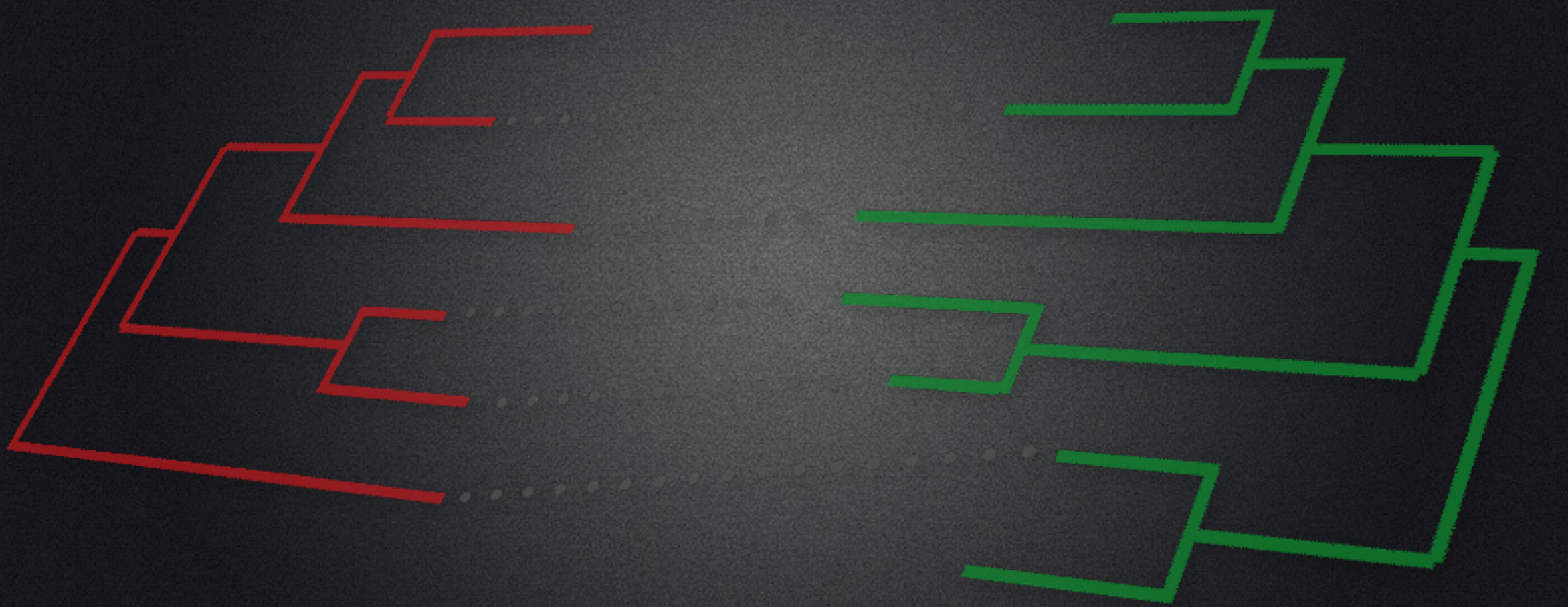
Protein Evolution Analysis: on the Use of Phylogenetic Trees

Studying molecular co-evolution



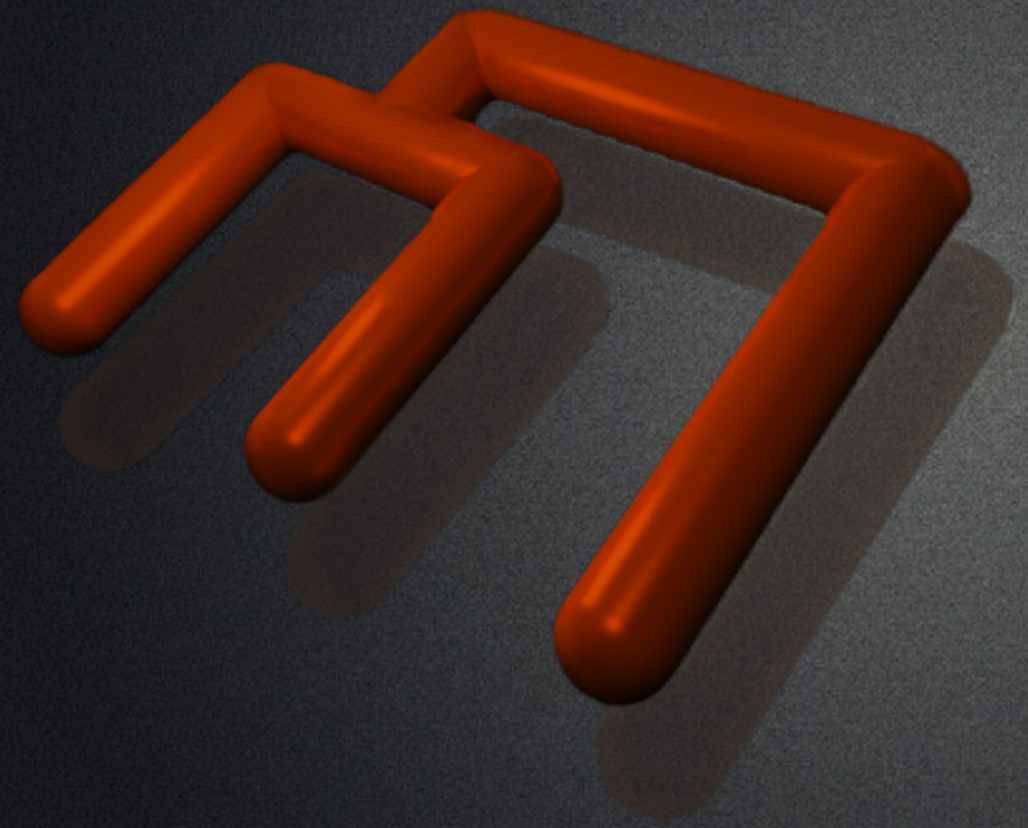
David Ochoa

EMBL-EBI

Outline

- The basics
- Practical aspects
- Applications

EMBL-EBI

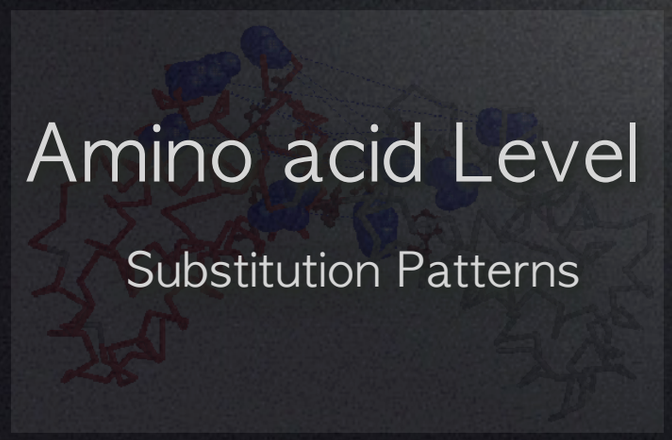


1. The basics

EMBL-EBI

Co-evolution

"reciprocal evolutionary change in interacting species" Thompson 1994



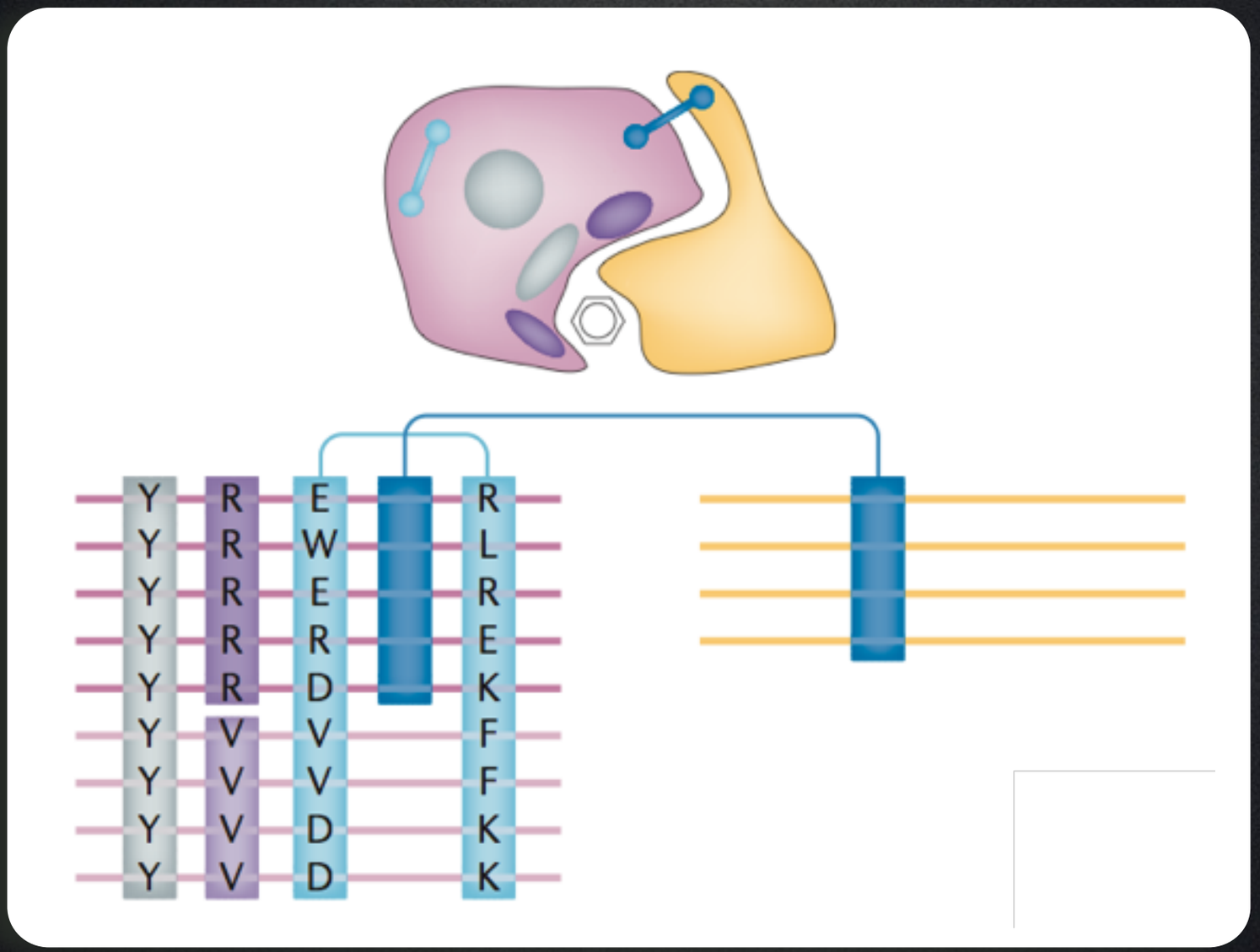
Ecological Factors
Trophic networks
Mutualisms
Symbioses
…

Protein Structure
Protein Function
Protein Networks
…

Local Structural Effects
Amino acid Contact Networks
Functional Sites
Protein-Protein Interfaces
…

EMBL-EBI

Residue co-evolution

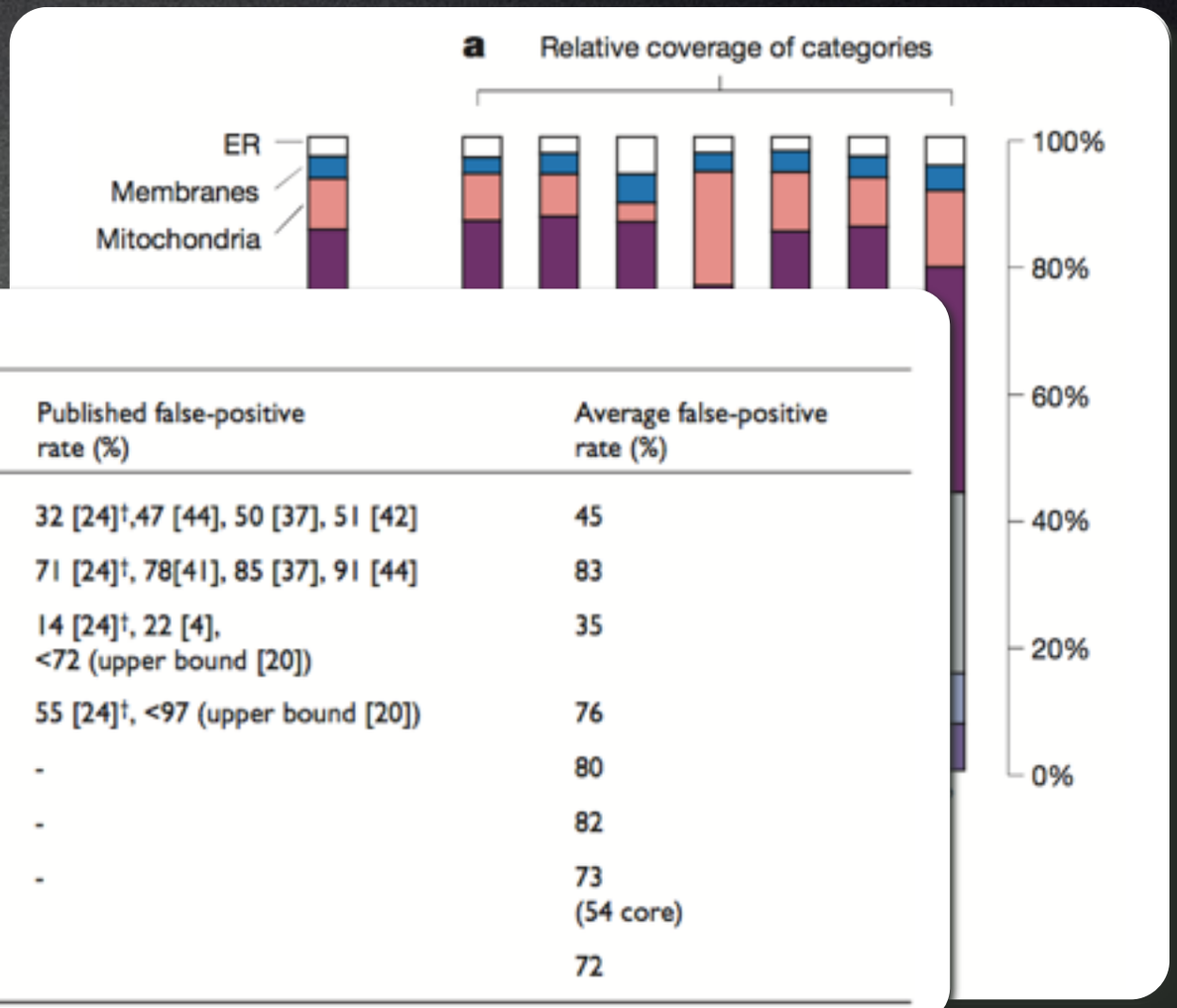


de Juan, D., Pazos, F. & Valencia, A. Emerging methods in protein co-evolution. Nat Rev Genet – (2013). doi:10.1038/nrg3414

EMBL-EBI

Historical background

- High False-positives Rates
- Pr
- Pr



Yeast protein-interaction assay false-positive rates: yeast datasets

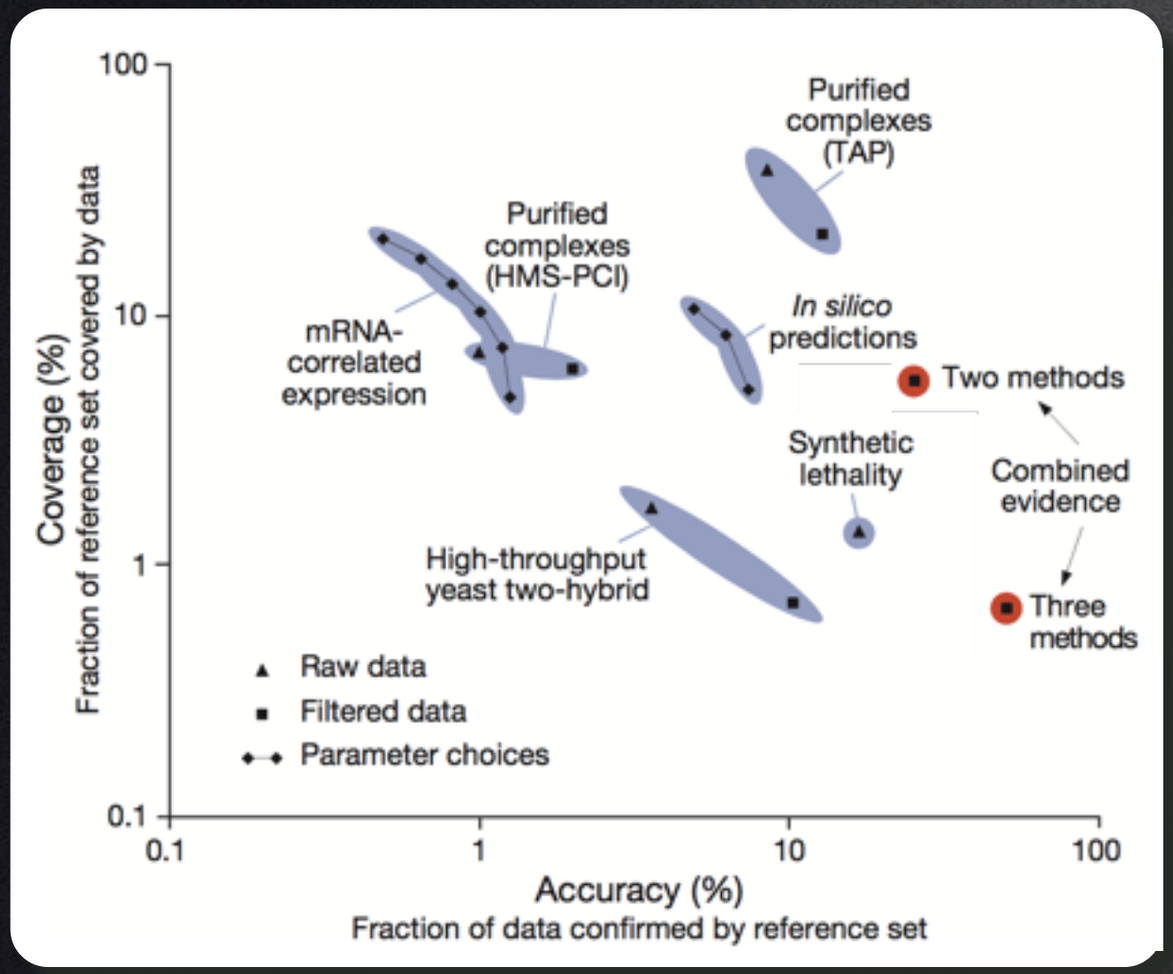
Dataset	Number of interactions	Derived false-positive rate* (%)	Published false-positive rate (%)	Average false-positive rate (%)
Uetz *et al.* [35]	854	46 [32]	32 [24]†,47 [44], 50 [37], 51 [42]	45
Ito [36]	4,393	89 [32]	71 [24]†, 78[41], 85 [37], 91 [44]	83
Gavin *et al.* [16]	3,180	68 [32]	14 [24]†, 22 [4], <72 (upper bound [20])	35
Ho *et al.* [17]	3,618	83 [32], 81, 82, 80	55 [24]†, <97 (upper bound [20])	76
Jansen *et al.* [22]	15,922	81, 79	-	80
Gavin *et al.* [27]	18,137	78, 82, 86‡	-	82
Krogan *et al.* [28]	14,317 (7,123 core)	75, 79, 66‡ (59, 65, 37‡ core)	-	73 (54 core)
Overall	51,419			72

Hart, G. T., Ramani, A. K., Marcotte, E. M. (2006) How complete are current yeast and human protein-interaction networks? Genome Biol 7, 120.

