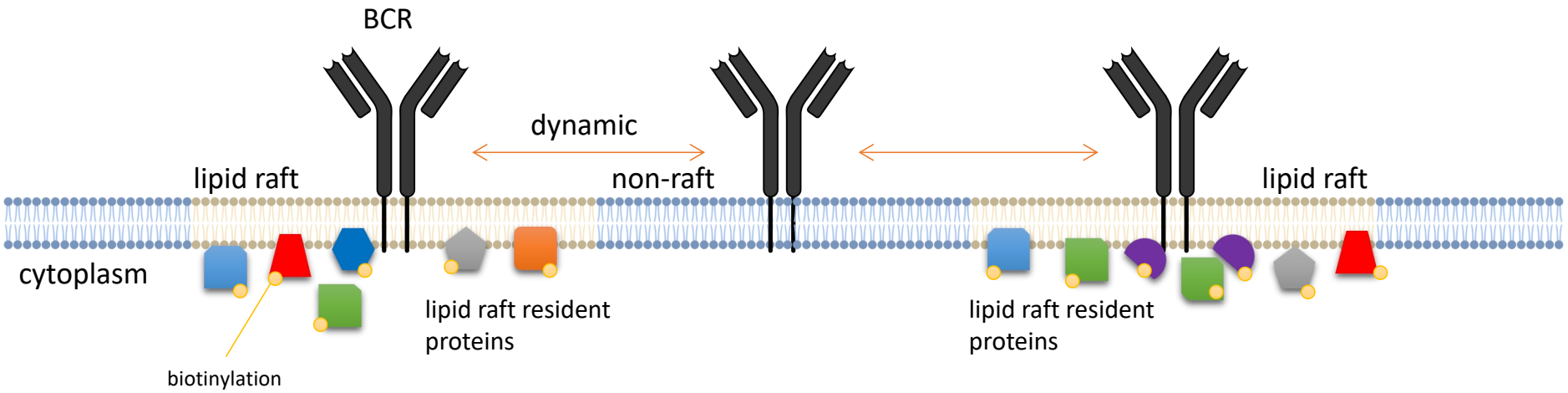
AutoCoEv

13th September 2021

Screen proteins identified in lipid rafts after BCR activation

Novel players and large-scale protein dynamics of BCR activation revealed by APEX2 proximity labelling of lipid rafts

Luqman O. Awoniyi[1,2], Marika Runsala[1,2], Vid Šustar[1], Sara Hernández-Pérez[1,2], Alexey V. Sarapulov[1,2], Petar Petrov[1,2] and Pieta K. Mattila[1,2,*]



Identification

Screen for novel protein-protein (functional) interactions

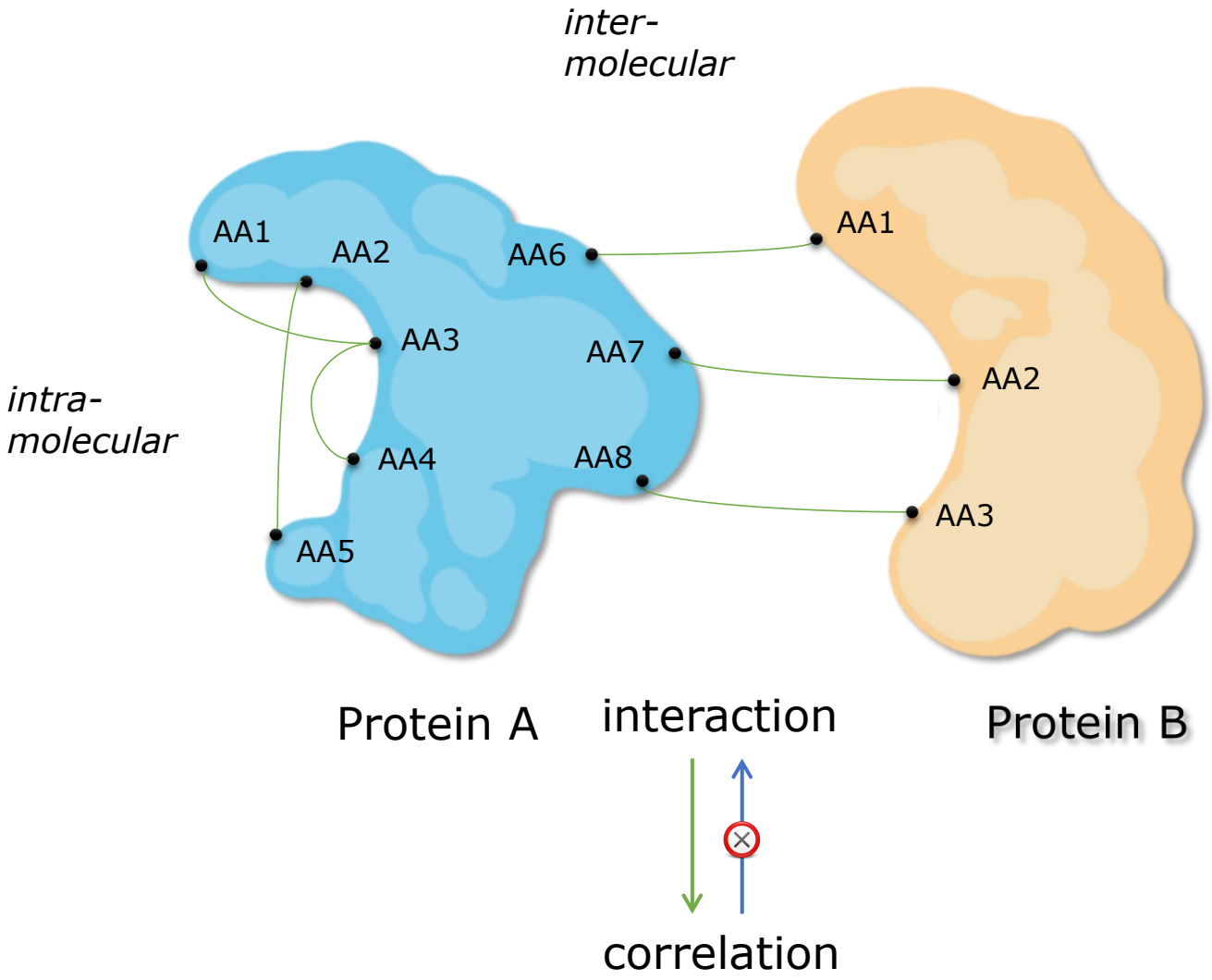
Co-evolution

Evolution of proteins is influenced by structural and functional constraints between amino acids

