What to do with a billion years of evolution

Melissa DeBiasse

University of Florida Whitney Laboratory for Marine Bioscience

@MelissaDeBiasse melissa.debiasse@gmail.com melissadebiasse.weebly.com

Acknowledgements

University of Florida Ryan Lab

PI: Joseph Ryan

Alexandra Hernandez

Jessica Whelpley

DEEPC

PI: Steven Haddock • Monterey Bay Aquarium Research Institute

Darren Schultz

Jacob Winnikoff

PI: Erik Thuesen • The Evergreen State College

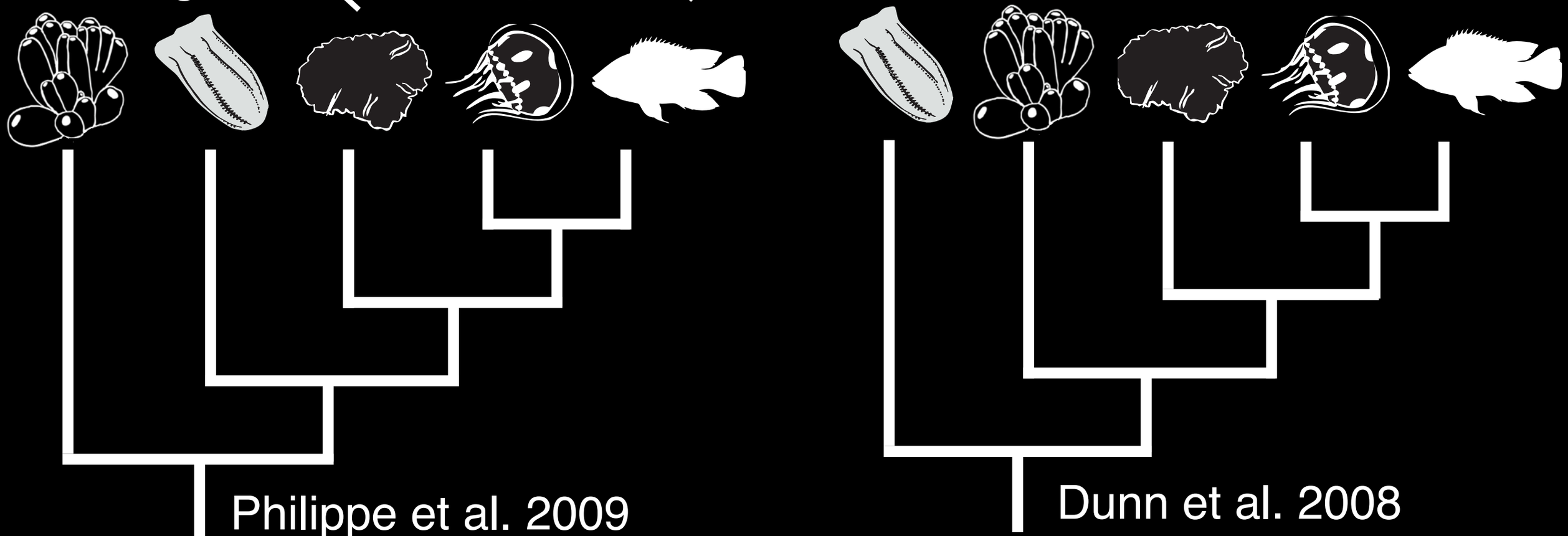
Telissa Wilson

Warren Francis • University of Southern Denmark

Funding

NSF Dimensions of Biodiversity (1542679, 1542673, 1542597)

Porifera
Ctenophora
Placozoa
Cnidaria
Bilateria



Philippe et al. 2009

Pick et al. 2010

Nosenko et al. 2013

Pisani et al. 2015

Simion et al. 2017

Feuda et al. 2017

Dunn et al. 2008

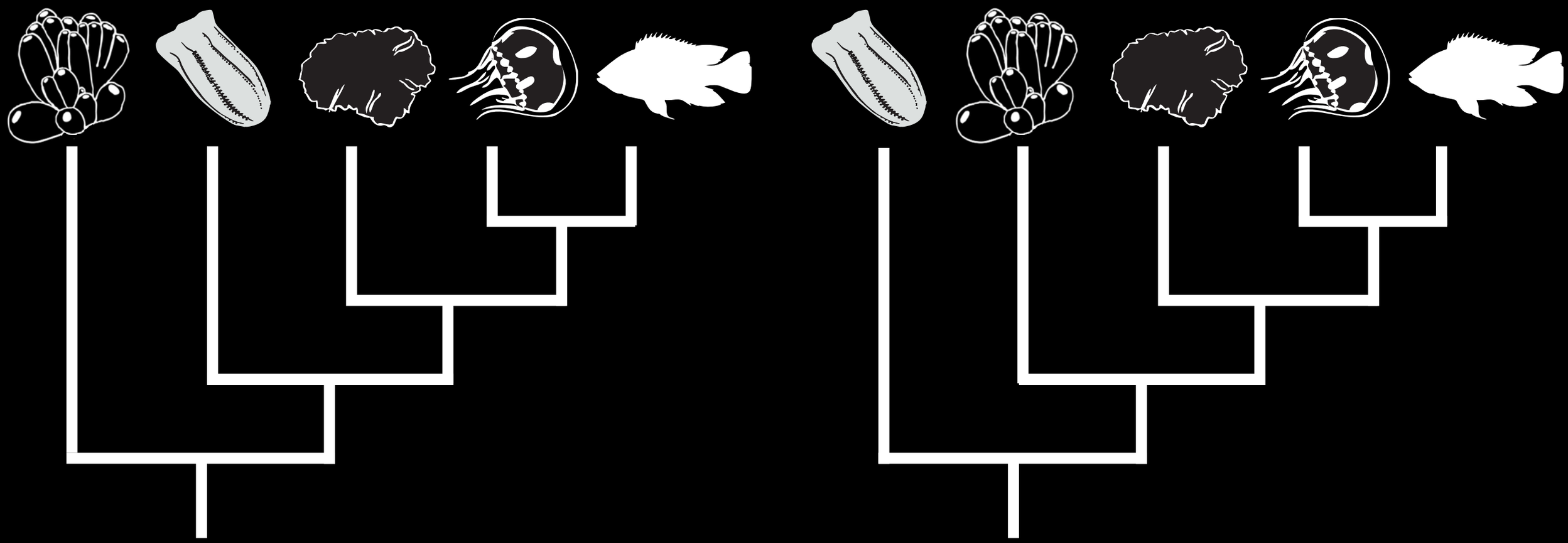
Ryan et al. 2013, Hejnol et al. 2009

Moroz et al. 2014, Chang et al. 2015

Whelan et al. 2015, Torruella et al. 2015

Borowiec et al. 2015, Arcila et al. 2017

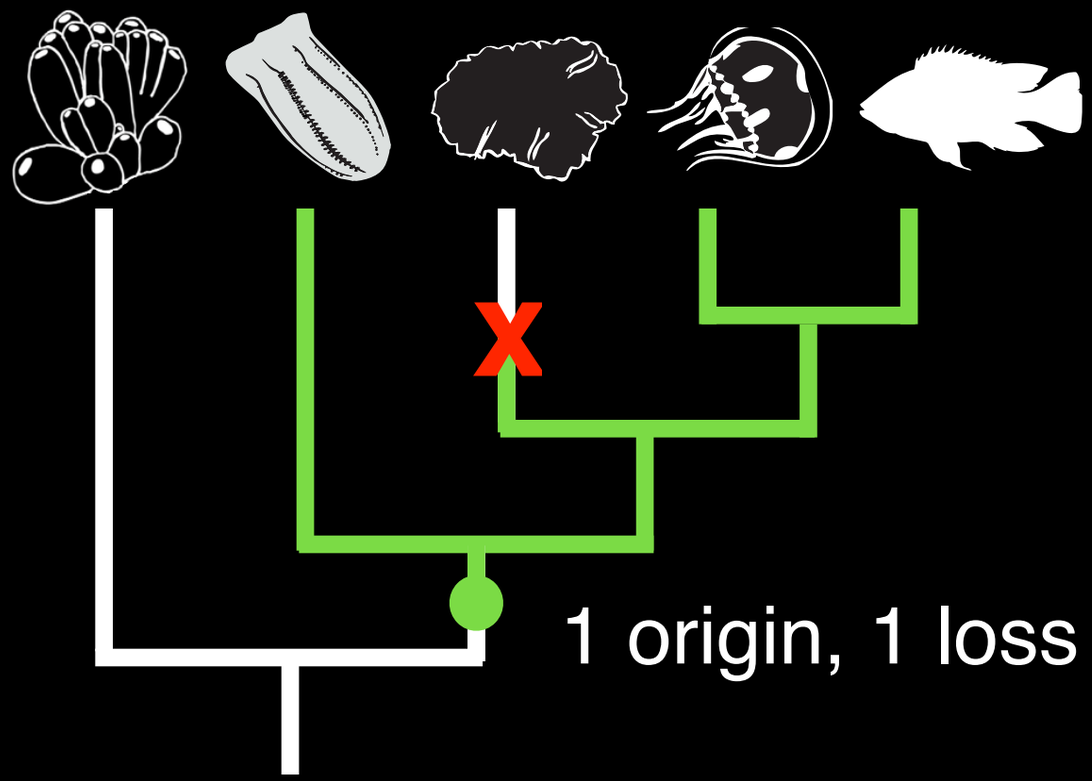
Whelan et al. 2017, Shen et al. 2017

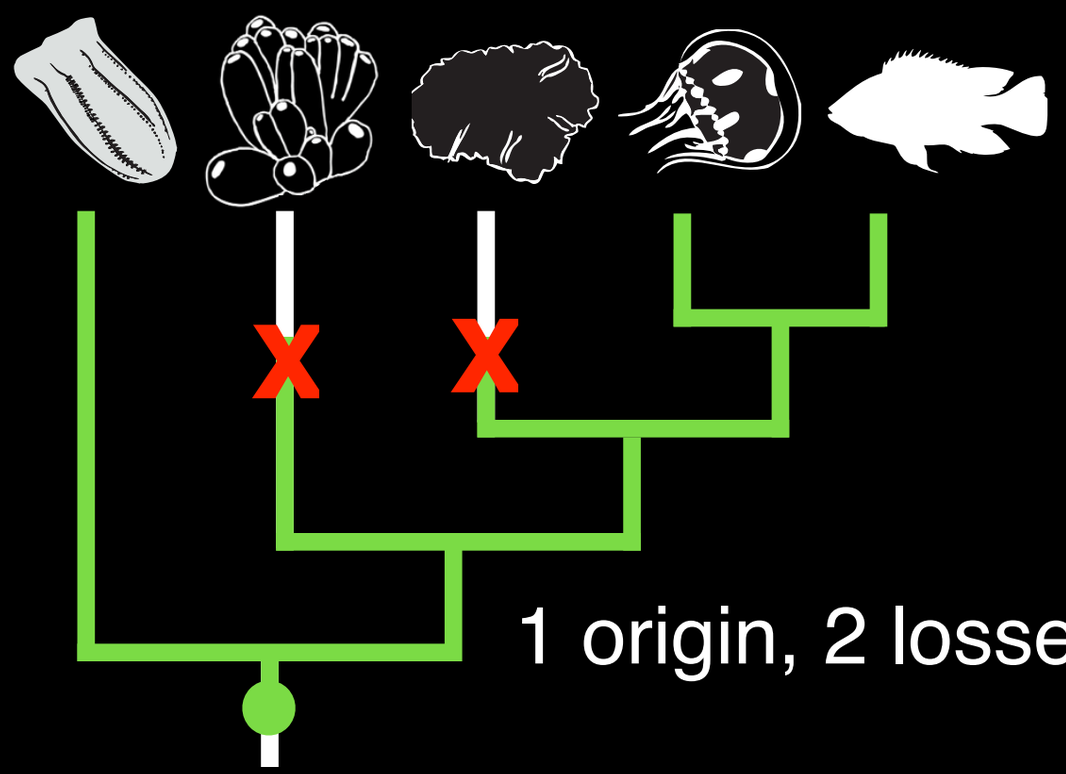


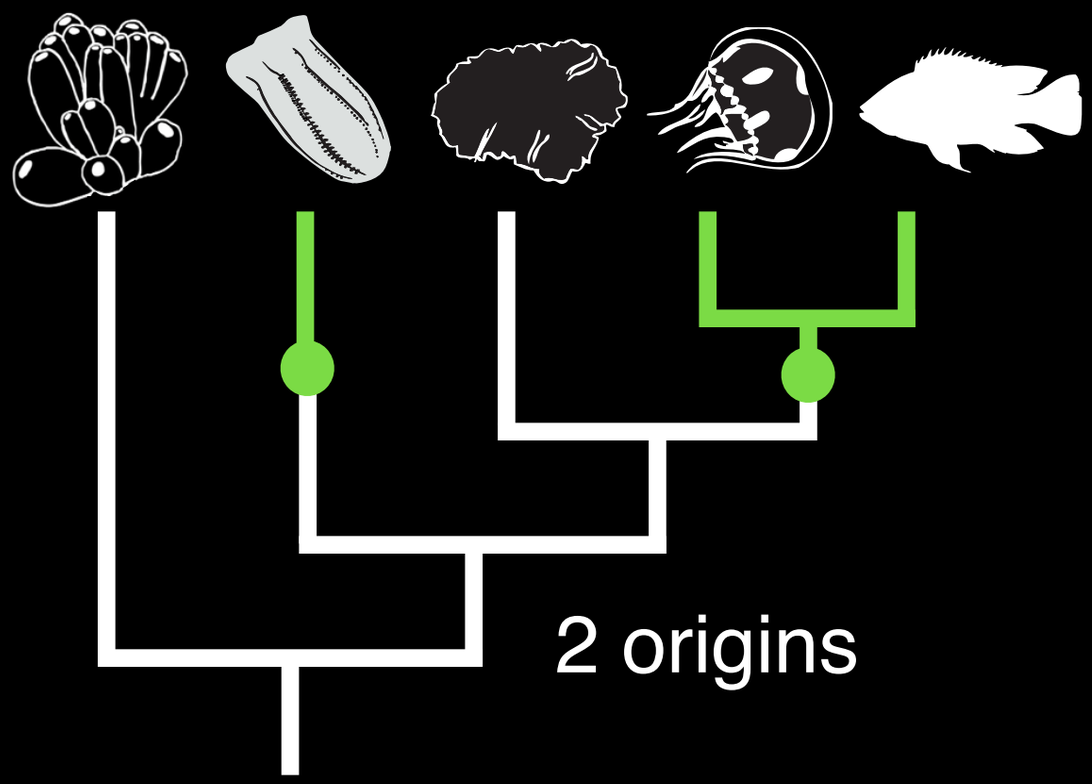
Who cares?

