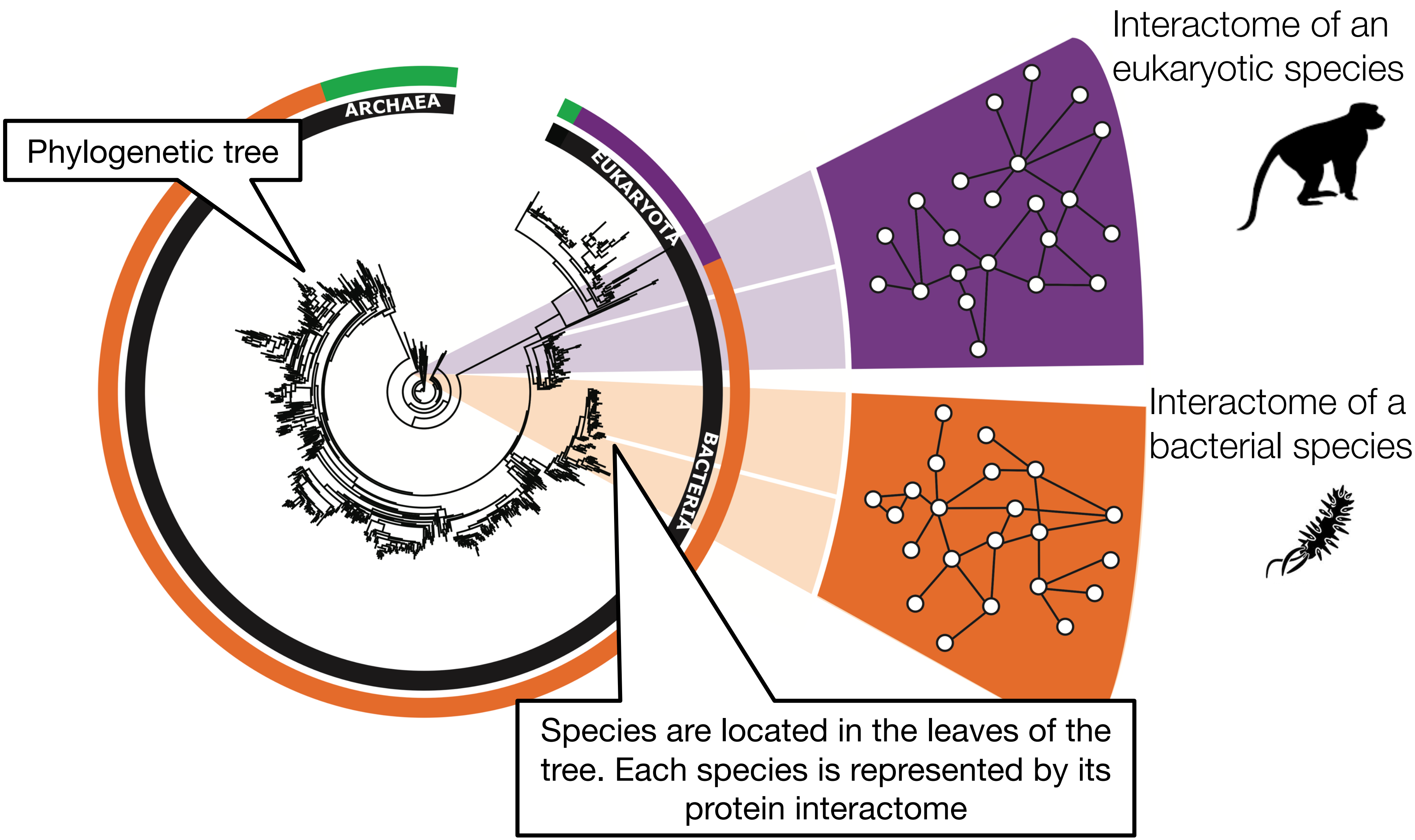
- 1,840 networks: 1,539 bacteria, 111 archaea, 190 eukarya
- 1,450,633 nodes: Species' proteins
- 8,762,166 edges: Physical protein-protein interactions (PPIs)

- 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

Proceedings of the National Academy of Sciences, 116, 10:4426-4433, 2019

Tree of Networks



Modeling Task

Data:

Tree of networks

Task: How interactomes respond to protein network failures affect and how that response changes over time:

- Protein network failures can occur through:
  - Removal of a protein (e.g., nonsense mutation)
  - Disruption of a PPI (e.g., environmental factors, such as availability of resources)
- Resilience to network failures is critical:
  - Breakdown of proteins affects the exchange of biological information in the cell [Huttlin et al., Nature'17]
  - Failures can fragment the interactome and lead to cell death and disease [Chen et al., Nat. Genet.'18]

Proceedings of the National Academy of Sciences, 116, 10:4426-4433, 2019

How to characterize resilience to network failures?

Define **interactome resilience** measure:

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

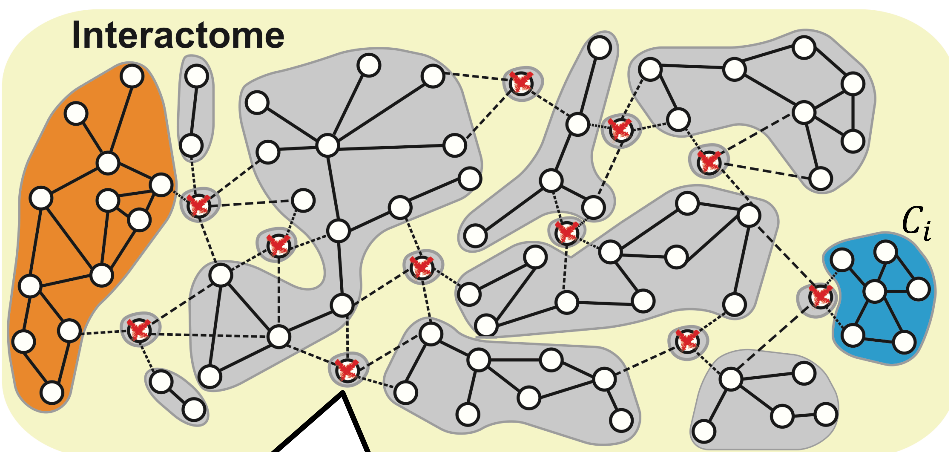
Resilience measure has three key elements:

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

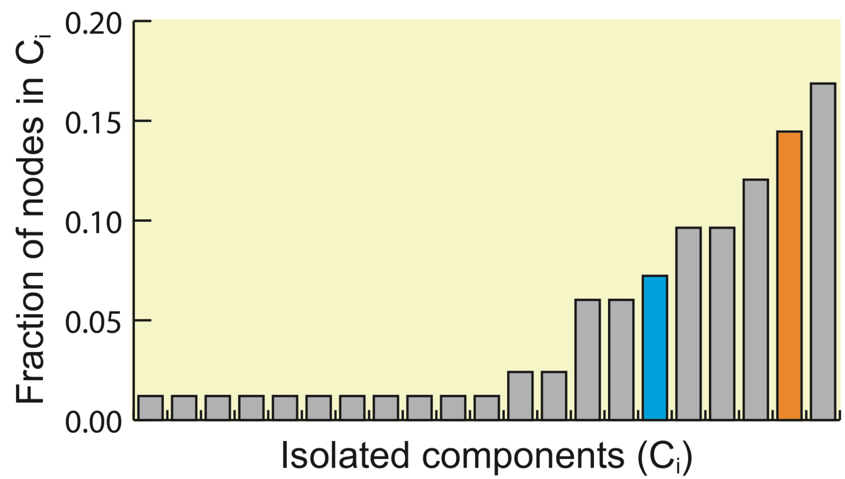
Proceedings of the National Academy of Sciences, 116, 10:4426-4433, 2019