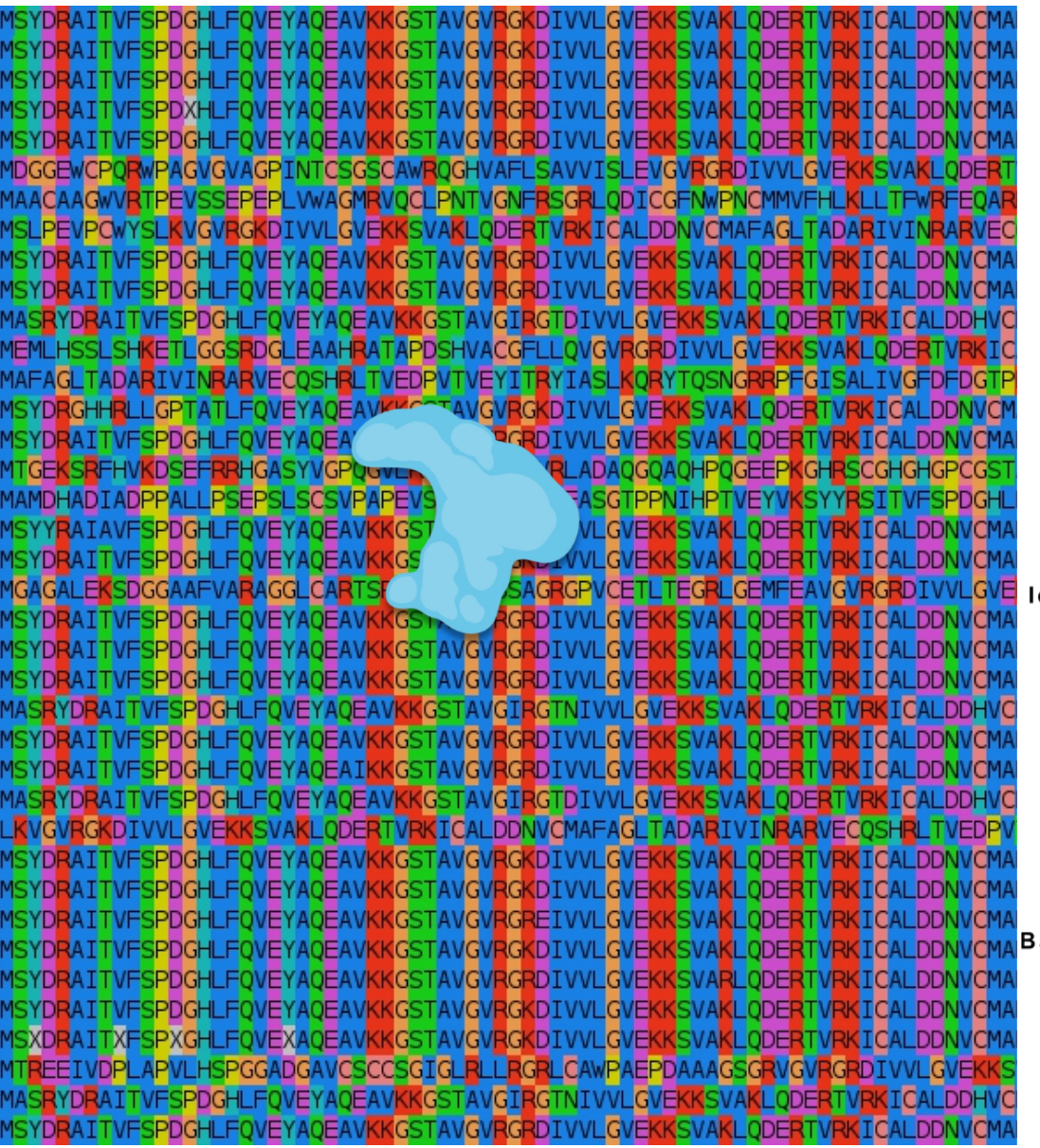
Changes happen in a synchronized (concerted) manner

Detecting co-evolution is a sign of functional dependence

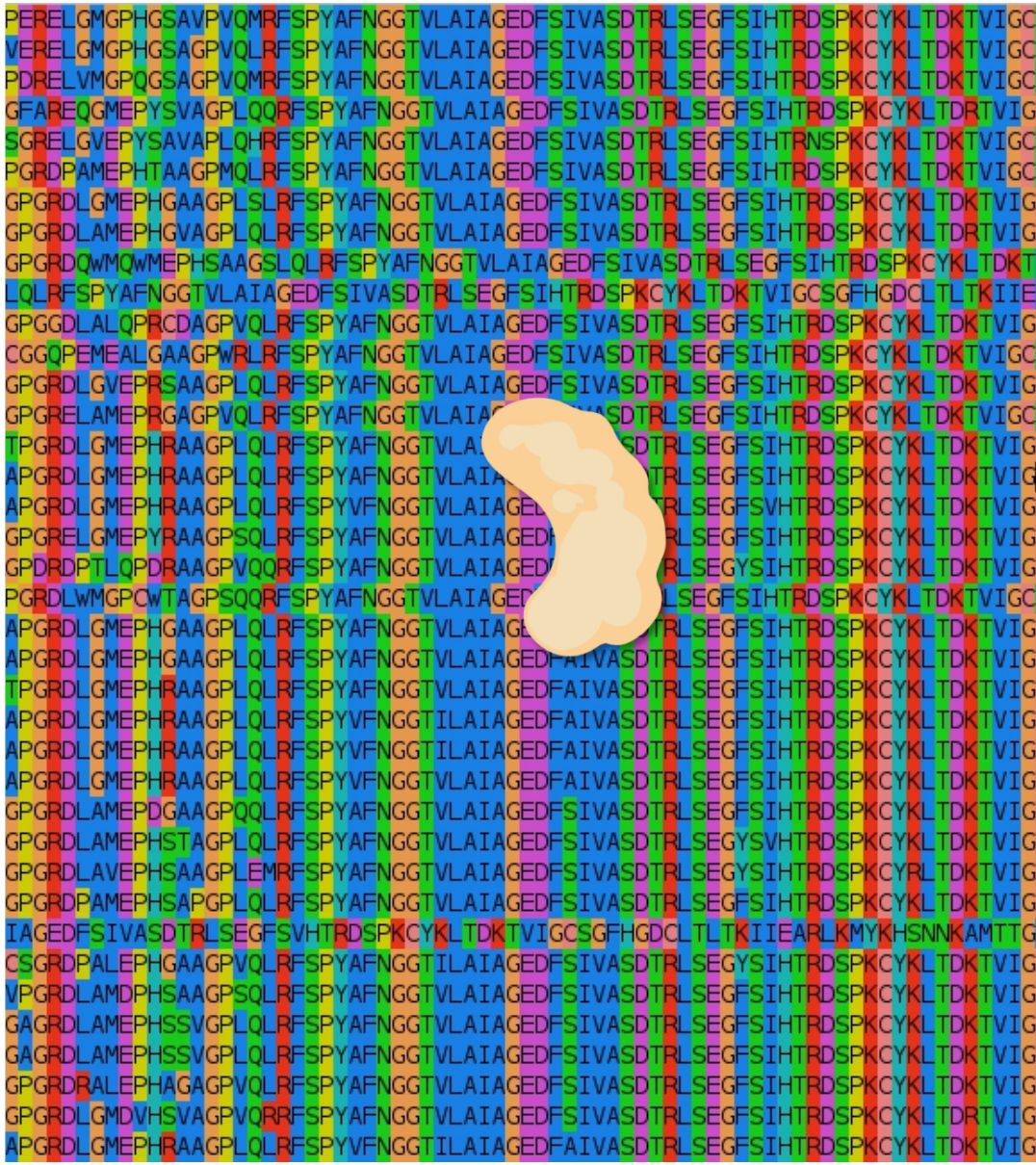
Relations between proteins can be extrapolated from the evolutionary history of their genes, *in silico*