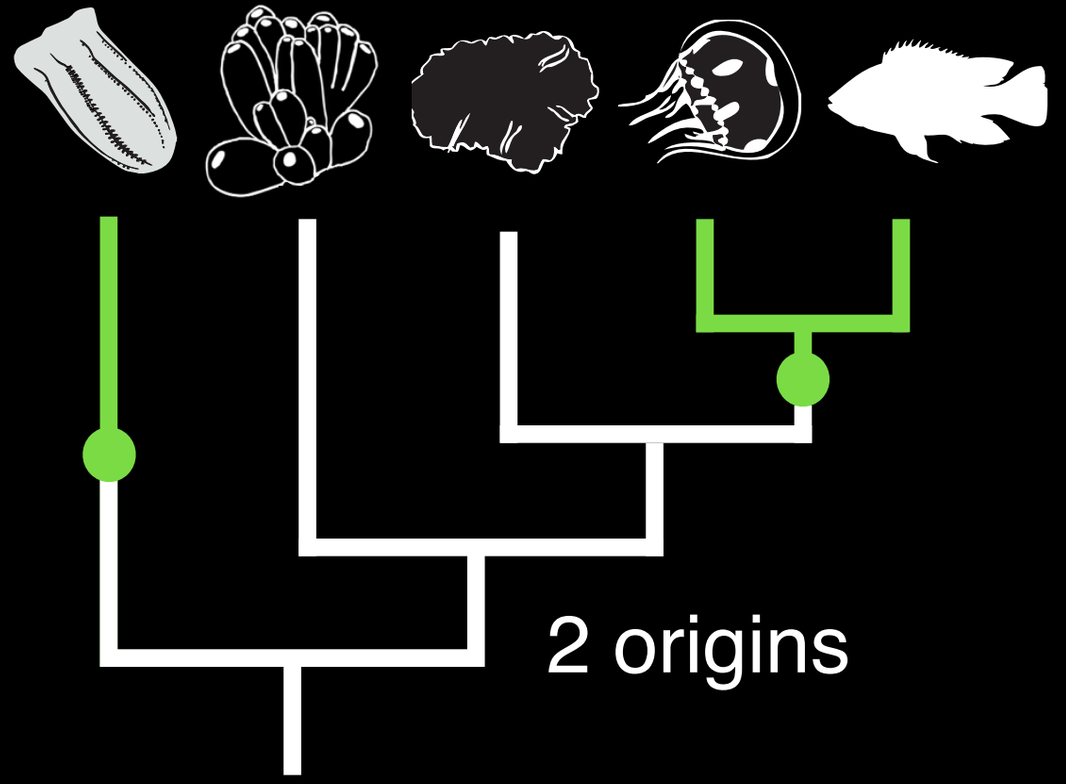
Accurate species trees are important tools
for understanding evolutionary history

Evolution of neural cell types









Evolution of neural cell types

Species trees estimated over different time scales

species complex in cardinals

10 mya

<u>shallow</u>

introgression

rapid radiation

RADseq/UCEs

nucleotide alignments

gene tree / species tree methods

relationships among insects

500 mya

origin of multicellularity

>1 billion mya

past ---------- time ---------- present

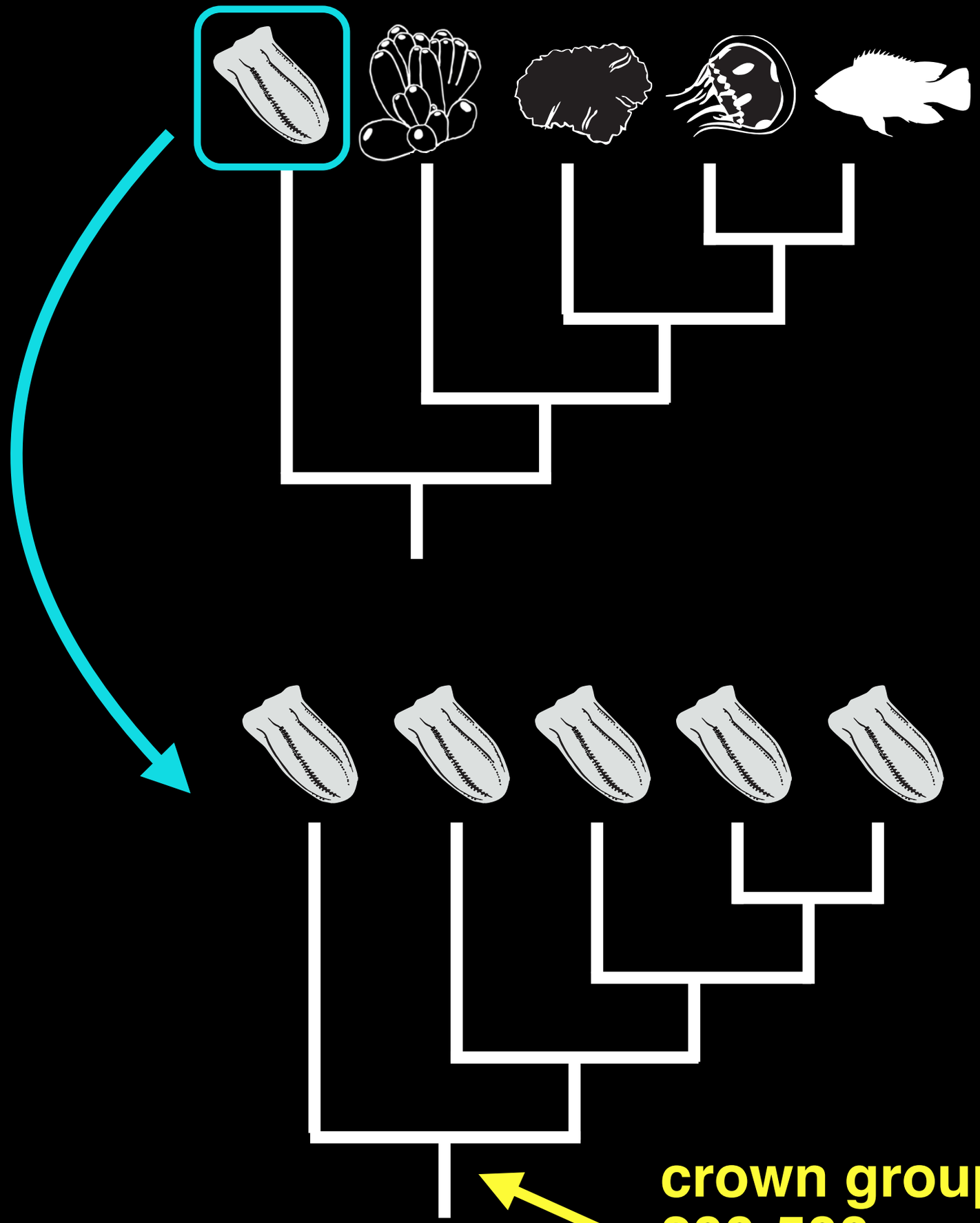
<u>deep</u>

ancient splits

saturation

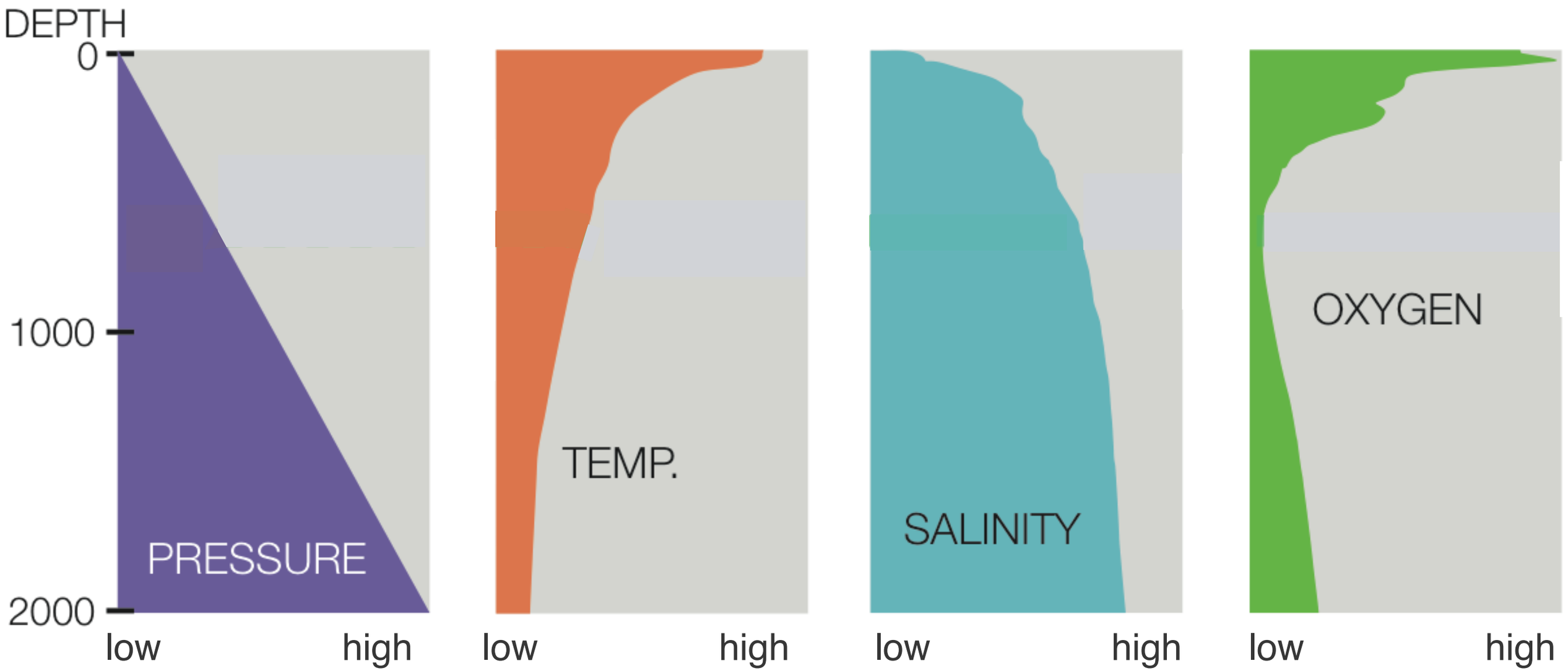
compositional heterogeneity

methods???


crown group
200-500 mya

Dimensions of Biodiversity

DEEPC:
Diversity, Ecology, EcoPhysiology of Ctenophores



Dimensions of Biodiversity

DEEPC:
Diversity, Ecology, EcoPhysiology of Ctenophores

Comprehensive ctenophore species phylogeny

Reconstruct ancestral depths

Identify evolutionary depth transitions

Test loci for convergent evolution to depth

Link genotype to phenotype 

PHYLOTOCOL

Promoting transparency and overcoming bias in phylogenetics

OUTLINE DECISIONS BEFORE RUNNING ANALYSES

TRANSPARENCY ACCOUNTABILITY

reproducability
collaboration
education

Joseph Ryan & Melissa DeBiasse

https://peerj.com/preprints/26585/

https://github.com/josephryan/phylotocol

UF UNIVERSITY of FLORIDA

NSF #1542597

Data processing & analysis pipeline

RNA-Seq

Data processing & analysis pipeline

RNA-Seq

Assemble & translate transcriptome

CCGGAACGAGCACAGCCA

↓

PGRAQP

Data processing & analysis pipeline

RNA-Seq

Assemble & translate transcriptome

Filter non-target sequences

>Trinity_1234
PGRAQPPGRAQPPGRAQP
PVRNQADCHAQIPGMAFTI
PGRAQPPGRAQPPGRAQP
PVRNQADCHAQIPGMAFTI

BLAST database

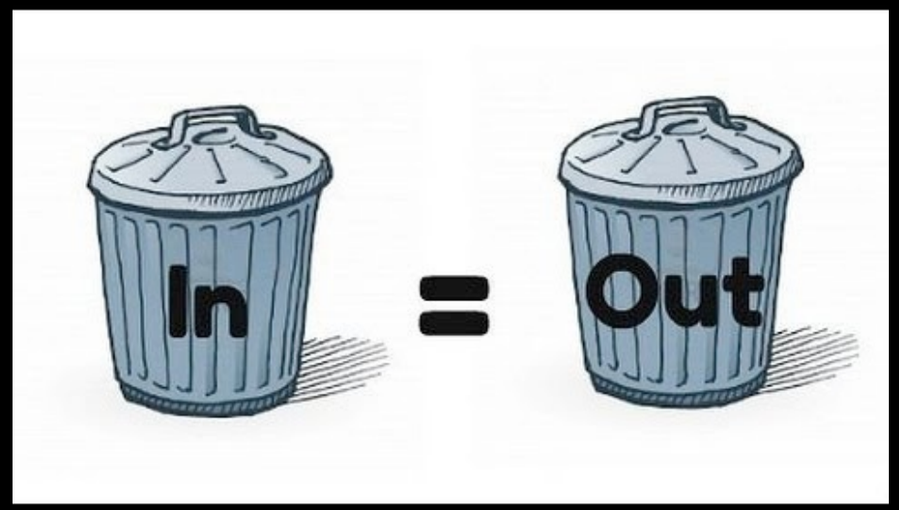
Data processing & analysis pipeline

RNA-Seq

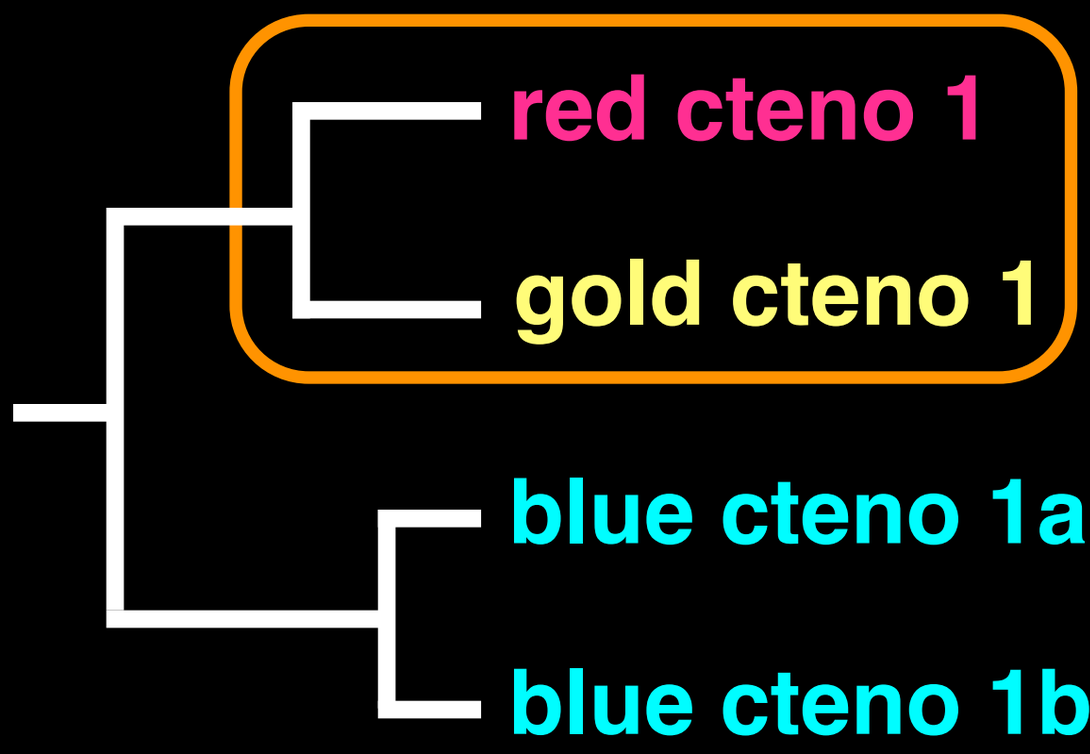
Assemble & translate transcriptome

Filter non-target sequences

Identify orthologous loci



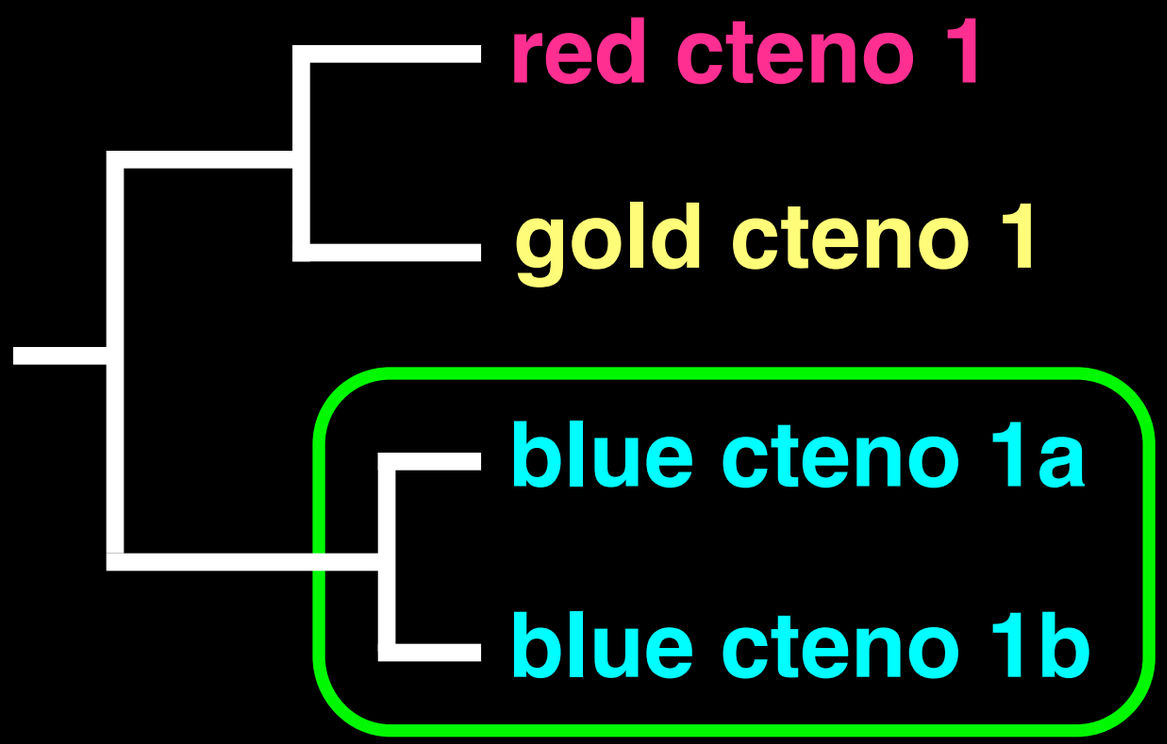
Orthologs: evolved from a common ancestor by speciation