Mering, von, C. et al. Comparative assessment of large-scale data sets of protein-protein interactions. Nature 417, 399–403 (2002).

EMBL-EBI

Possible Solutions



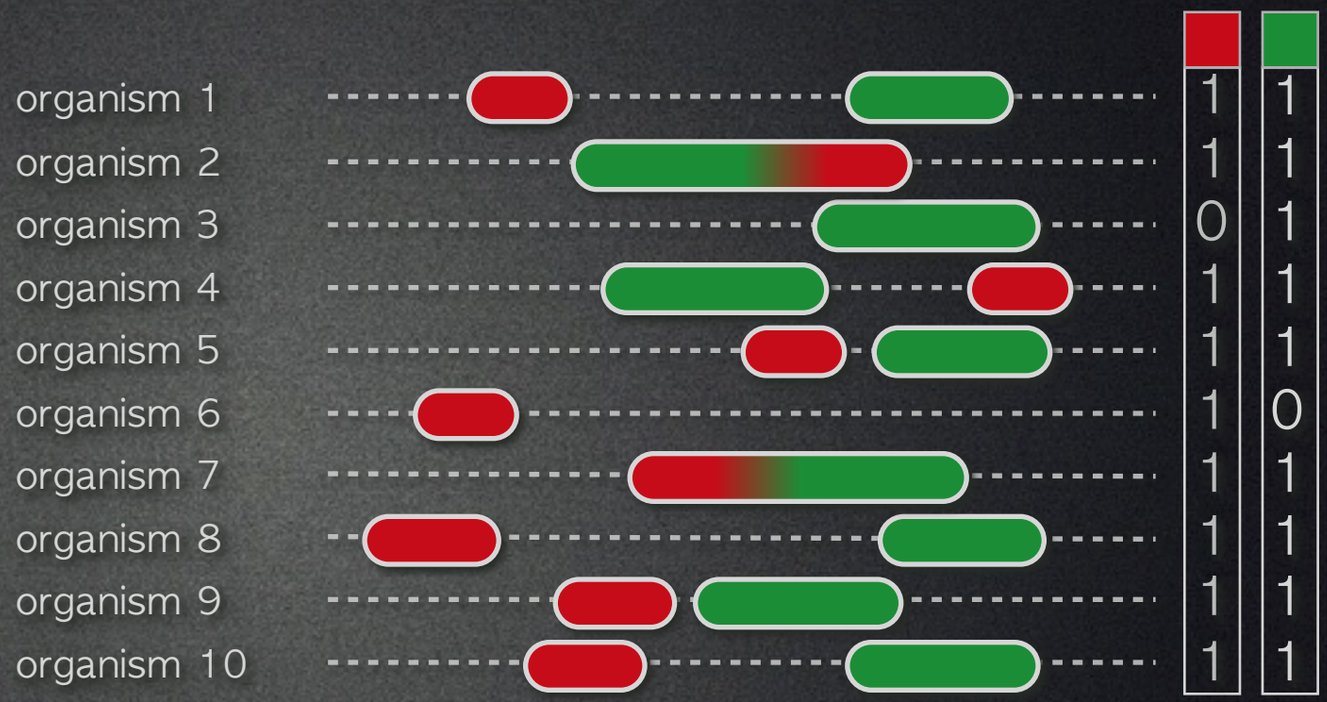
- Repeated screenings
- Combining evidences
- Confidence evaluation

Mering, von, C. et al. Comparative assessment of large-scale data sets of protein-protein interactions. Nature 417, 399–403 (2002).

EMBL-EBI

Computational Methods

- Gene Fusion Events
- Conservation of Gene Neighborhood
- Similarity of Phylogenetic Profiles
- Similarity of Phylogenetic Trees



Enright et al. Protein interaction maps for complete genomes based on gene fusion events. Nature (1999) vol. 402 (6757) pp. 86-90

Overbeek et al. Use of contiguity on the chromosome to predict functional coupling. In Silico Biol (1999) vol. 1 (2) pp. 93-108

Dandekar et al. Conservation of gene order: a fingerprint of proteins that physically interact. Trends Biochem Sci (1998) vol. 23 (9) pp. 324-328

Pellegrini et al. Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. Proc Natl Acad Sci U S A (1999) vol. 96 (8) pp. 4285-4288

Date y Marcotte. Discovery of uncharacterized cellular systems by genome-wide analysis of functional linkages. Nat Biotechnol (2003) vol. 21 (9) pp. 1055-1062

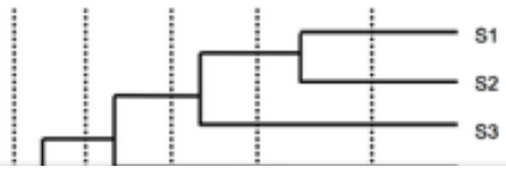
EMBL-EBI

organism 1	1	1
organism 2	1	1
organism 3	0	1
organism 4	1	1
organism 5	1	1
organism 6	1	0
organism 7	1	1
organism 8	1	1
organism 9	1	1
organism 10	1	1

Local Profiles: Domains

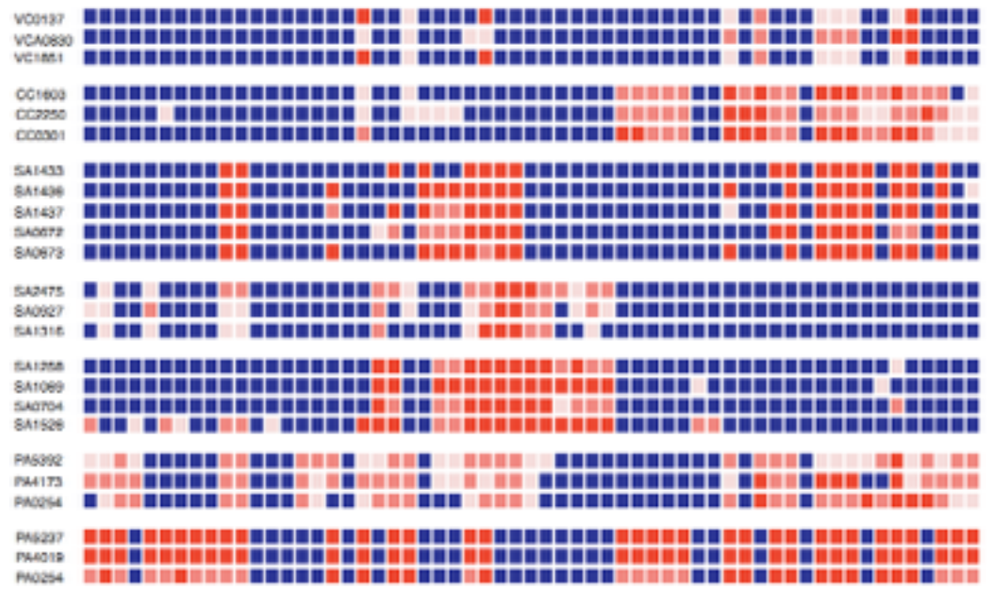
Pagel, P., Wong,
phylogenetic p

Selection of Organisms



Sun, J., Li, Y. &
protein intera
Res Commu

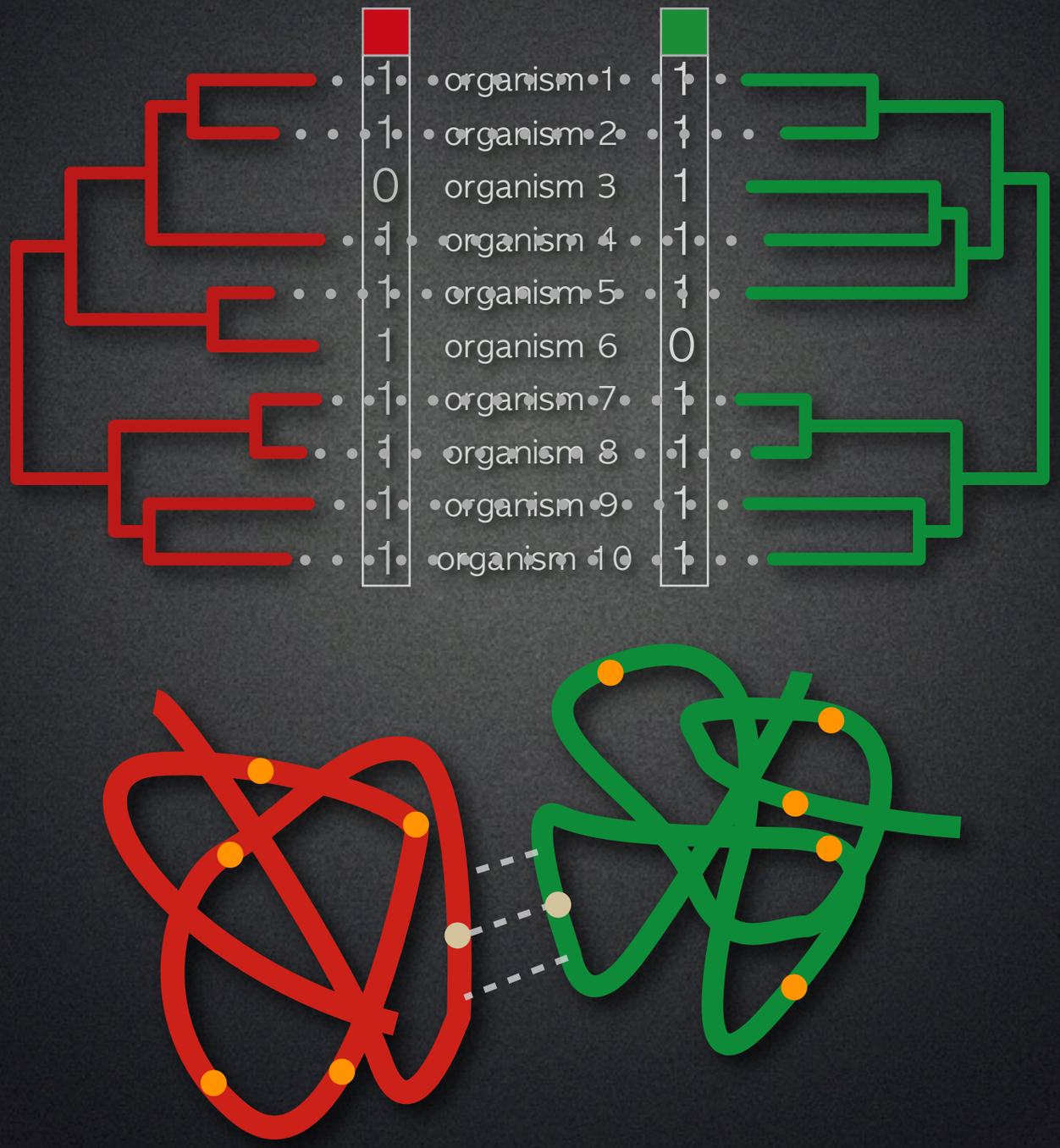
Quantitative Profiles



Date, S. V. & Marcotte, E. M. Discovery of uncharacterized cellular systems by genome-wide analysis of functional linkages. Nat Biotechnol 21, 1055–1062 (2003)

EMBL-EBI

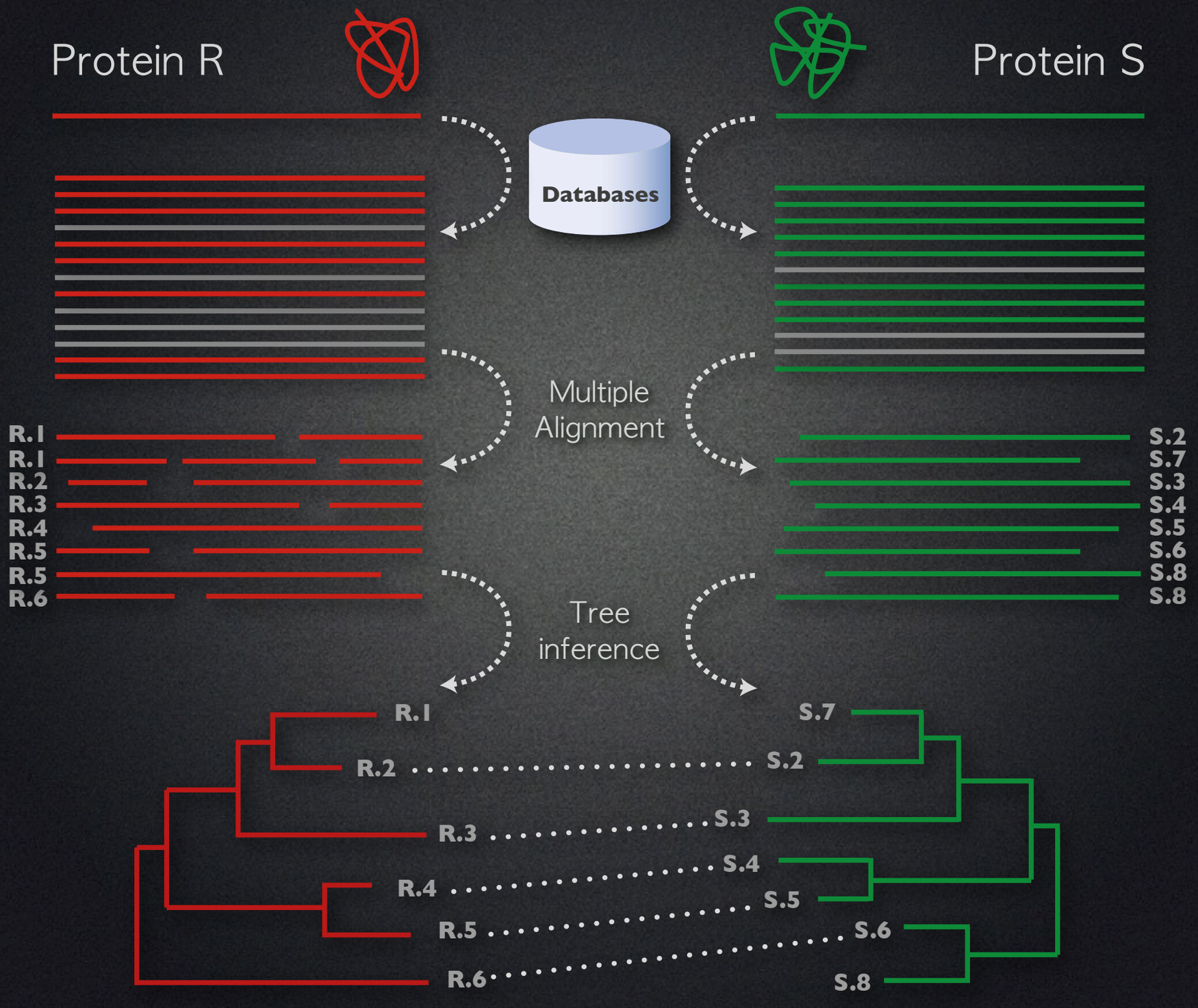
Similarity of Phylogenetic Trees



Fryxell, K. J. The coevolution of gene family trees. Trends Genet 12, 364–369 (1996).

