Domain architecture conservation in orthologs

Kristoffer Forslund, Isabella Pekkari, Erik Sonnhammer
Web: http://www.sbc.su.se   Contact: krifo@sbc.su.se

Orthology   **Paralogy**

i-i: inparalogs   **i-cni:** closest same-species outparalogs
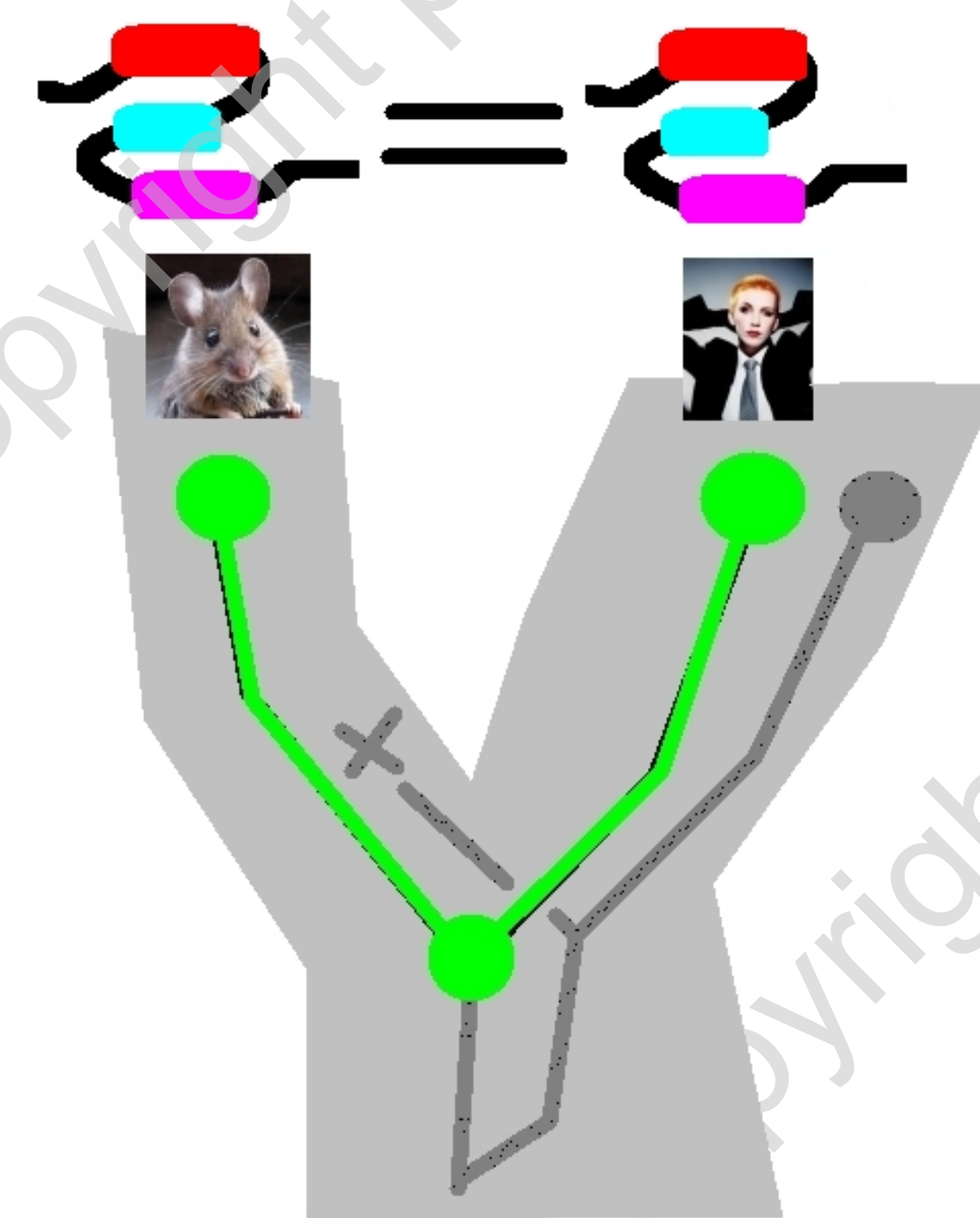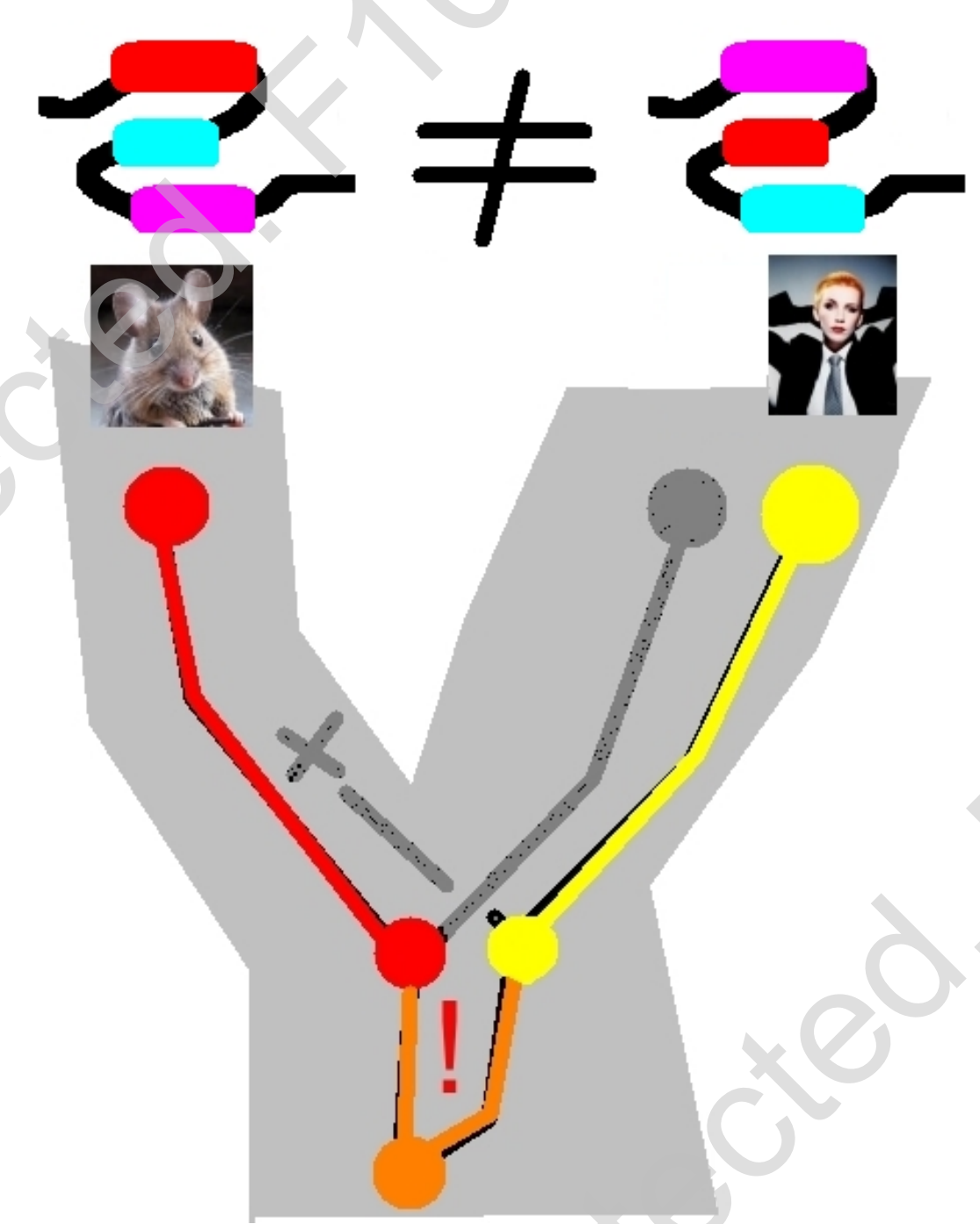**o-o:** orthologs   **o-cno:** closest cross-species outparalogs

Objectives

As sequences diverge, higher relative conservation of orthologous architectures is expected if:

• Gene duplications relax selective pressure through redundancy

• Domain architectures are functionally important

This has not previously been evaluated systematically
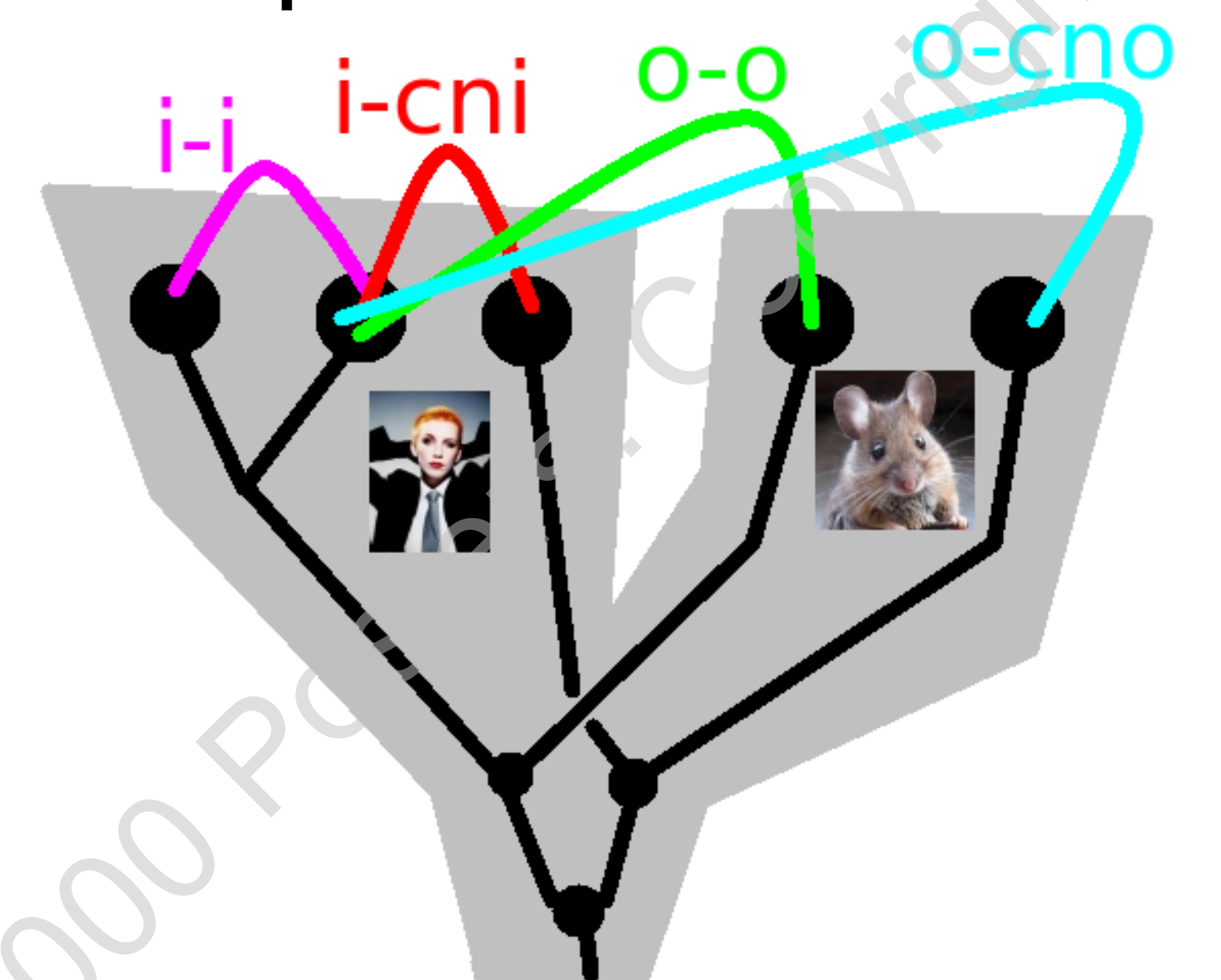
Materials