Simulate a 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



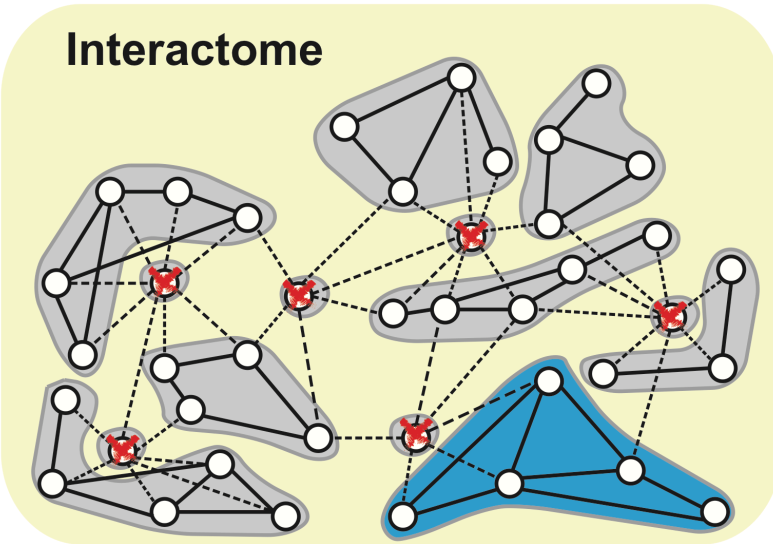
$$H = -\sum_{C_i} \frac{C_i}{N} \log \frac{C_i}{N}$$

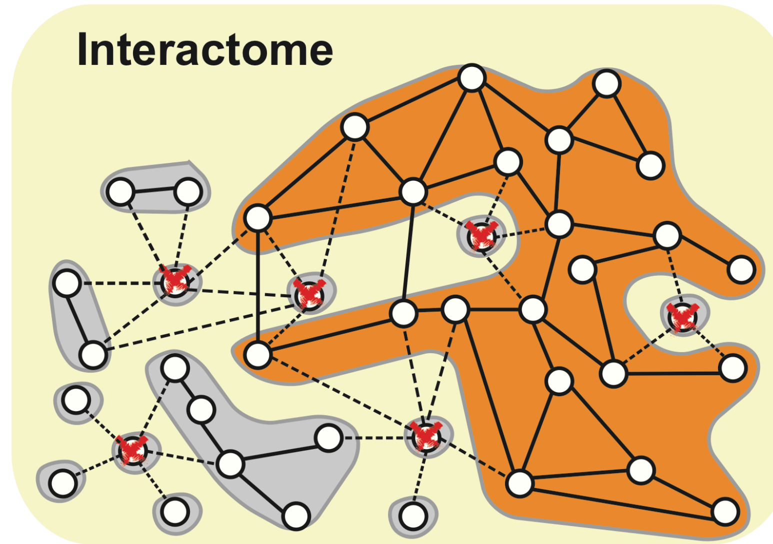
C_i/N is fraction of all nodes N in isolated component C_i

Removed node Isolated component Removed PPI

Proceedings of the National Academy of Sciences, 116, 10:4426-4433, 2019

Fragmentation: Example



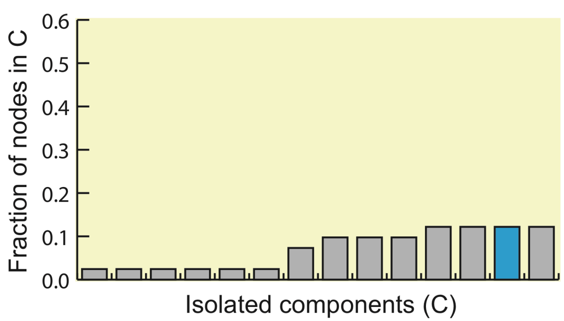


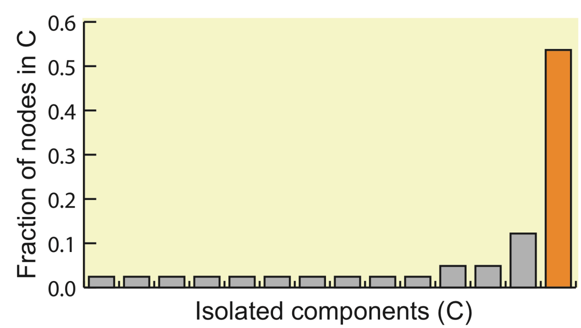
High entropy

Many small isolated components, all of approximately the same size

Low entropy

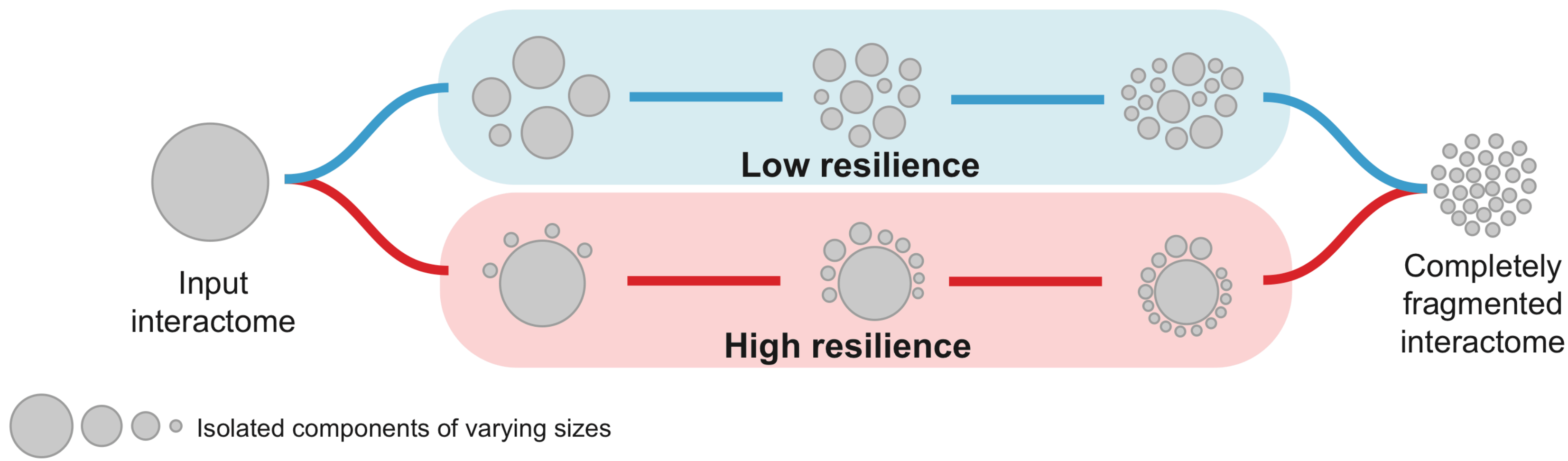
Large isolated component, only a few small broken-off components