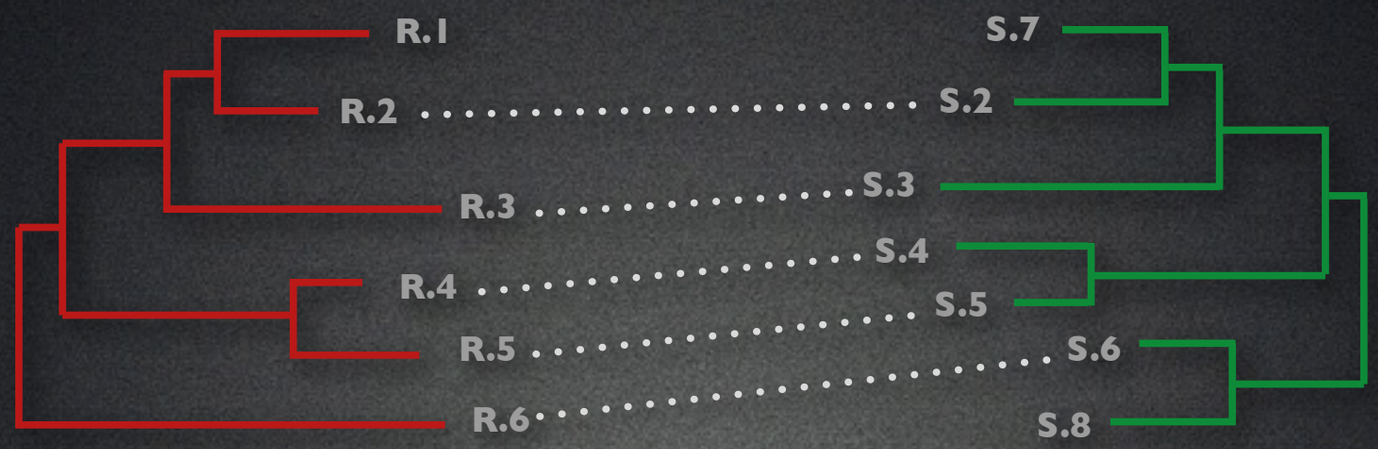
EMBL-EBI

MirrorTree



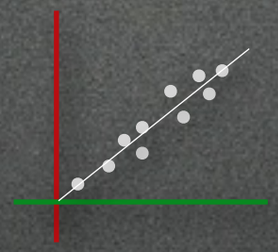
EMBL-EBI

MirrorTree



Distances matrix

	R1	R2	R3	R4	R5	R6
R1						
R2						
R3						
R4						
R5						
R6						



Pearson Correlation

$$r = \frac{\sum_{i=1}^{n}(R_i - \bar{R})(S_i - \bar{S})}{\sqrt{\sum_{i=1}^{n}(R_i - \bar{R})^2}\sqrt{\sum_{i=1}^{n}(S_i - \bar{S})^2}}$$

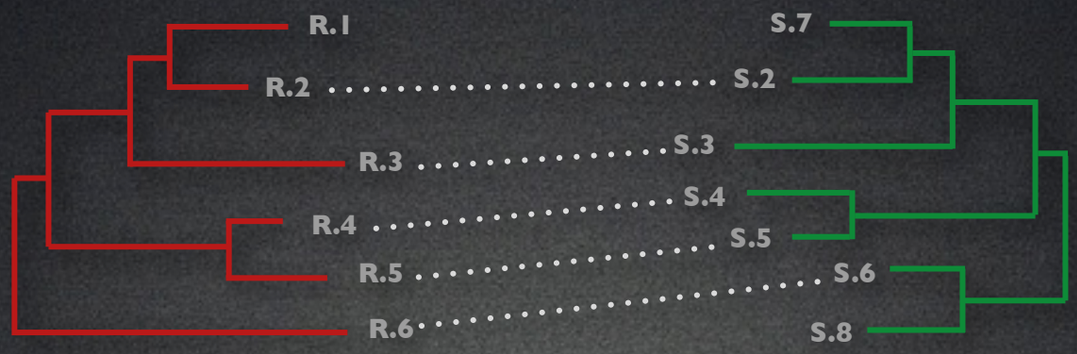
Distances matrix

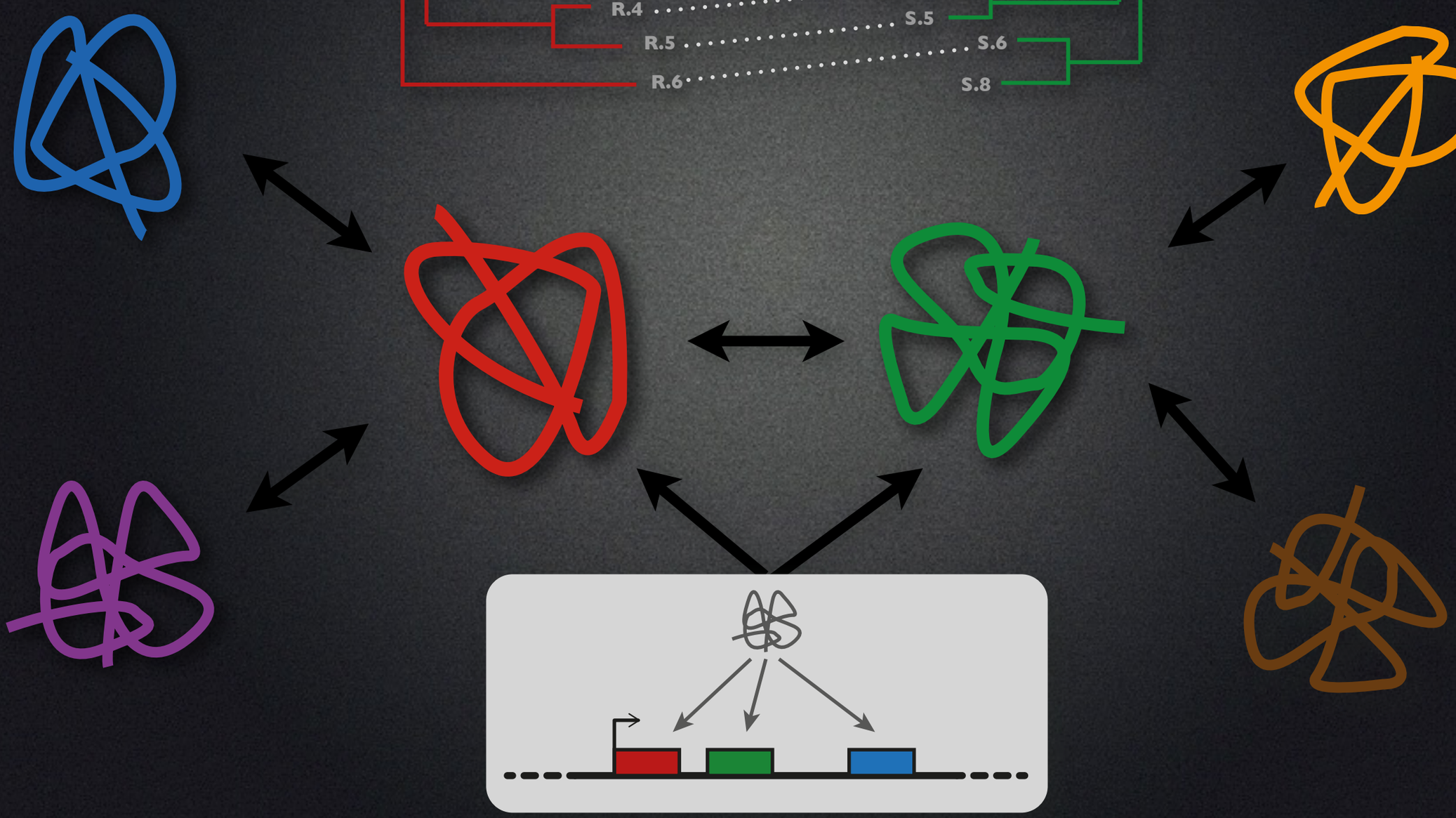
	S2	S3	S4	S5	S6	S7	S8
S2							
S3							
S4							
S5							
S6							
S7							
S8							

Pazos, F. & Valencia, A. Similarity of phylogenetic trees as indicator of protein-protein interaction. *Protein Eng* **14**, 609–614 (2001)

EMBL-EBI

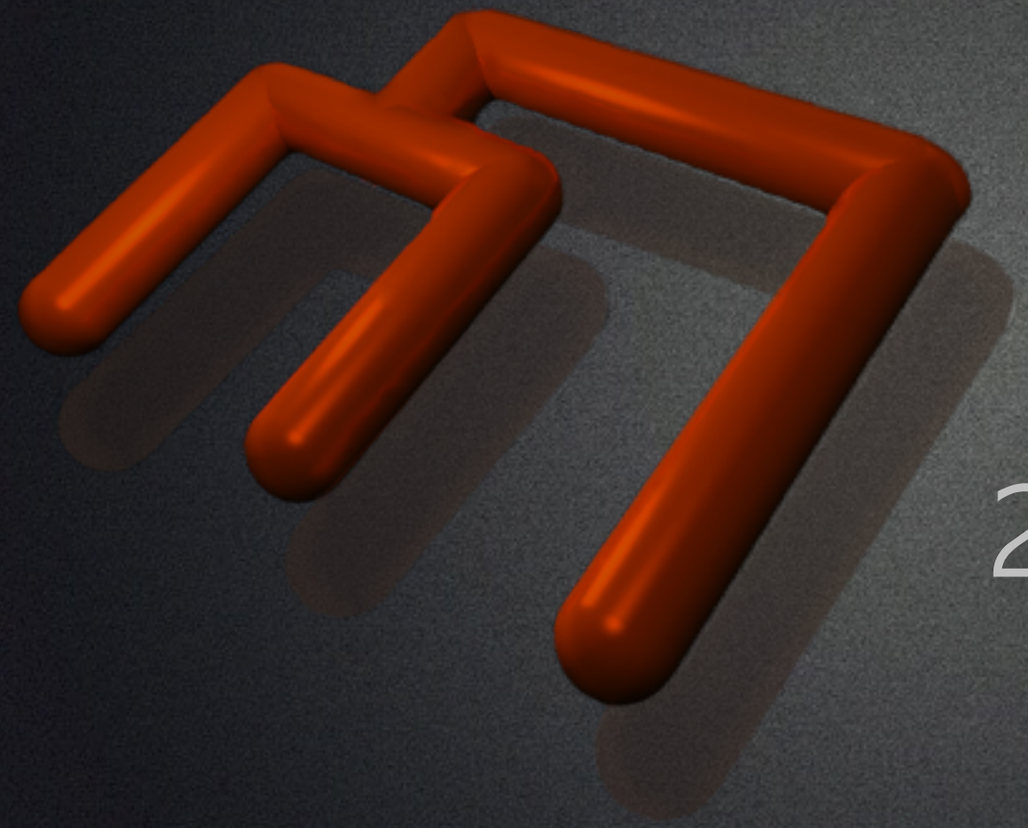
Possible causes of observed similarity





Pazos, F. & Valencia, A. Protein co-evolution, co-adaptation and interactions. EMBO J 27, 2648–2655 (2008).

EMBL-EBI

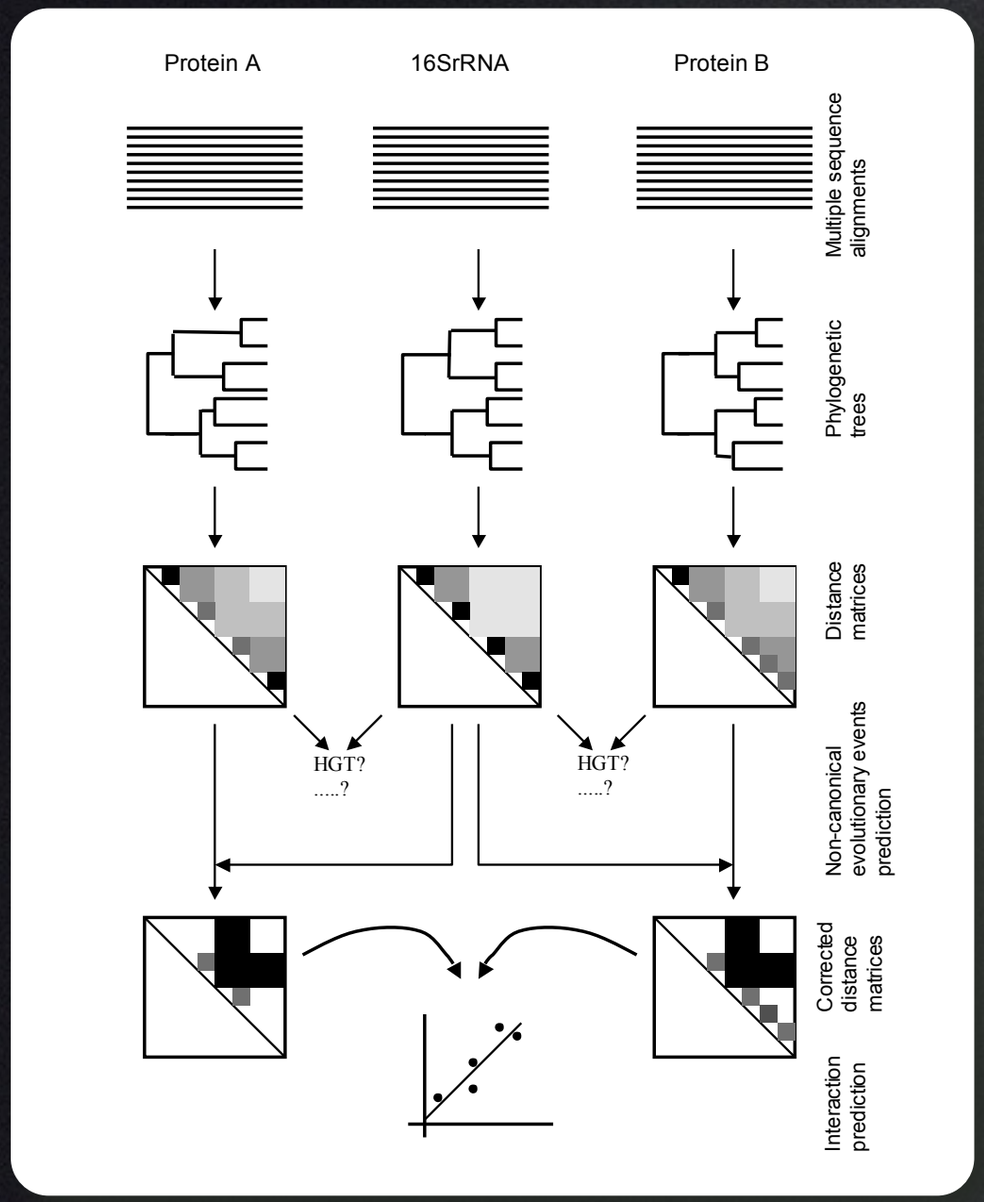


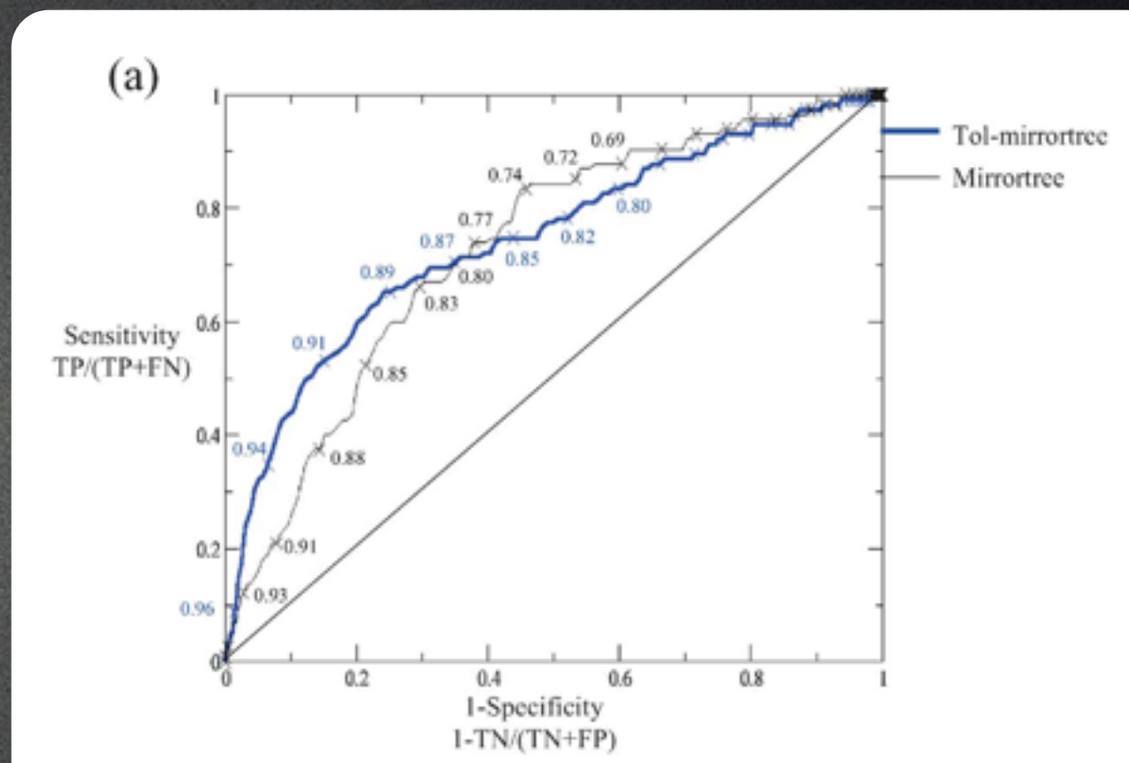
2.Practical aspects

Ochoa, D. & Pazos, F. Practical aspects of protein co-evolution. Molecular Medicine 2, (2014).

EMBL-EBI

TOL-MirrorTree



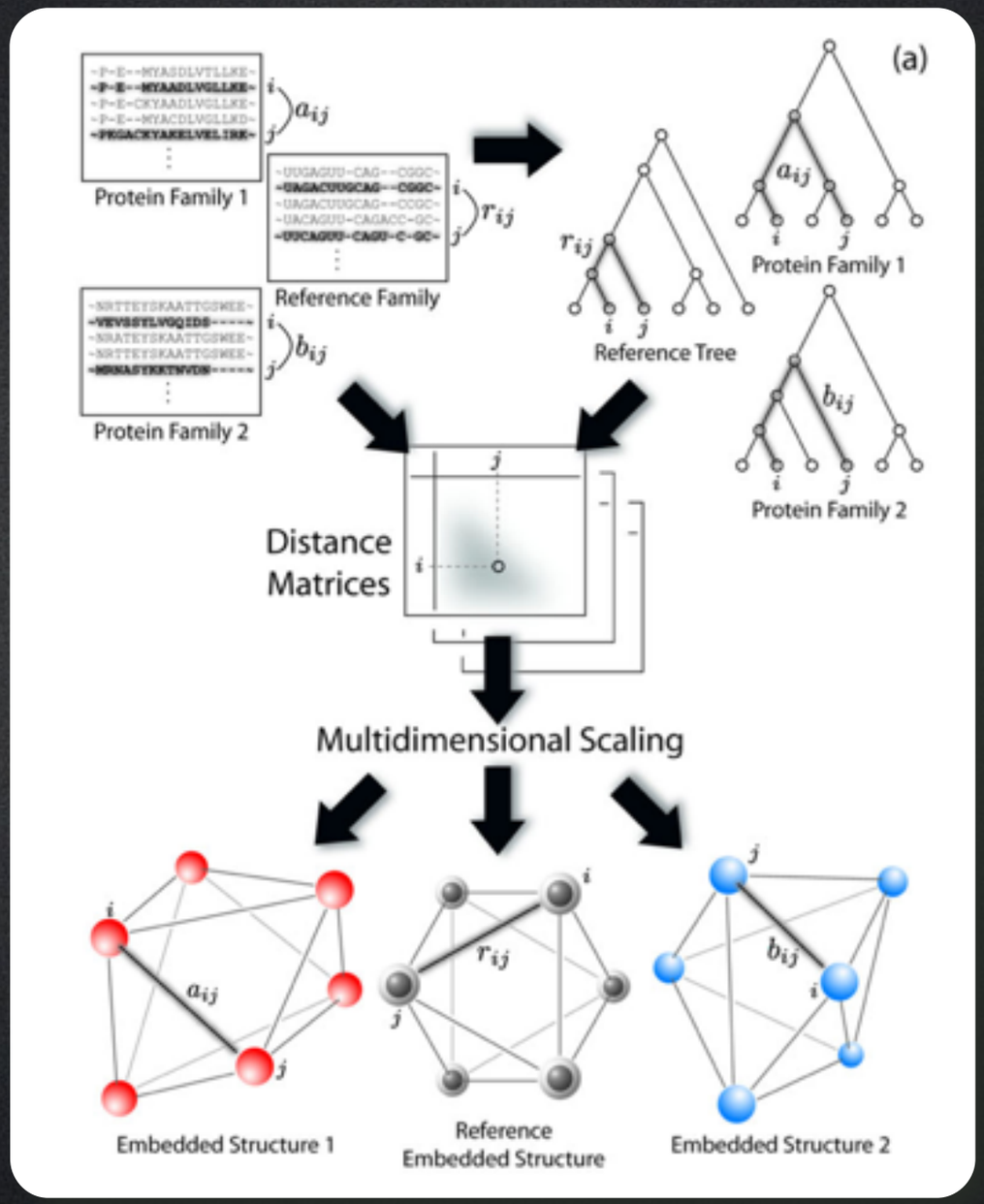


Sato, T., Yamanishi, Y., Kanehisa, M. & Toh, H. The inference of protein-protein interactions by co-evolutionary analysis is improved by excluding the information about the phylogenetic relationships. Bioinformatics 21, 3482–3489 (2005).