Large-scale dataset of complete proteomes

• Orthology inferences from *InParanoid*

• Domain architectures from *Pfam*

VS

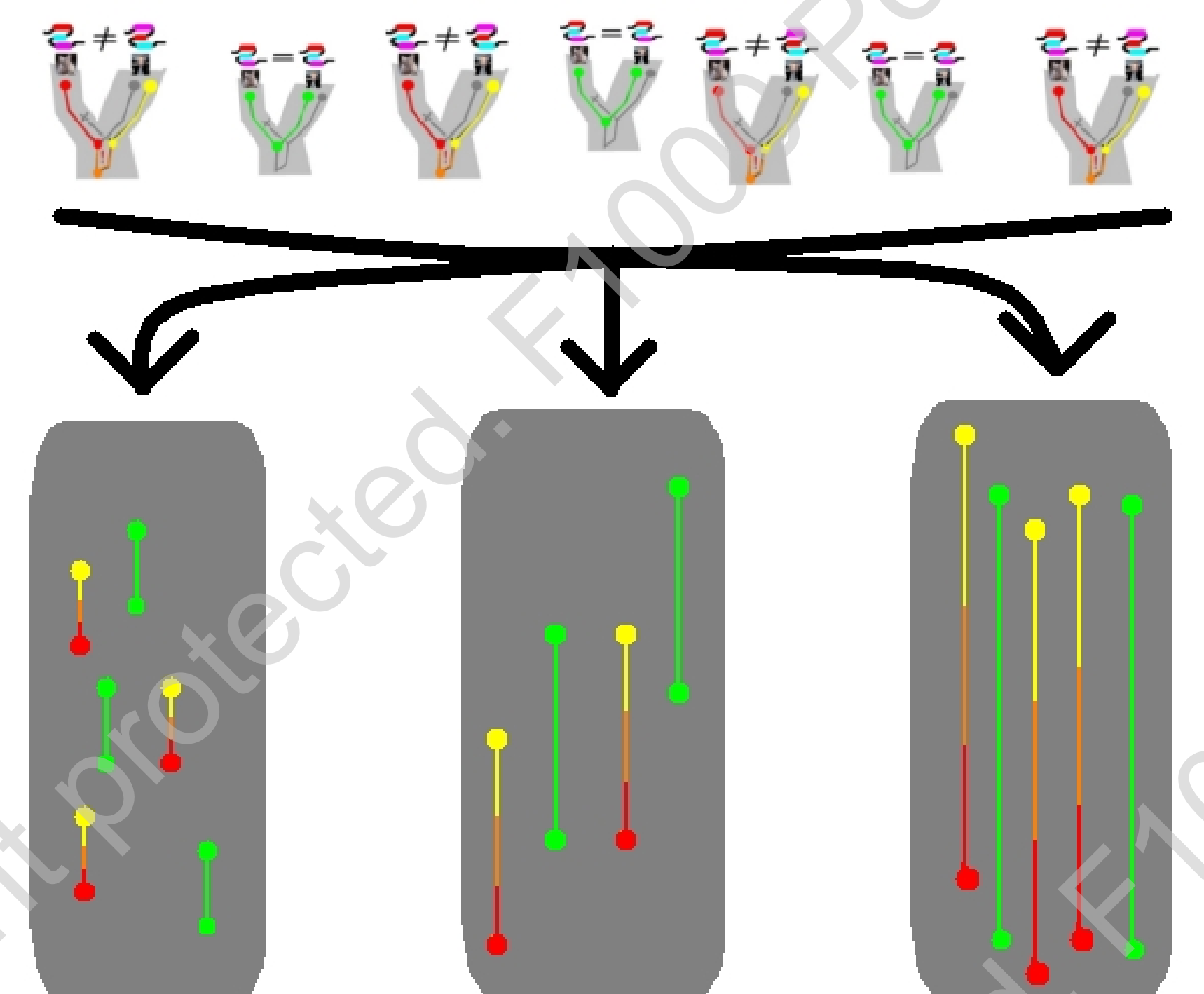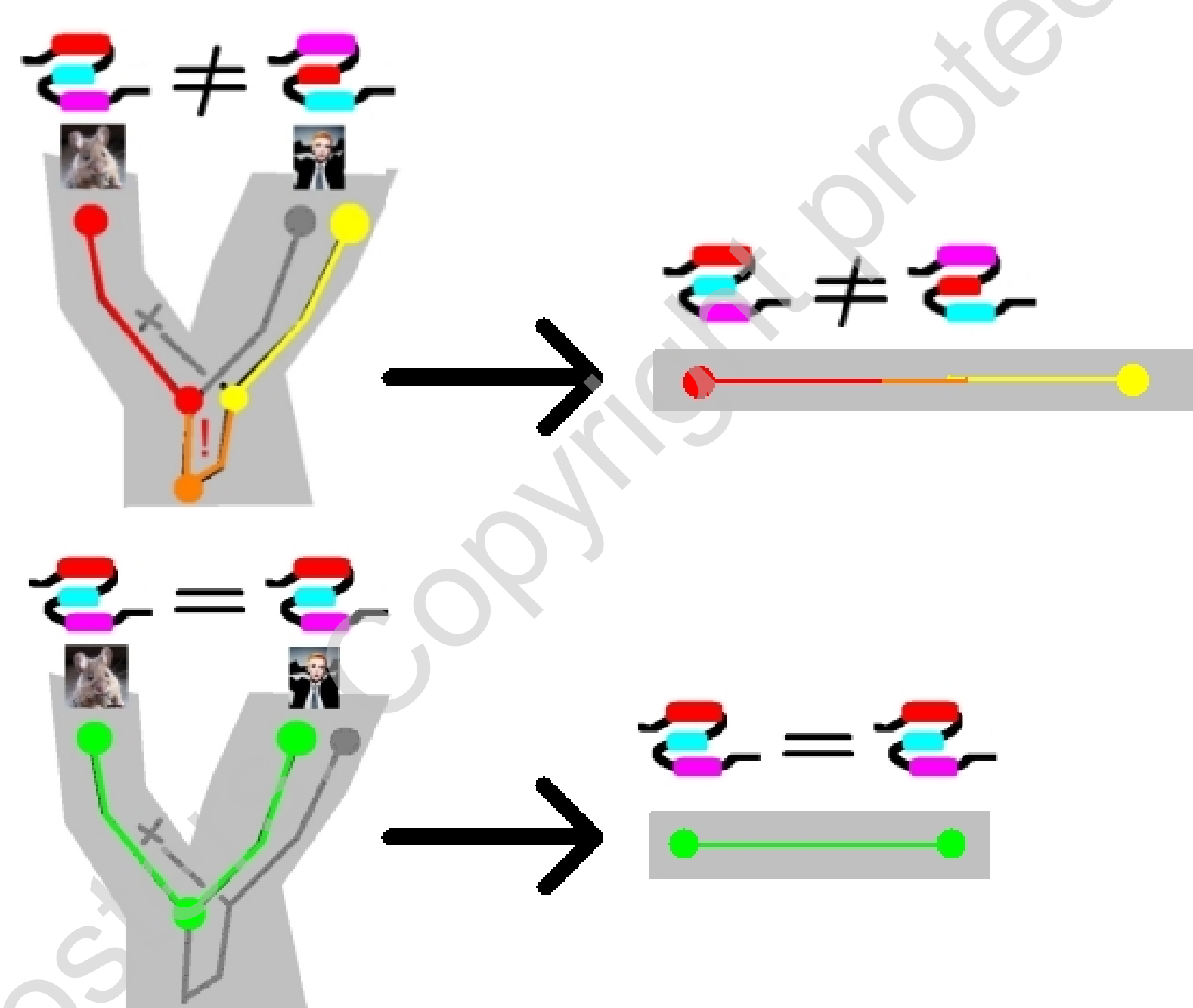We consider four types of protein pairs:

i-i   i-cni   o-o   o-cno

Methods

• **DA-score:** Domain architecture similarity measure, taken as fraction of aligned domains

• **Sequence divergence:** JC-corrected sequence identity

• Divide pairs into sequence divergence bins

• Test for significant DA-score differences within each bin

Outcome

Orthologous proteins are significantly more similar in domain architecture than in- or outparalogs at the same evolutionary distance