

Conservation Genetics in the Tropics

Assembly of mitochondrial genomes

Miguel Camacho Sánchez

October 19th, 2021

miguelcamachosanchez@gmail.com

miguelcamachosanchez.weebly.com



Conservation Genetics in the Tropics

Assembly of mitochondrial genomes

Practical sessions

Preprocessing

- Quality filtering
- Adapter / primer removal

Assembly

- assemble and alignment sequences

Evolutionary
PopGen
analysis

- structure
- diversity
- phylogeny

Content

assembly

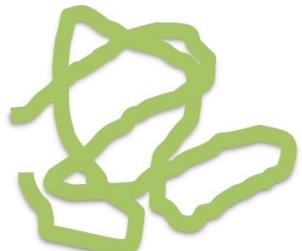
- *de novo*
- reference based

annotations

alignment

Assembly

Aligning and merging shorter fragments from a longer DNA sequence in order to reconstruct the original sequence.



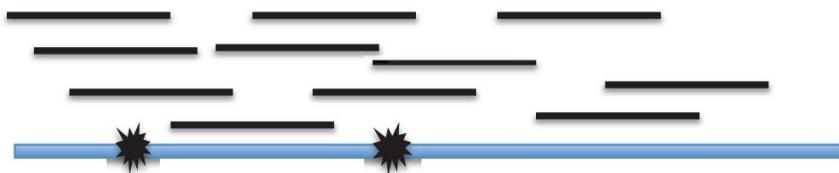
Genomic
DNA

Next-generation
DNA sequencing

... CATTCACTAG ...
... GGTAGTTAG ...
... TATAATTAG ...
... AGCCATTAG ...
... GGTAAACTAG ...
... CGTACCTAG ...

millions-billions of *reads*
~30-1000 nucleotides

Reference based



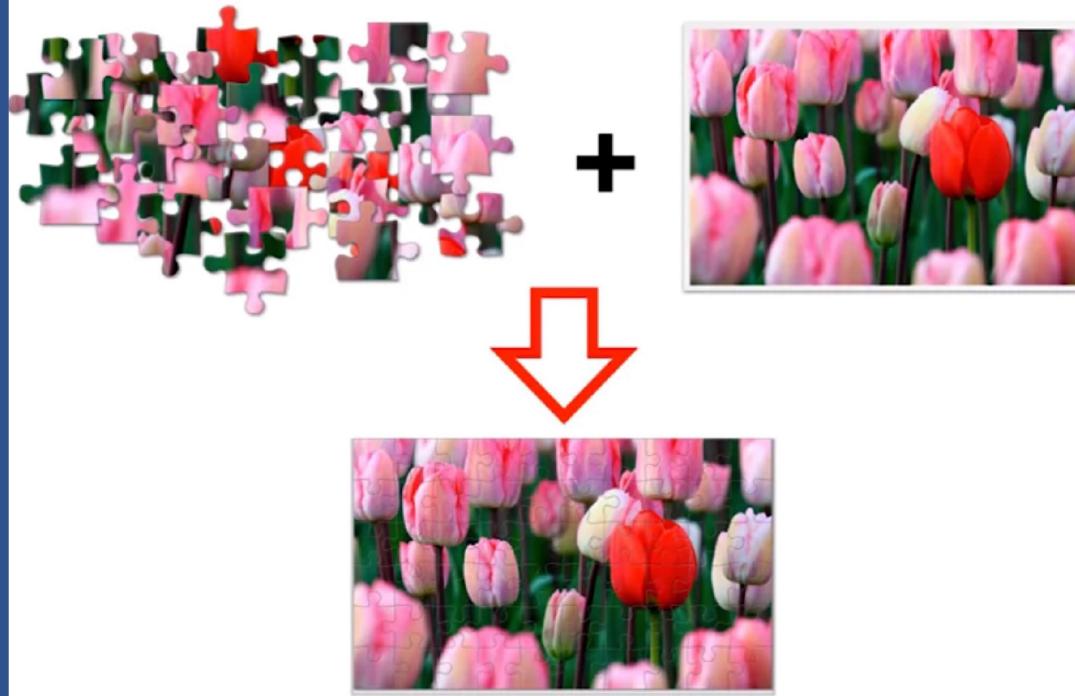
Align reads to *reference genome* and identify variants