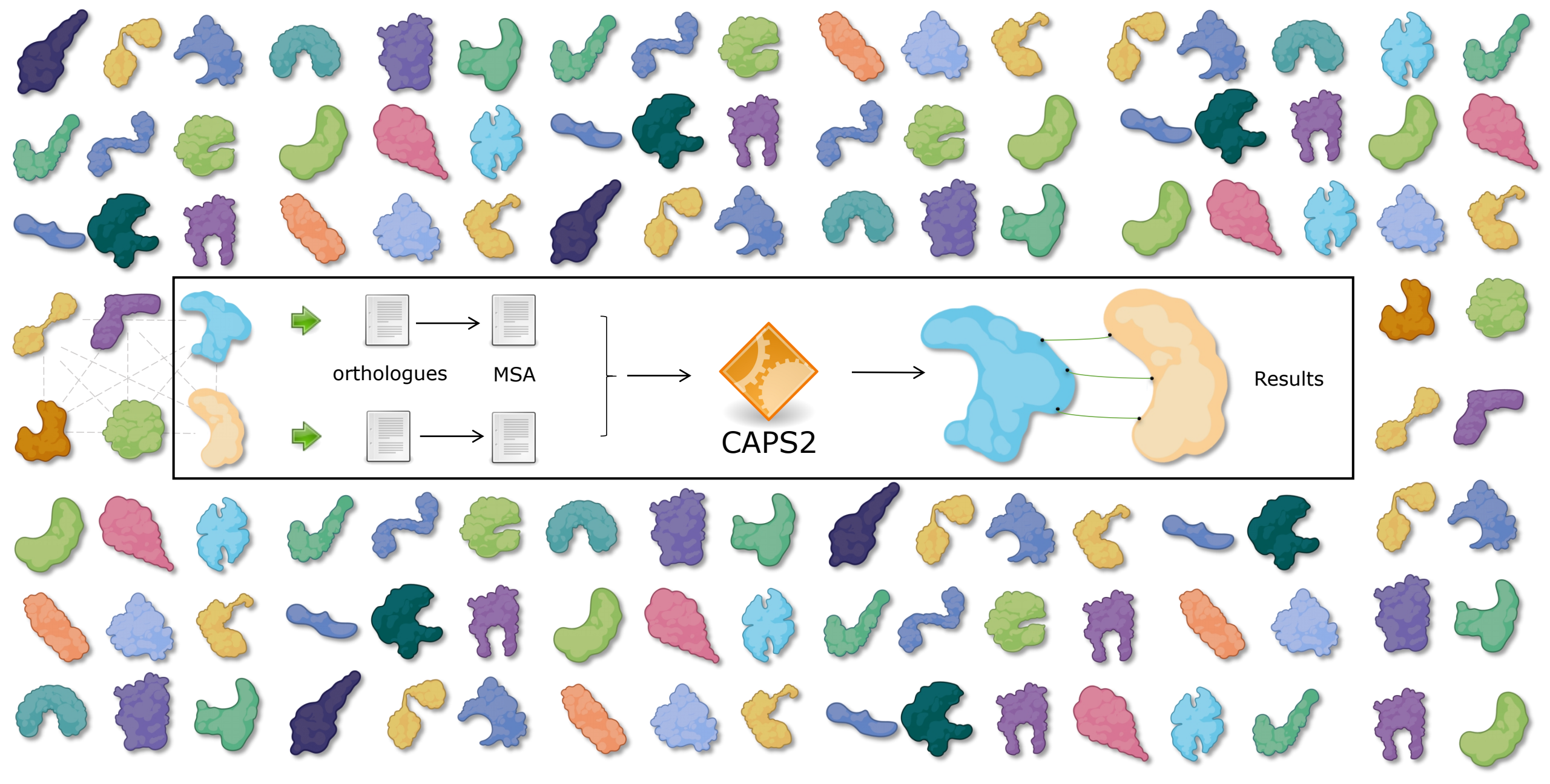
Co-evolution tests require protein orthologous sequences



