



Gene extraction & alignment for comparative genomics

Summer RCN Mentoring 2021
Part 1 &2

OUTLINE

Part 1&2: Gene Extraction

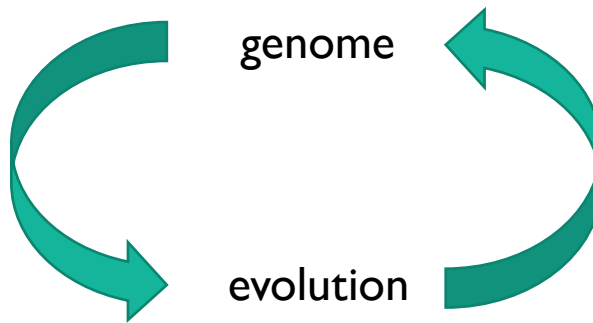
- Why do comparative genomics?
- Gene Models
- Sequence Extraction
- BLAST (Part 2 is if we get time)
- Homology
- Translation

Part 3: Multiple Alignment

- Codon aware alignment
- Alignment quality control

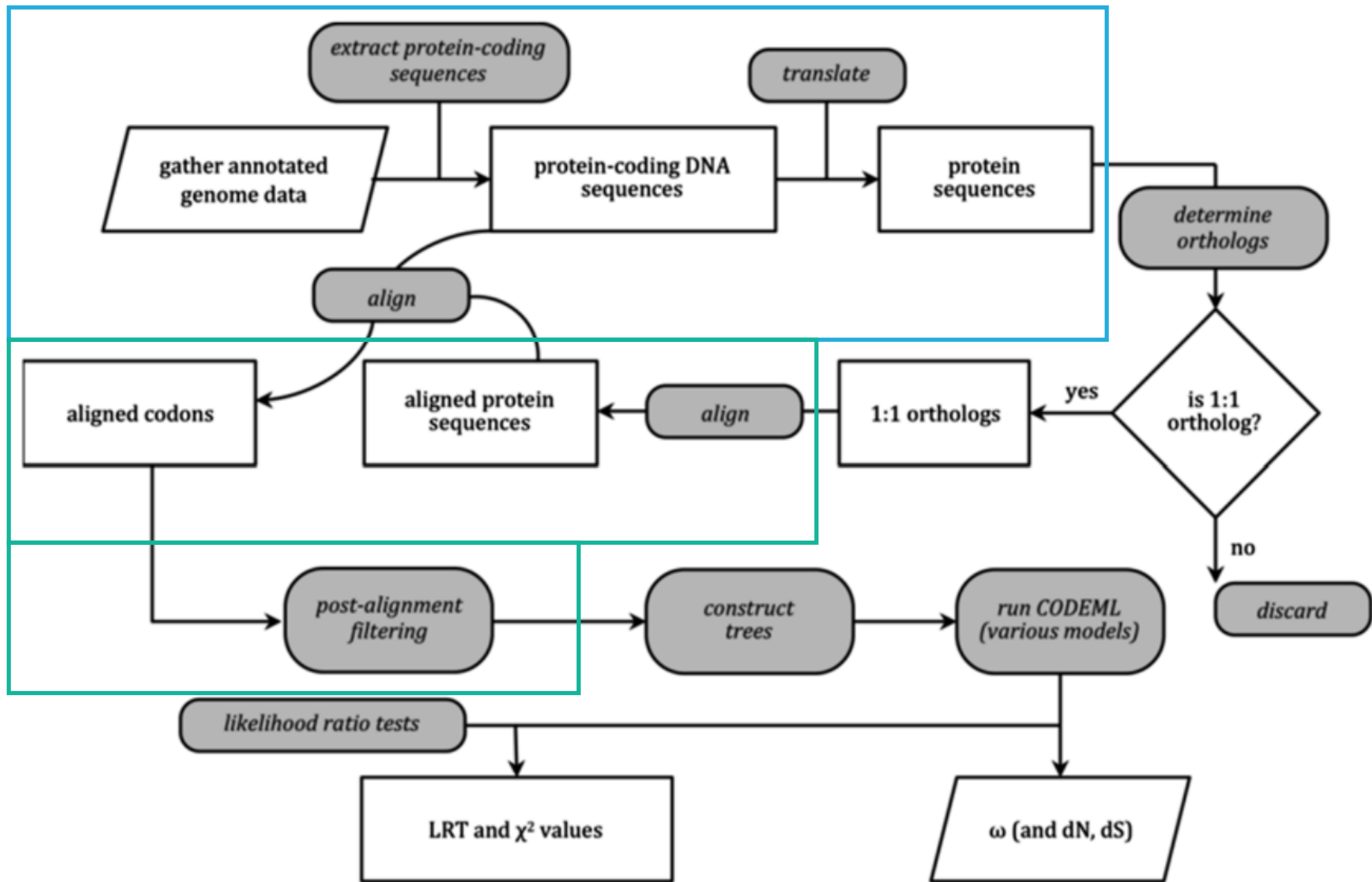
QUESTIONS IN COMPARATIVE GENOMICS

- What genes are under selection in a phylogenetic tree?
- How have genes or other genome features evolved across species?