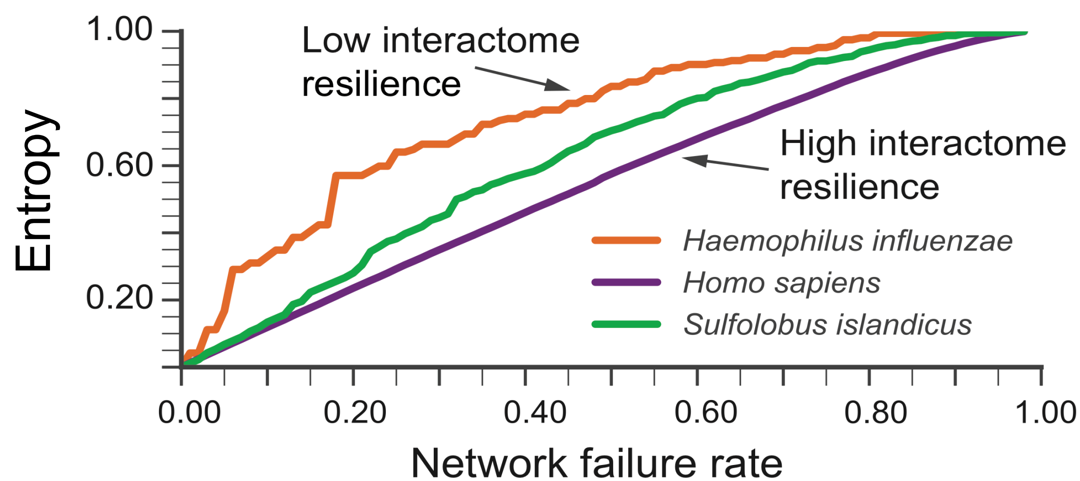


Proceedings of the National Academy of Sciences, 116, 10:4426-4433, 2019

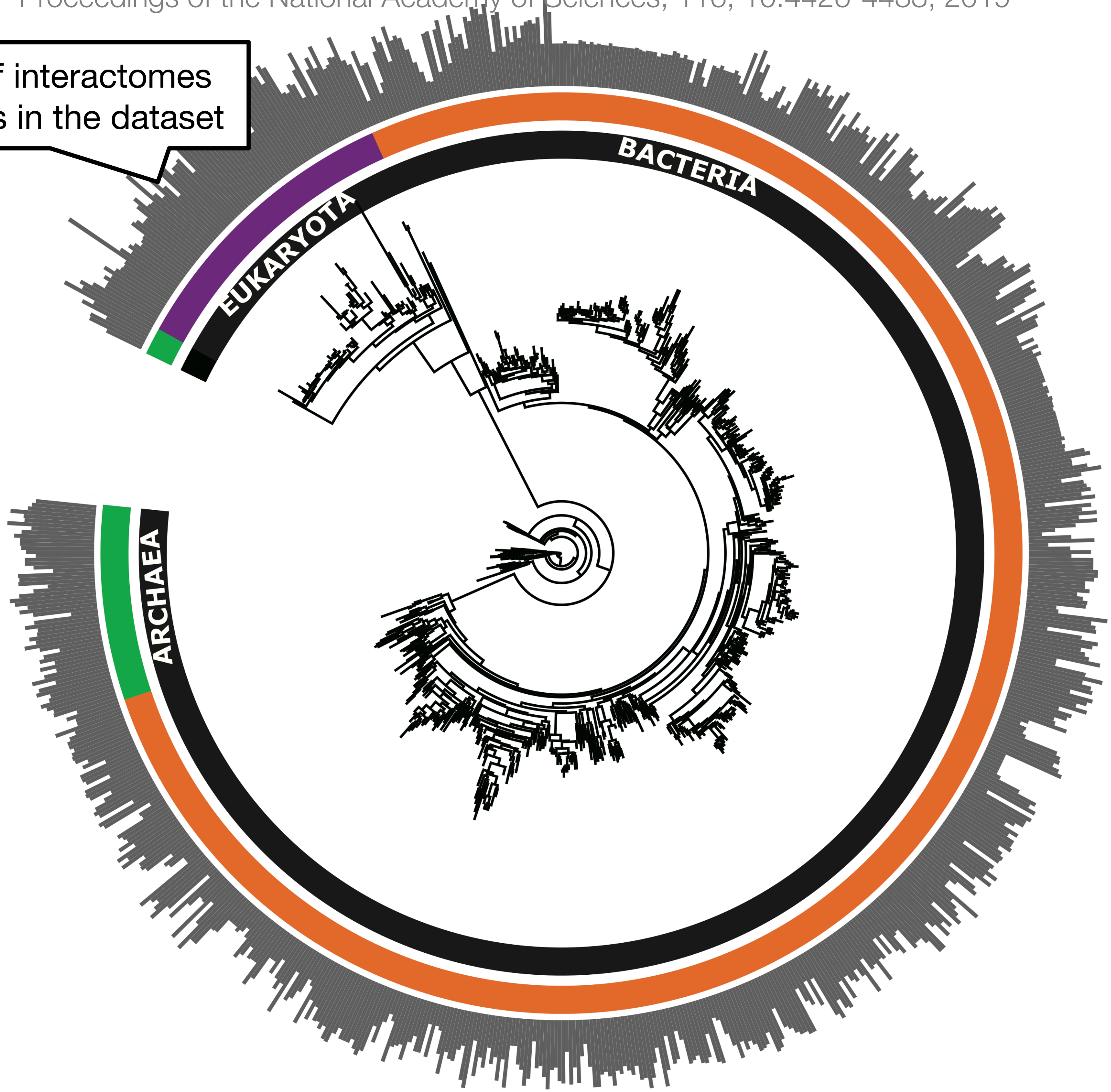
Resilience: Fragmentation integrated across all possible failures



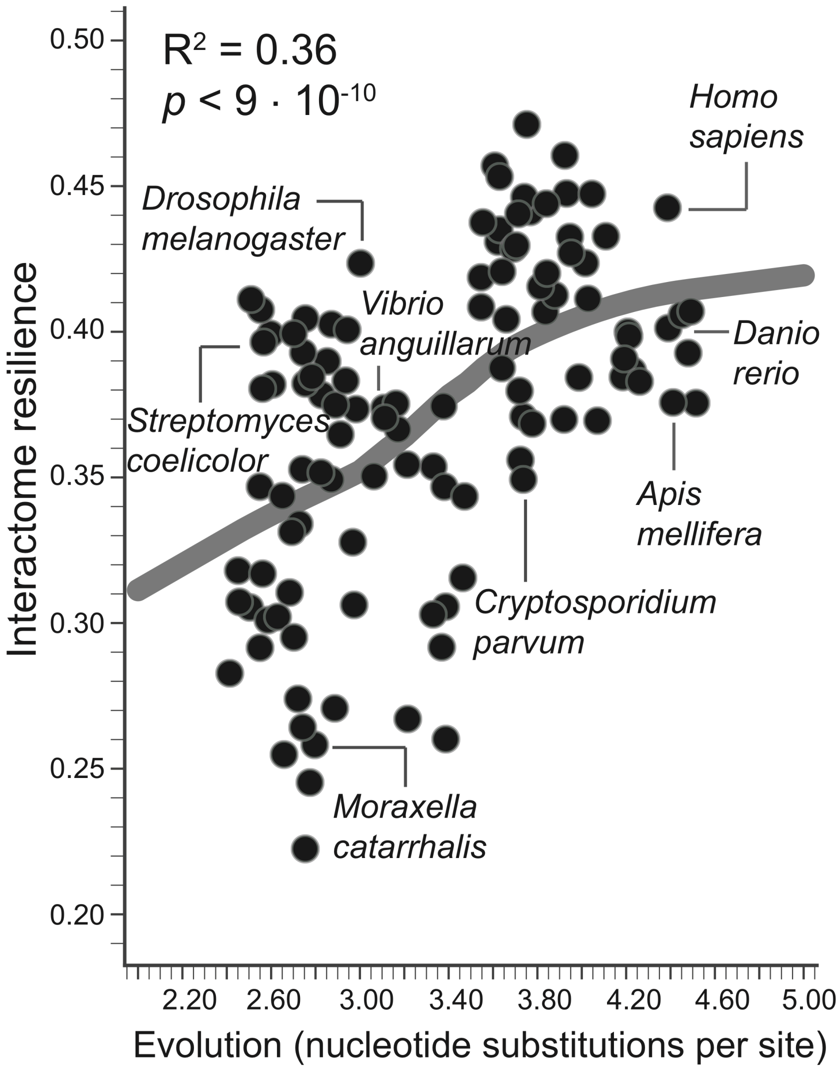
E.g., resilience for three species:



Resilience of interactomes for all species in the dataset



Evolution leads to resilience



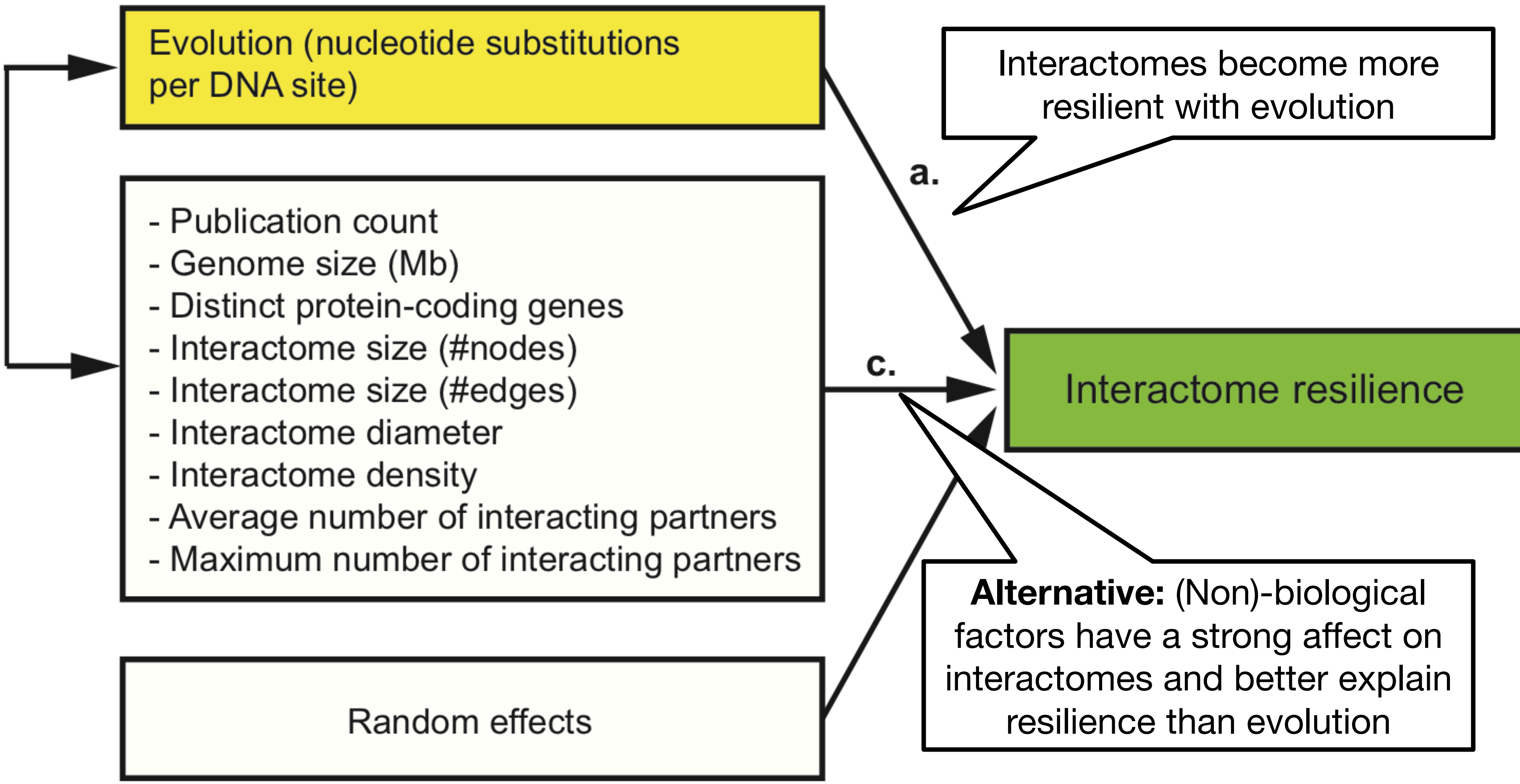
Protein interactomes become **more resilient with evolution**

- **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, 116, 10:4426-4433, 2019

Is this finding due to data biases?

Causal model: Alternative hypotheses for the relationship between evolution and resilience



Proceedings of the National Academy of Sciences, 116, 10:4426-4433, 2019

Findings are not due to data biases

Findings are:

- Consistent across taxonomic groups
- Robust to network data quality and network size
- Consistent across different types of assays