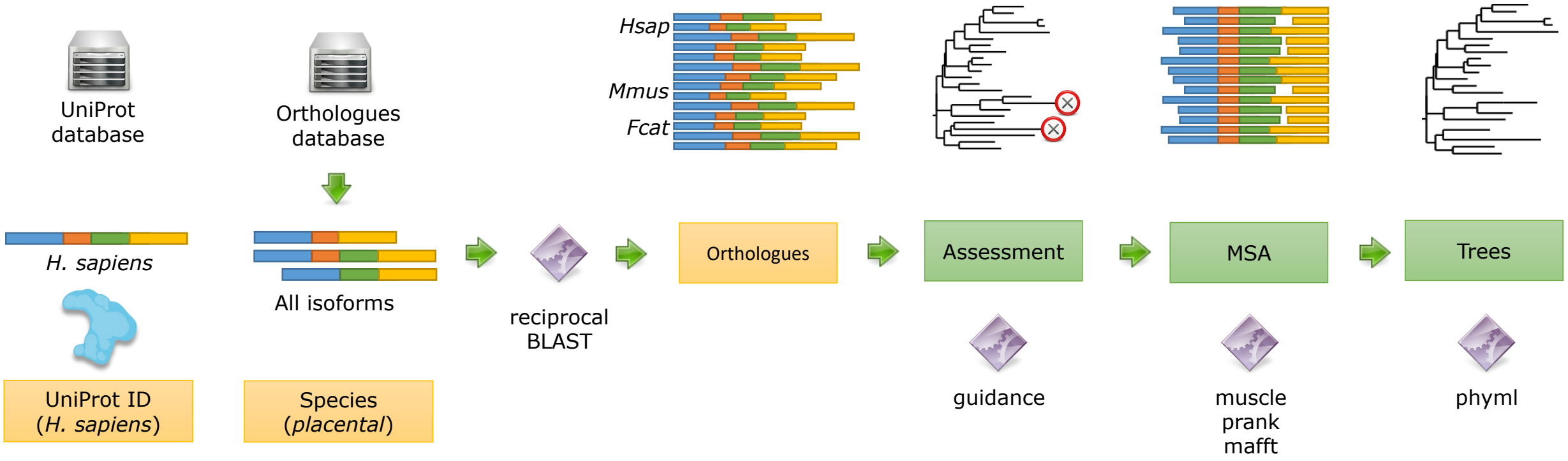
Echinops_telfairi
Elephas_maximus
Loxodonta_cyclotis
Loxodonta_africana
Procavia_capensis
Trichechus_manatus
Choloepus_hoffmanni
Dasypus_novemcinctus
Tupaia_glis
Galeopterus_variegatus
Carlito_syrichta
Pongo_abelii
Homo_sapiens
Pan_paniscus
Pan_troglodytes
Gorilla_gorilla
Nomascus_leucogenys
Nasalis_larvatus
Rhinopithecus_roxellana
Chlorocebus_sabaeus
Macaca_fascicularis
Macaca_mulatta
Papio_anubis
Saimiri_boliviensis
Callithrix_jacchus
Microcebus_murinus
Otolemur_garnettii
Ictidomys_tridecemlineatus
Dipodomys_ordii
Cricetulus_griseus
Mus_musculus
Rattus_norvegicus
Cavia_porcellus
Heterocephalus_glaber
Ochotona_princeps
Oryctolagus_cuniculus
Myotis_lucifugus
Pteropus_vampyrus
Sus_scrofa
Ovis_aries
Bison_bison
Bos_taurus
Tursiops_truncatus
Balaenoptera_acutorostrata
Vicugna_pacos
Ceratotherium_simum
Equus_caballus
Manis_pentadactyla
Ailuropoda_melanoleuca
Mustela_putorius
Canis_lupus
Felis_catus
Sorex_araneus



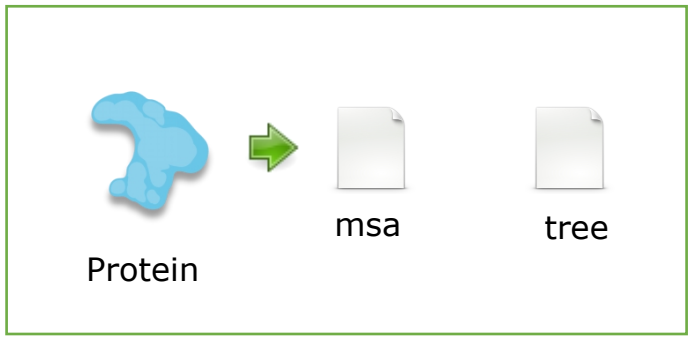
CAPS2 program tests protein pairs for co-evolving amino acids