***De novo* assembly**



Construct genome sequence from overlaps between reads

Reference based



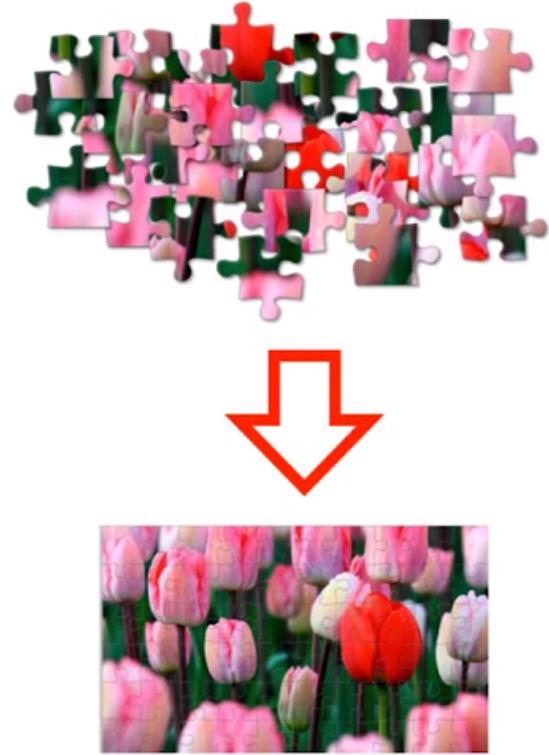
Linear relation between data and computation time

Reference with **high homology** is required

Robust to structural variation

Simple

De novo



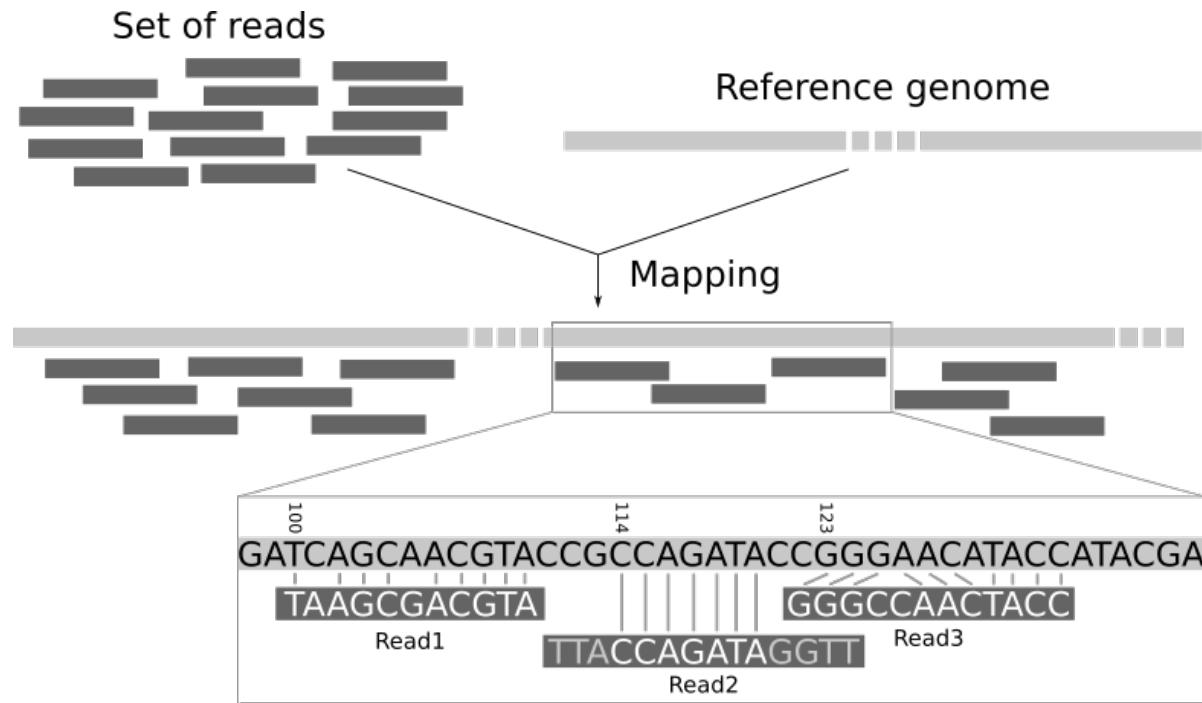
Exponential relation between data and computation time