Orthologs

Orthologs: evolved from a common ancestor by speciation

Paralogs: produced through a duplication event

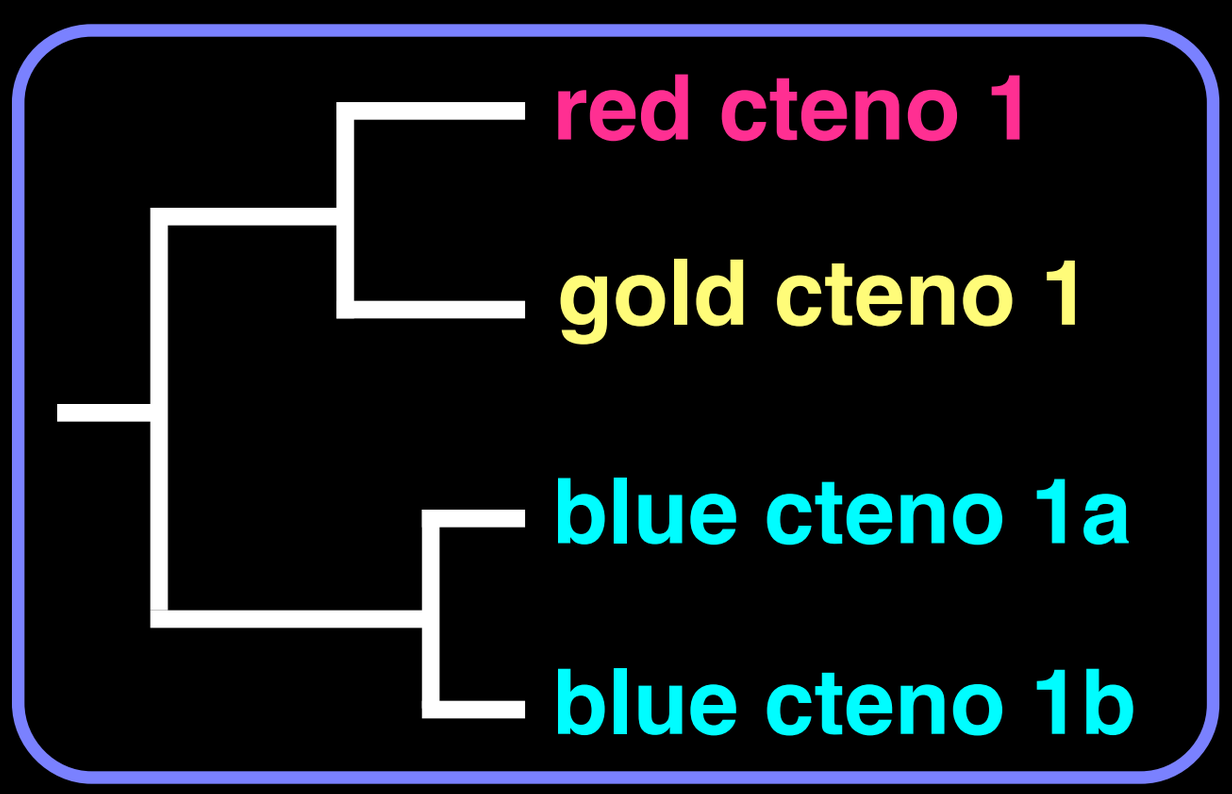


Paralogs

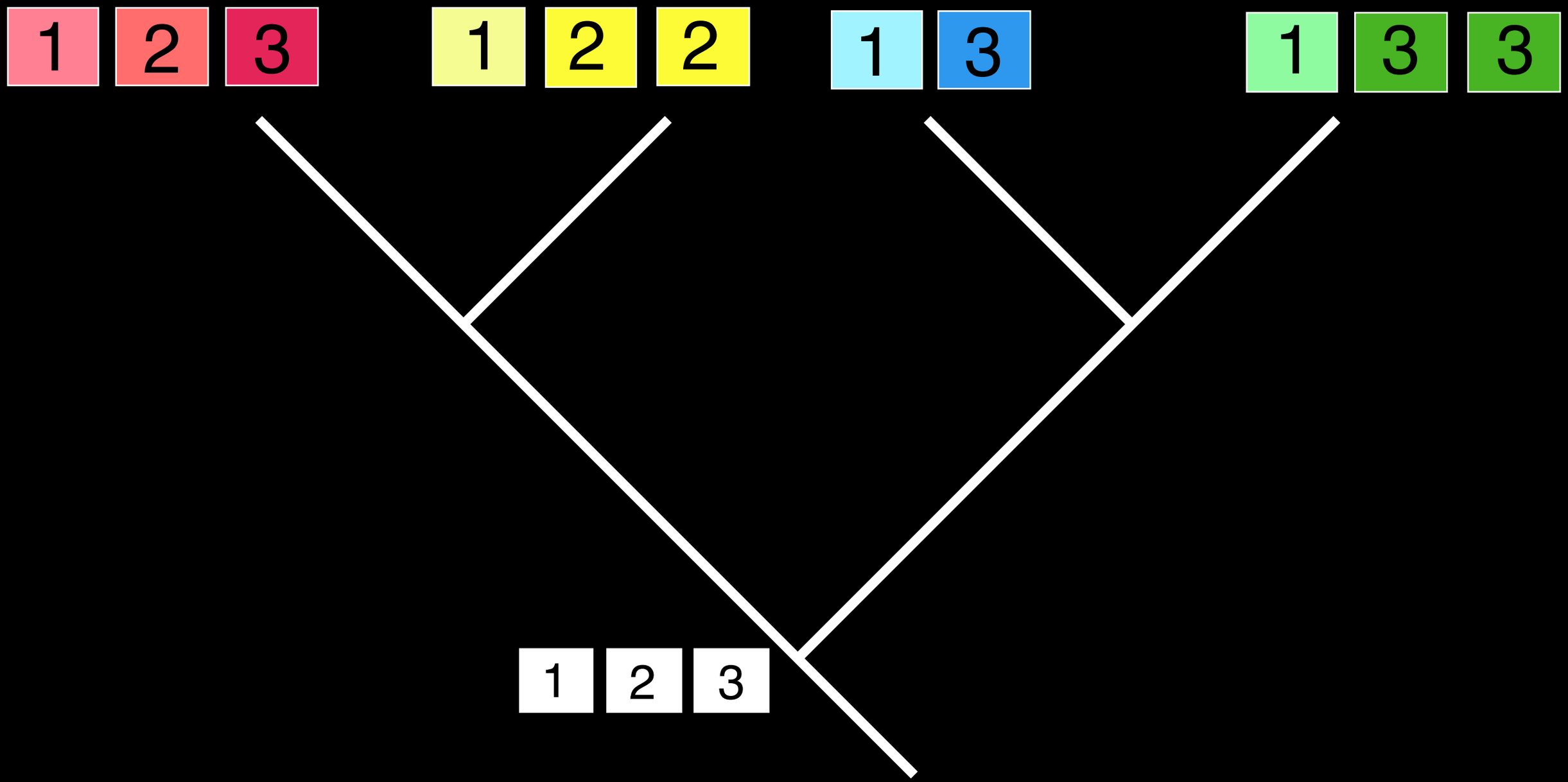
Orthologs: evolved from a common ancestor by speciation

Paralogs: produced through a duplication event

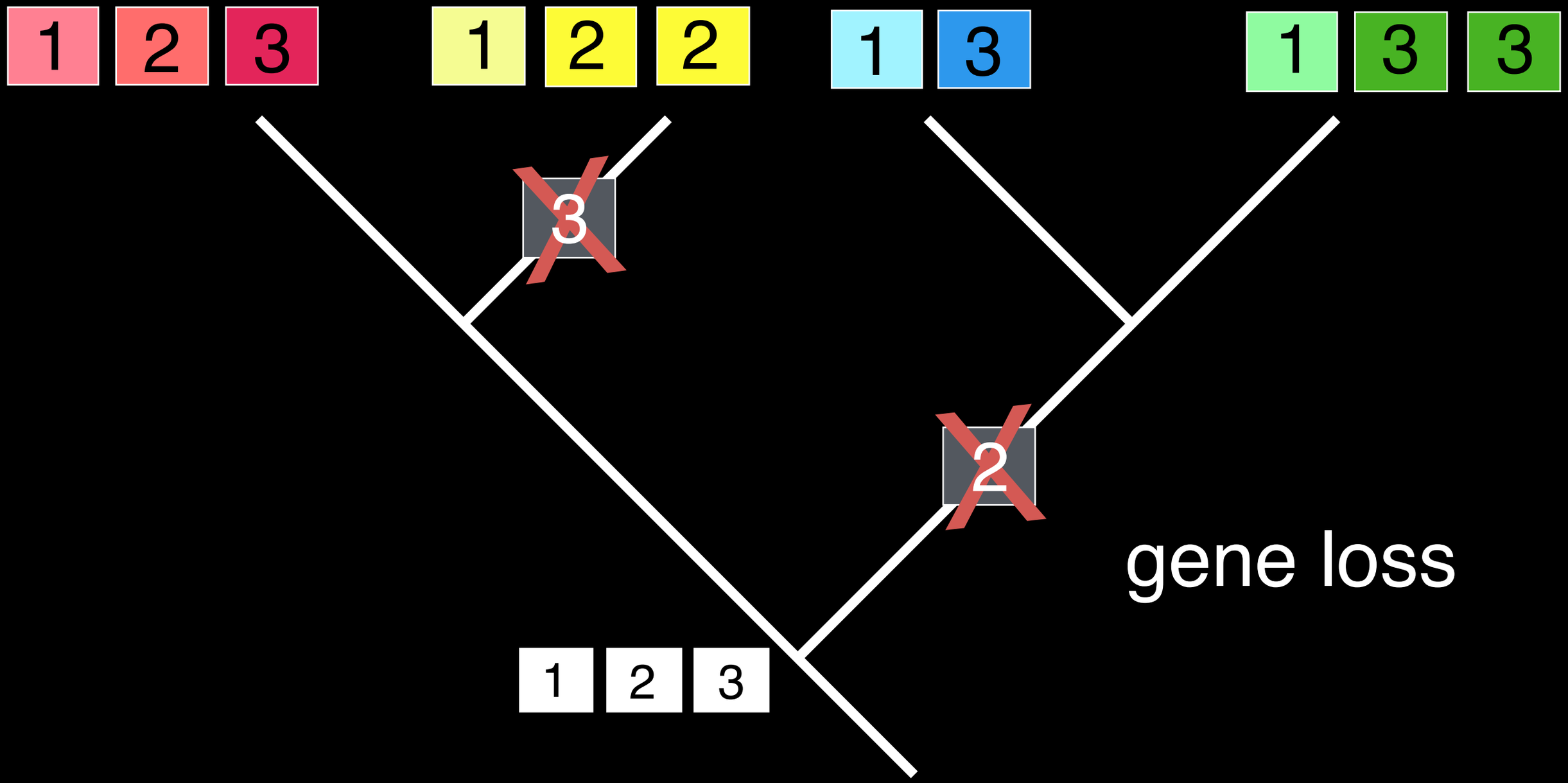
Orthogroup: orthologs and paralogs from MRCA