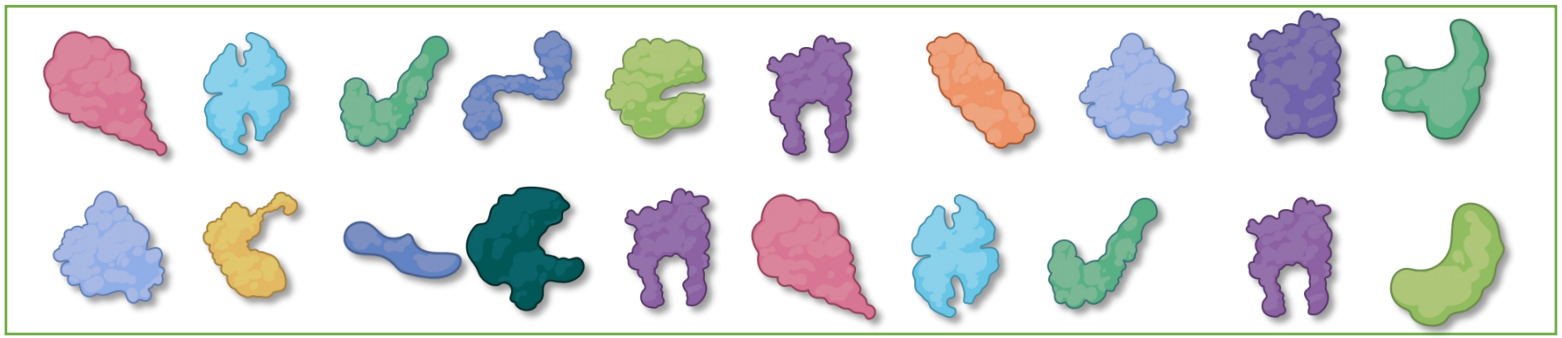
Pre-run steps



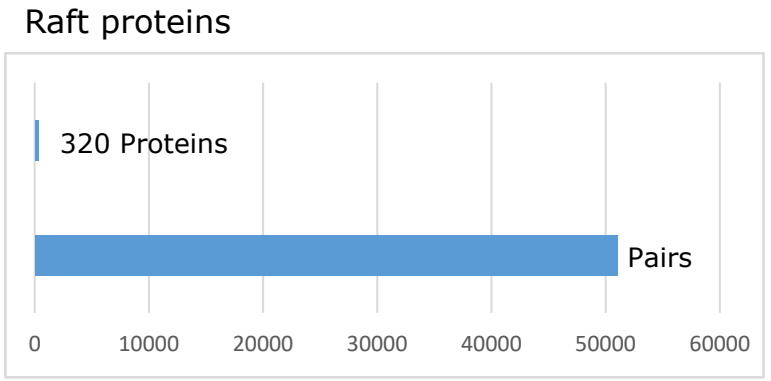
End result

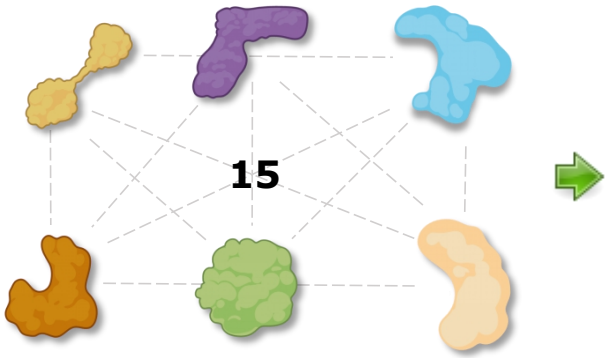


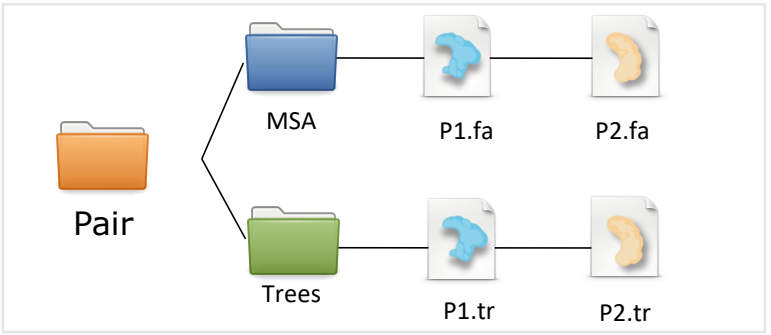
Do the same for the rest!



Run in parallel







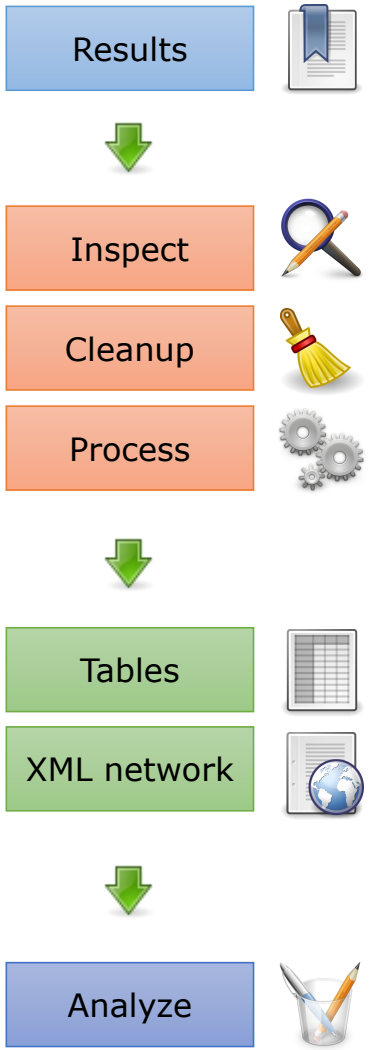
6-core CPU

parallel

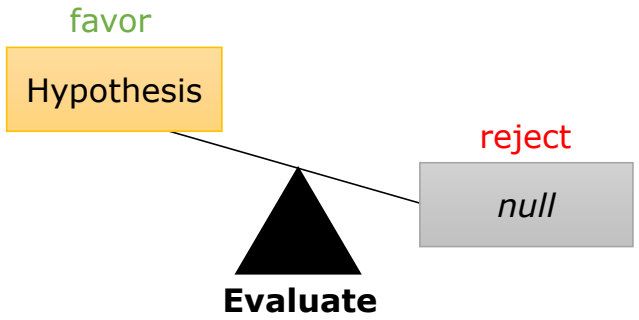
CAPS2

Speed up!

Post-run and multiple hypothesis testing



Hypothesis: there is co-evolution b/n AA
null hypothesis: there is no co-evolution



p-value: the probability of error when rejecting the ***null*** hypothesis

p = 0.05: there is 5% chance that rejecting the ***null*** hypothesis was wrong

Type I error: a true ***null*** hypothesis is rejected

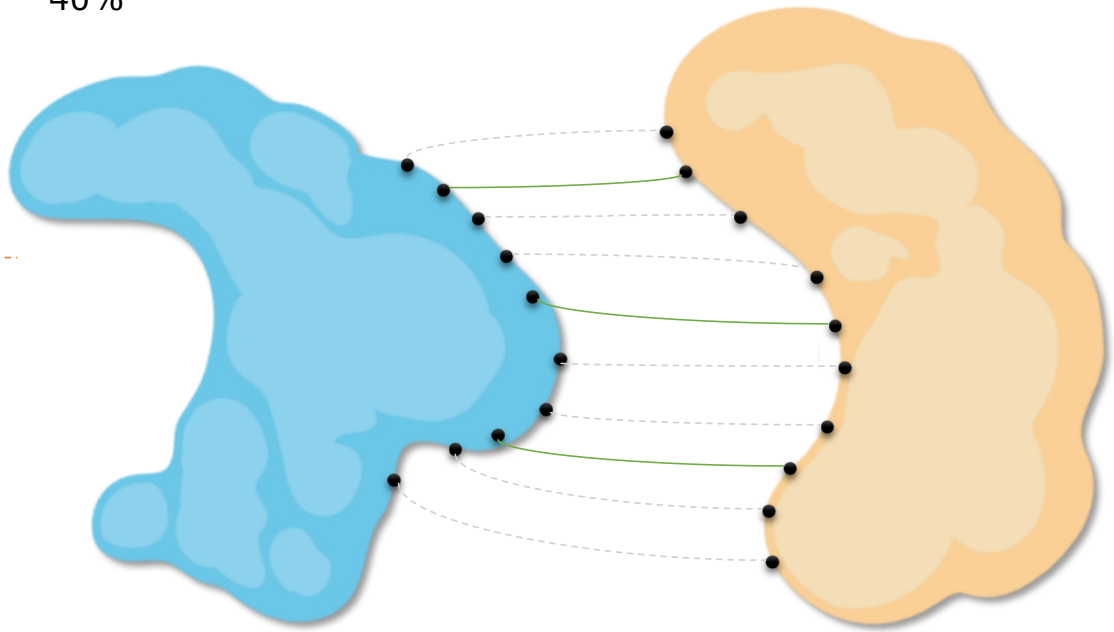
Type II error: a false null hypothesis is not rejected

Multiple testing problem: the more results you have, the more likely is one of them is false

Probability (**P**) of observing at least one result due to chance.
10 results of p = 0.05