No reference required

Sensitive to structural variation

More **complex** steps

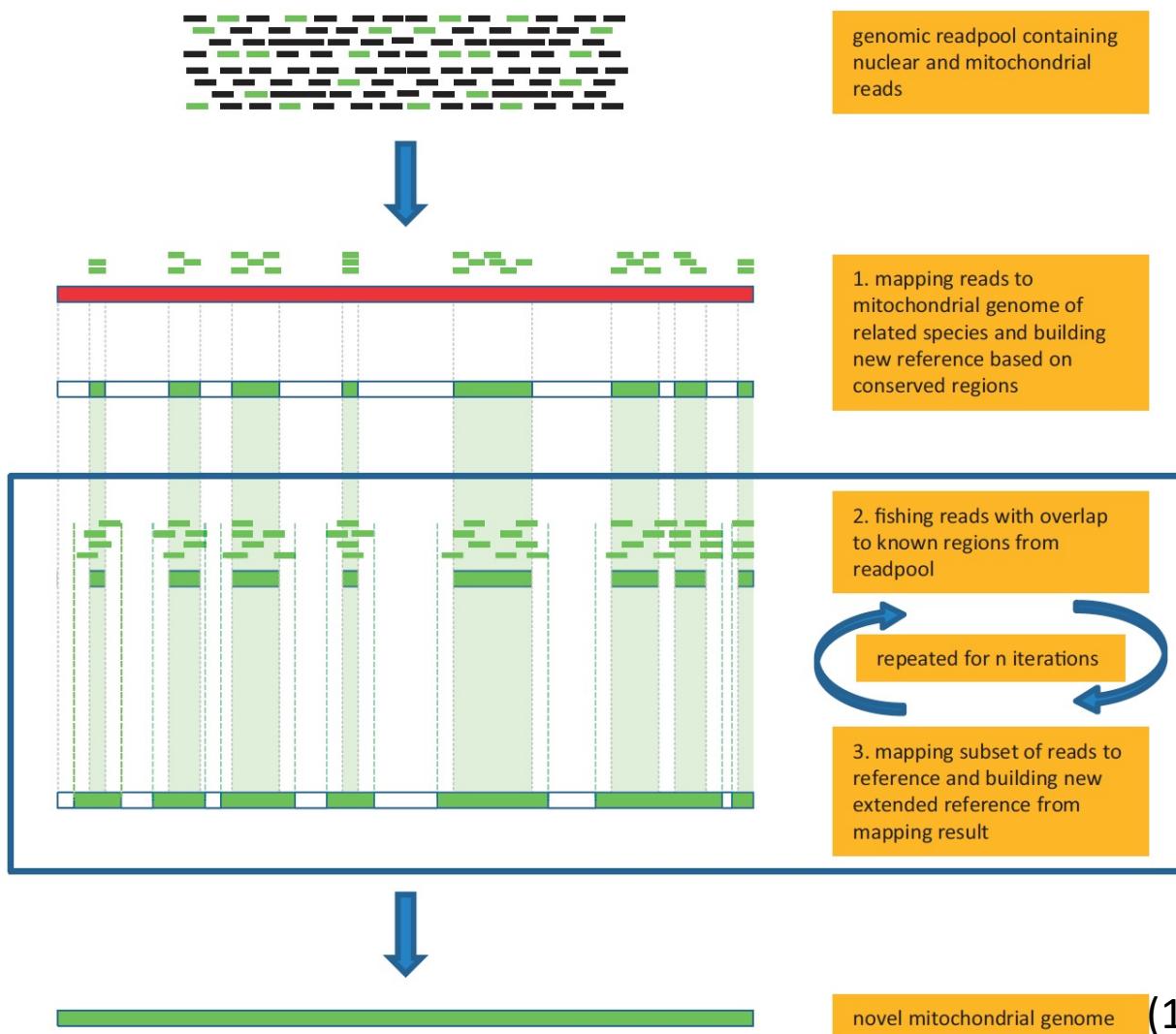
Reference based



Algorithm: [BWA mem](#)

[10.1093/bioinformatics/btp324](https://doi