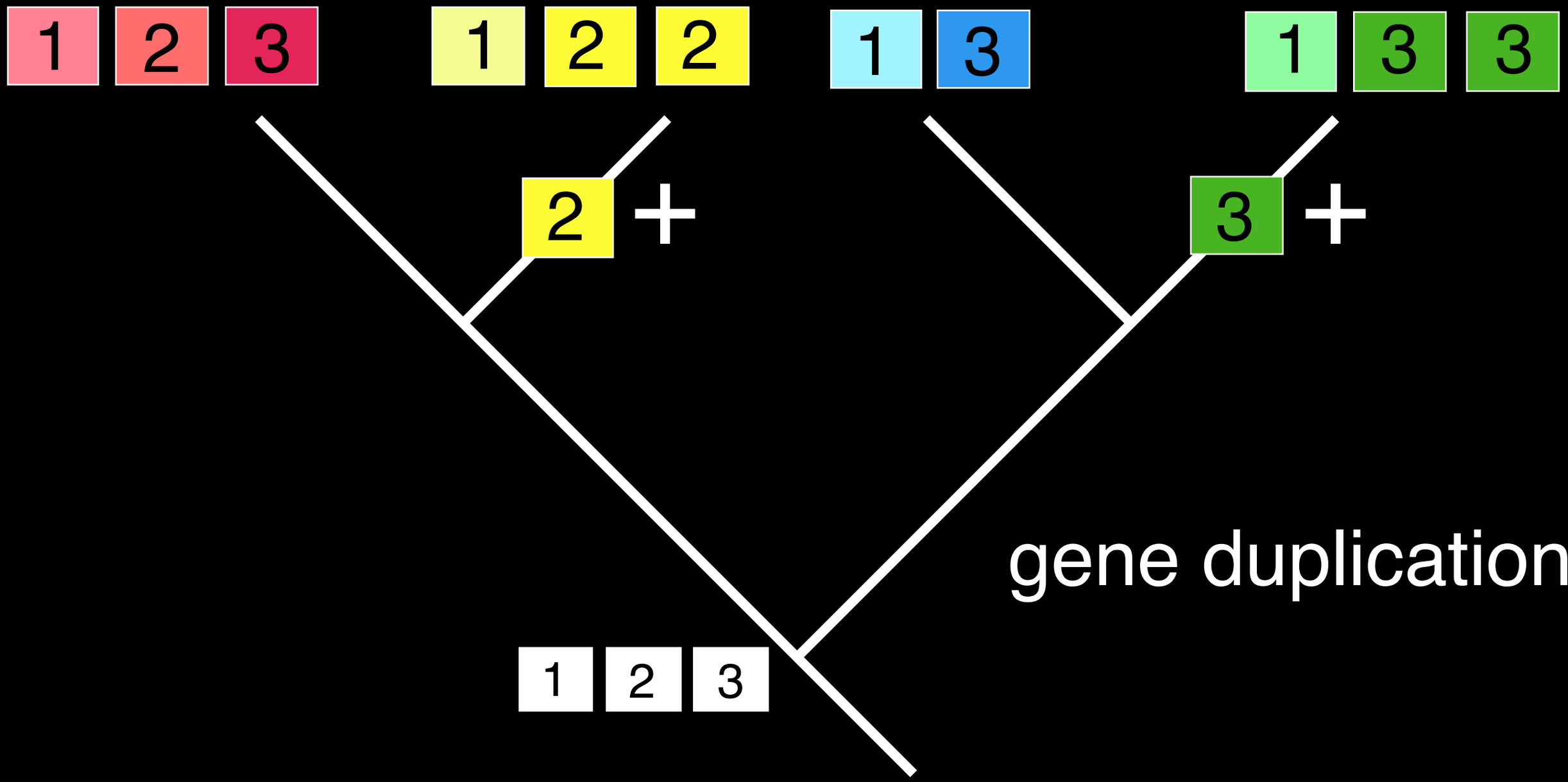
Orthogroup



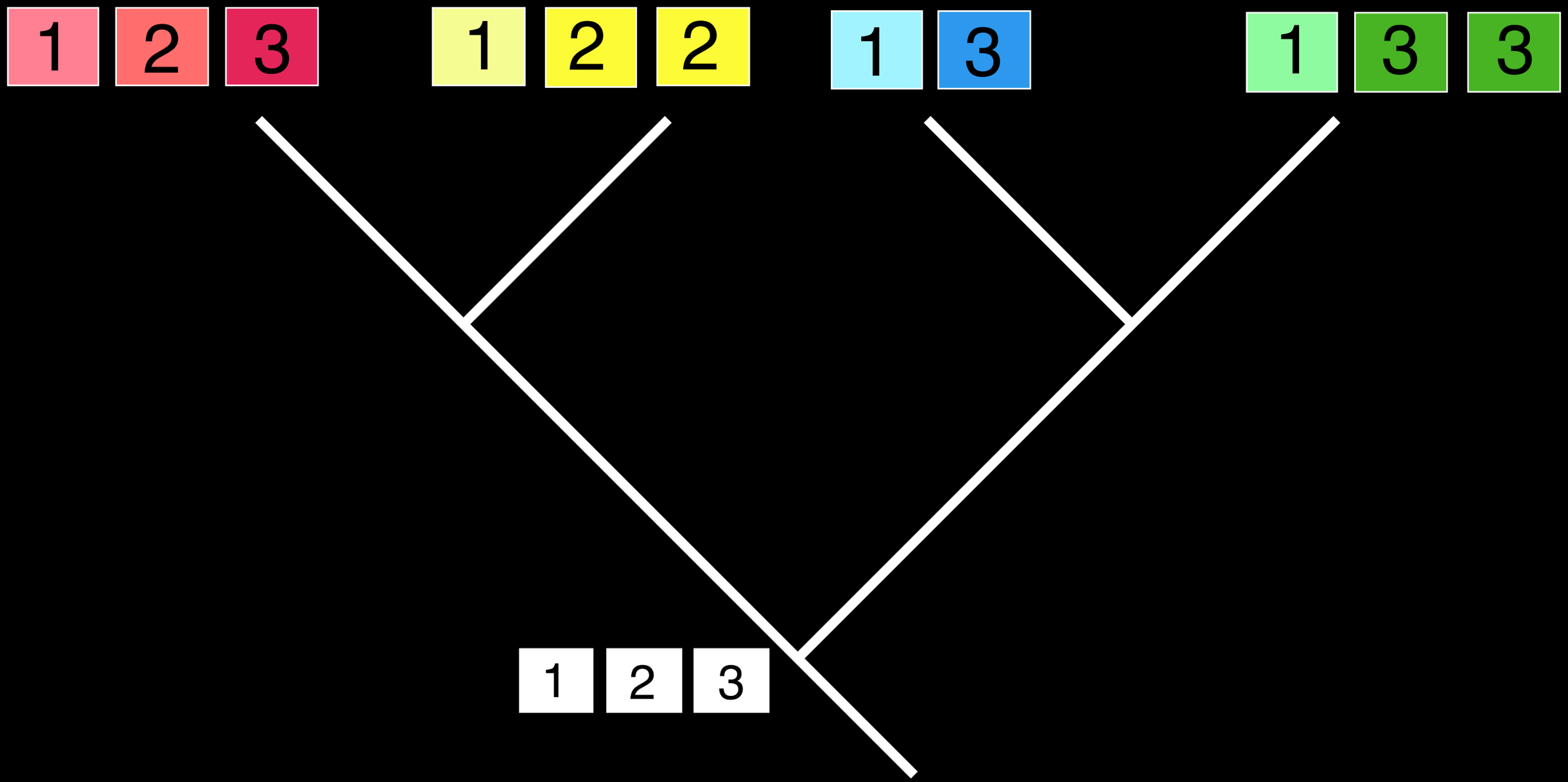
colors = species
numbers = genes



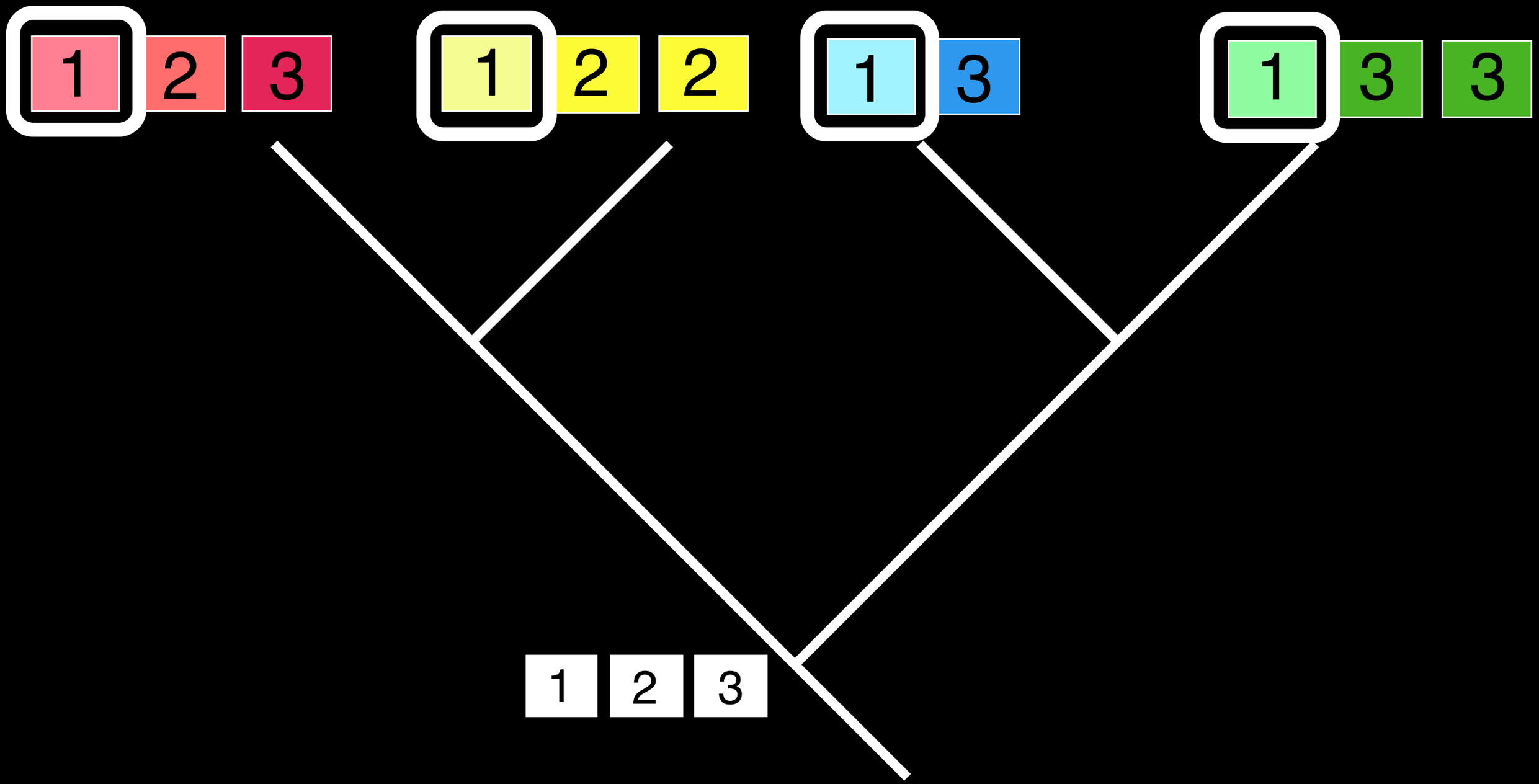
colors = species
numbers = genes

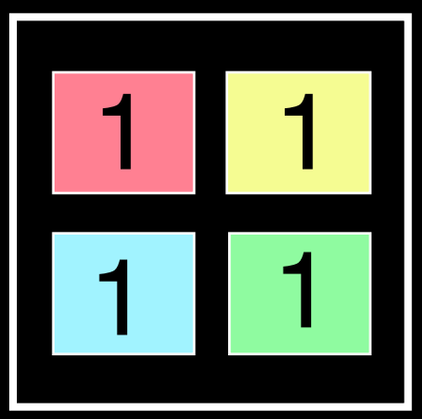


colors = species
numbers = genes

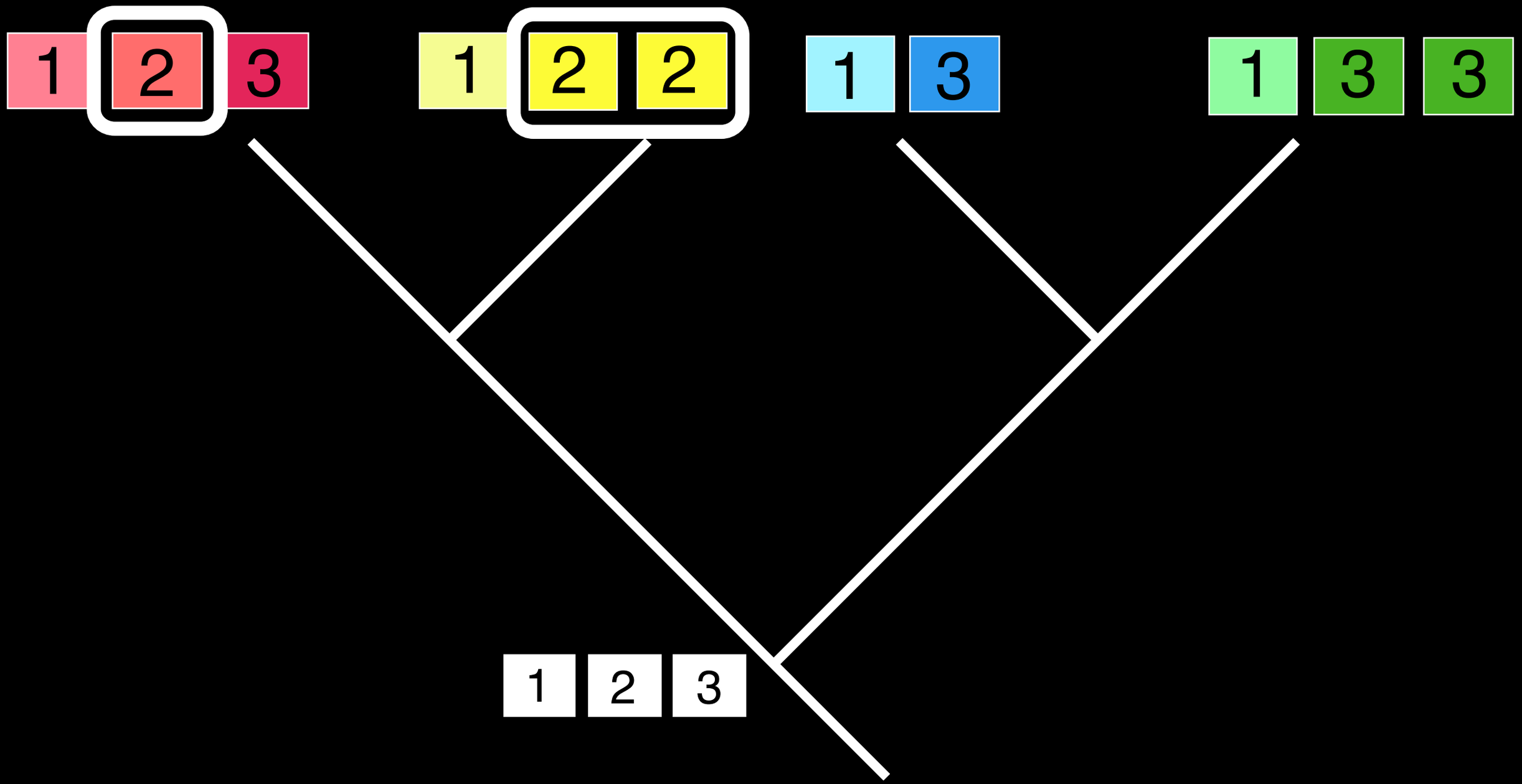


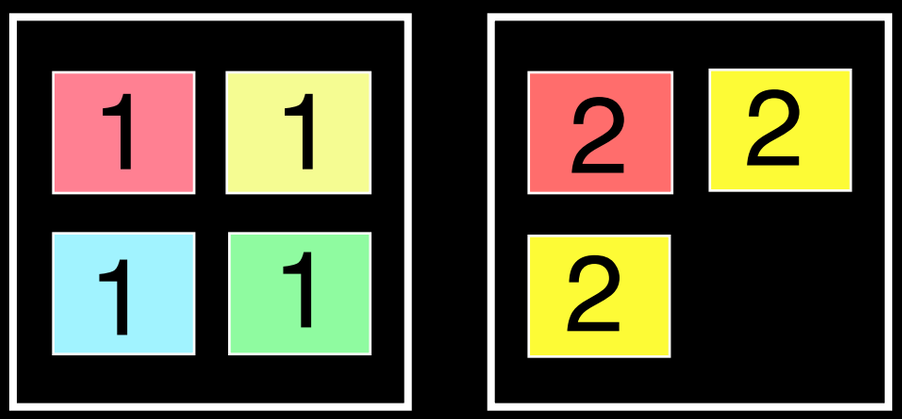
what are the ideal orthogroups?