Pazos, F., Ranea, J. A. G., Juan, D. & Sternberg, M. J. E. Assessing protein co-evolution in the context of the tree of life assists in the prediction of the interactome. J Mol Biol 352, 1002–1015 (2005).

EMBL-EBI

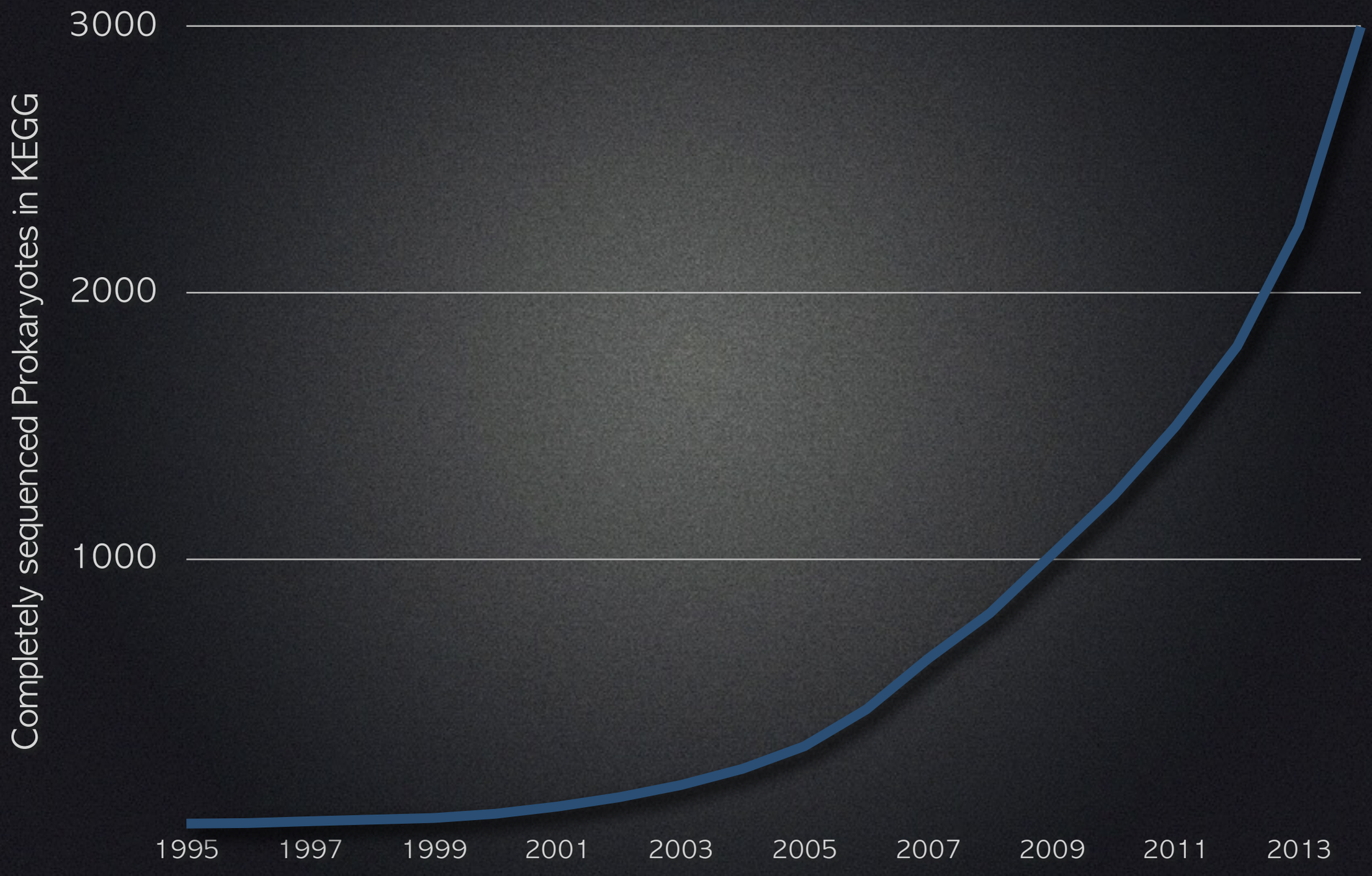
Tree of life corrections



Choi, K. & Gomez, S. M. Comparison of phylogenetic trees through alignment of embedded evolutionary distances. BMC Bioinformatics 10, 423 (2009).

EMBL-EBI

Sequence explosion

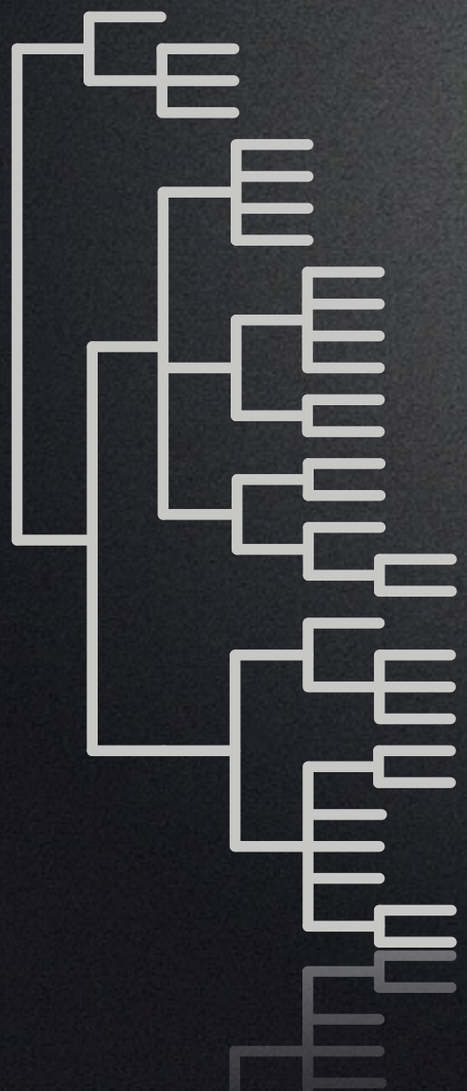


EMBL-EBI

The redundancy problem

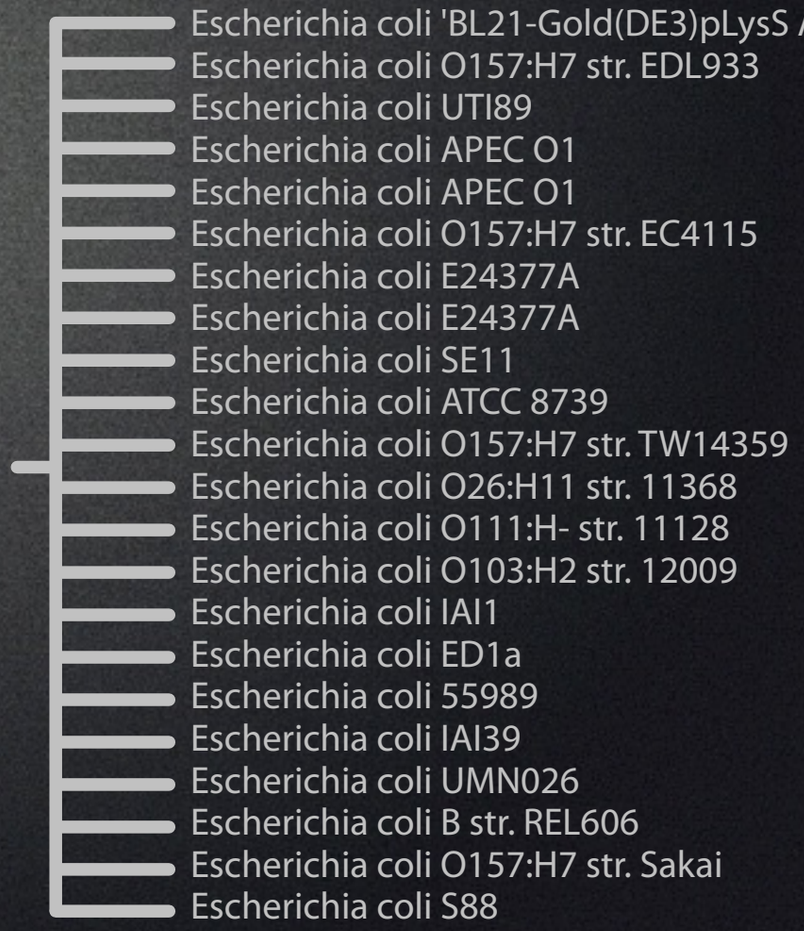
2001

70 sequenced prokaryotes

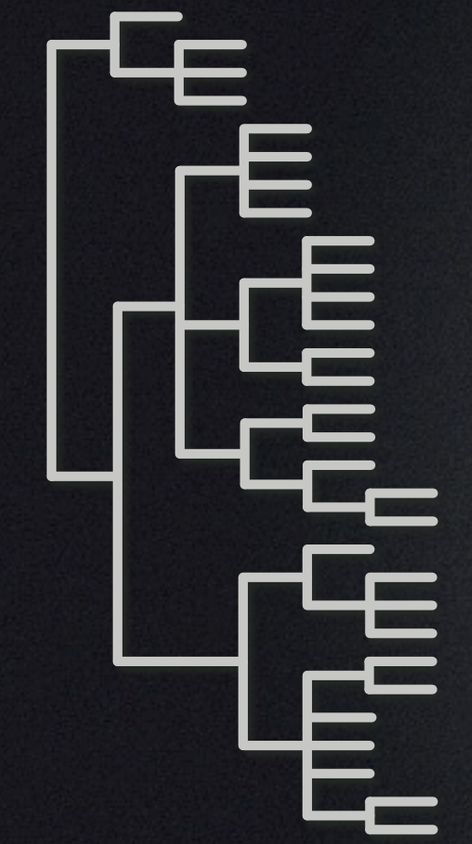


2014

~3,000 sequenced prokaryotes



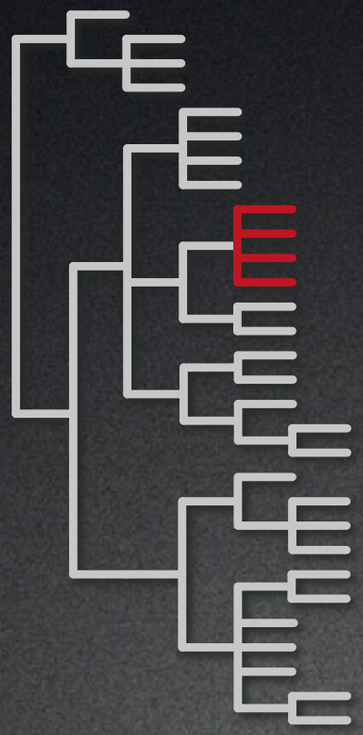
EMBL-EBI



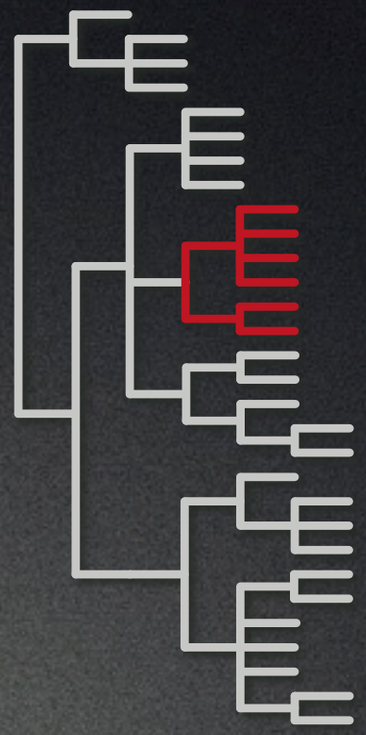
NCBI taxonomy tree

Nearest

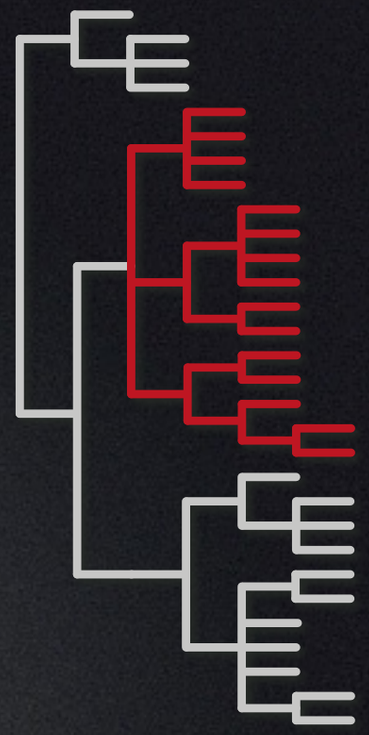
Level



Nearest 1

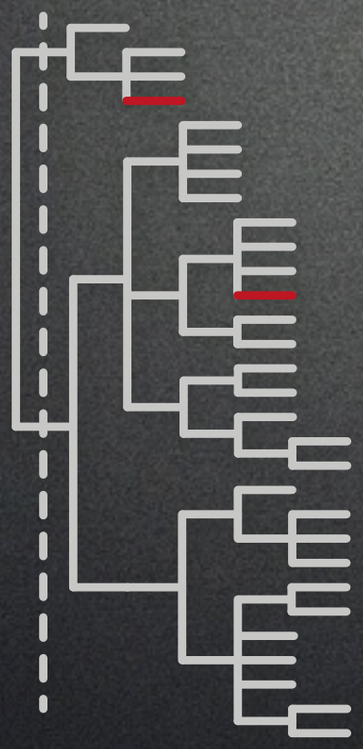


Nearest 2

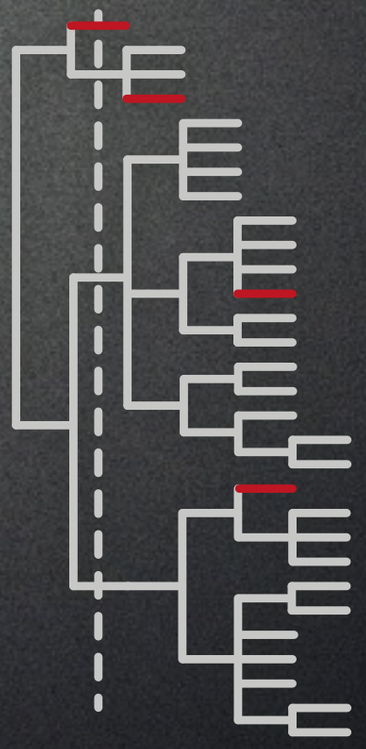


Nearest 3

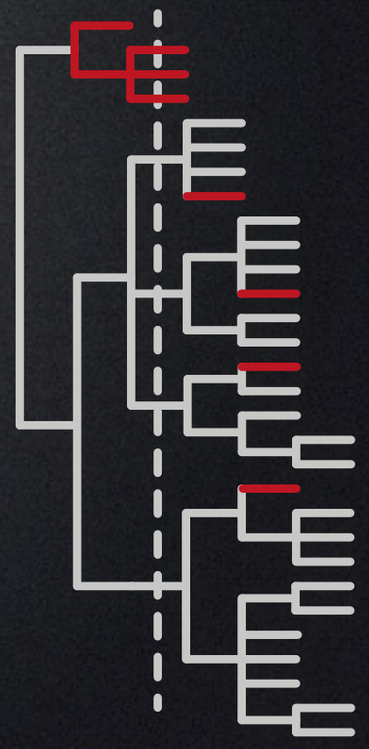
...



Level 1



Level 2



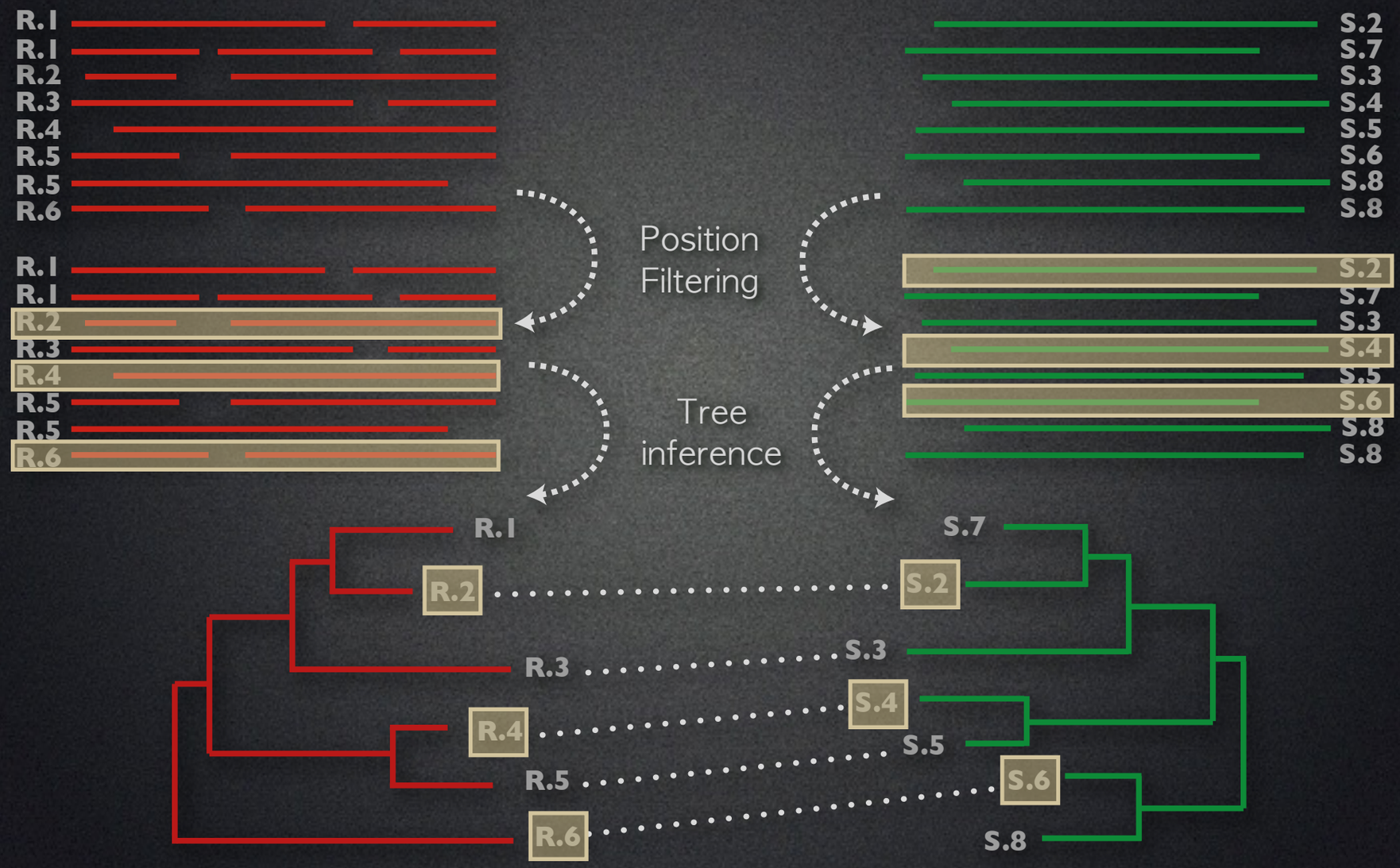
Level 3

...