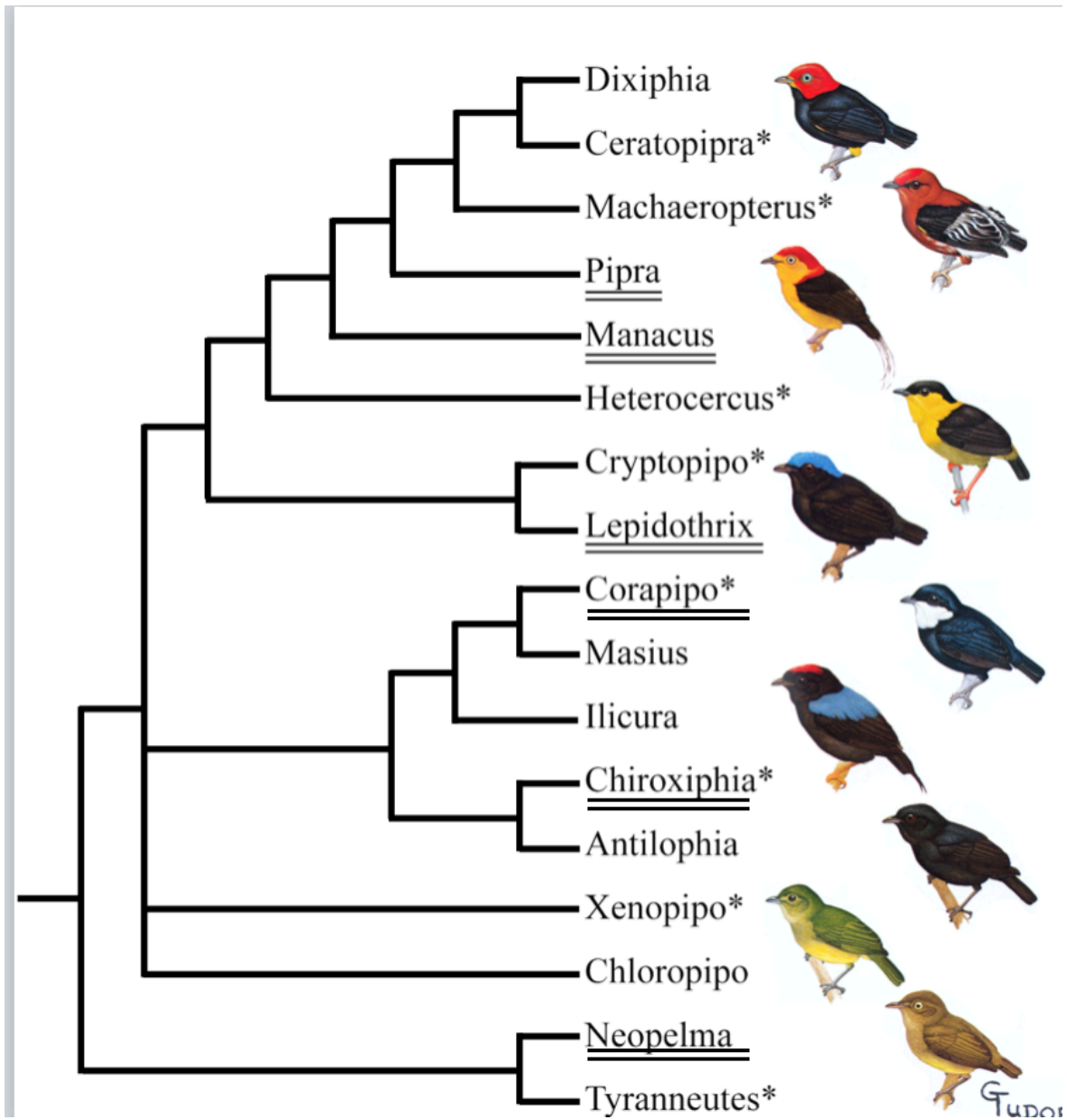


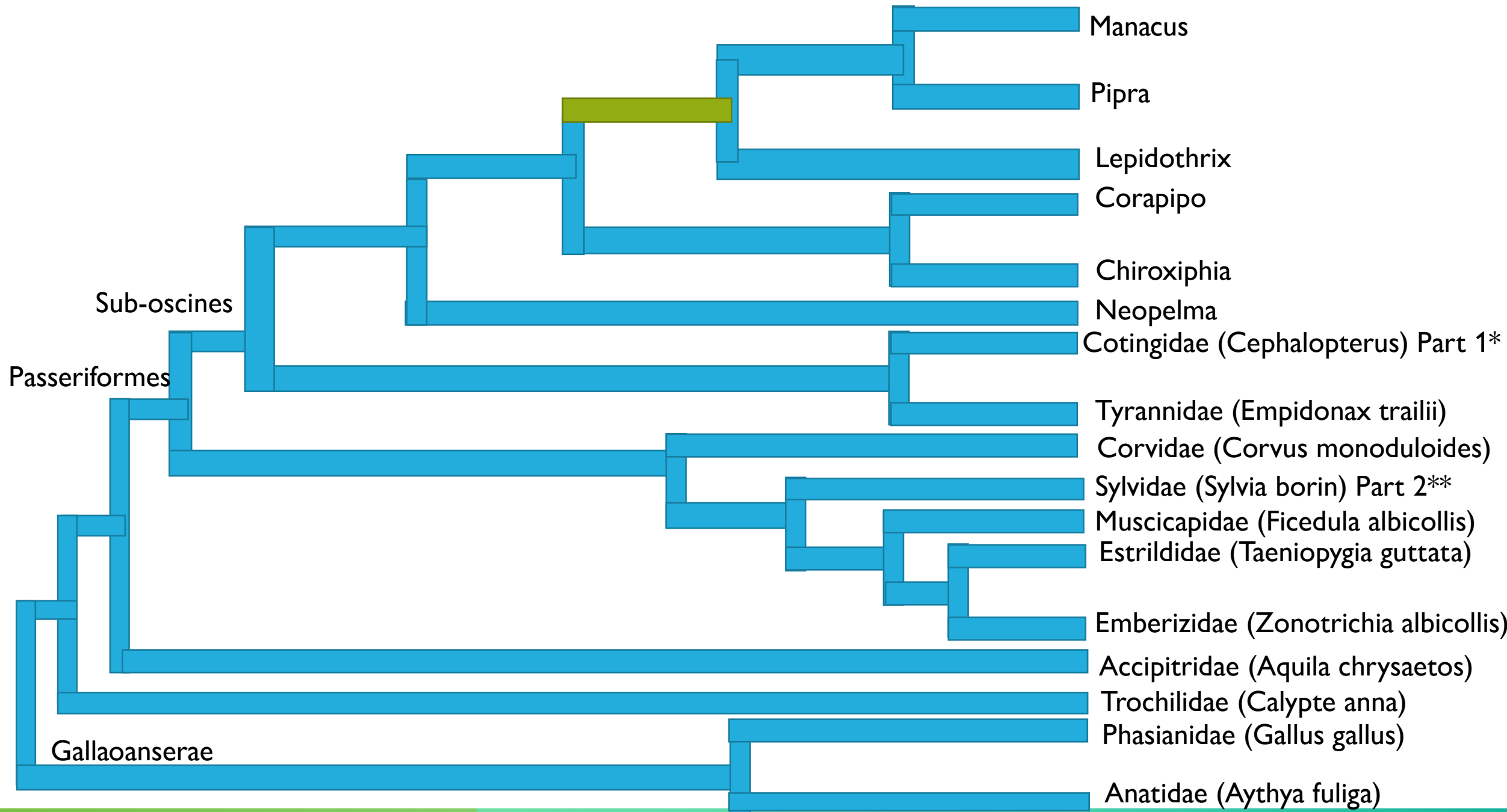
Part
1 & 2