.org/10.1093/bioinformatics/btp324)

baiting and iterative mapping approach



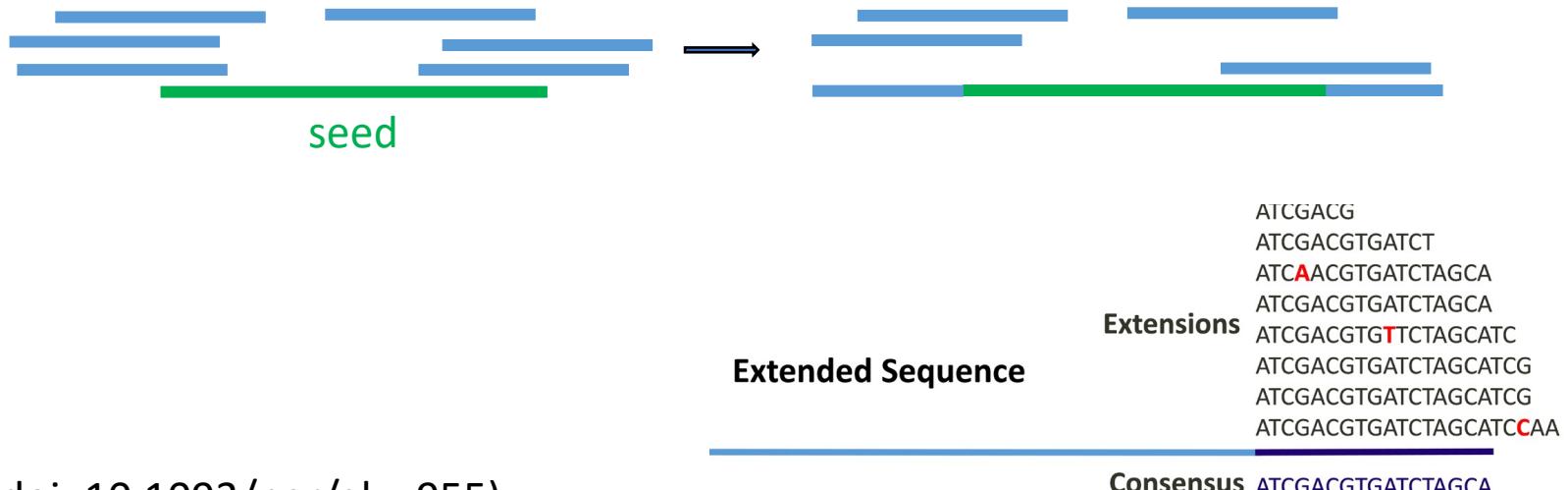


- Iterative mapping works very well for species within the same genus.
- Paid version.