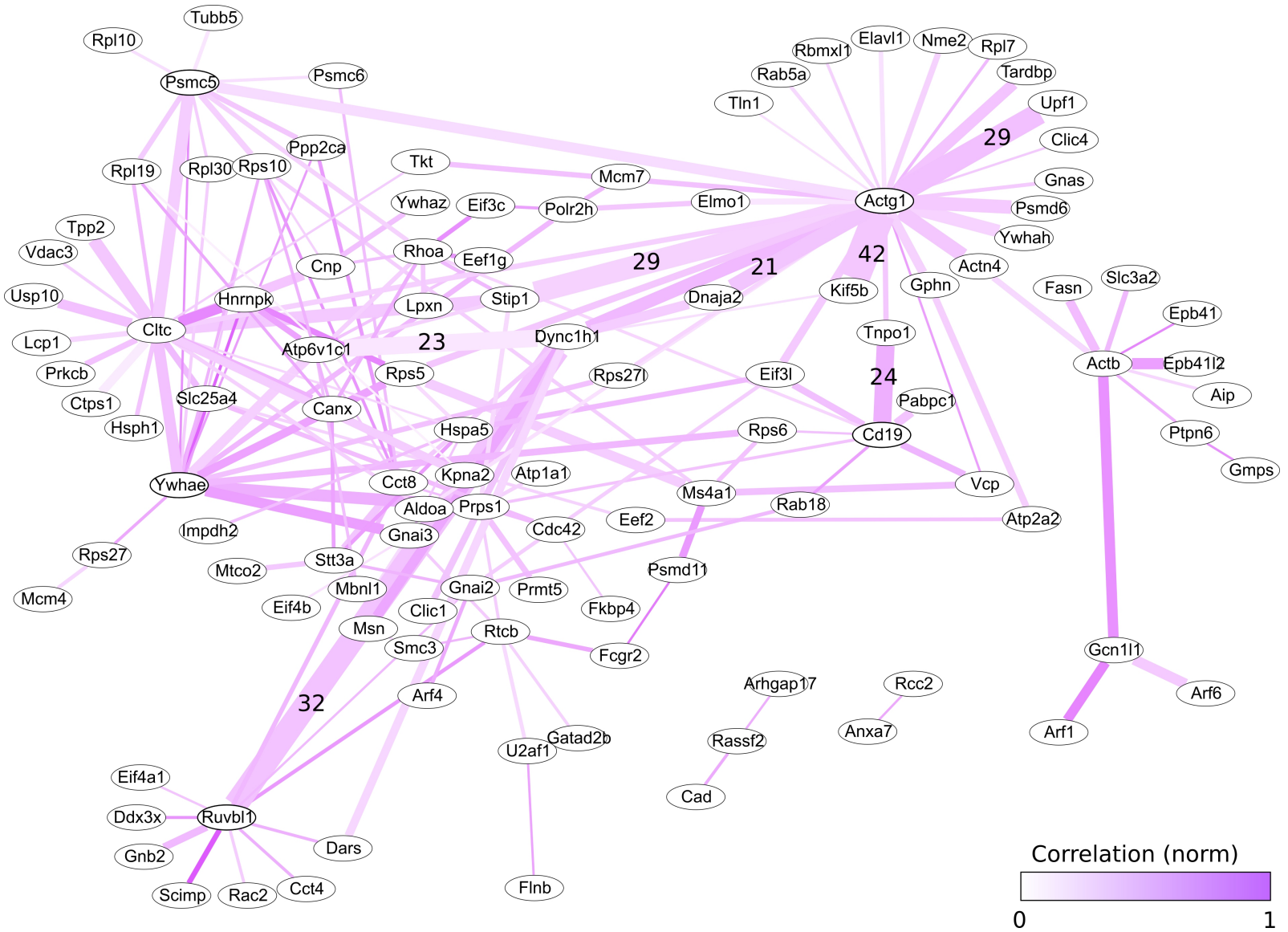
P = 1 - (1 - 0.05)10 = 0.4

40%



Bonferroni correction:
p-value * number of results = adjusted p-value

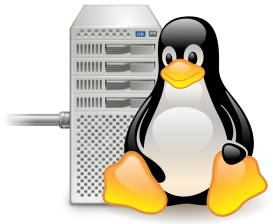
Anticipated results



AutoCoEv is a bash script

BASH

Software packages



Linux workstation

Patch for verbose output (C++)

BLAST

CAPS2

Datamash

Exonerate

Gblocks

Guidance

MAFFT

MUSCLE

Parallel

PhyML

PRANK

R

SeqKit

squizz

TreeBeST

Availability online: preprint, script, manual and VM

bioRxiv

AutoCoEv – a high-throughput *in silico* pipeline for predicting novel protein-protein interactions

Petar B. Petrov[1,2*], Luqman O. Awoniyi[1,2], Vid Šuštar[1] and Pieta K. Mattila[1,2*]

https://www.biorxiv.org/content/10.1101/2020.09.29.315374v3

autocoev Public

AutoCoEv – a high-throughput in silico pipeline for predicting novel protein-protein interactions

Shell 1

https://github.com/mattilalab

SlautoCoEv

Shared by: Petar Petrov

Current path: SlautoCoEv ZIP

Name	Size	Last Update	Operations
account.txt	50 bytes	2021-03-28	
SlautoCoEv.ova	2.5 GB	2020-12-19	