Part 3



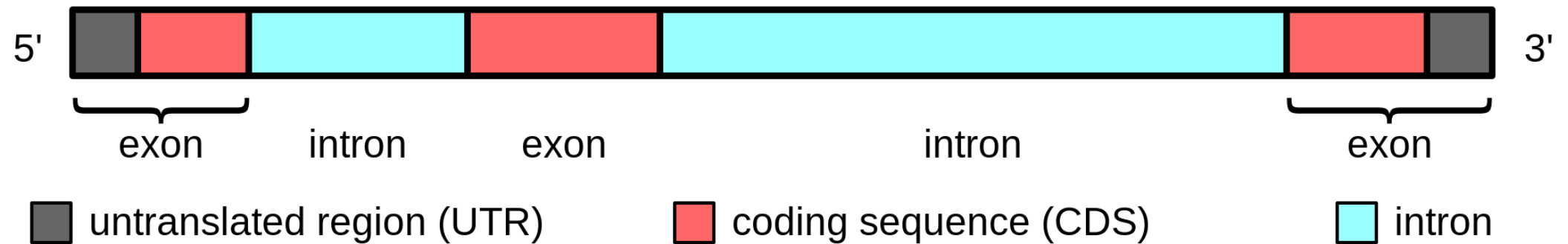
From Jeffares et al (2015) A Beginners Guide to Estimating the Non-synonymous to Synonymous Rate Ratio of all Protein-Coding Genes in a Genome