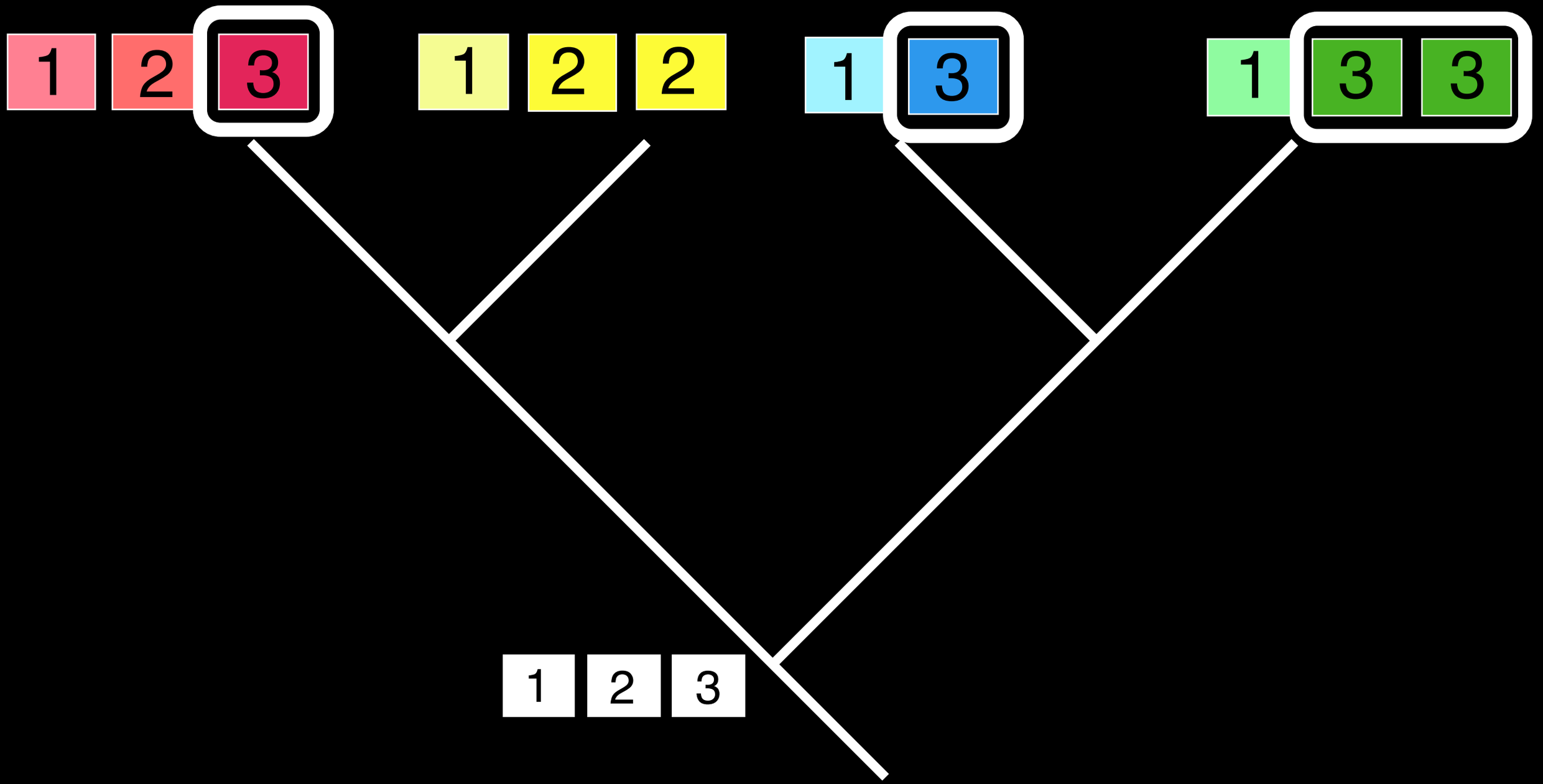


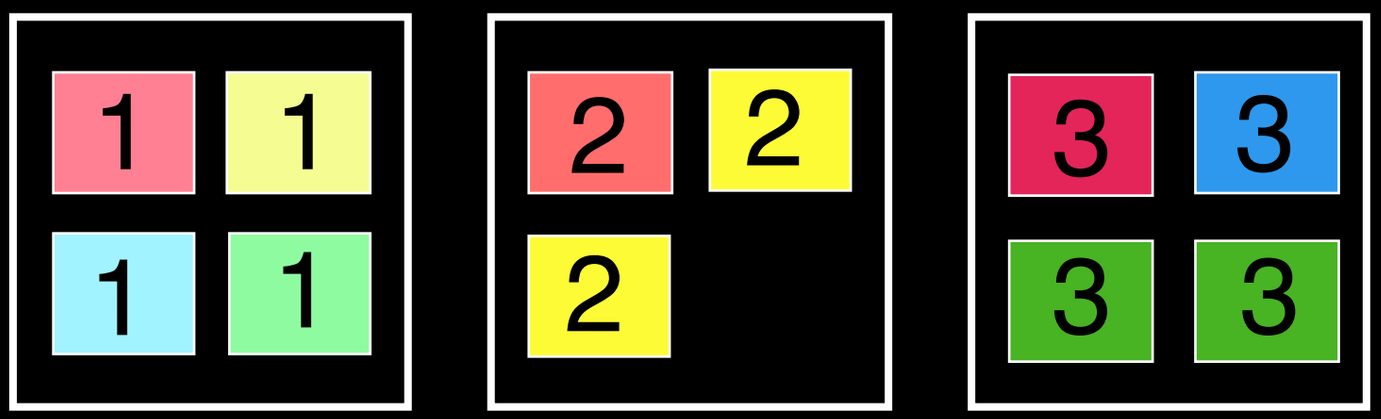
orthogroup 1



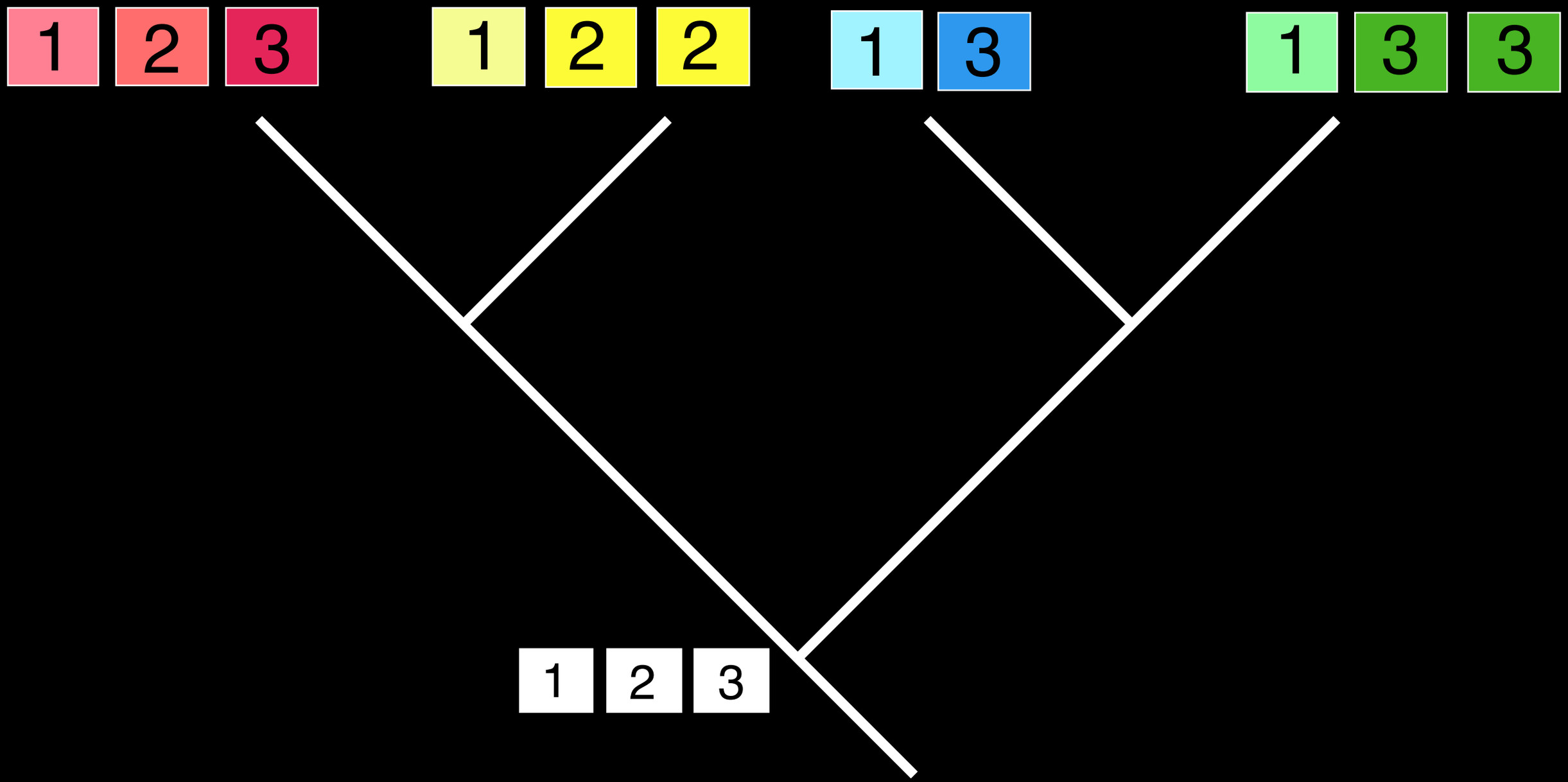


orthogroup 2



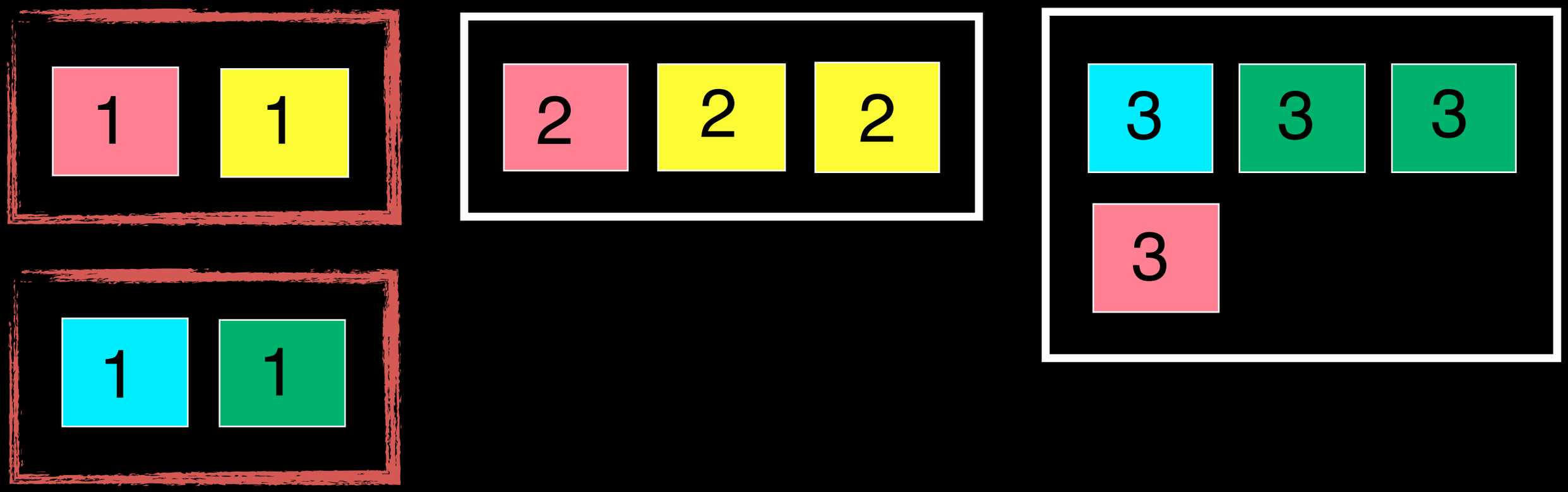


orthogroup 3



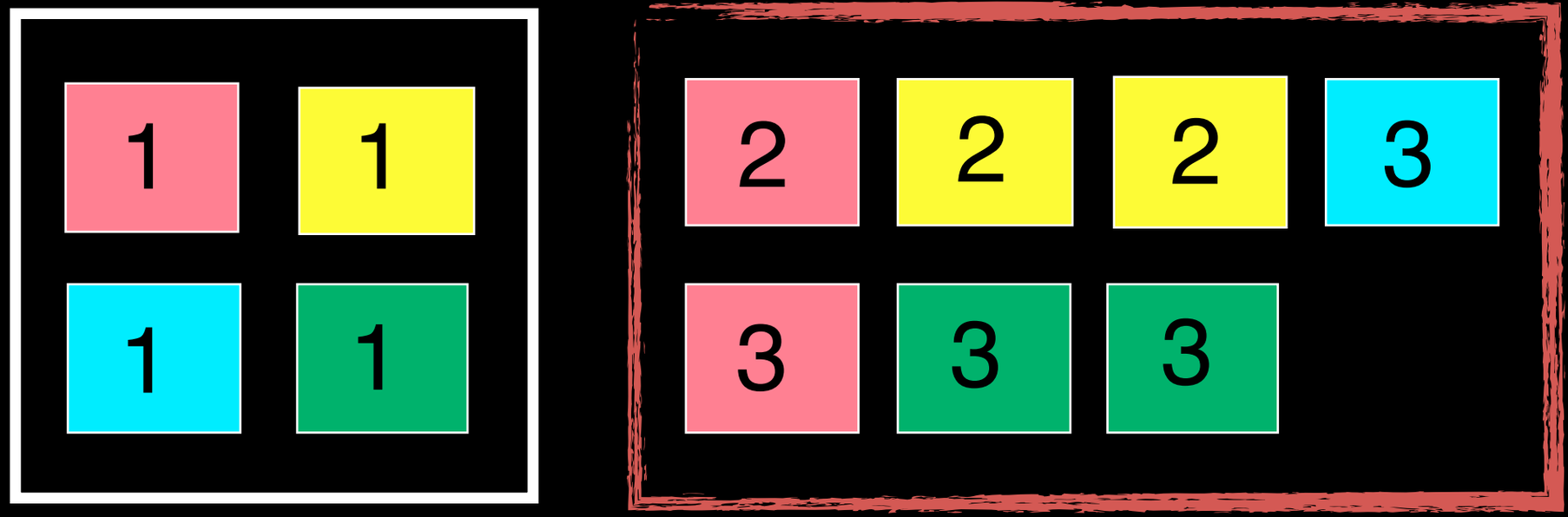
identifying orthology complicated by gene dup. & loss, rate variation, HGT, contamination, etc