https://seafile.utu.fi/d/c73b5164095f4fafb23d/

Download VirtualBox 6.1

https://www.virtualbox.org/

Current stage: species and example protein complexes

CORUM

The comprehensive resource of mammalian protein complexes

https://mips.helmholtz-muenchen.de/corum/

20S

ARC

COG

COX

DNA synth

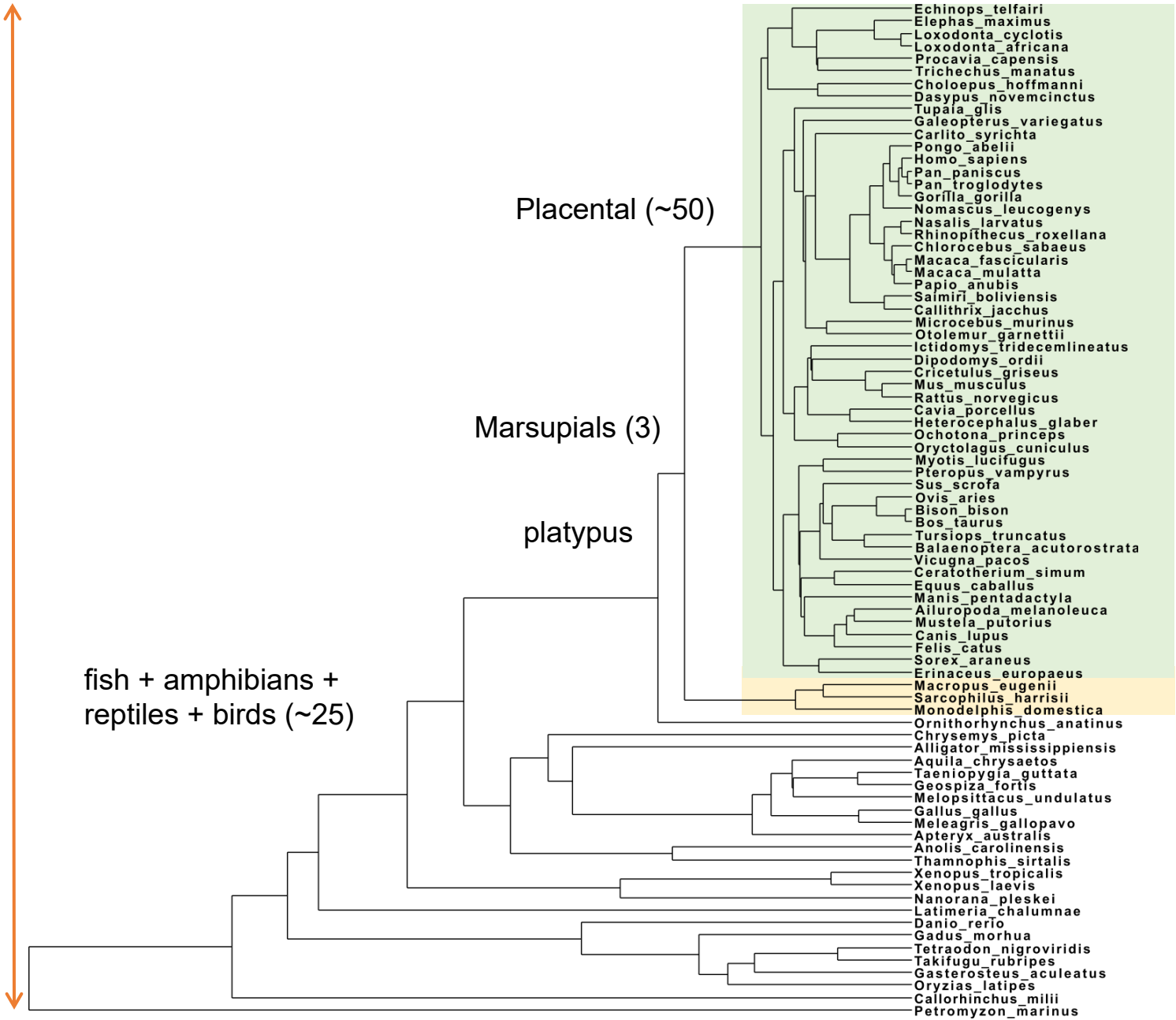
F1F10

IL12

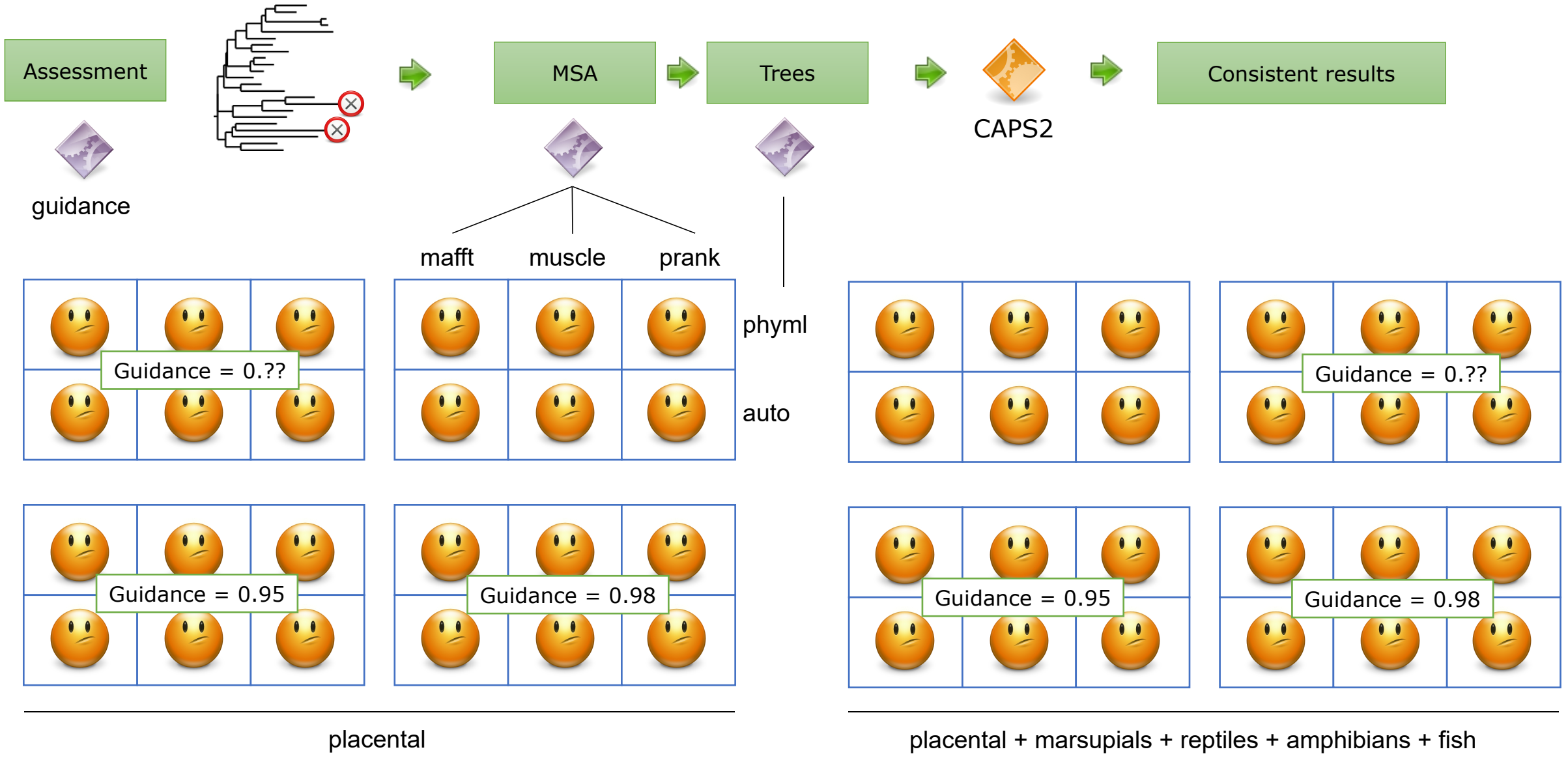
IL

eIF3

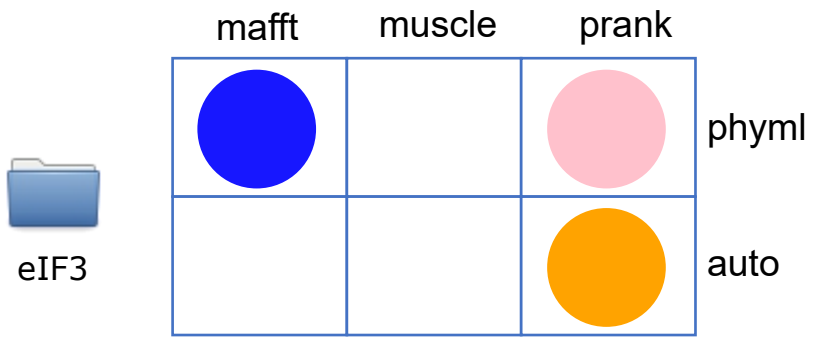
VERTEBRATES

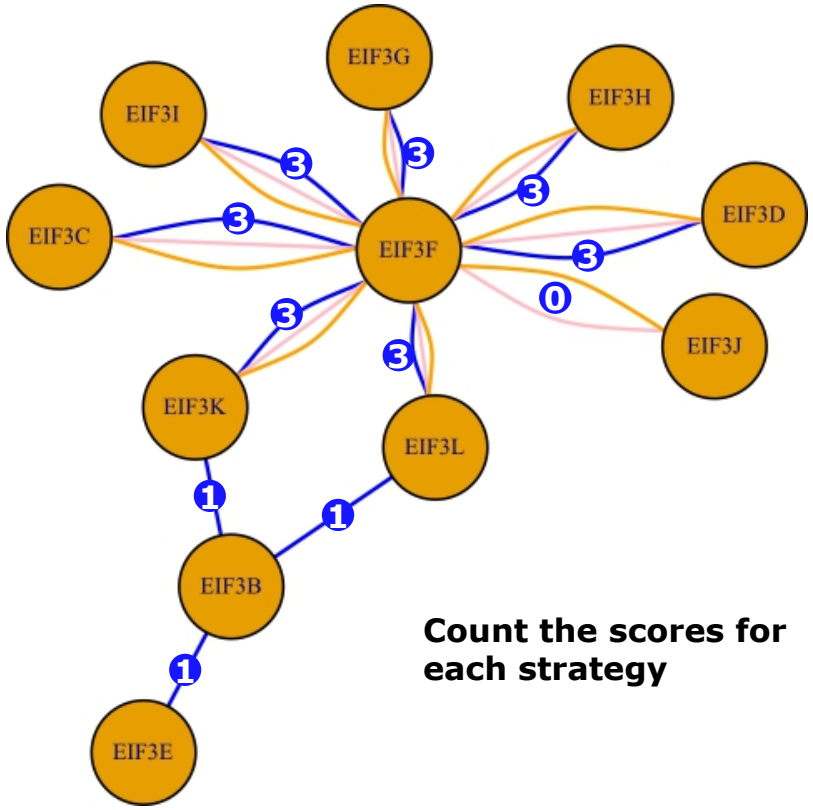


Current stage: strategies



Rank the strategies





Count the scores for each strategy

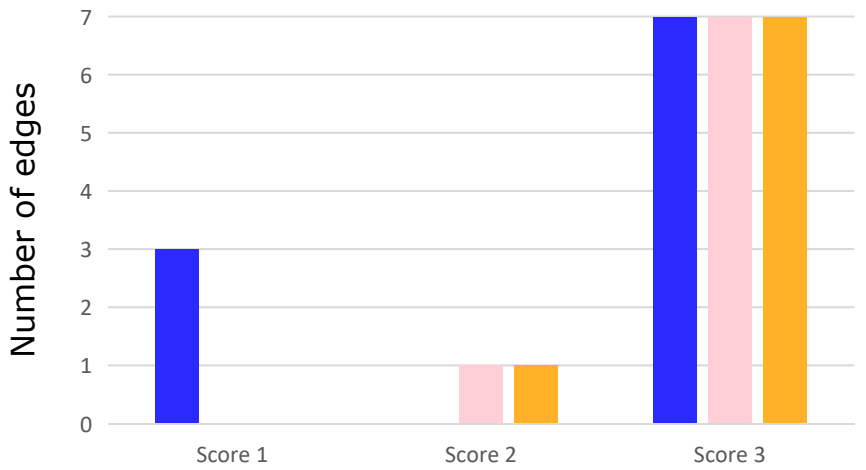
How well the results of one strategy concord with the others.

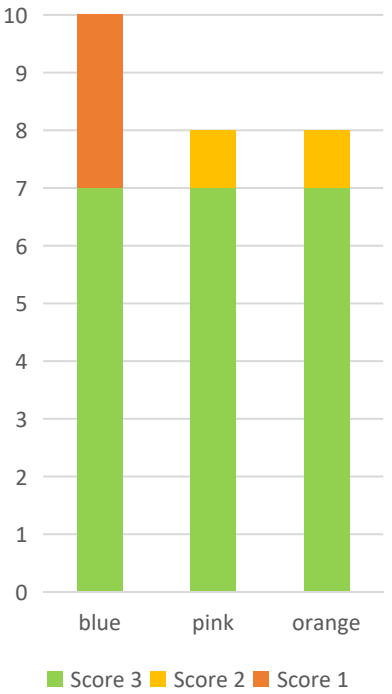
Look for a strategy with:

- High number of edges, also detected by the others.
- Low number of edges, not detected by the others (edges that are unique to the particular strategy only).

Assign scores per network:

- 1: single network edge
- 2: overlaps with 1 other network
- 3: overlaps with 2 other networks





To do

Improve the R script for networks analyses (Luqman)

Analyse the networks from example complexes

Make plots, prepare them as supplementary figures where comparisons between strategies and other conditions are clear

Determine which strategy would be best, including
- species (placental, all?)
- guidance cutoff

Run the big analyses for the 320 lipid raft resident proteins

Update manual

Update manuscript and resubmit