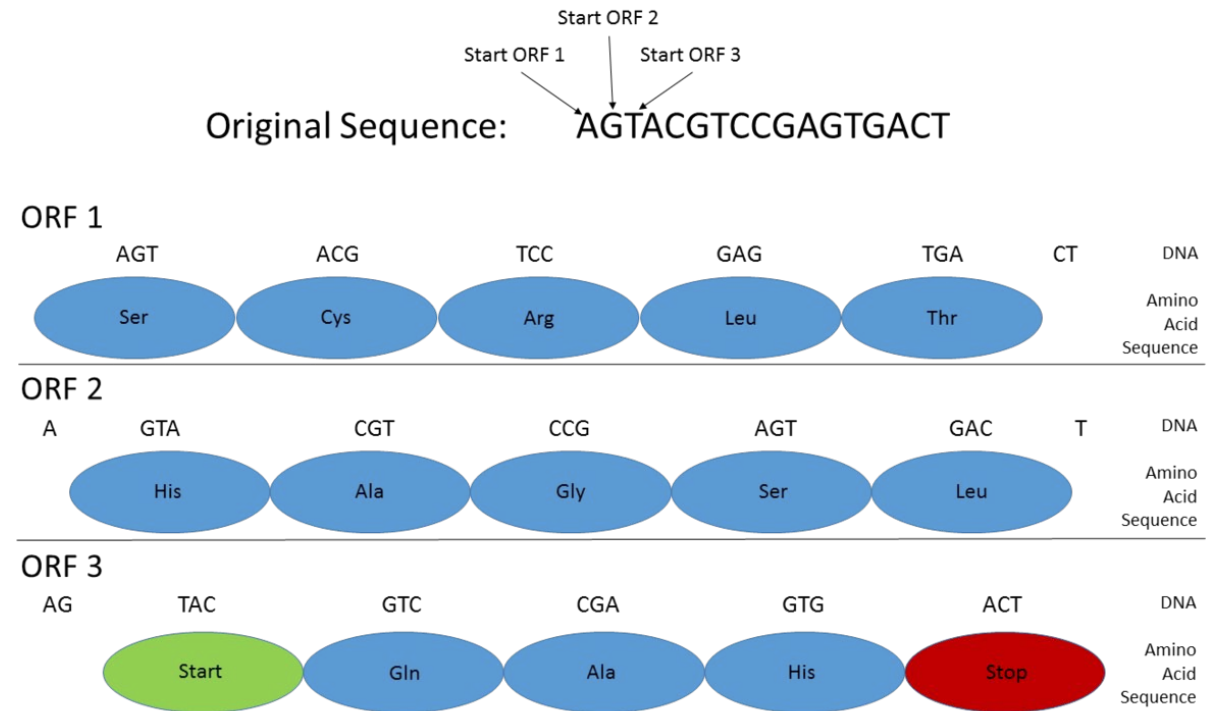
STRUCTURE OF A GENE

- How to use the genome browser in NCBI
- Understanding gene models in NCBI



GENE PREDICTION

- Promoters
- ORFs + length of ORF
- RNAseq
- Sequence comparison with other species



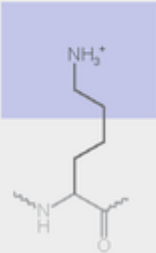
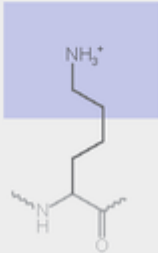
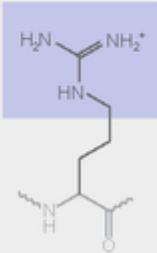
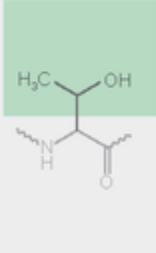
SEQUENCE EXTRACTION

- CDS region only
- FASTA file
- Include header information that is useful for you.

BLAST

- Web service
- Commandline code

TRANSLATION

	No mutation	Point mutations			
		Silent	Nonsense	Missense	
				conservative	non-conservative
DNA level	TTC	TTT	ATC	TCC	TGC
mRNA level	AAG	AAA	UAG	AGG	ACG
protein level	Lys	Lys	STOP	Arg	Thr
					