Herman, D. et al. Selection of organisms for the co-evolution-based study of protein interactions. BMC Bioinformatics 12, 363 (2011).

EMBL-EBI

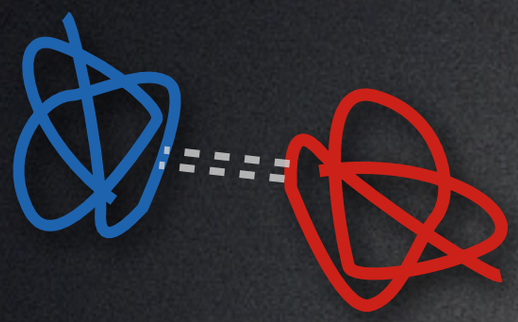
Adapted workflow



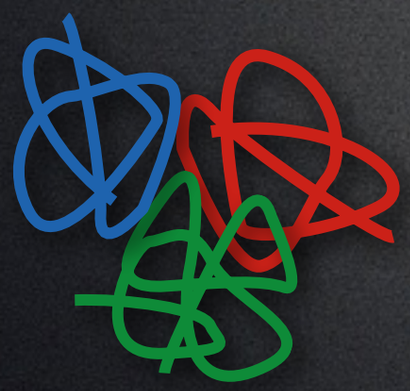
EMBL-EBI

Redundancy effect

EMBL-EBI



recent



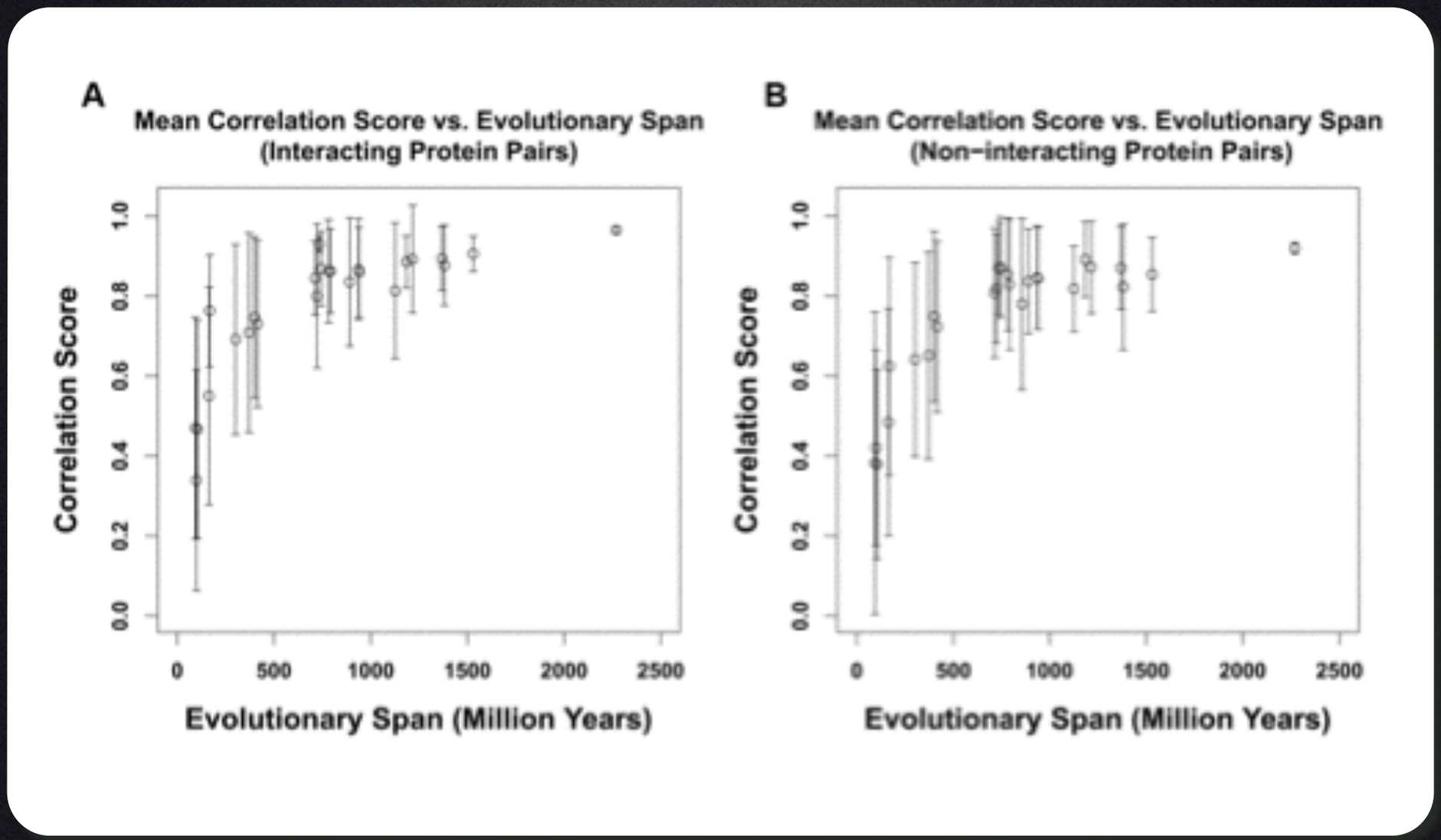
old

	AUCs	
	Level9 (=all)	Nearest2
MINE_ECOLI	0.12	**0.83**
PABA_ECOLI	0.28	**0.96**
DHAS_ECOLI	0.17	**0.81**
GSHB_ECOLI	0.3	**0.93**
DPO3A_ECOLI	**0.7**	0.11
DPO3B_ECOLI	**0.64**	0.22
RPOB_ECOLI	**0.82**	0.48
RPOA_ECOLI	**0.81**	0.48
ZNUB_ECOLI	**1**	0.36
ZNUC_ECOLI	**0.99**	0.41
ZNUA_ECOLI	**0.98**	0.79

Herman, D. et al. Selection of organisms for the co-evolution-based study of protein interactions. BMC Bioinformatics 12, 363 (2011).

EMBL-EBI

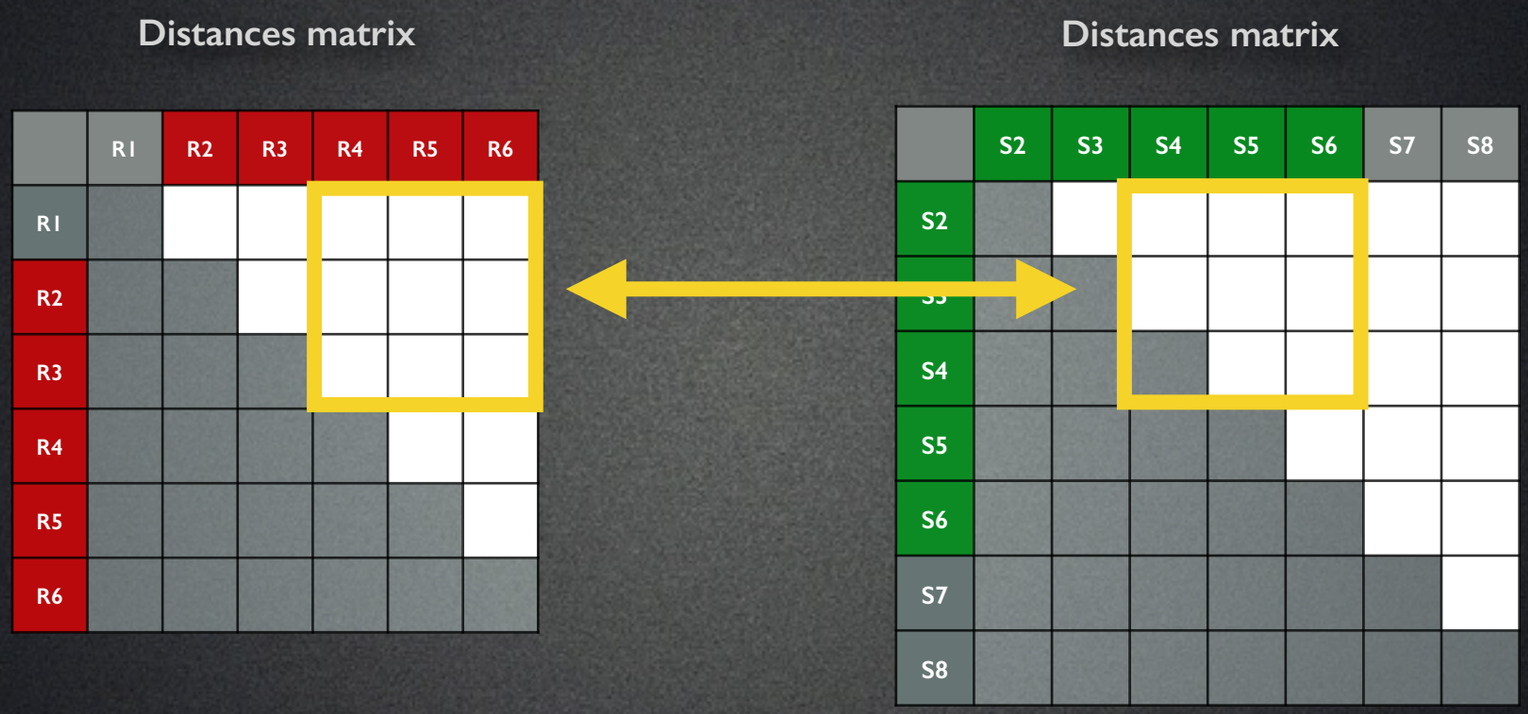
Correcting by evolutionary span



Zhou, H. & Jakobsson, E. Predicting Protein-Protein Interaction by the Mirrortree Method: Possibilities and Limitations. PLoS ONE 8, e81100 (2013).

EMBL-EBI

MMM

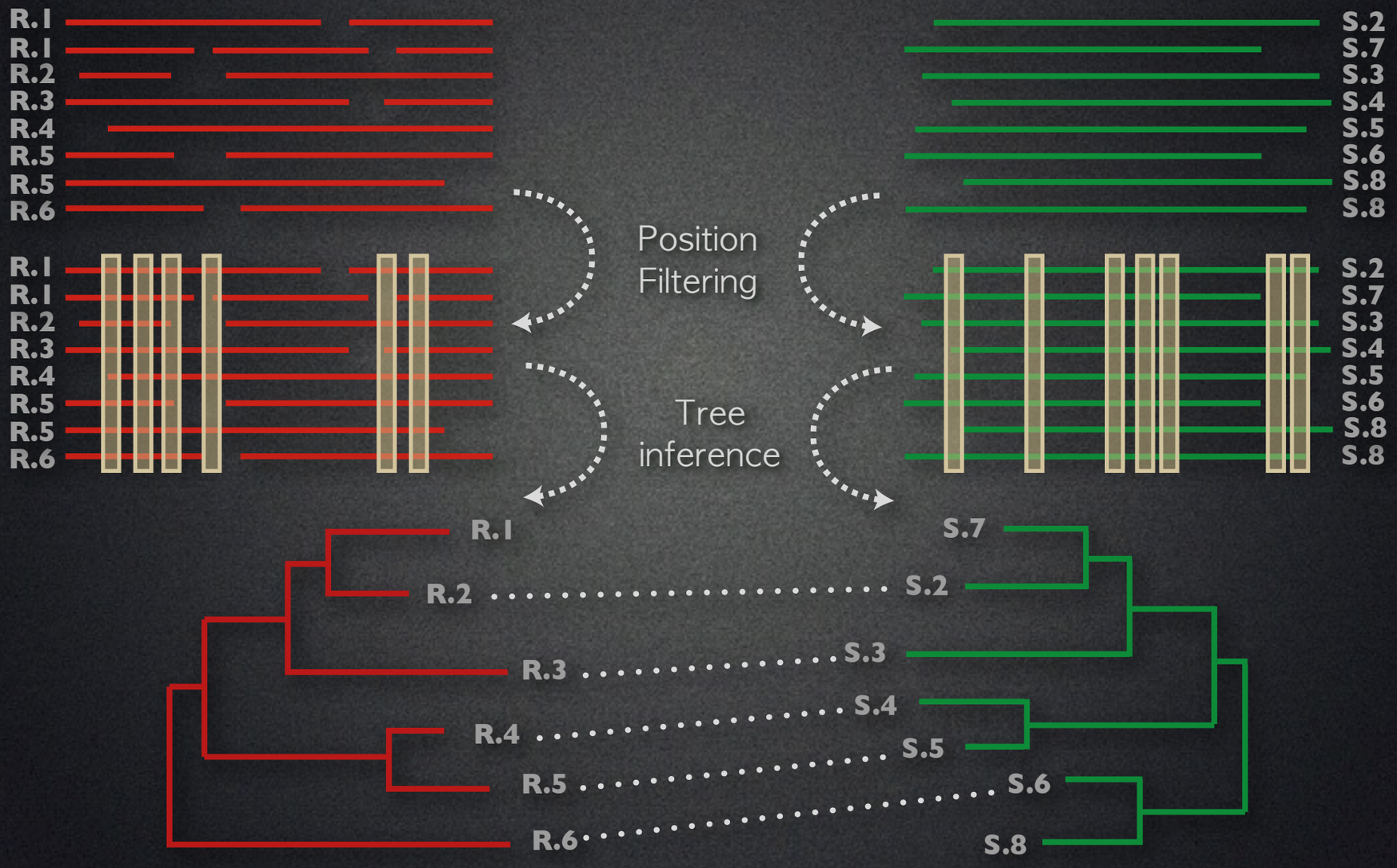


http://www.uhnresearch.ca/labs/tillier/MMMWEBvII/MMMWEBvII.php

Tillier, E. R. M. & Charlebois, R. L. The human protein coevolution network. Genome research 19, 1861–1871 (2009)