Splitting



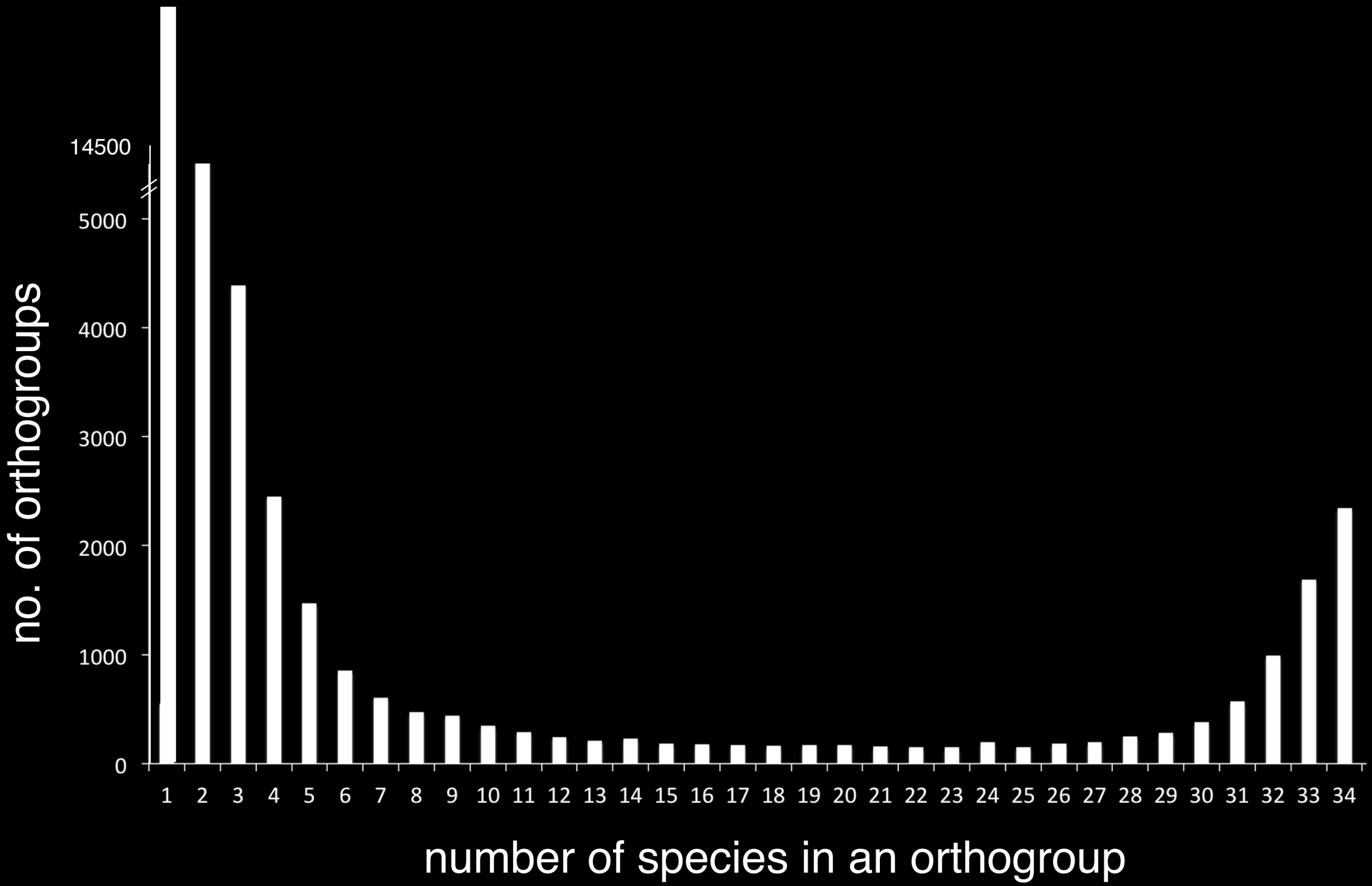
colors = species
numbers = genes

Lumping

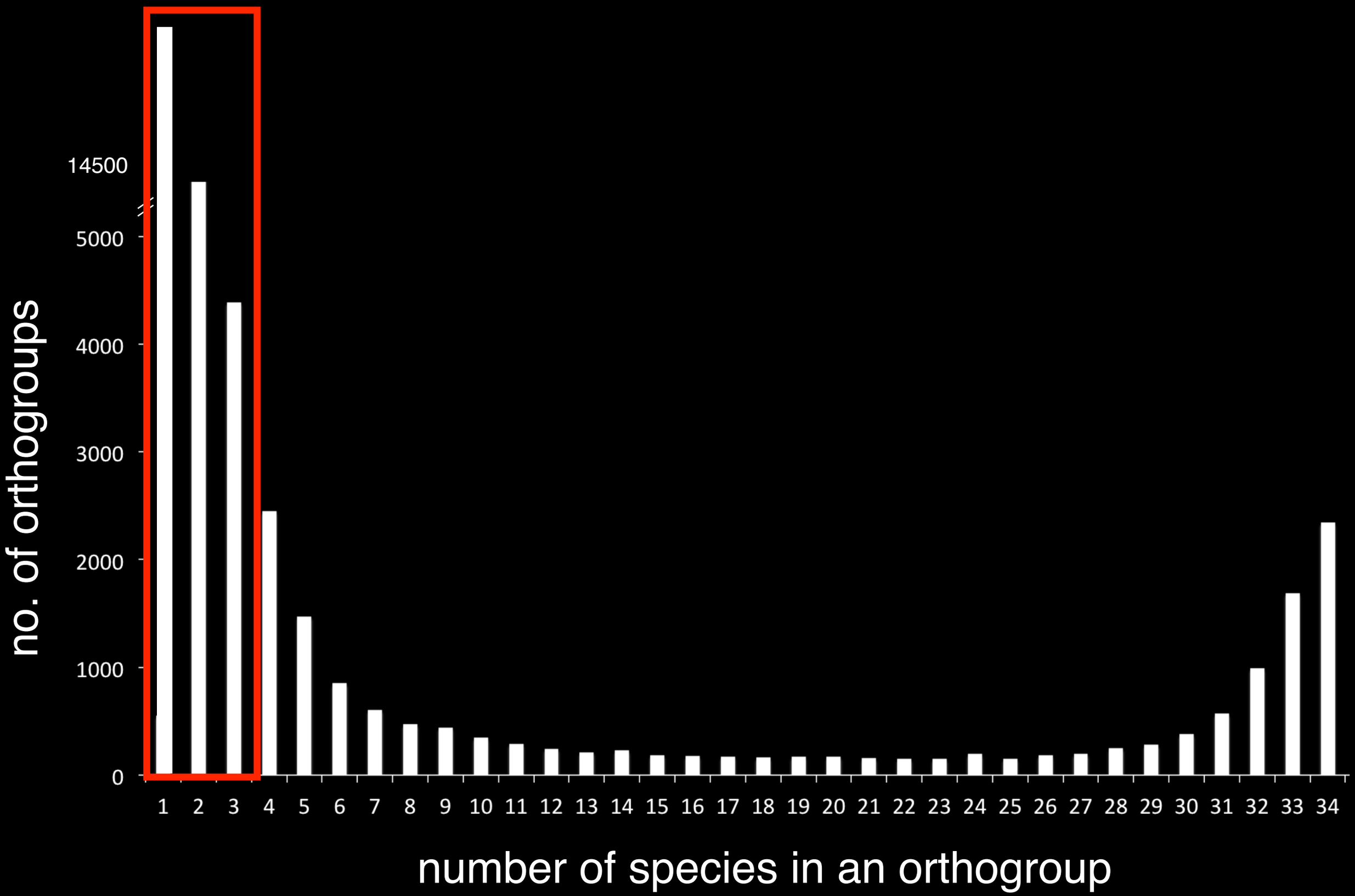


colors = species
numbers = genes

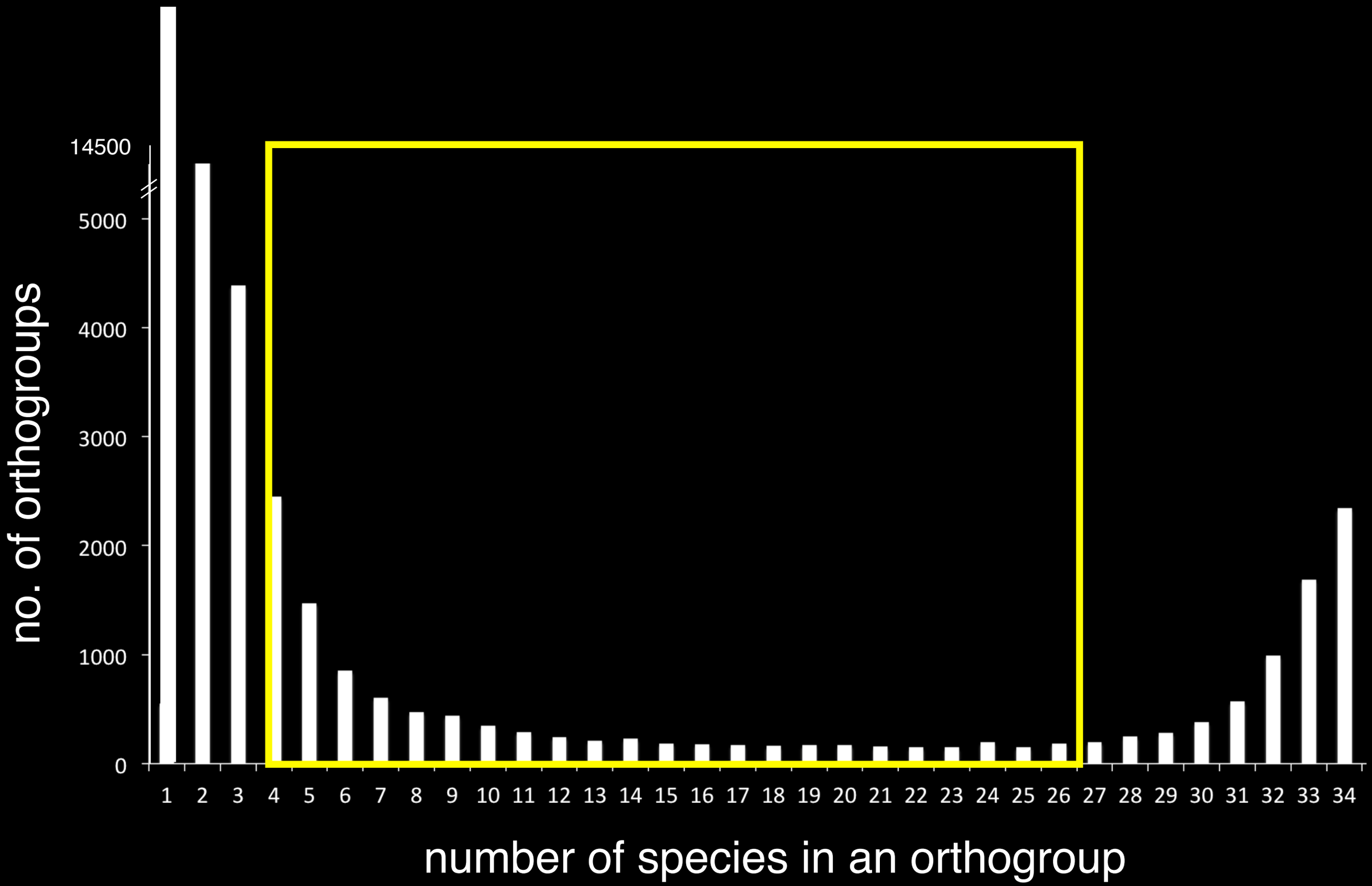
OrthoFinder



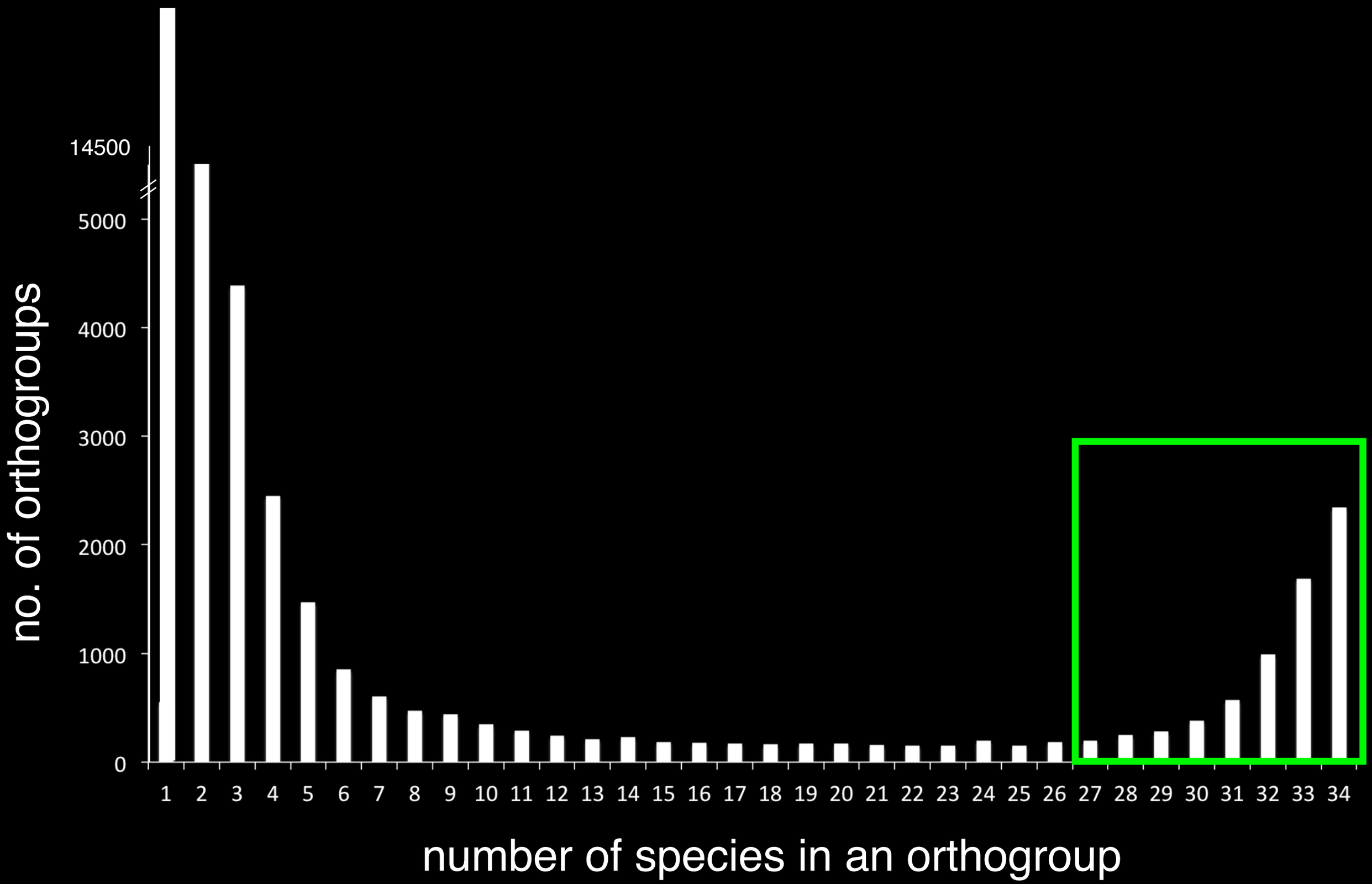
Splitting + translation artifacts



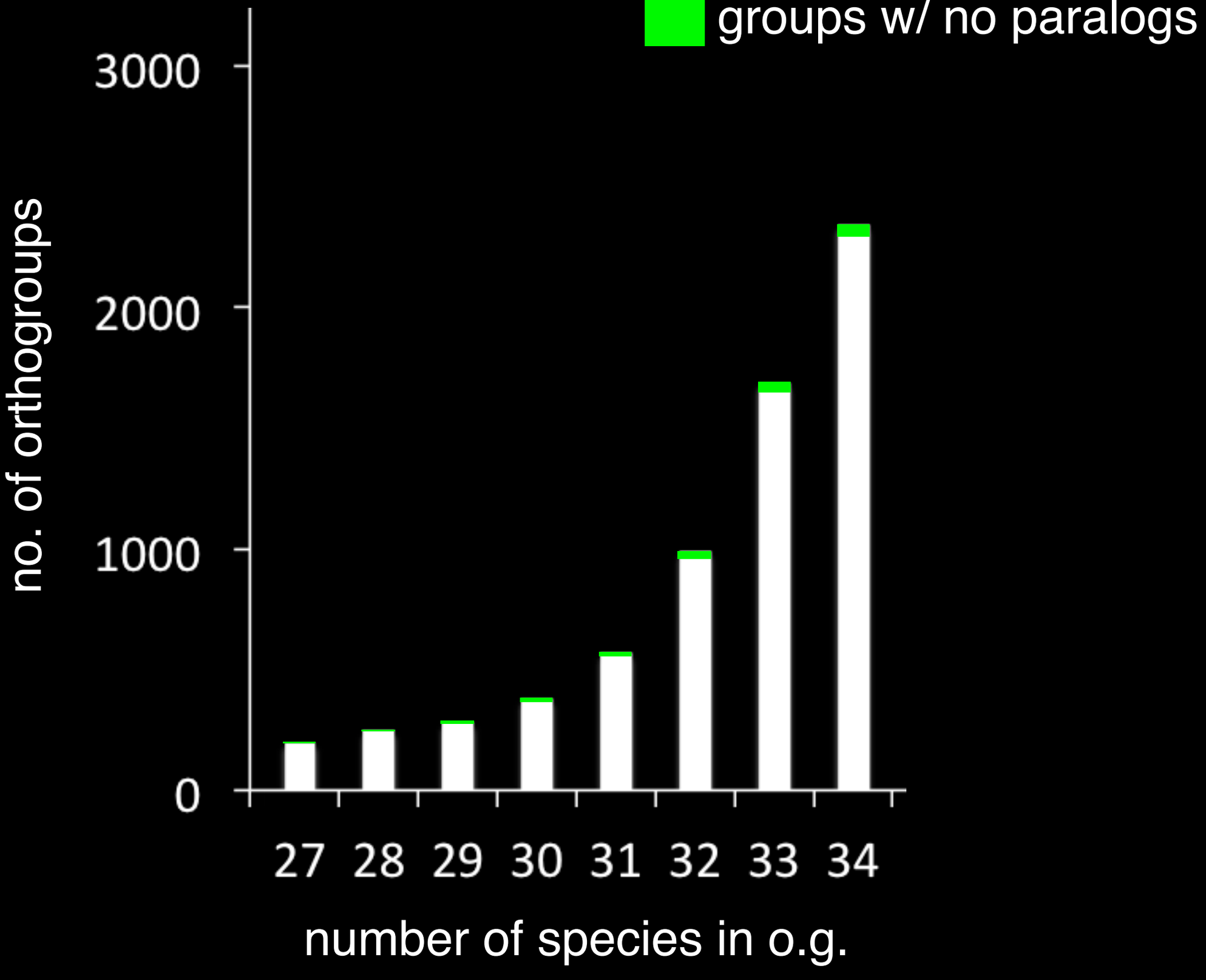
Splitting + gene loss



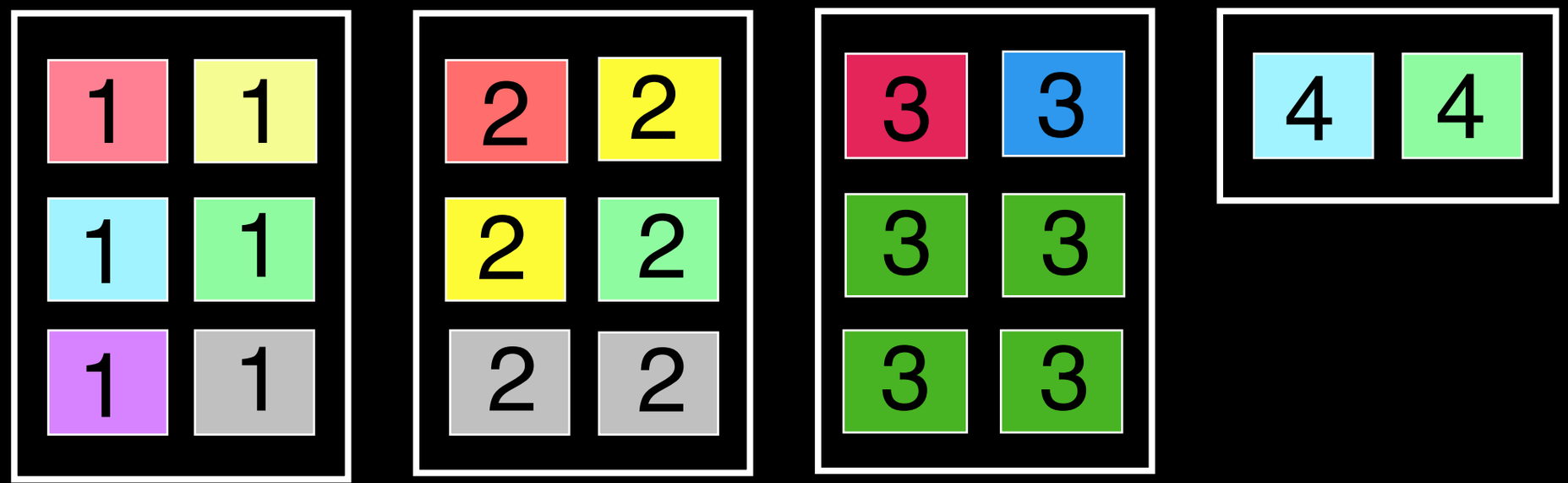
Lumping + single-copy orthogroups



OrthoFinder



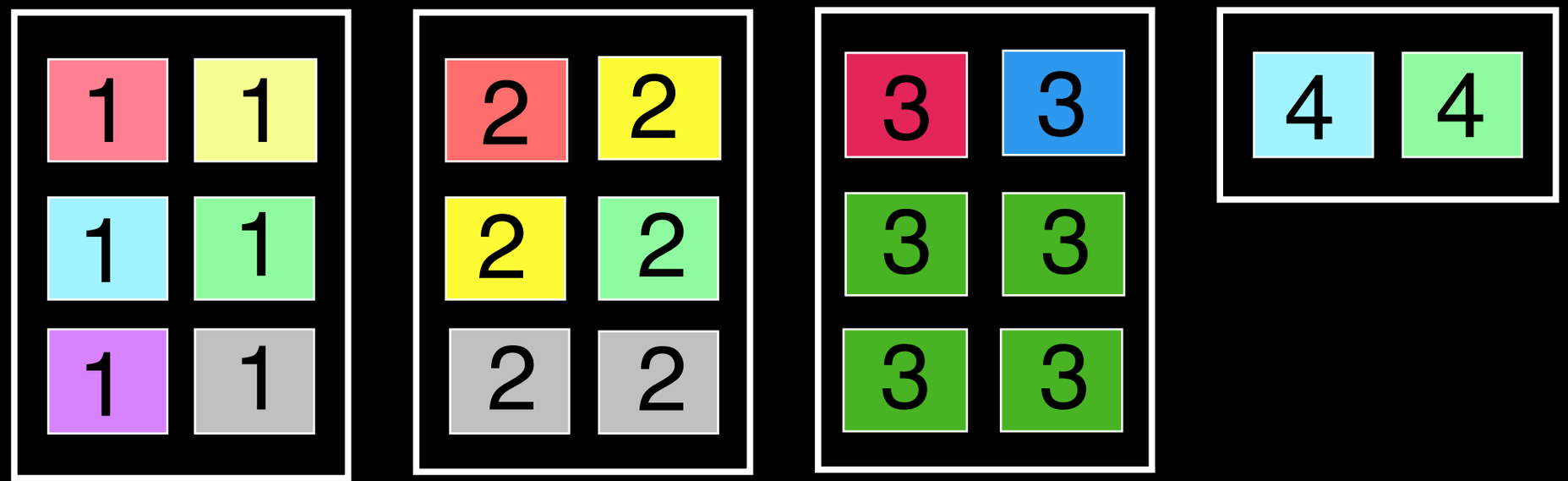
Orthogroup filtering



colors = species
numbers = genes

Orthogroup filtering

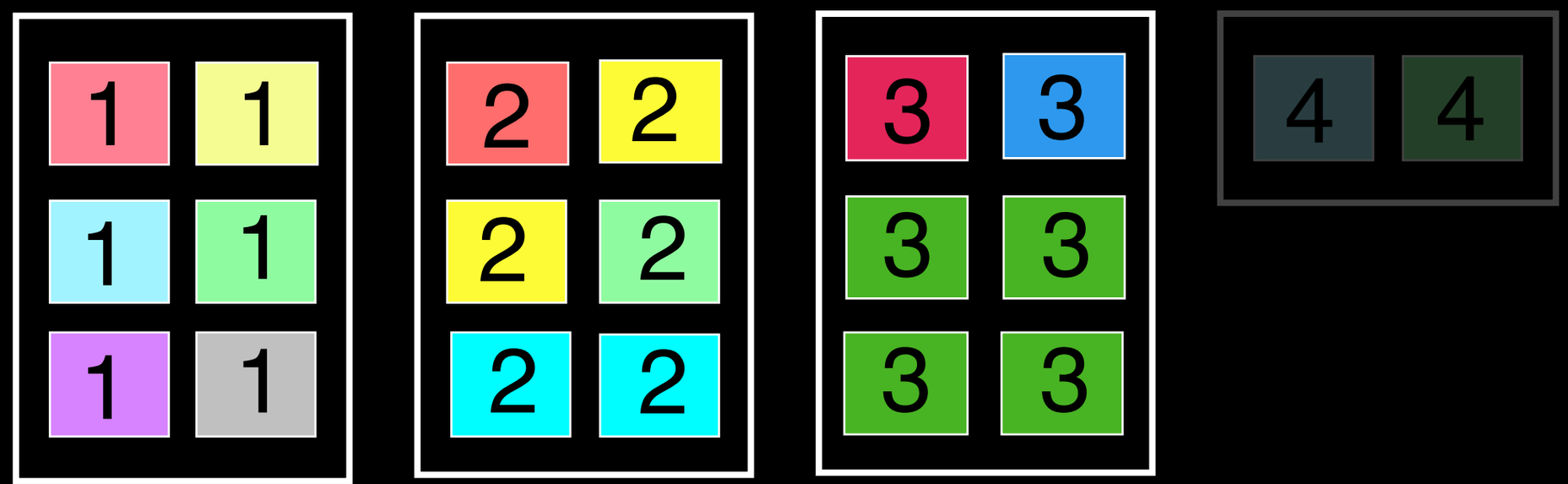
Minimum number of taxa (80%)



colors = species
numbers = genes

Orthogroup filtering

Minimum number of taxa (80%)