				basic	polar

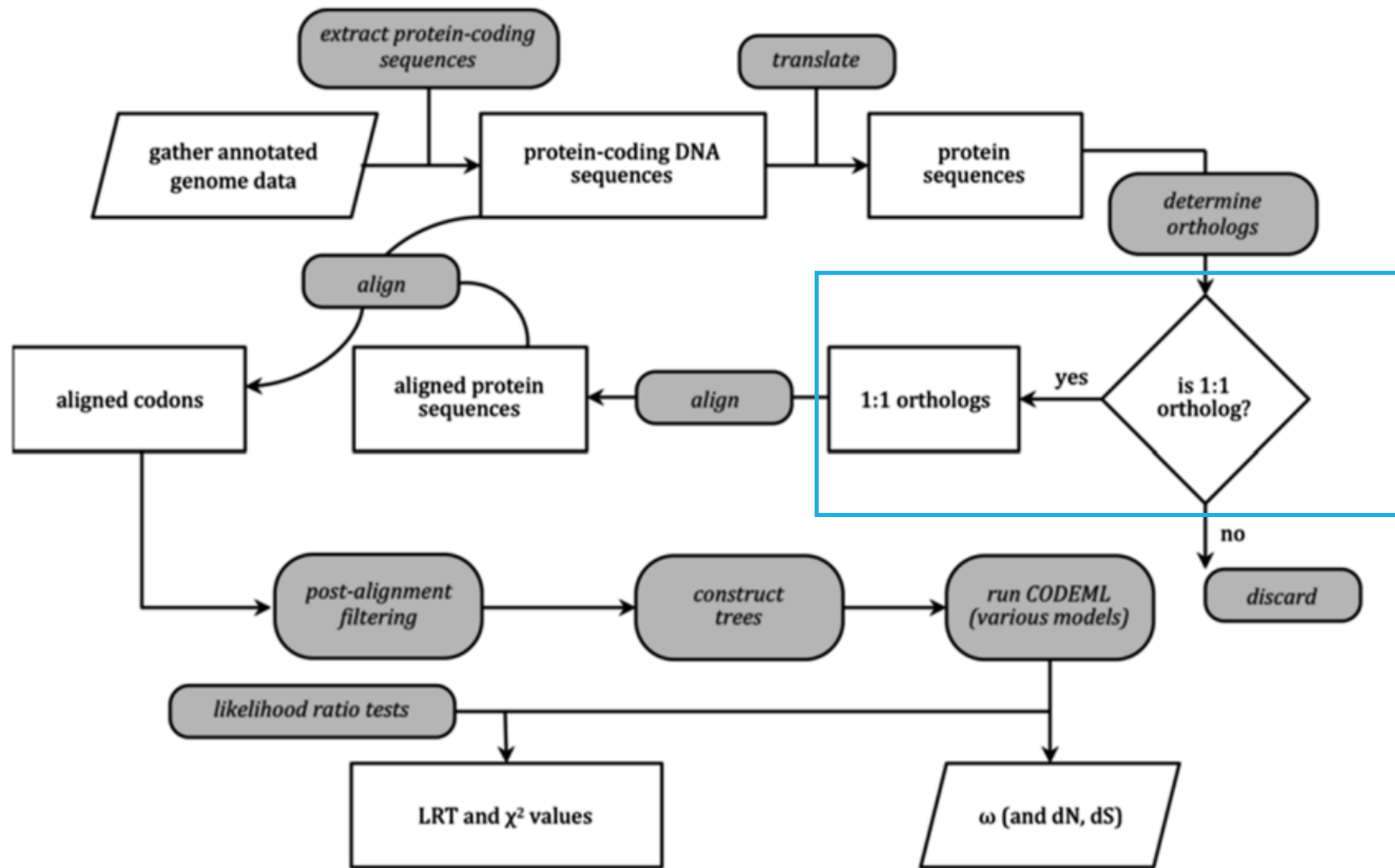
TRANSLATION

- Commandline: use EMBOSS' transeq function
- GUI:
 - MEGA
 - translatorX (<http://translatorx.co.uk/>)