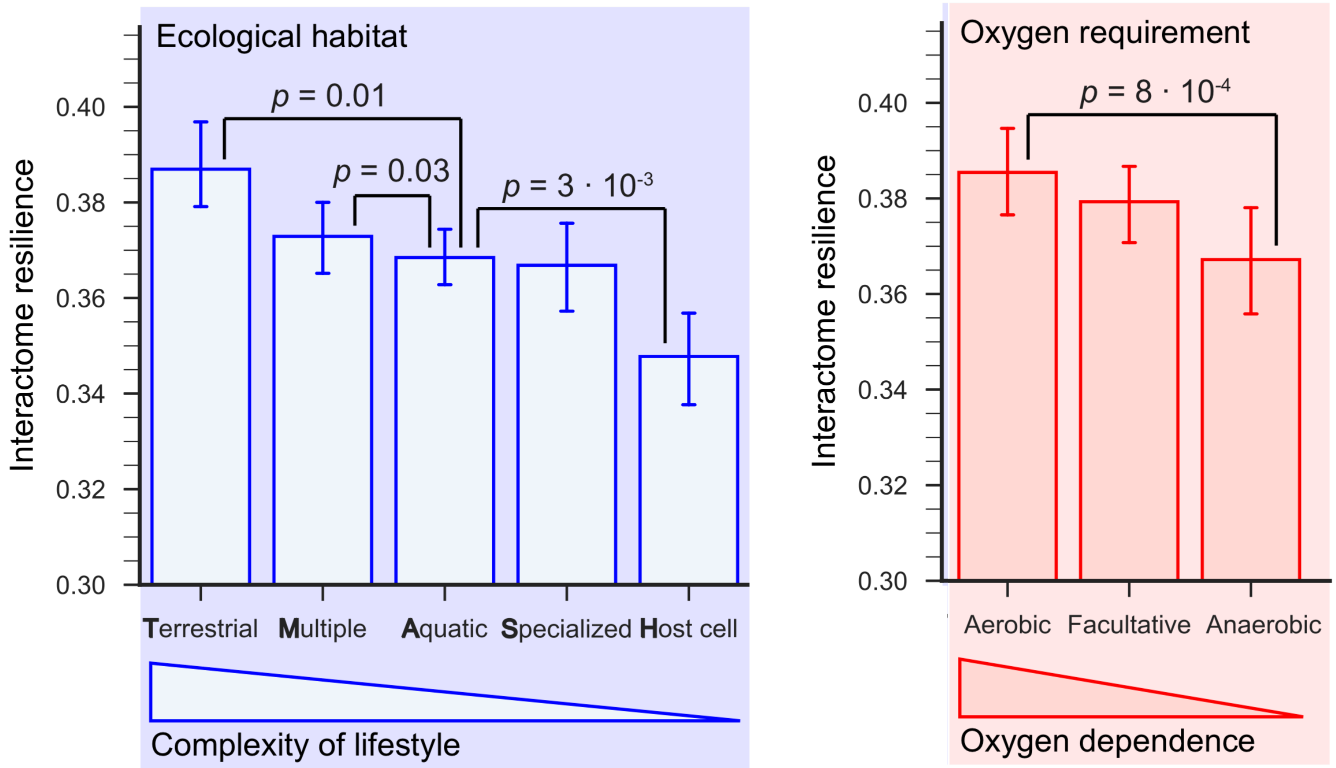
Findings are not due to confounding:

- Genome, e.g., genome size, protein-coding genes
- Networks, e.g., hub nodes, broad-tailed degree distributions, number of interactions in each species
- Investigative bias, e.g., much-studied proteins and species

Results indicate key findings will still hold when more protein interaction data become available

Proceedings of the National Academy of Sciences, 116, 10:4426-4433, 2019

Resilience is beneficial

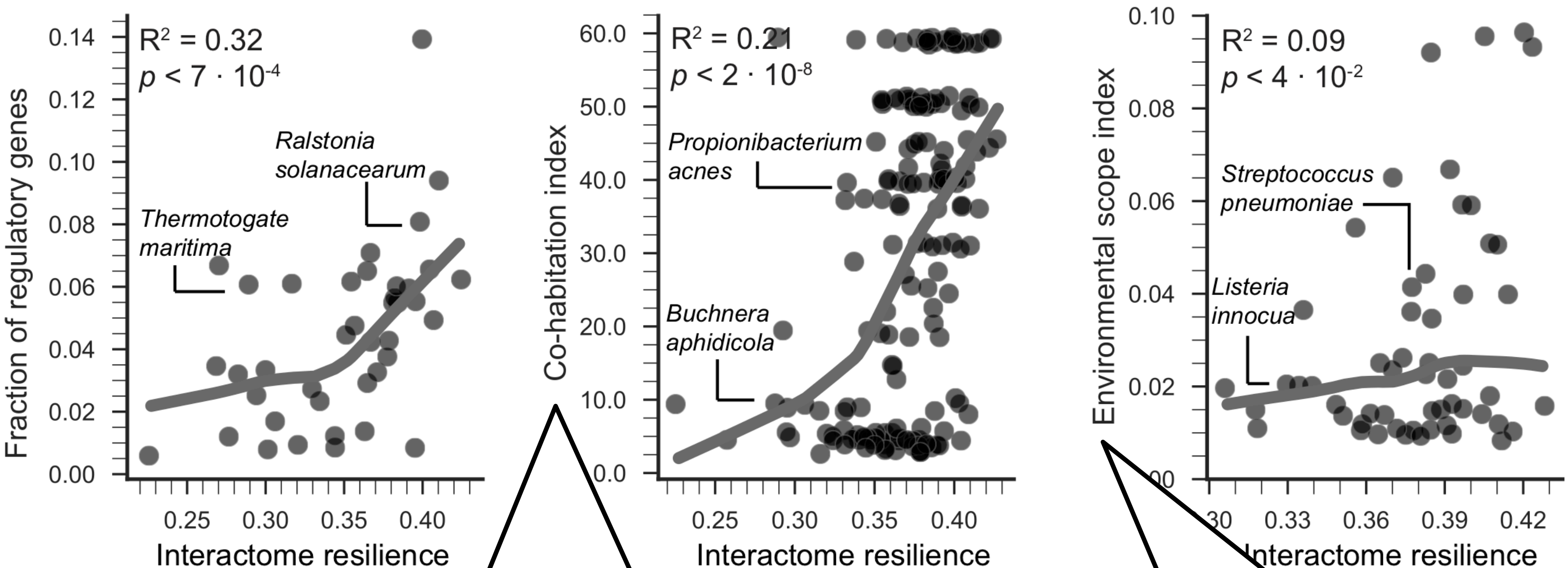


Organisms with **more resilient interactomes** can survive in **more complex, diverse**, and **competitive habitats**

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

Proceedings of the National Academy of Sciences, 116, 10:4426-4433, 2019

Resilience is beneficial



Level of competition in environments in which species live

Diversity of environments in which species live

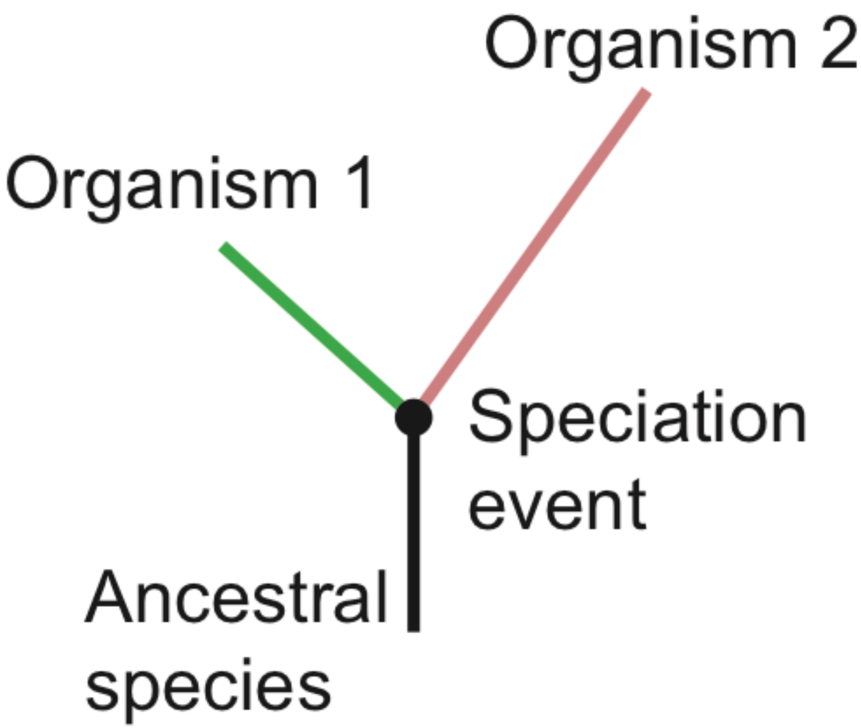
Organisms with **more resilient interactomes** survive in **more complex, diverse**, and **competitive habitats**