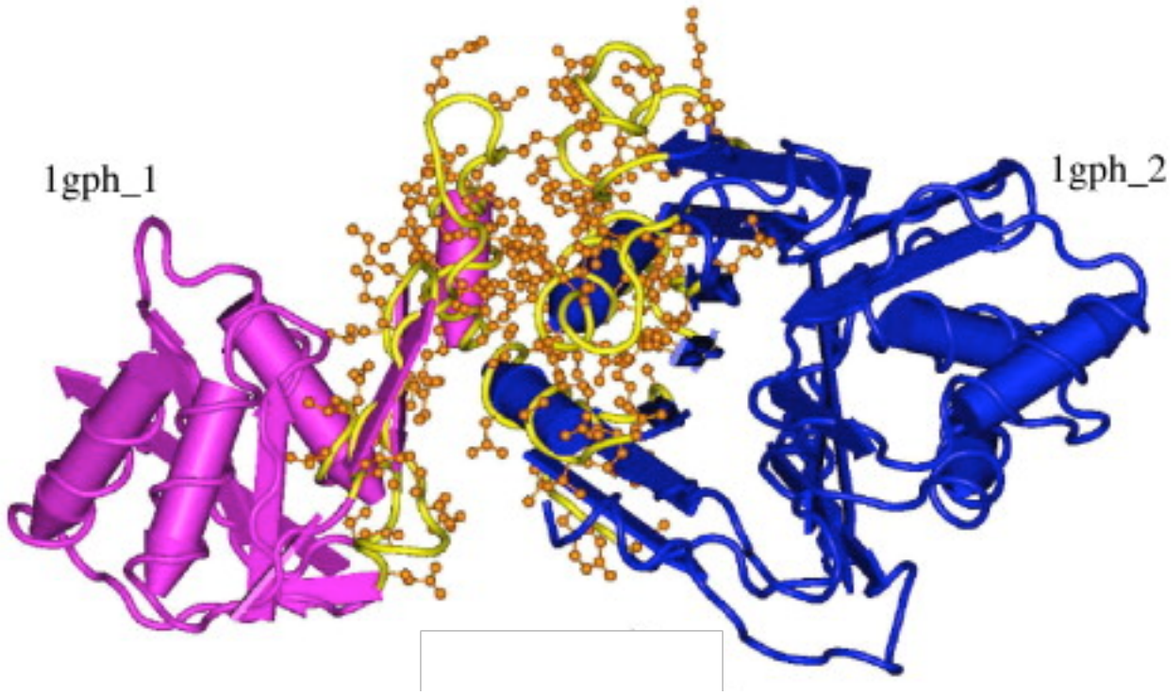
EMBL-EBI

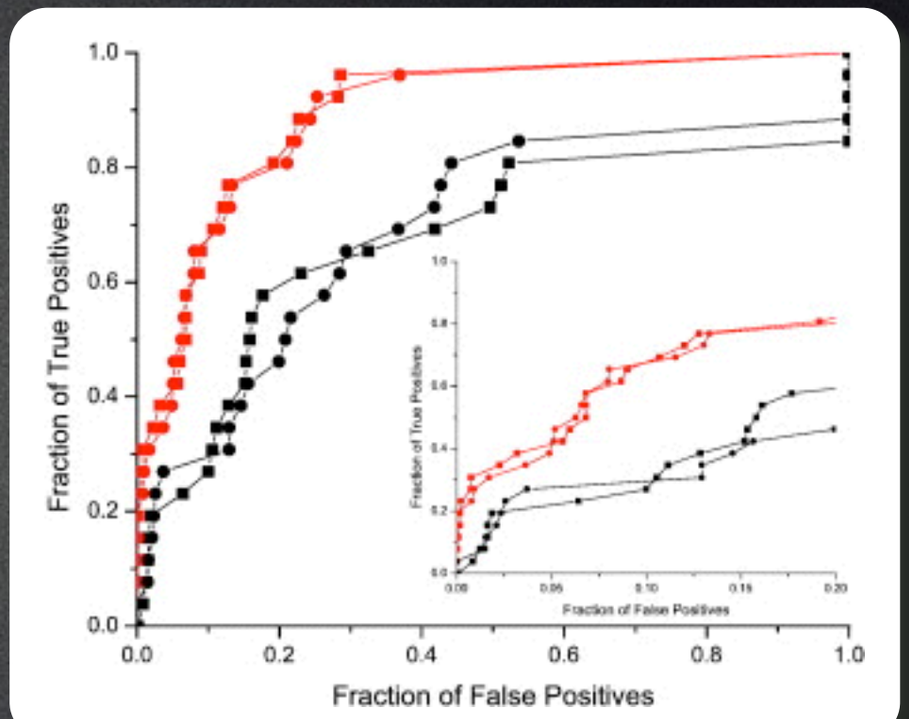
Adapted workflow



EMBL-EBI

Residue filtering





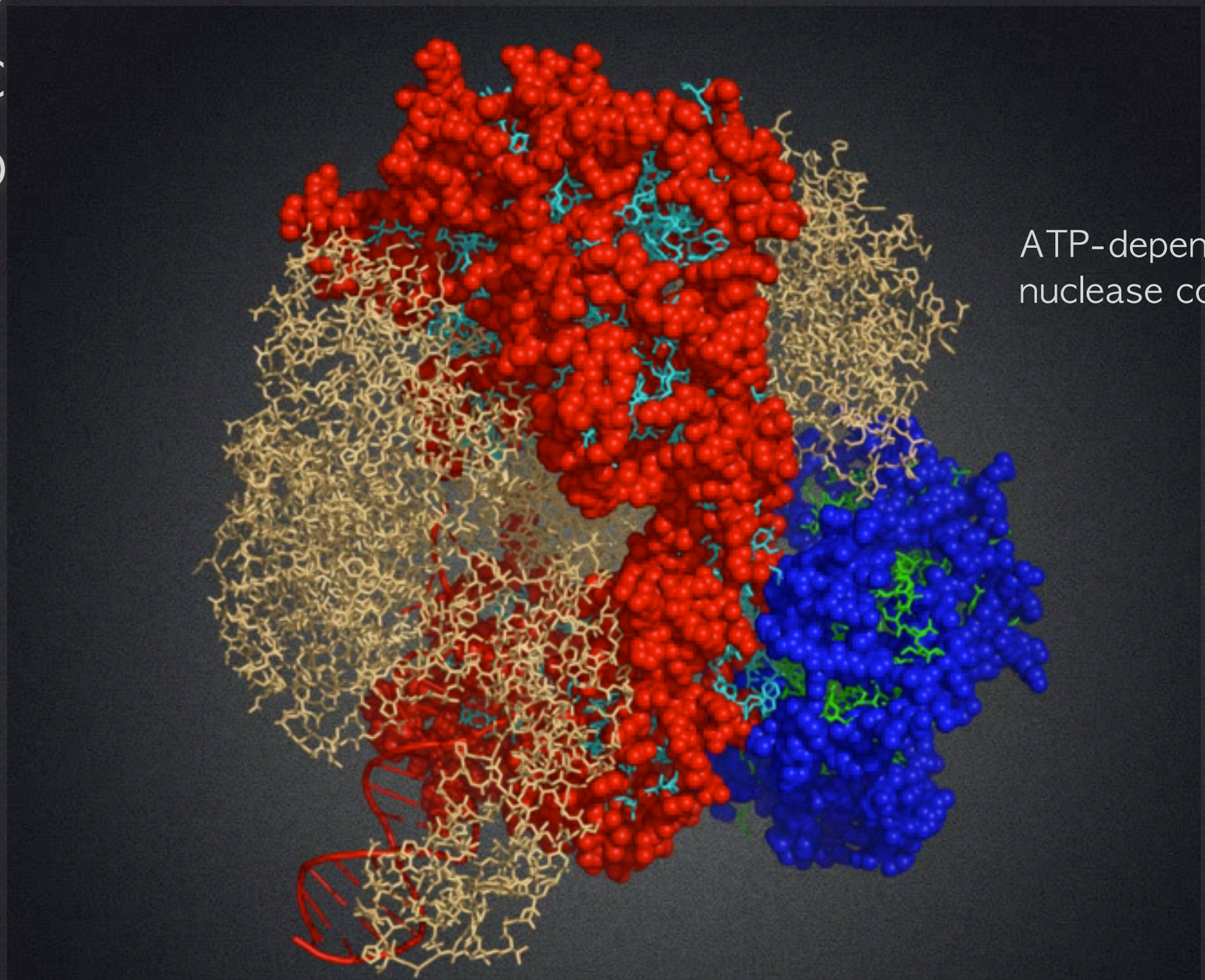
Kann, M. G., Shoemaker, B. A., Panchenko, A. R. & Przytycka, T. M. Correlated evolution of interacting proteins: looking behind the mirrortree. J Mol Biol 385, 91–98 (2009).

EMBL-EBI

RecB

RecC

RecD



pACC2

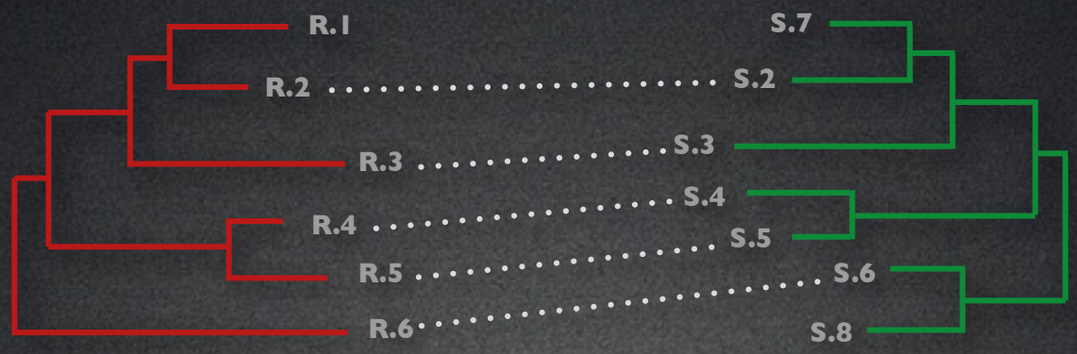
ATP-dependent helicase/
nuclease complex

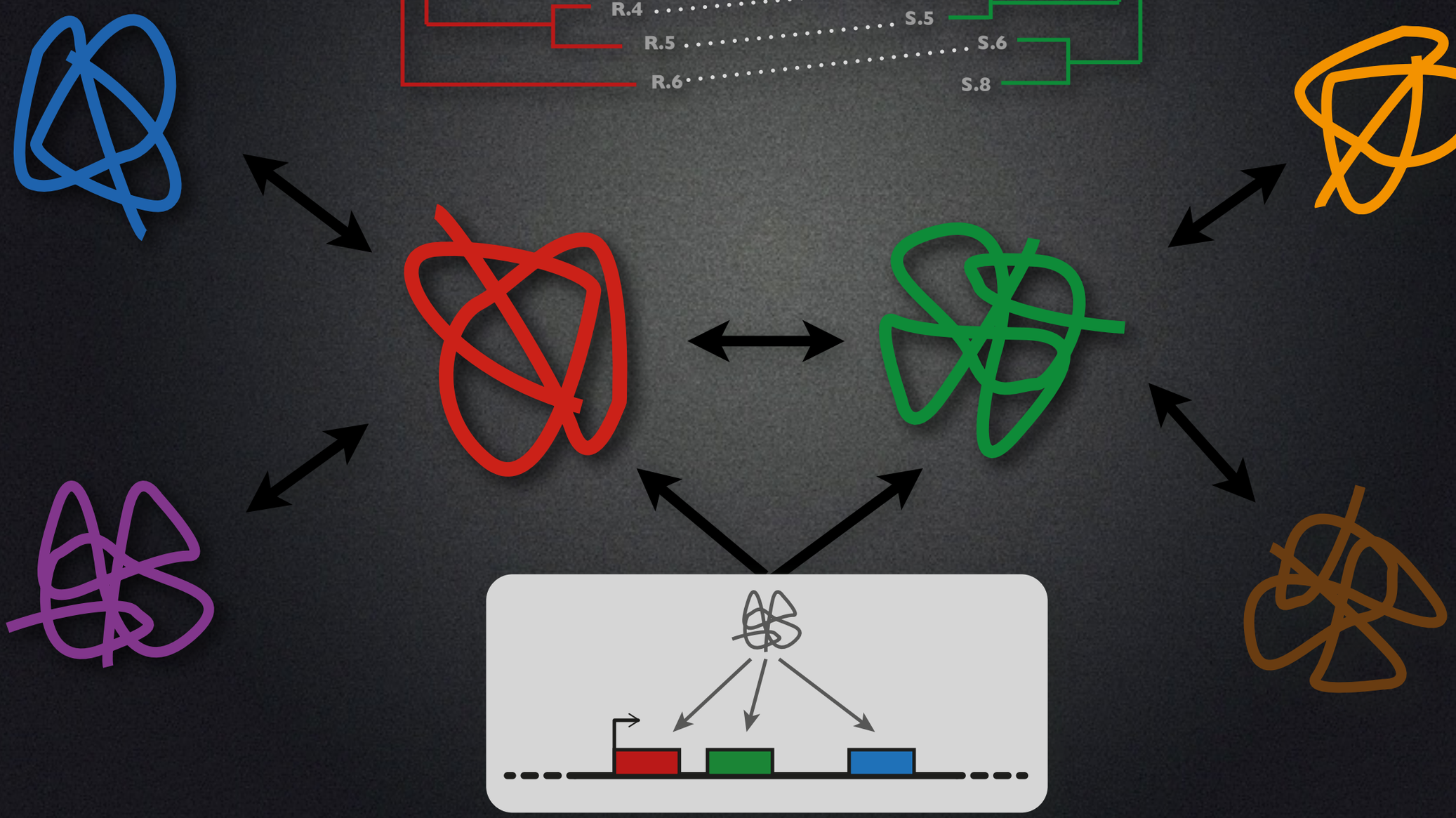
		ALL	eRIA0	eRIA3	pACC2	pACC12	pACC50
✔	RecC	0.701	0.768	Not significant	0.739	0.75	0.806
✘	RecB	0.427	Not significant	Not significant	Not significant	Not significant	Not significant

Ochoa, D., García-Gutiérrez, P., Juan, D., Valencia, A. & Pazos, F. Incorporating information on predicted solvent accessibility to the co-evolution-based study of protein interactions. Molecular bioSystems 9, 70–76 (2013).

EMBL-EBI

Possible causes of observed similarity

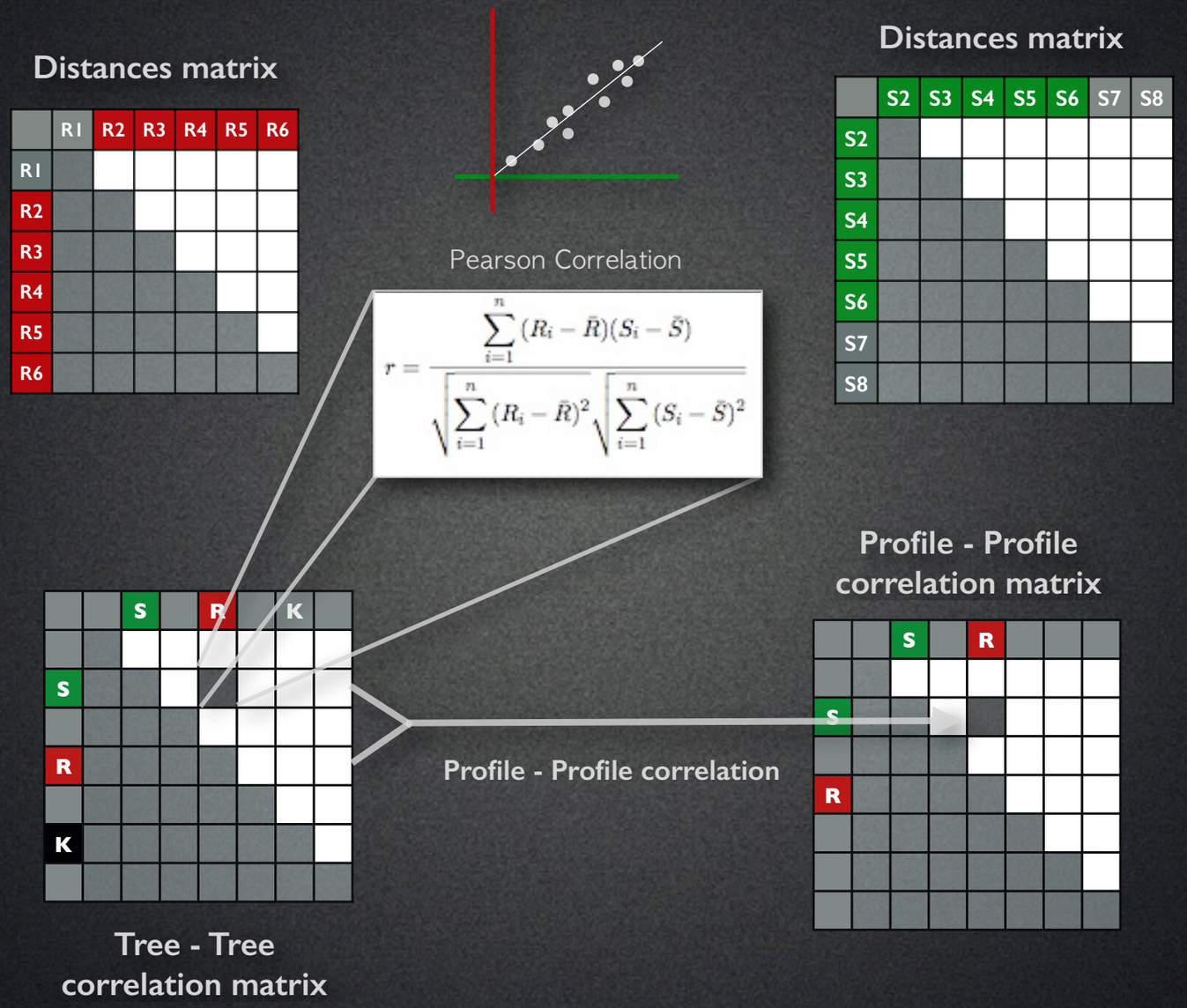




Pazos, F. & Valencia, A. Protein co-evolution, co-adaptation and interactions. EMBO J 27, 2648–2655 (2008).

EMBL-EBI

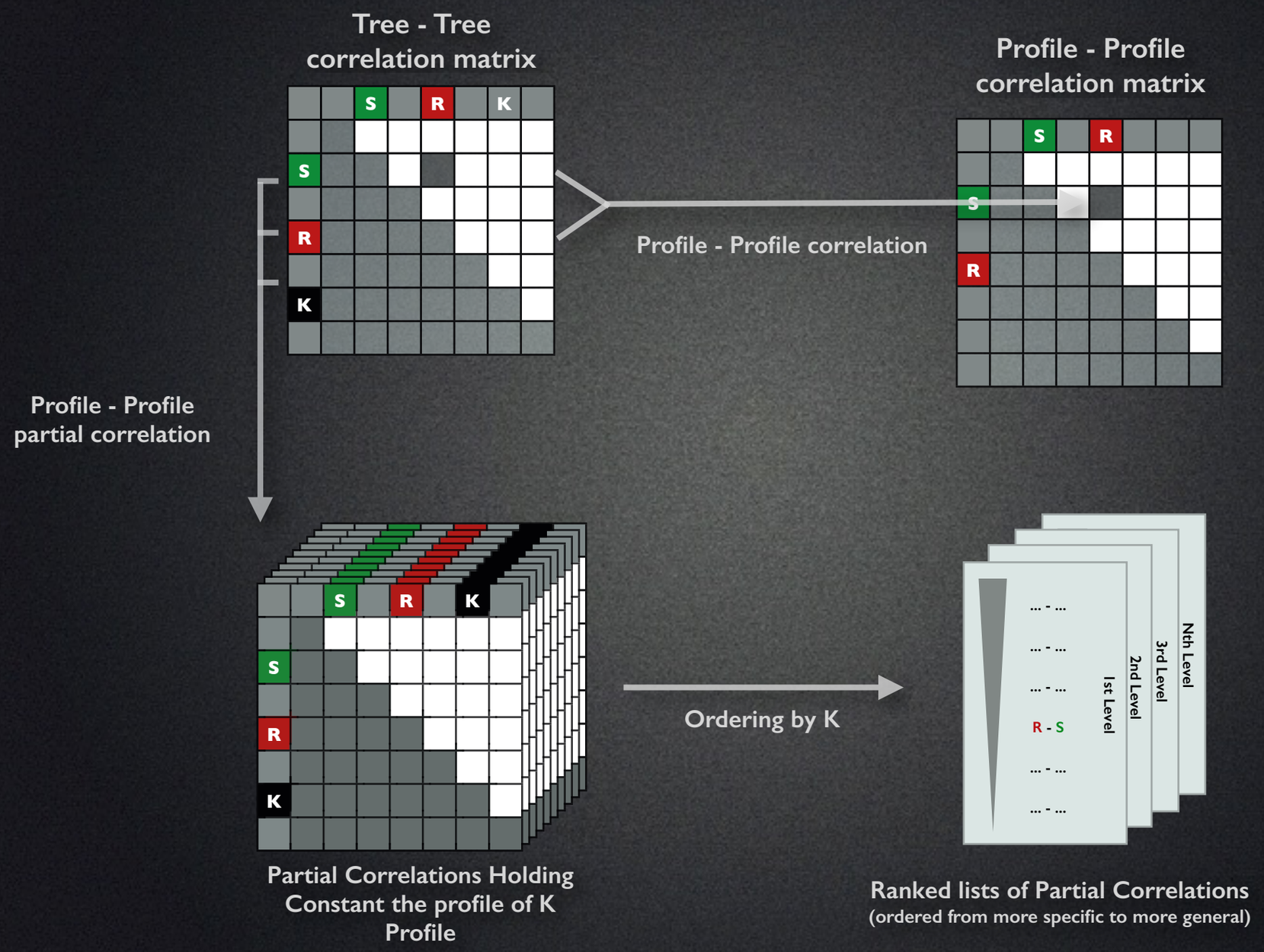
Profile Correlation



Juan, D., Pazos, F. & Valencia, A. High-confidence prediction of global interactomes based on genome-wide coevolutionary networks. *Proc Natl Acad Sci U S A* **105**, 934–939 (2008)