BLAST ON UNANNOTATED GENOME

- Tutorial Part 2
- Vertebrate Genomes Project
- Garden Warbler





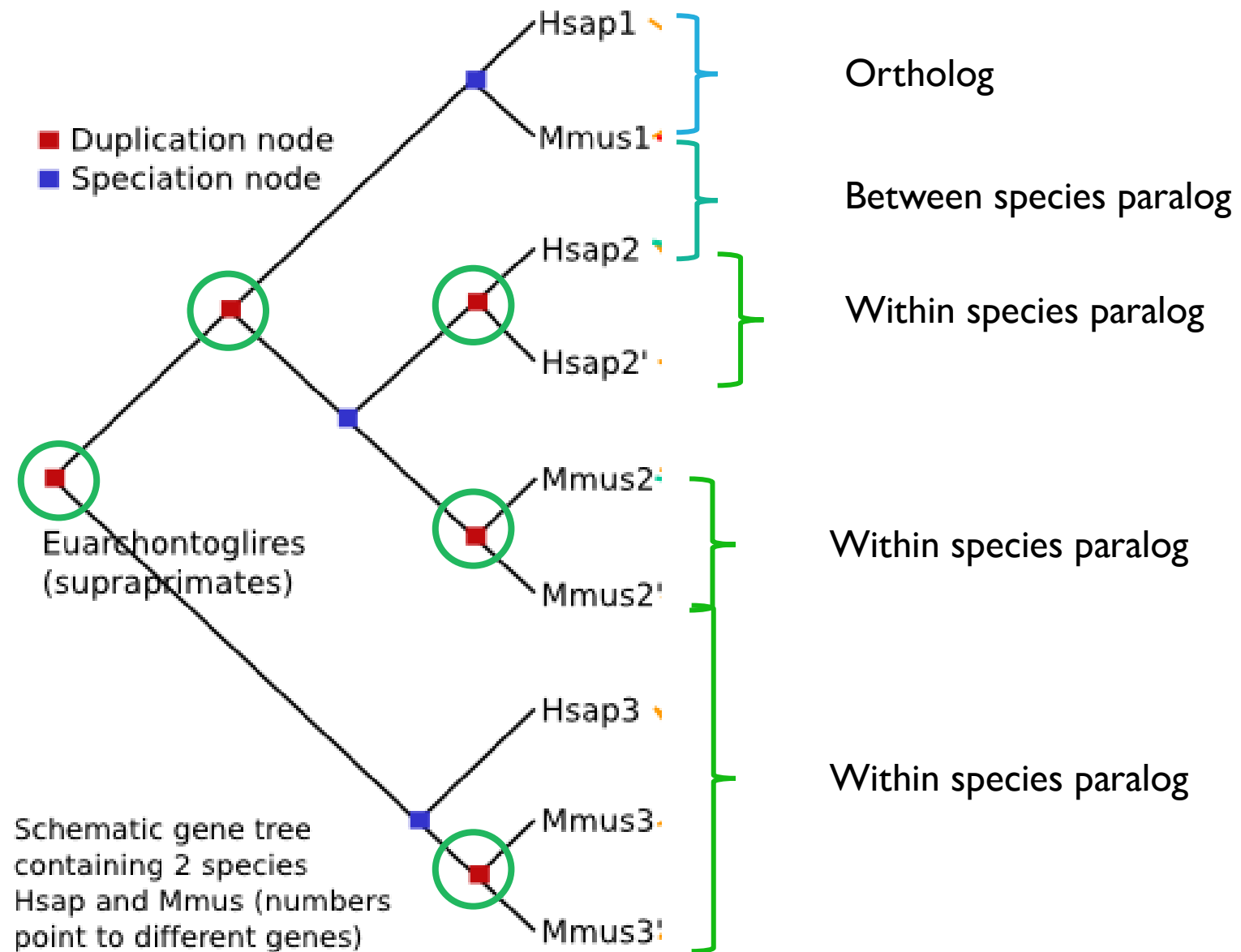
From Jeffares et al (2015) A Beginners Guide to Estimating the Non-synonymous to Synonymous Rate Ratio of all Protein-Coding Genes in a Genome