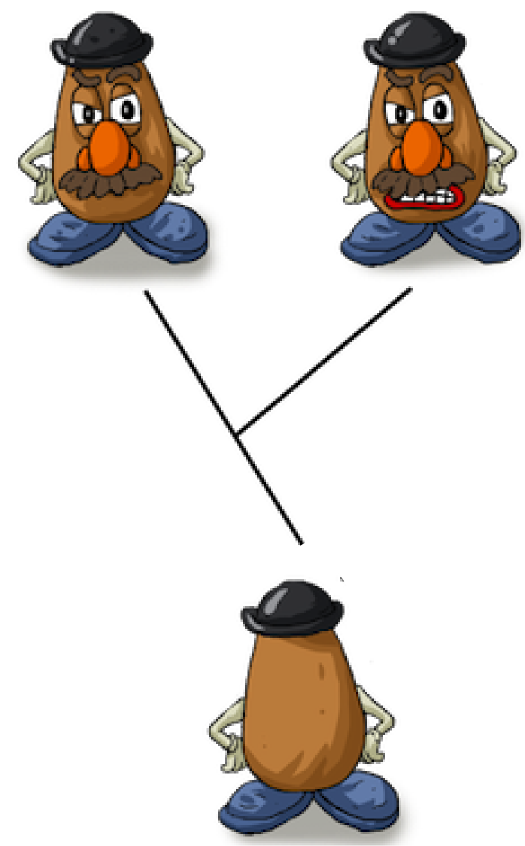
How does resilience arise through changes in network topology?

Goal: Identify mechanisms that explain how local network changes lead to increased interactome resiliency