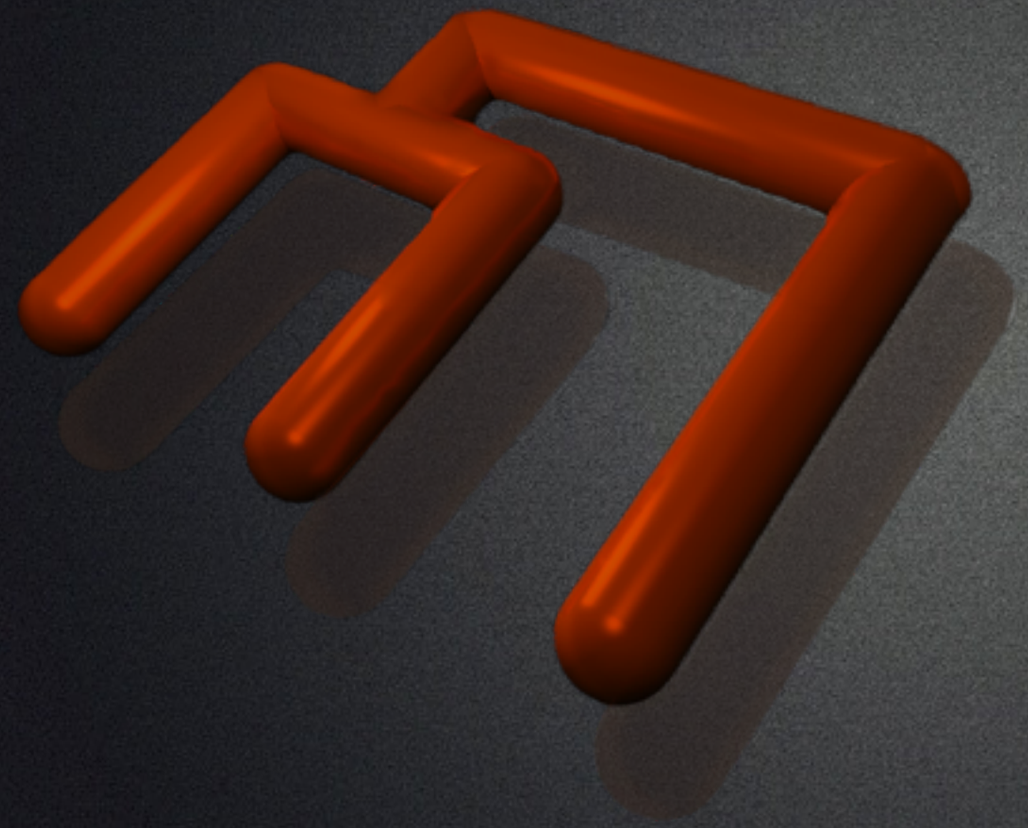
EMBL-EBI

ContextMirror



Juan, D., Pazos, F. & Valencia, A. High-confidence prediction of global interactomes based on genome-wide coevolutionary networks. *Proc Natl Acad Sci U S A* **105**, 934–939 (2008)

EMBL-EBI



3. Applications

EMBL-EBI

Ligand-receptor interactions

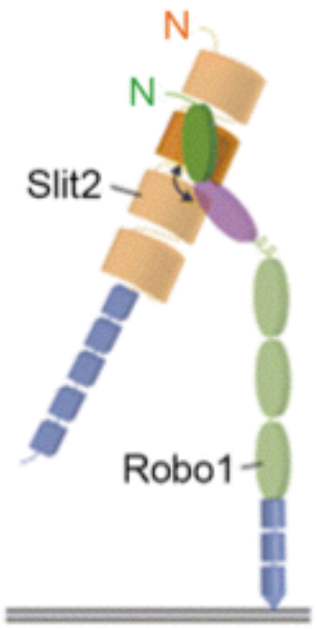


Table 1. Pearson's correlation coefficient of evolutionary distances between Slit and Robo in vertebrates.

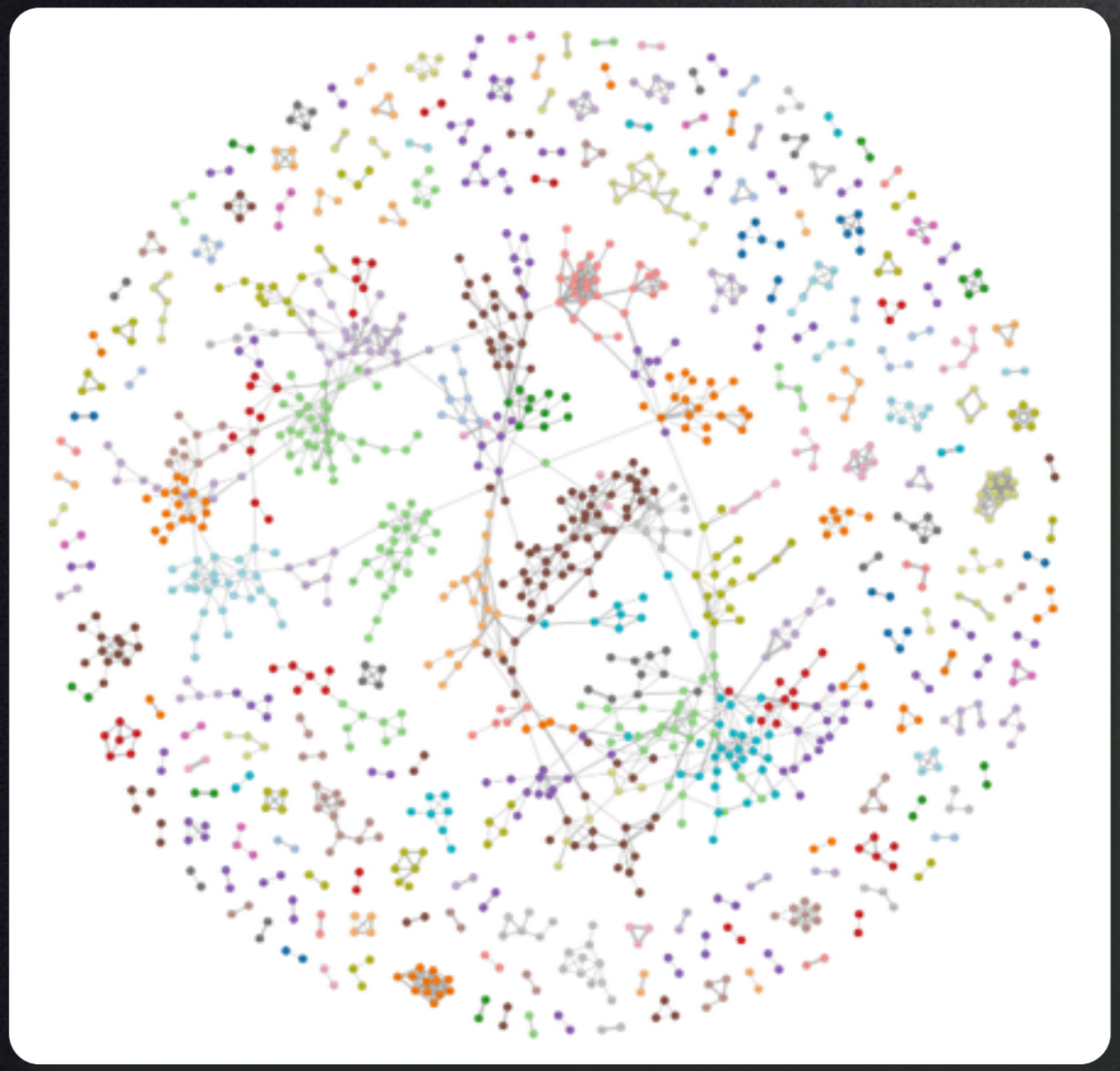
	Slit1	Slit2	Slit3	Gapdh
Robo1	0.949**	0.991**	0.961**	0.790
Robo2	0.945**	0.980**	0.961**	0.819
Robo3	0.890**	0.738	0.850**	0.896
Gapdh	0.814	0.757	0.779	1

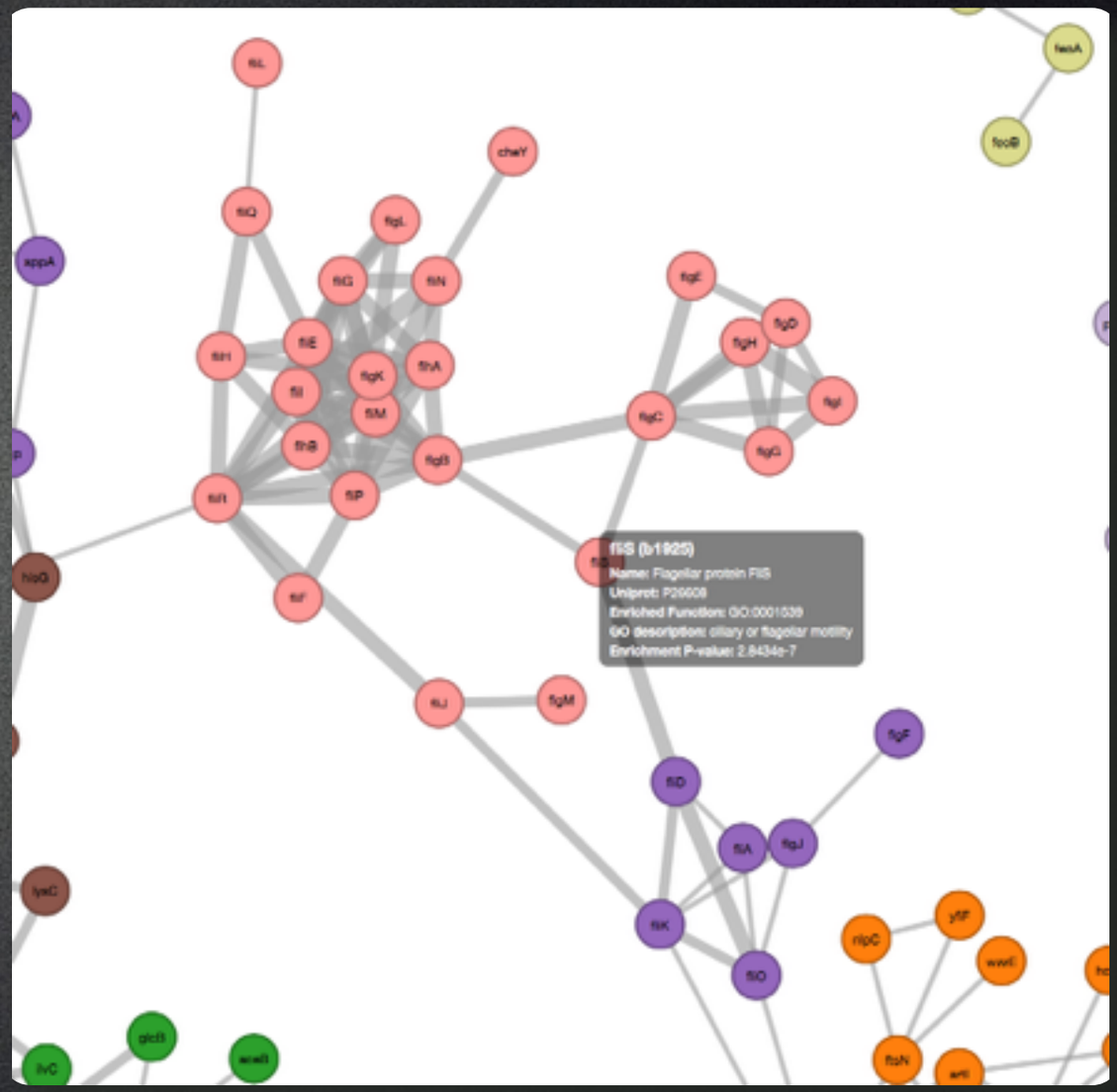
**the correlation value of Slit-Robo pair is significantly different from controls at 0.01 level.
doi:10.1371/journal.pone.0094970.t001

Yu, Q. et al. Coevolution of axon guidance molecule Slit and its receptor Robo. PLoS ONE 9, e94970 (2014).

EMBL-EBI

Genome-wide PPI prediction





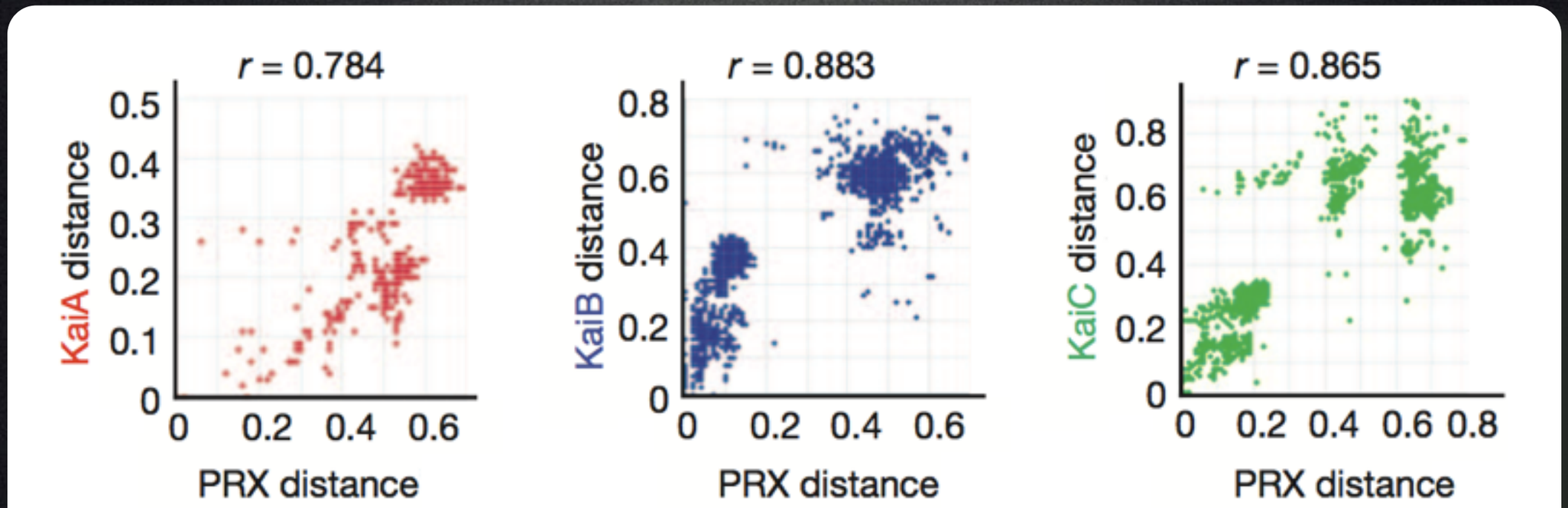
E.coli protein interactome

http://csbg.cnb.csic.es/colievolution

Ochoa, D., Juan, D., Valencia A., Pazos, F. Detection of significant protein co-evolution. Bioinformatics (under review)

EMBL-EBI

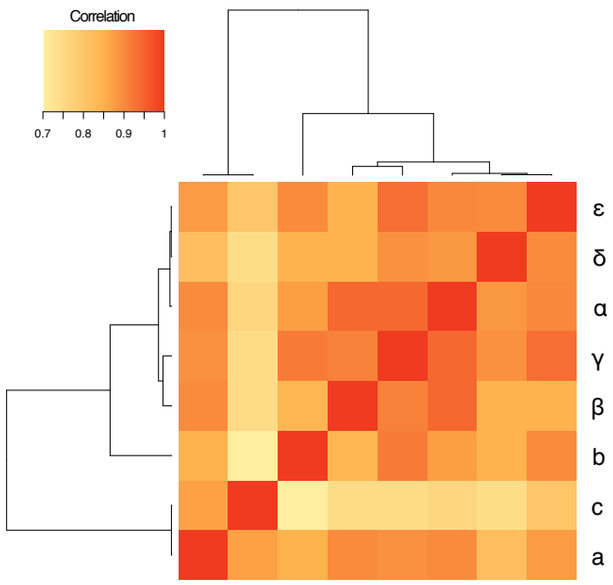
As functional evidence

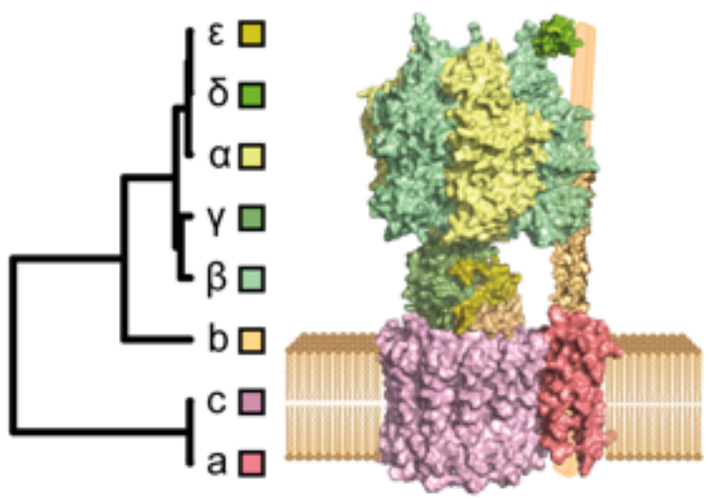


Edgar, R. S. et al. Peroxiredoxins are conserved markers of circadian rhythms. Nature 485, 459–464 (2012).