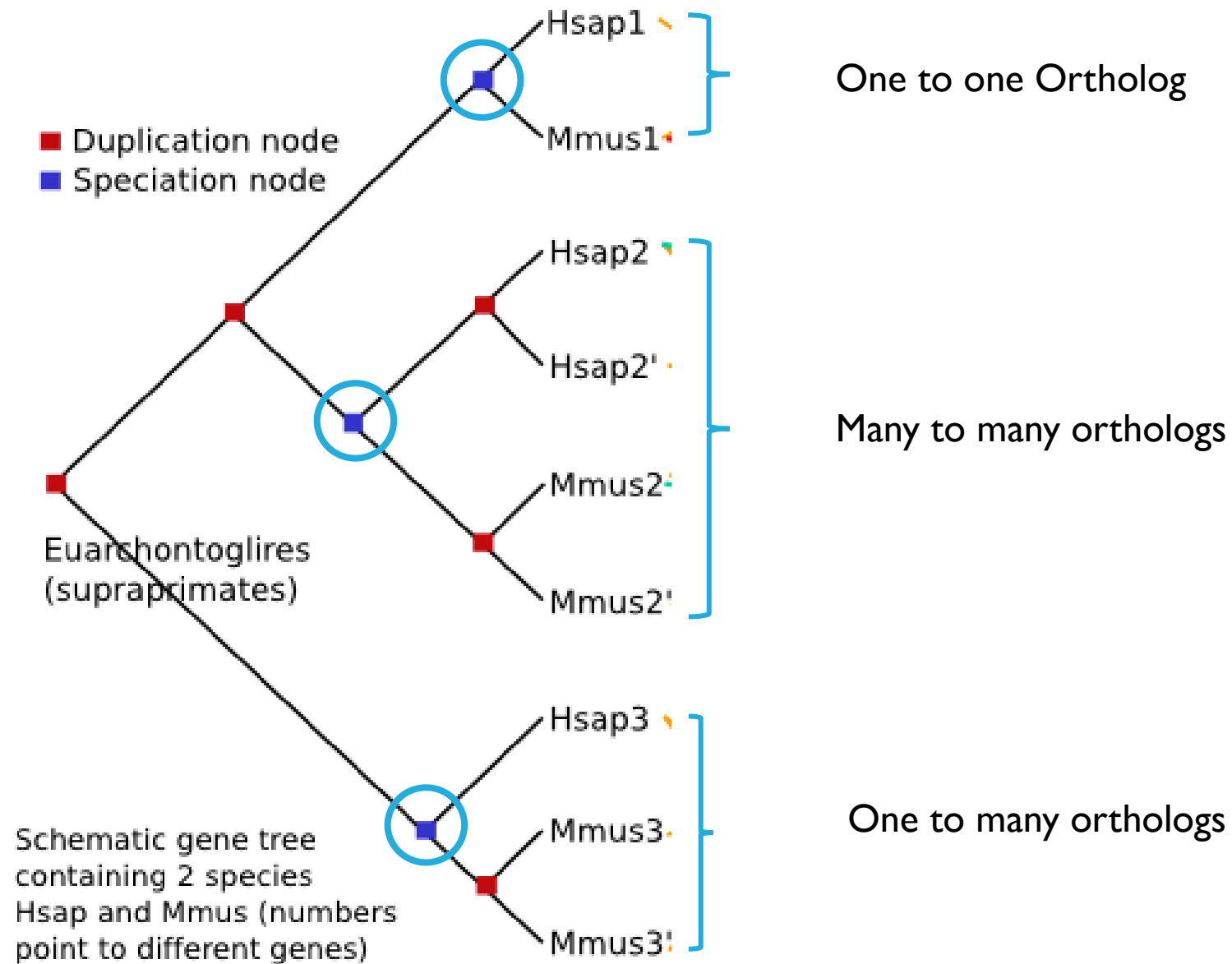
HOMOLOG Y

Homolog: Sequences that are descended from a common sequence ancestor

Paralogue/Paralog: sequences that have evolved by gene duplication

Orthologue/Ortholog: sequences derived from a common ancestor via speciation





FIND 1:1 ORTHOLOGOUS SEQUENCES

- NCBI orthology (demo)
- proteinOrtho (https://gitlab.com/paulklemm_PHD/proteinortho)
- Review of different tools here
 - Nichio et al (2017) New Tools in Orthology Analysis: A Brief Review of Promising Perspectives. *Frontiers in Genetics* 8: 165

TOMORROW

- Multiple Alignment & Quality Control

FURTHER READING

- Nichio et al (2017) New Tools in Orthology Analysis: A Brief Review of Promising Perspectives. *Frontiers in Genetics* 8: 165
- Vallender (2009) Bioinformatic approaches to identifying orthologs and assessing evolutionary relationships. *Methods* 49: 50-55