Proceedings of the National Academy of Sciences, 116, 10:4426-4433, 2019

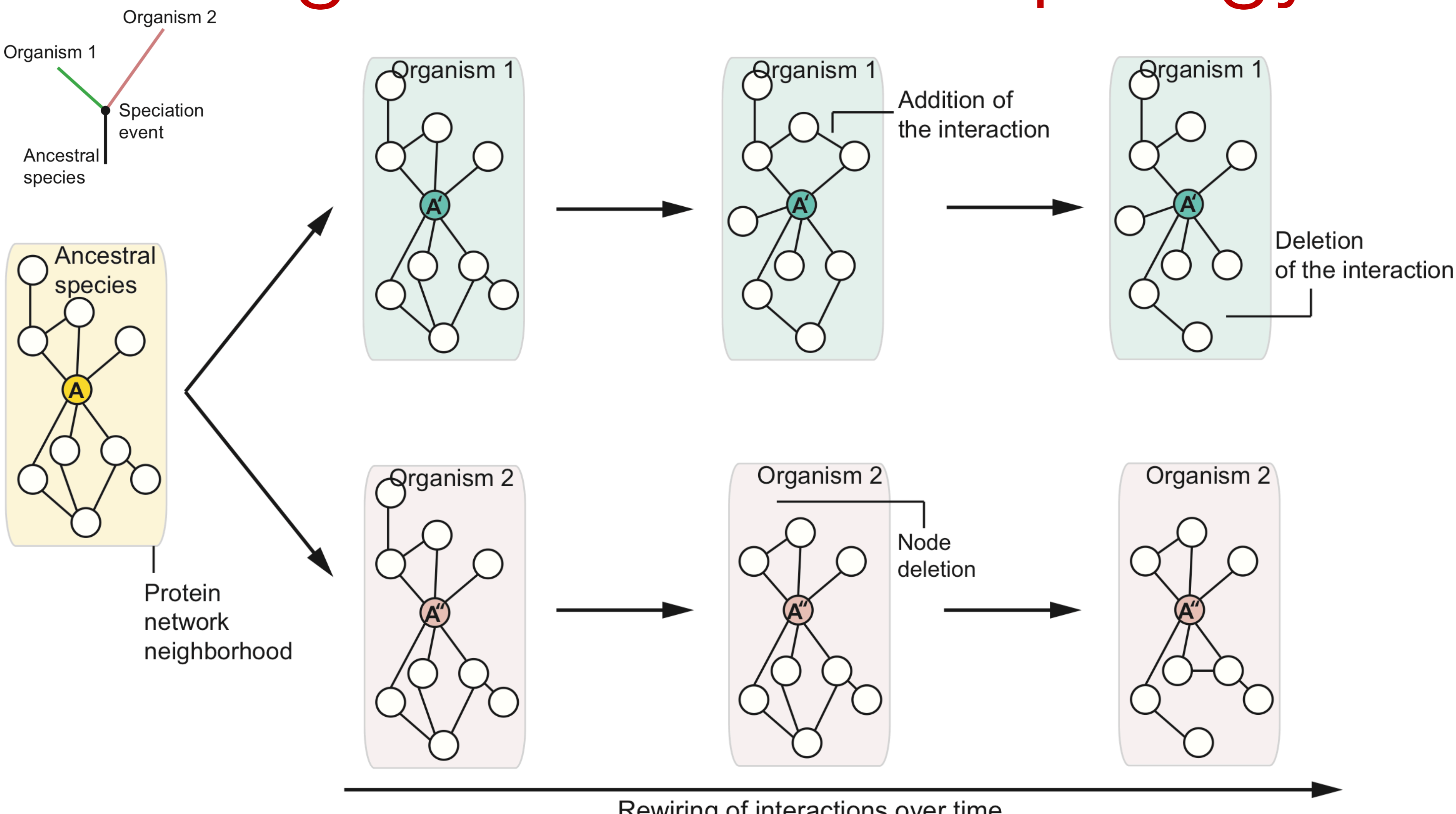
A species evolves into two new species…





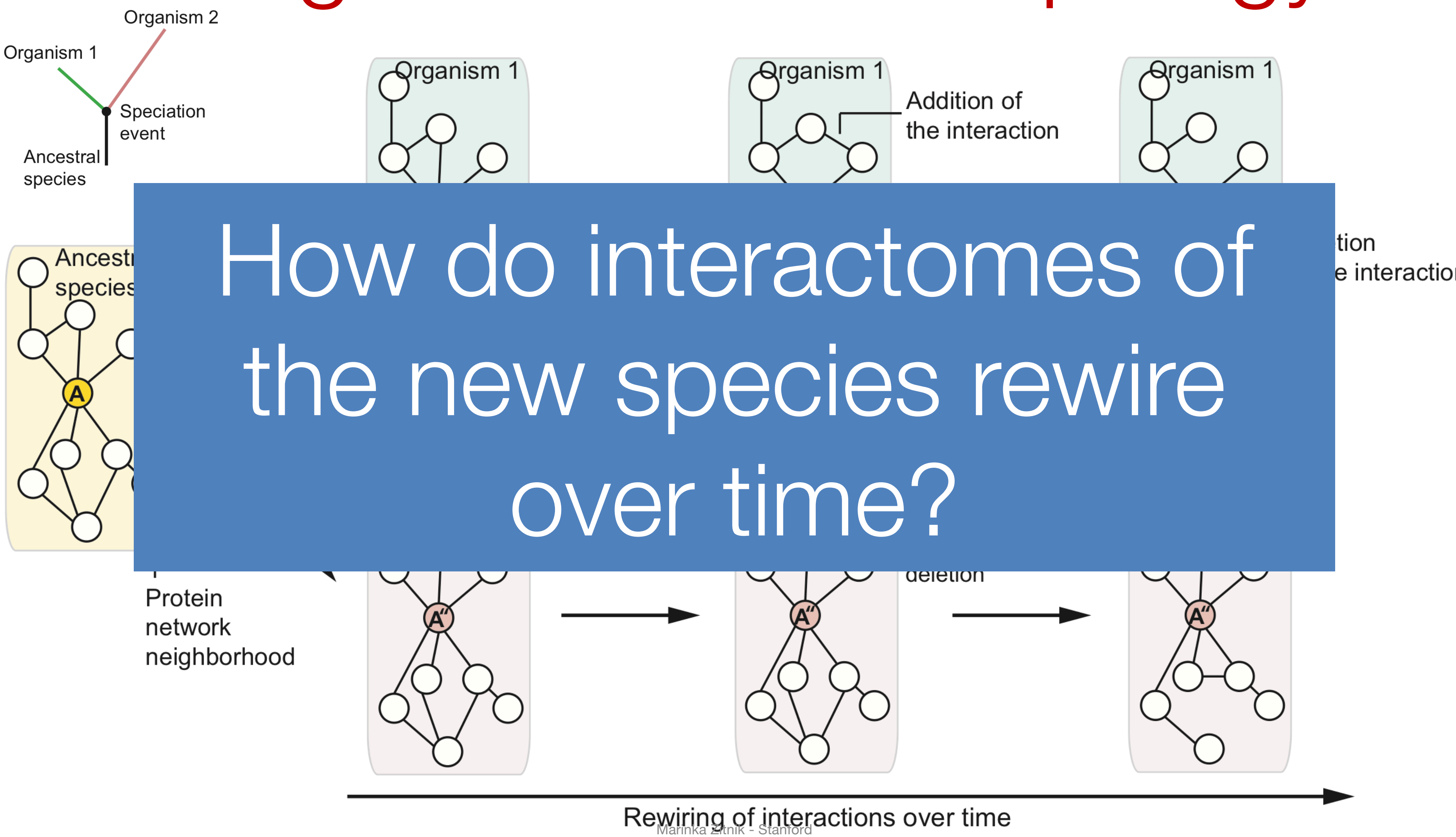
Proceedings of the National Academy of Sciences, 116, 10:4426-4433, 2019

How does resilience arise through changes in network topology?



Proceedings of the National Academy of Sciences, 116, 10:4426-4433, 2019

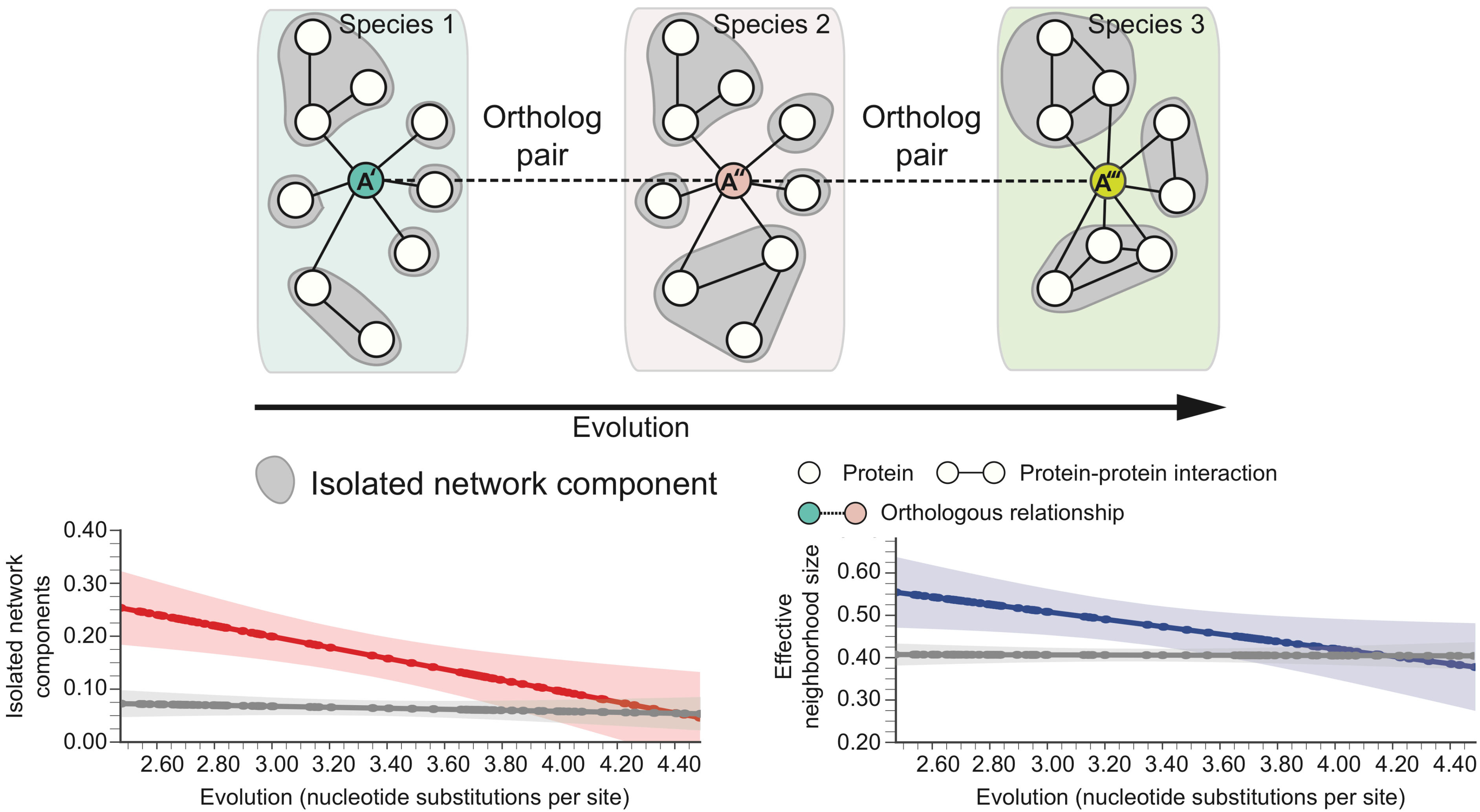
How does resilience arise through changes in network topology?



How do interactomes of the new species rewire over time?