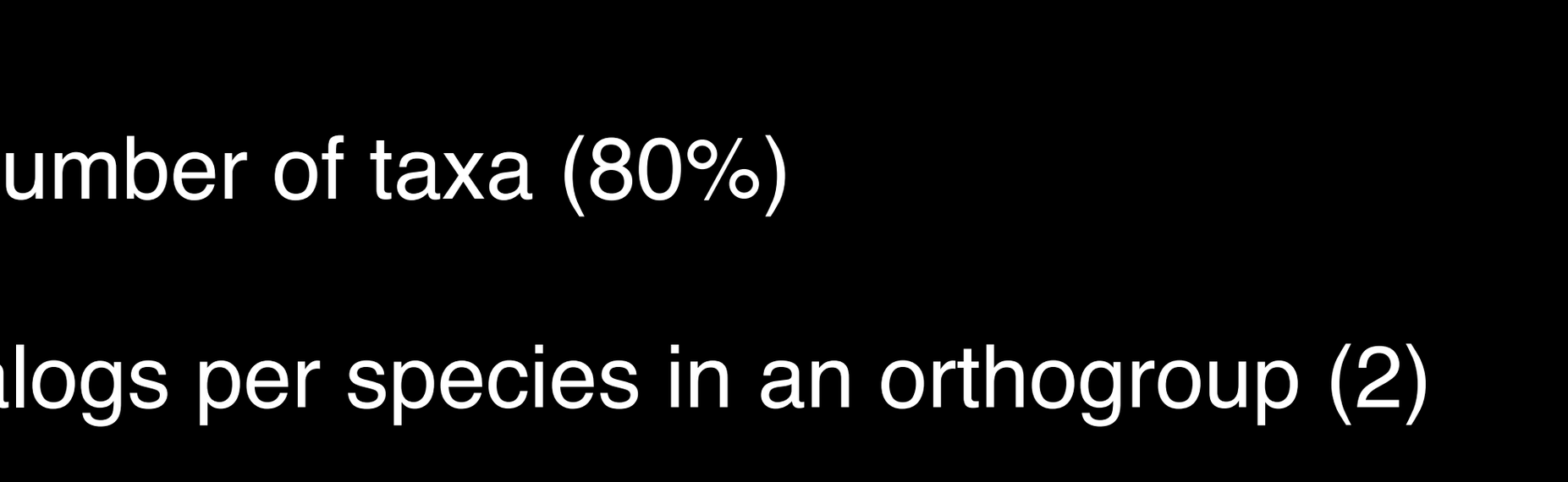


colors = species
numbers = genes

Orthogroup filtering

Minimum number of taxa (80%)

Max *n* paralogs per species in an orthogroup (2)

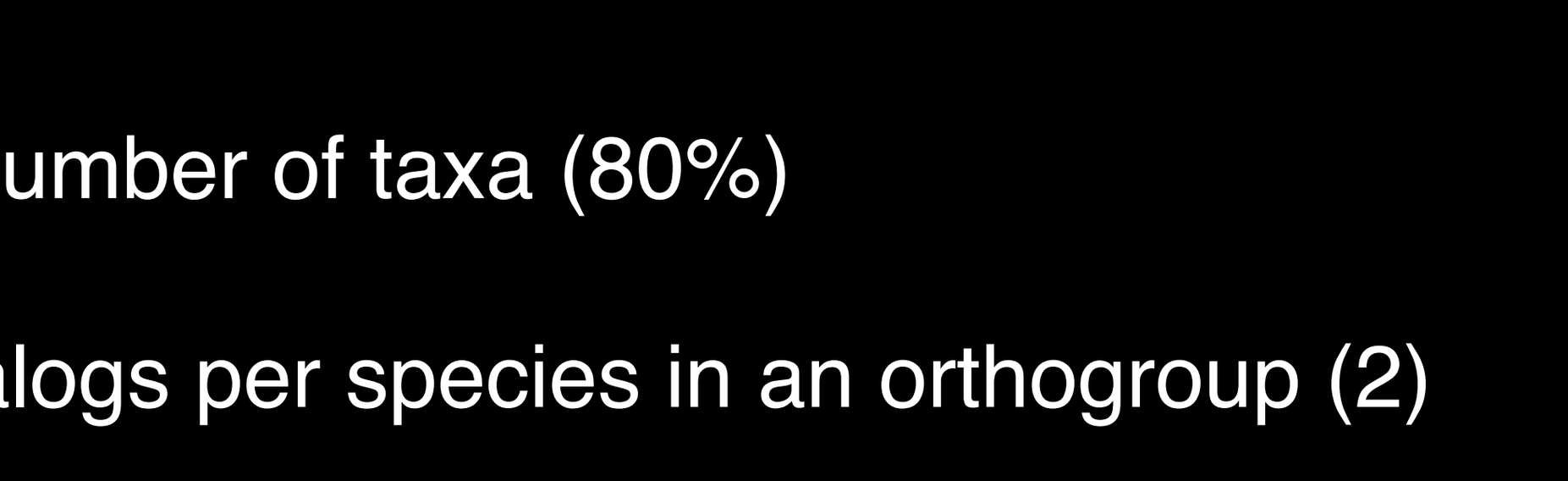


colors = species
numbers = genes

Orthogroup filtering

Minimum number of taxa (80%)

Max *n* paralogs per species in an orthogroup (2)

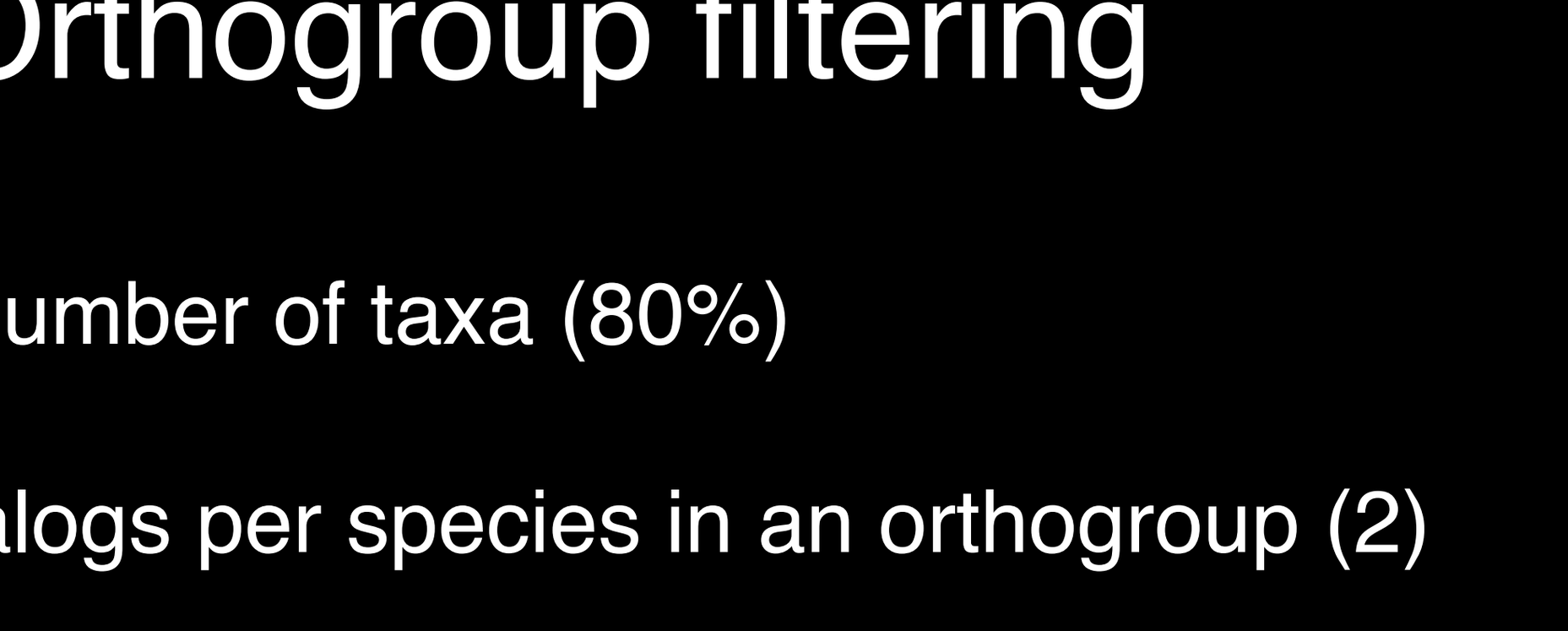


colors = species
numbers = genes

Orthogroup filtering

Minimum number of taxa (80%)

Max *n* paralogs per species in an orthogroup (2)



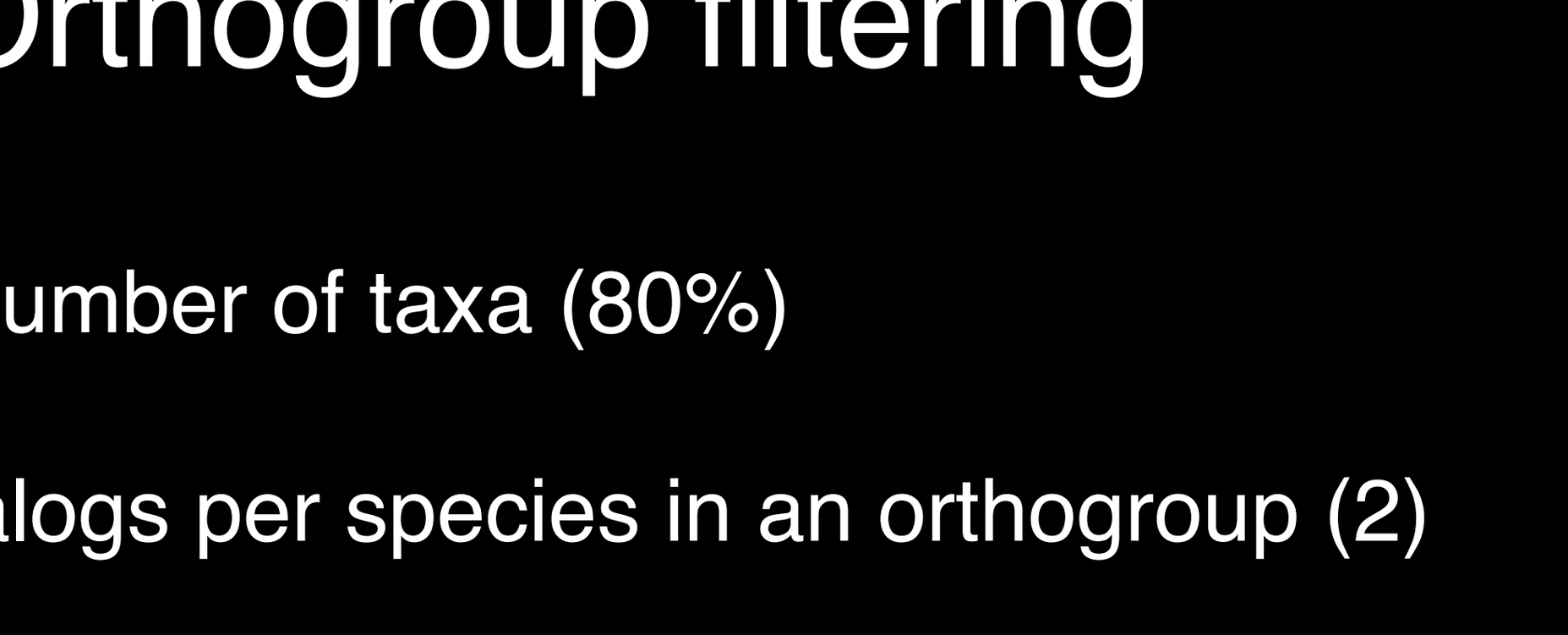
colors = species
numbers = genes

Orthogroup filtering

Minimum number of taxa (80%)

Max *n* paralogs per species in an orthogroup (2)

PhyloTreePruner



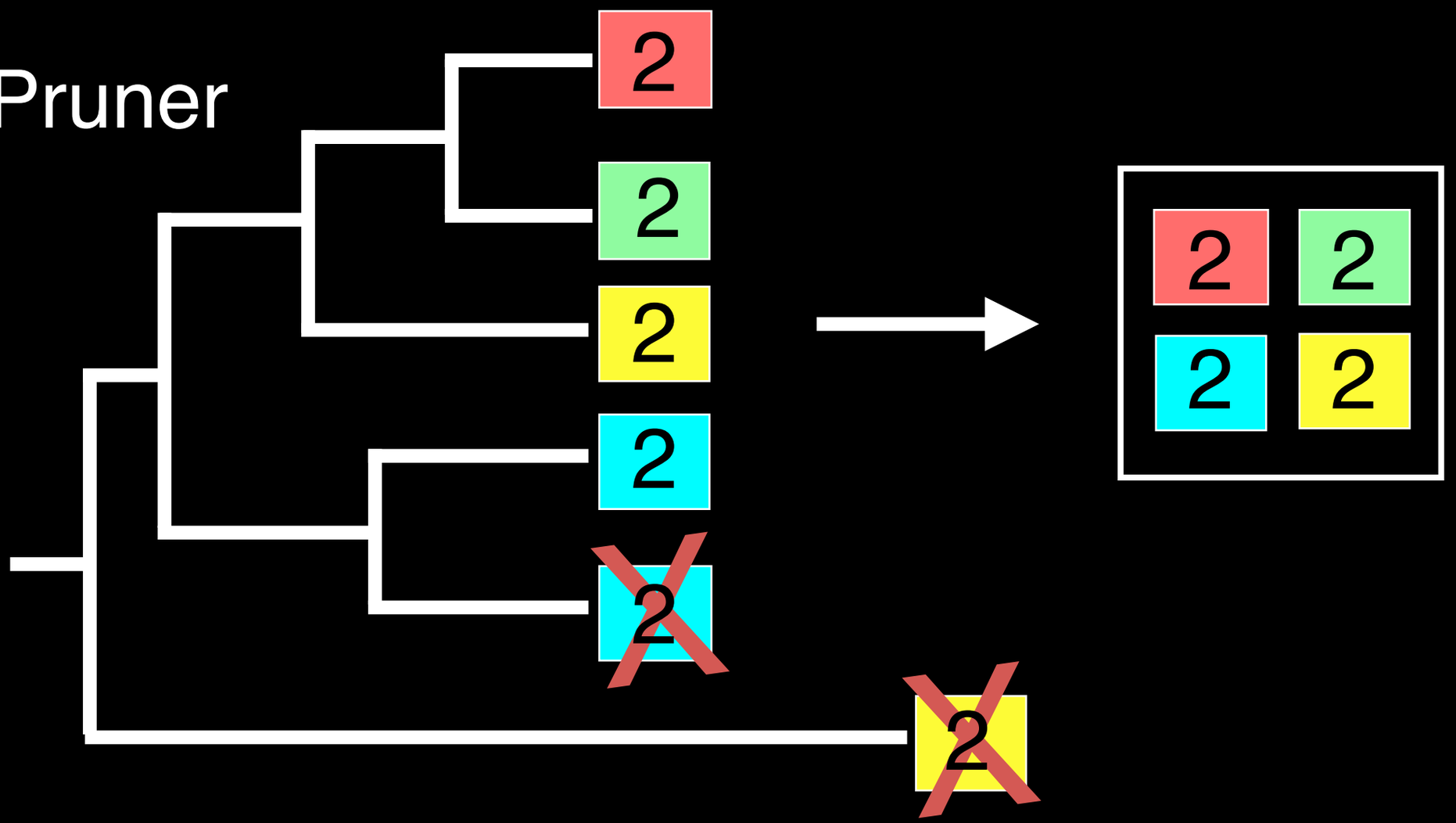
colors = species
numbers = genes

Orthogroup filtering

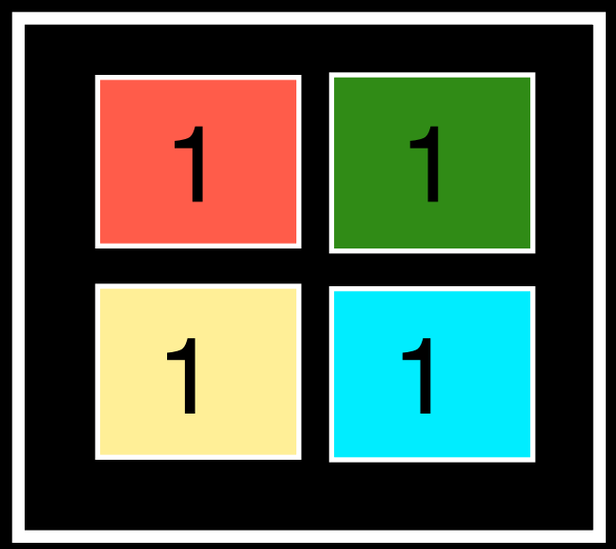
Minimum number of taxa (80%)