EMBL-EBI

Protein complexes



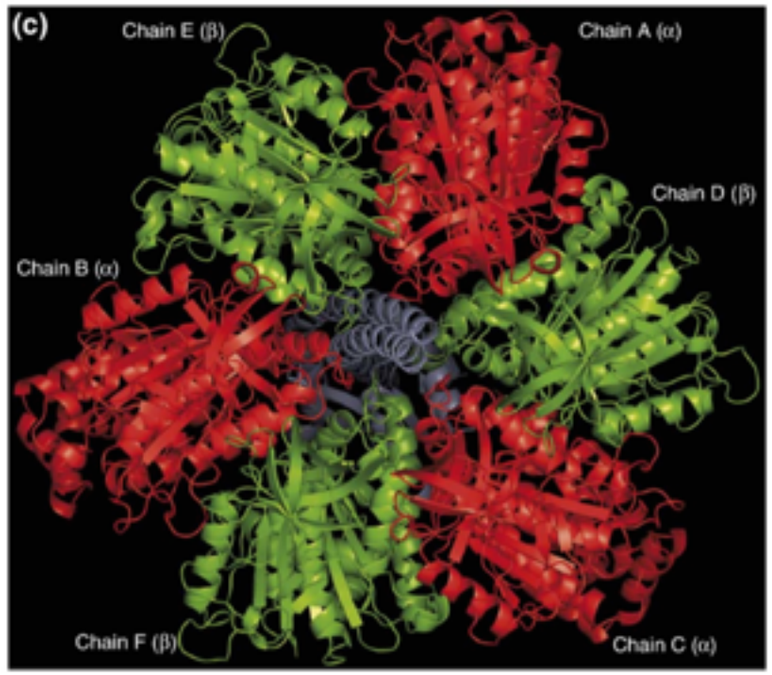


Ochoa, D., Juan, D., Valencia A., Pazos, F. Detection of significant protein co-evolution. Bioinformatics (under review)

YBL099w	YJR121w	Correlation	iPfam
PF00006	PF00006	0.95957039	Y
PF02874	PF00006	0.92390131	Y
PF00306	PF00306	0.89734590	Y
PF00006	PF02874	0.89692159	Y
PF02874	PF02874	0.88768393	Y
PF00006	PF00306	0.87369242	Y
PF00306	PF00006	0.86507957	Y
PF02874	PF00306	0.85735773	
PF00306	PF02874	0.84890155	

YBL099w	YBR039w	Correlation	iPfam
PF00306	PF00231	0.934749207	Y
PF00006	PF00231	0.928698402	Y
PF02874	PF00231	0.892046898	

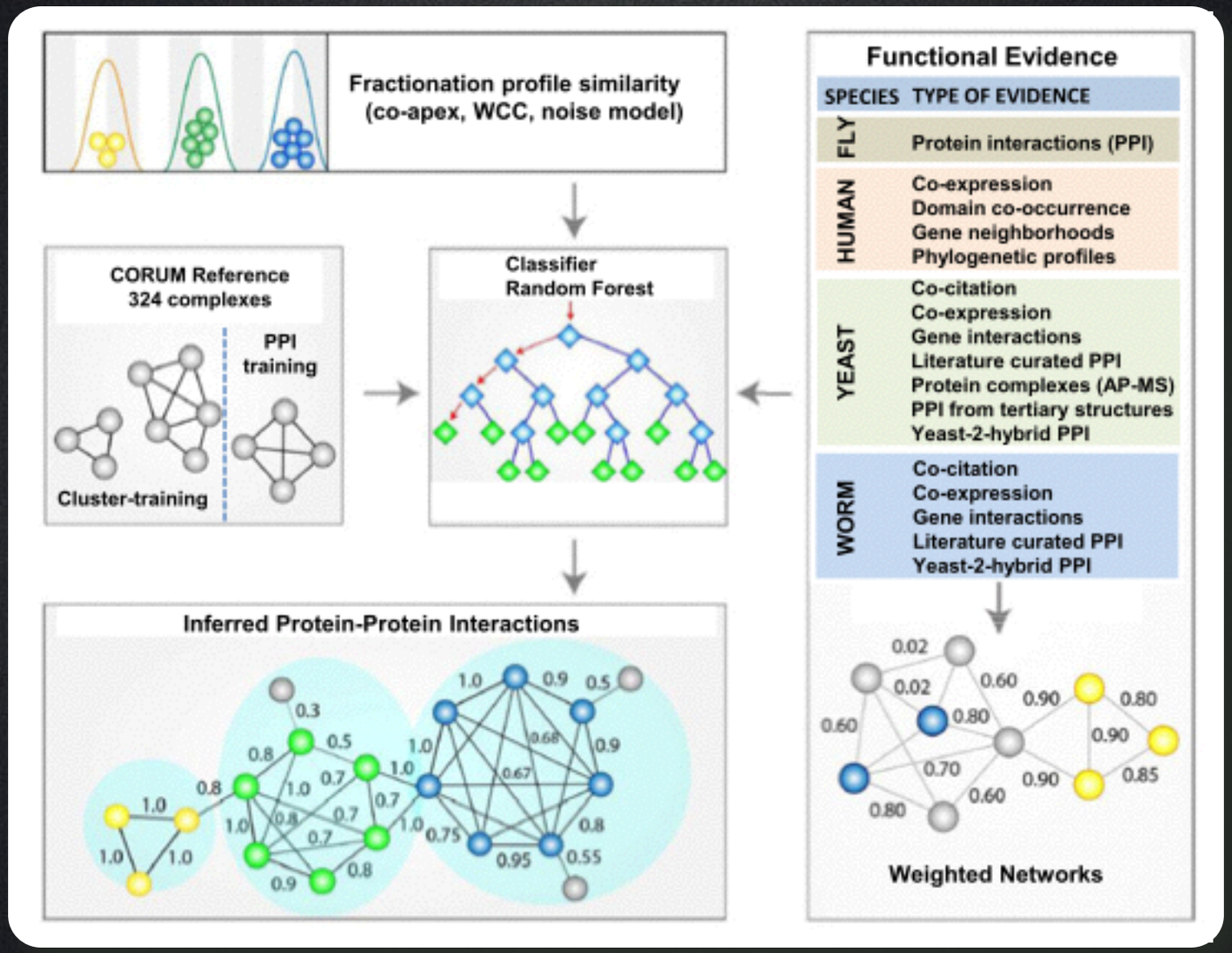
YJR121w	YBR039w	Correlation	iPfam
PF00306	PF00231	0.913245128	Y
PF00006	PF00231	0.872234447	Y
PF02874	PF00231	0.852998697	



Jothi, R., Cherukuri, P. F., Tasneem, A. & Przytycka, T. M. Co-evolutionary analysis of domains in interacting proteins reveals insights into domain-domain interactions mediating protein-protein interactions. J Mol Biol 362, 861–875 (2006).

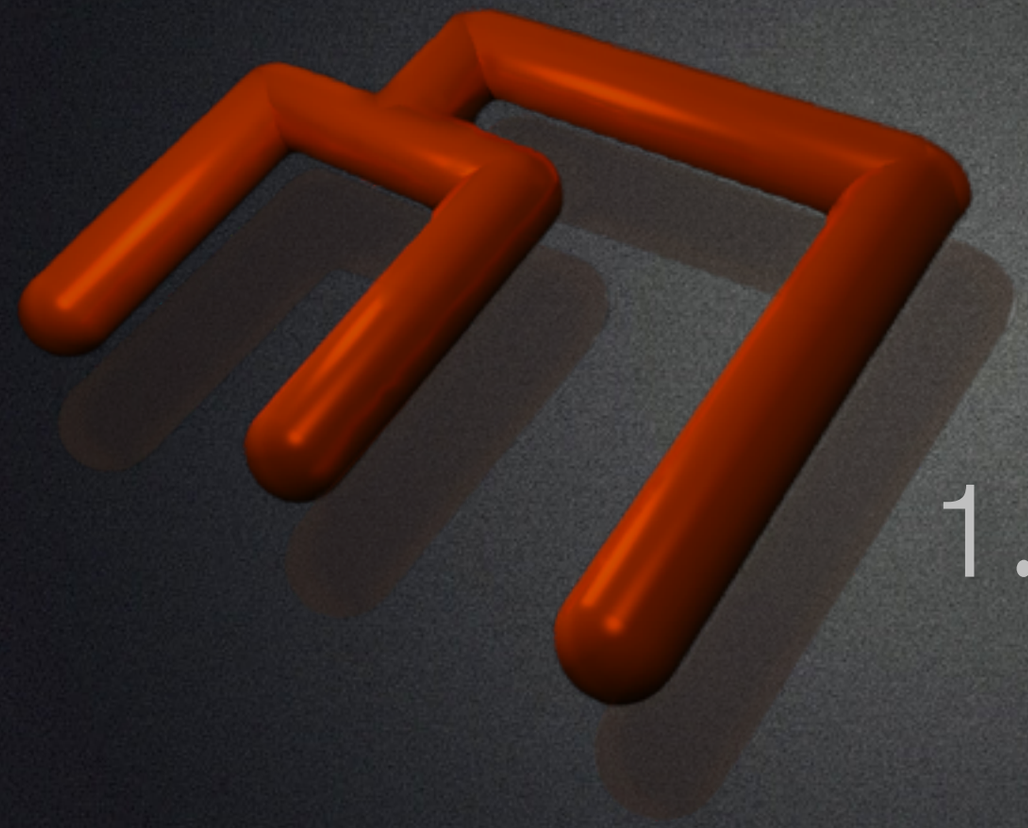
EMBL-EBI

Protein complexes



Havugimana, P. C. et al. A Census of Human Soluble Protein Complexes. Cell 150, 1068–1081 (2012).

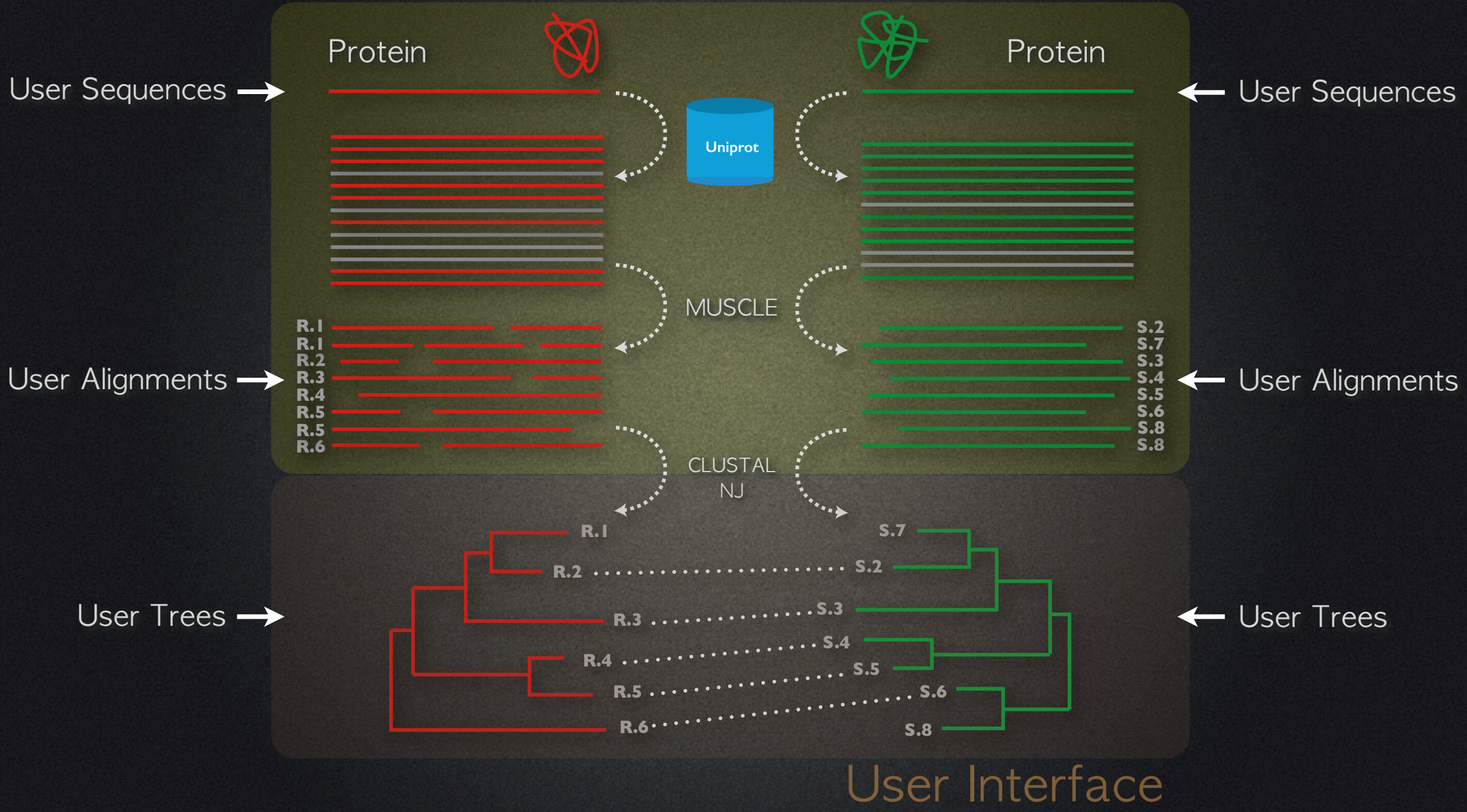
EMBL-EBI



1.MirrorTree Server

EMBL-EBI

Automatic Pipeline



User Interface

EMBL-EBI

MirrorTree Server

MTServer Help Tutorial

Welcome to *MirrorTree Server*

The Mirrortree Server allows to graphically and interactively studying the co-evolution of two protein families, and asses their interactions in an taxonomic context. The server accepts single sequences as input, although expert users can provide their own multiple sequence alignments or trees. Select below the point from which you want to start (single sequences, alignments or trees). See the "Help" page above for more information. *(Adobe Flash Player 10 is required for the interactive manipulation of the results)*

Ⓐ ***From sequences*** Ⓑ From multiple alignments Ⓒ From trees

From sequences

Provide the sequence of a member of each one of the two families (in ***FASTA format*** (***example***)). The system will look for orthologs in different species and generate the multiple sequence alignments and phylogenetic trees. The protocol implemented in the server is not intended to work with protein domains. You have to use whole-length sequences as input for the server. ***More info...***

Protein 1

Sequence (FASTA format):

SELECT

Protein 2

Sequence (FASTA format):

SELECT

Job Submission

Job name: (optional)

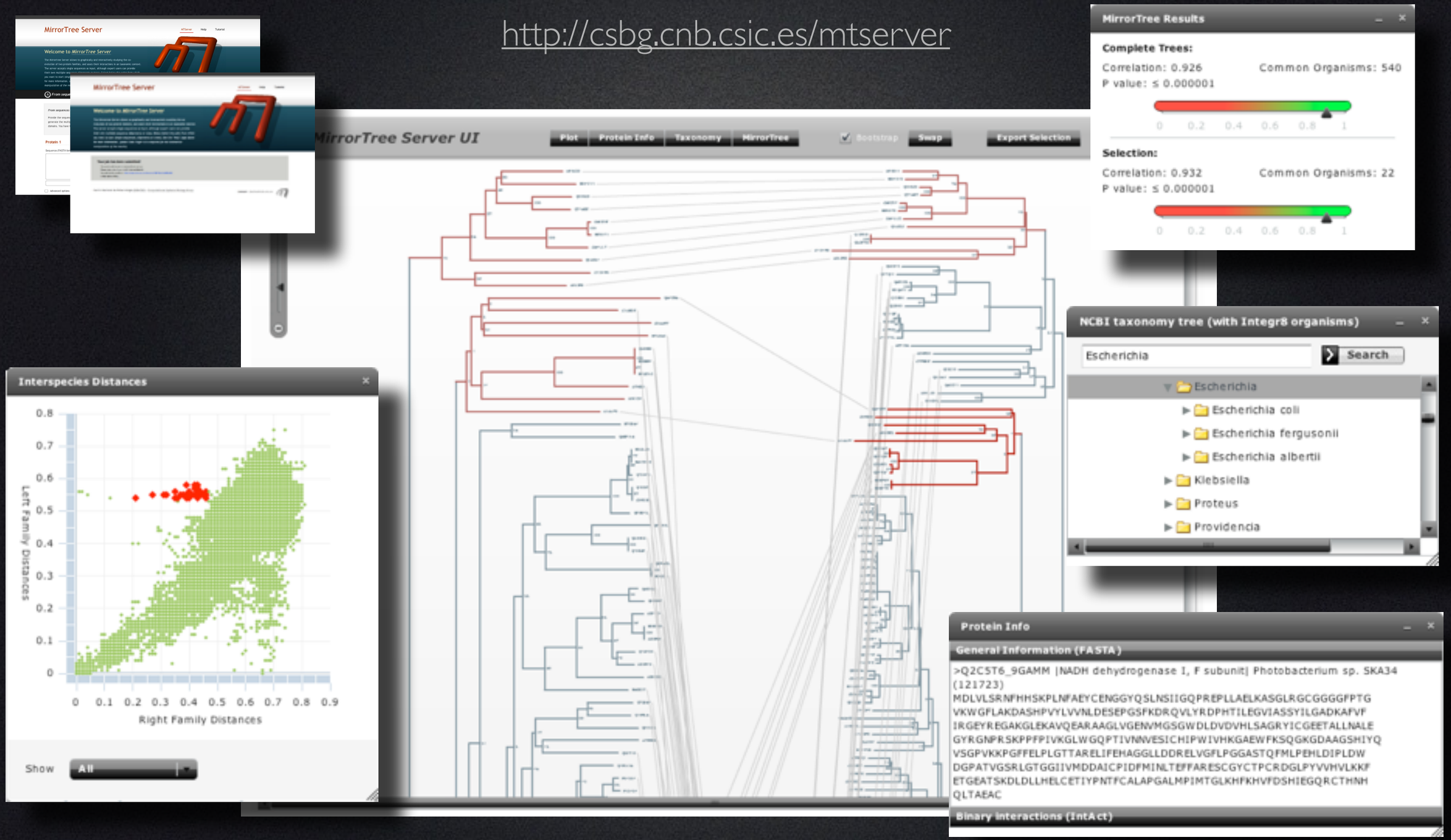
Email:

Send Reset

Advanced options

MirrorTree Server

http://csbg.cnb.csic.es/mtserver



Ochoa, D. & Pazos, F. Studying the co-evolution of protein families with the Mirrortree web server. *Bioinformatics* **26**, 1370–1371 (2010)

EMBL-EBI

MirrorTree Server Tutorial

http://csbg.cnb.csic.es/mtserver/eccb.html

EMBL-EBI