Proceedings of the National Academy of Sciences, 116, 10:4426-4433, 2019

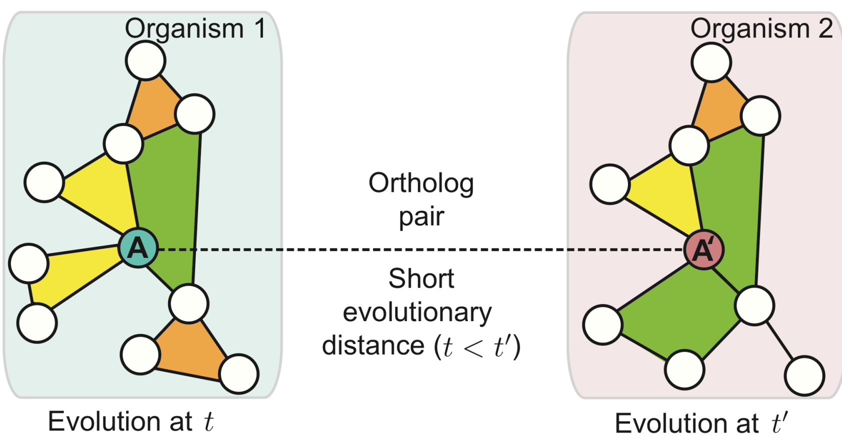
Resilience arises through gradual change of network topology

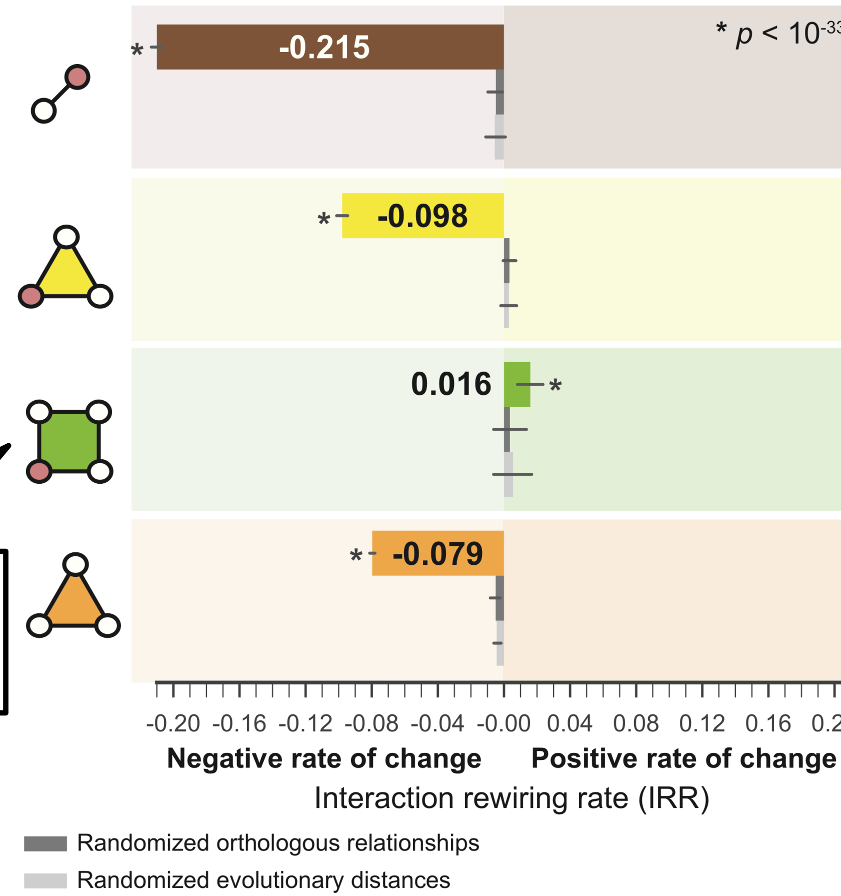


Proceedings of the National Academy of Sciences, 116, 10:4426-4433, 2019

Mechanism of Resilience

Rewiring of protein-protein interactions in local protein neighborhoods





Square network motifs become more common with evolution

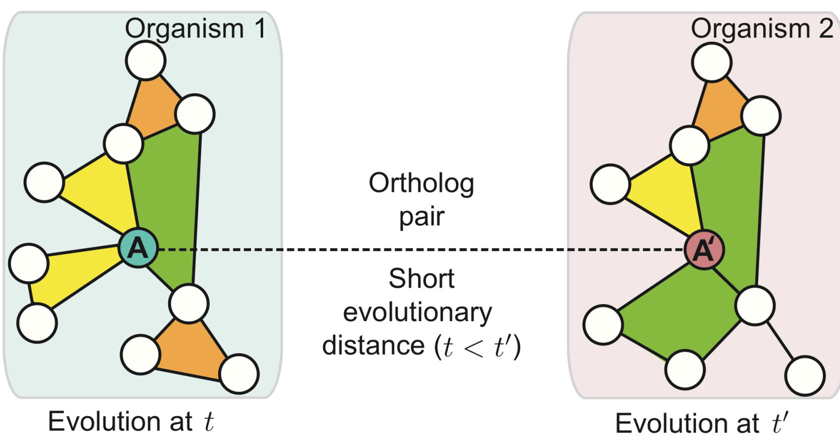
Protein   Protein-protein interaction

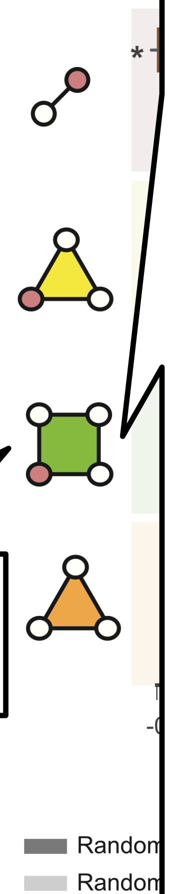
Orthologous relationship

Proceedings of the National Academy of Sciences, 116, 10:4426-4433, 2019

Mechanism of Resilience

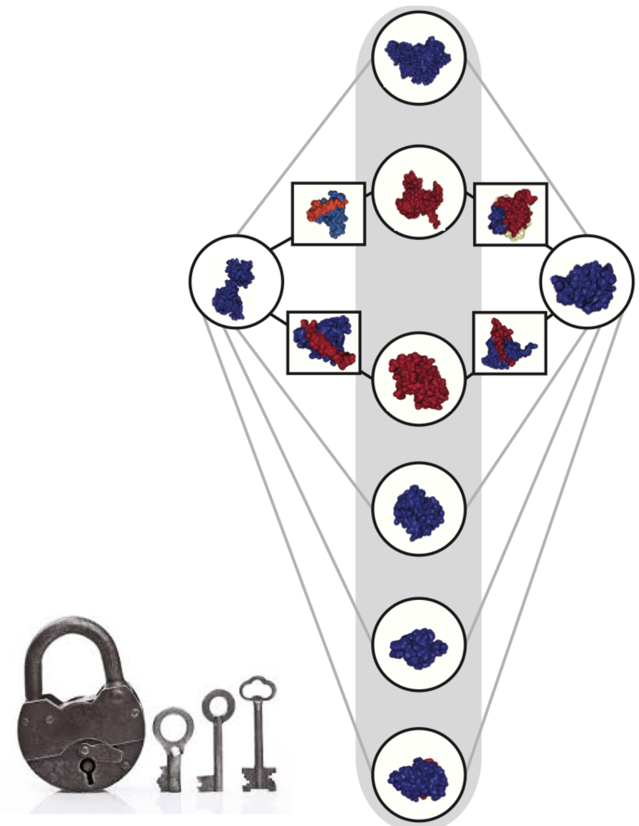
Rewiring of protein-protein interactions in local protein neighborhoods





Square network motifs become more common with evolution

Emergence of square network motifs: Proteins with similar interfaces share many neighbors, but do not interact with each other



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

HARVARD
UNIVERSITY

WE'RE HIRING!

Projects in Machine Learning, AI, and Large Biomedical Data

Research Problems: Precision Medicine & Health, Drug Discovery, Genomics, Network Biology, and many more

Hiring outstanding postdocs!

marinka@cs.stanford.edu