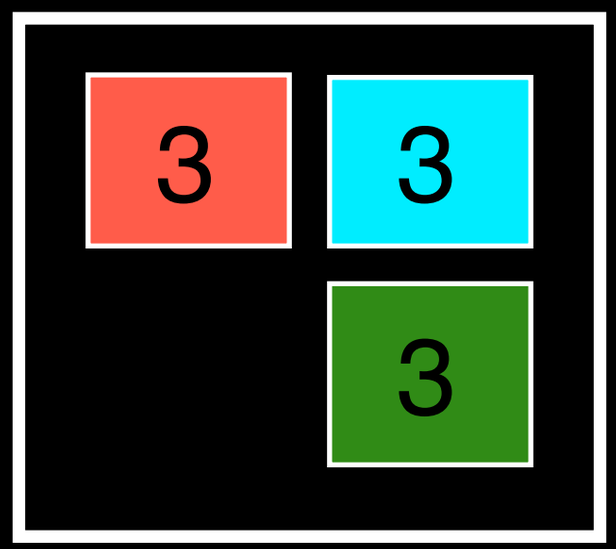
Max *n* paralogs per species in an orthogroup (2)

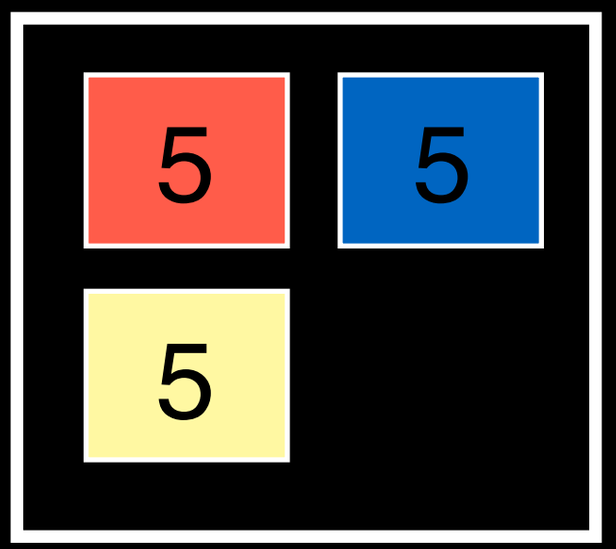
PhyloTreePruner

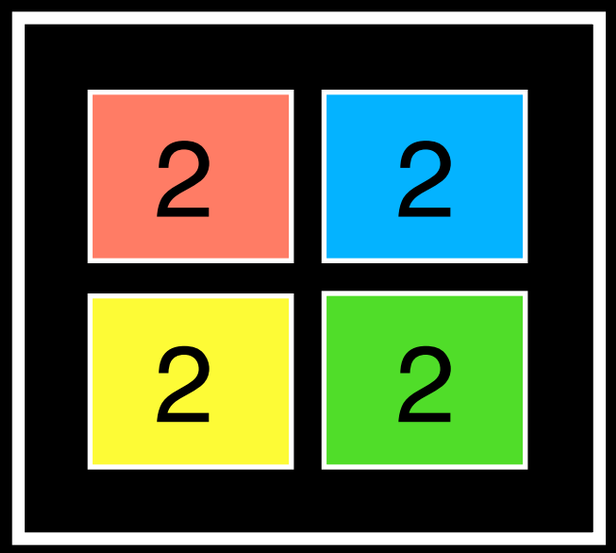


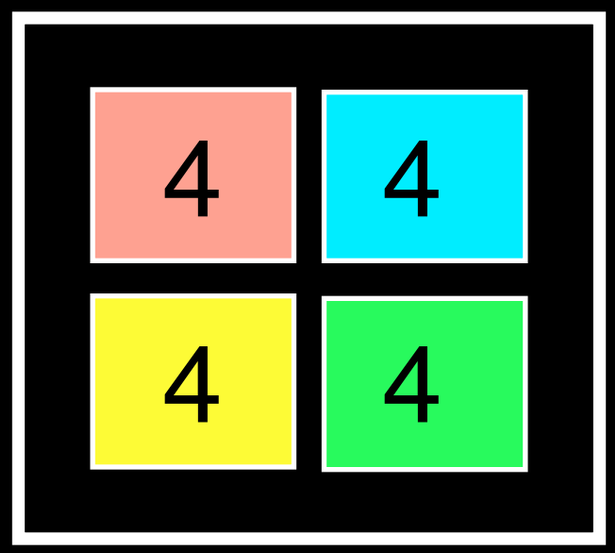
Orthogroup filtering

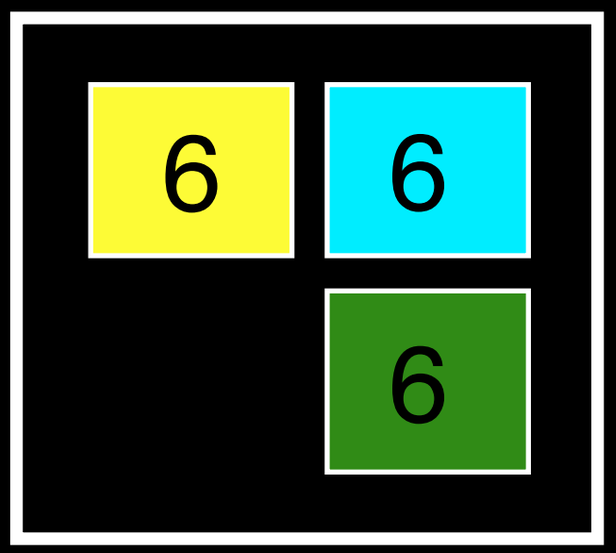












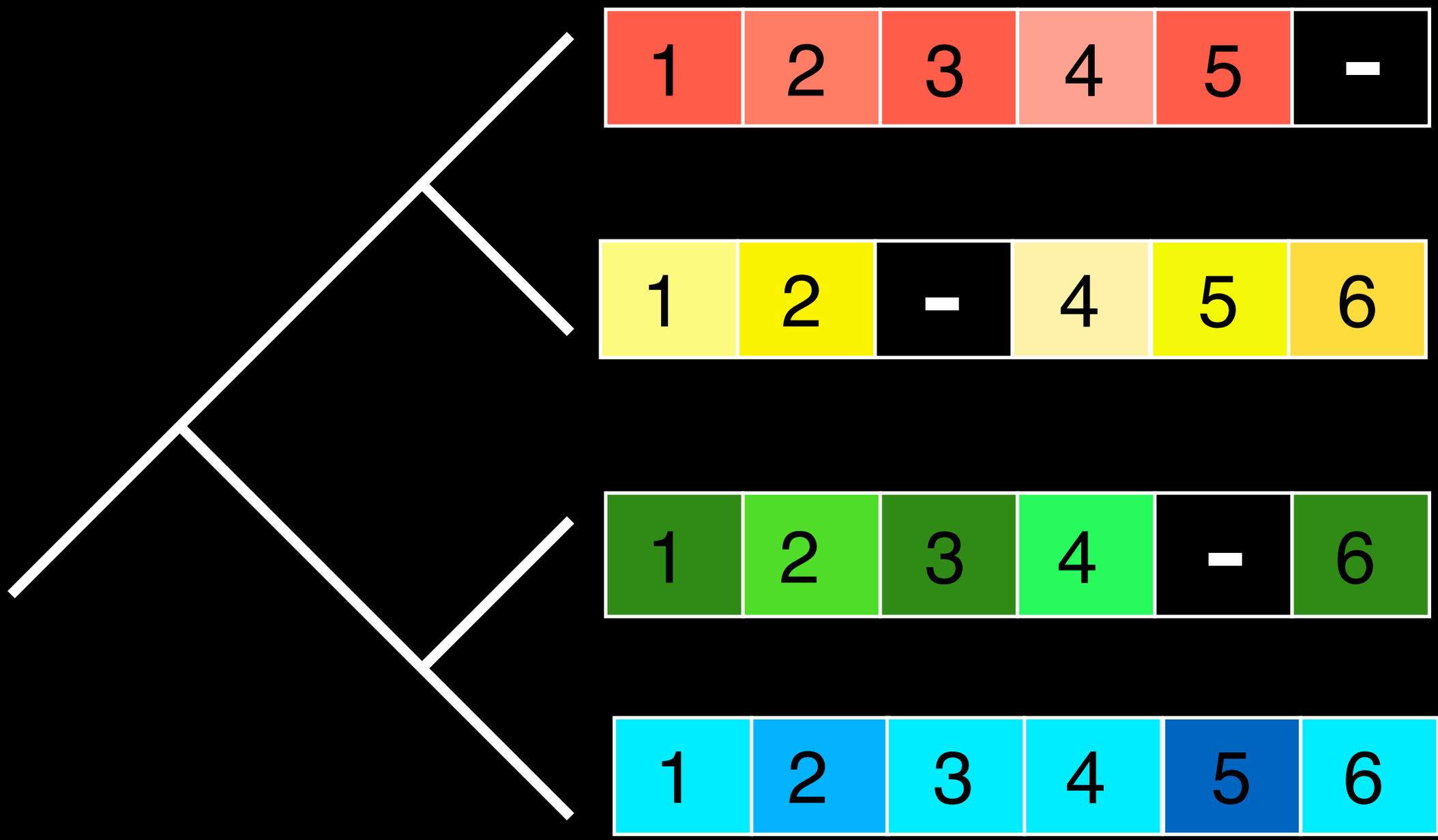
colors = species
numbers = genes

Align & concatenate single-copy orthogroups

1	2	3	4	5	-
1	2	-	4	5	6
1	2	3	4	-	6
1	2	3	4	5	6

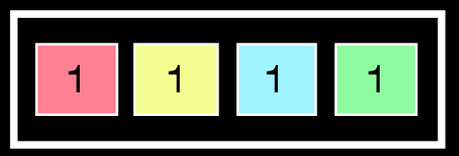
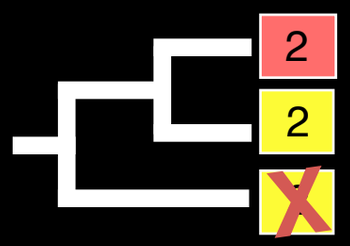
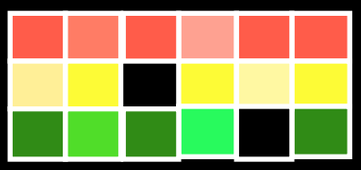
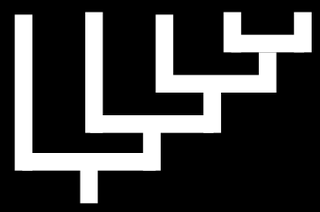
colors = species
numbers = genes

Use data matrix to infer species phylogeny

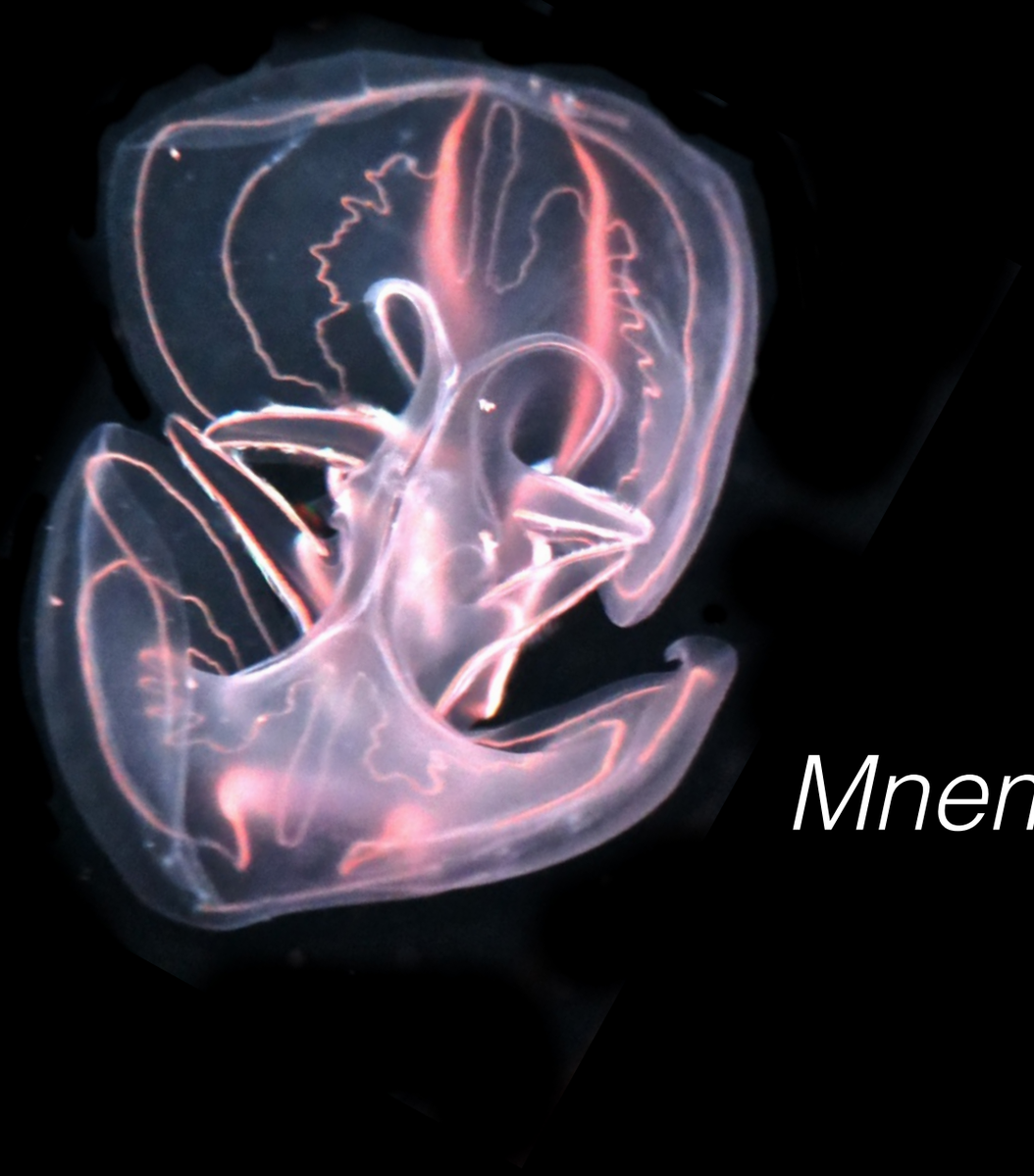


multi-locus concatenated species tree

PIPELINE

Illumina	**RNA-Seq**	
Trinity	**Assemble**	>Trinity_1357 CCGGAACGAGCACAGCCA
TransDecoder	**Translate**	>Trinity_1357_pep PGRAQPWLIAPQRAWGLQR
Alien Index	**Filter**	BLAST database metazoa non-metazoa
Orthofinder	**Orthogroups**	
PhyloTreePruner	**Paralogs**	
custom script	**Concatenate**	
RAxML/IQ-TREE/PB	**Tree**	

Thank you! Questions?



Mnemiopsis leidyi

photo: Joseph Ryan