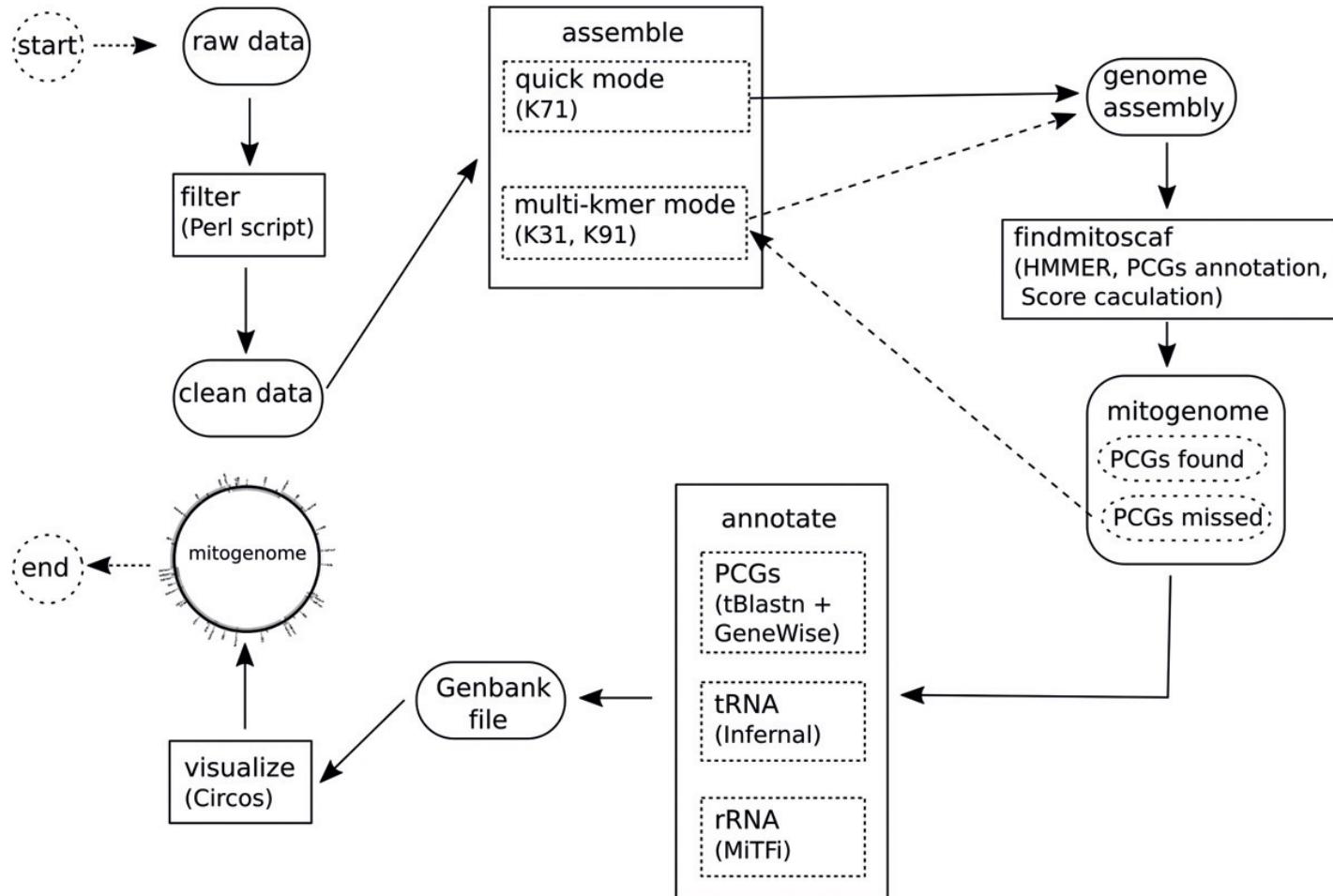
<https://www.geneious.com/>

NOVOplicity: seed-extend based assembler

- Organelles: circular genomes.
- Compatible with paired-end reads.
- Recommended > 30x for retrieving a single contig.
- Starts with a **seed**: a conserved sequence. It can be from a distant species.
- The seed is iteratively extended bidirectionally until:
 - > 1 consensus extension (the is split in 2)
 - a circular genome is formed



mitoZ



(10.1093/nar/gkz173)

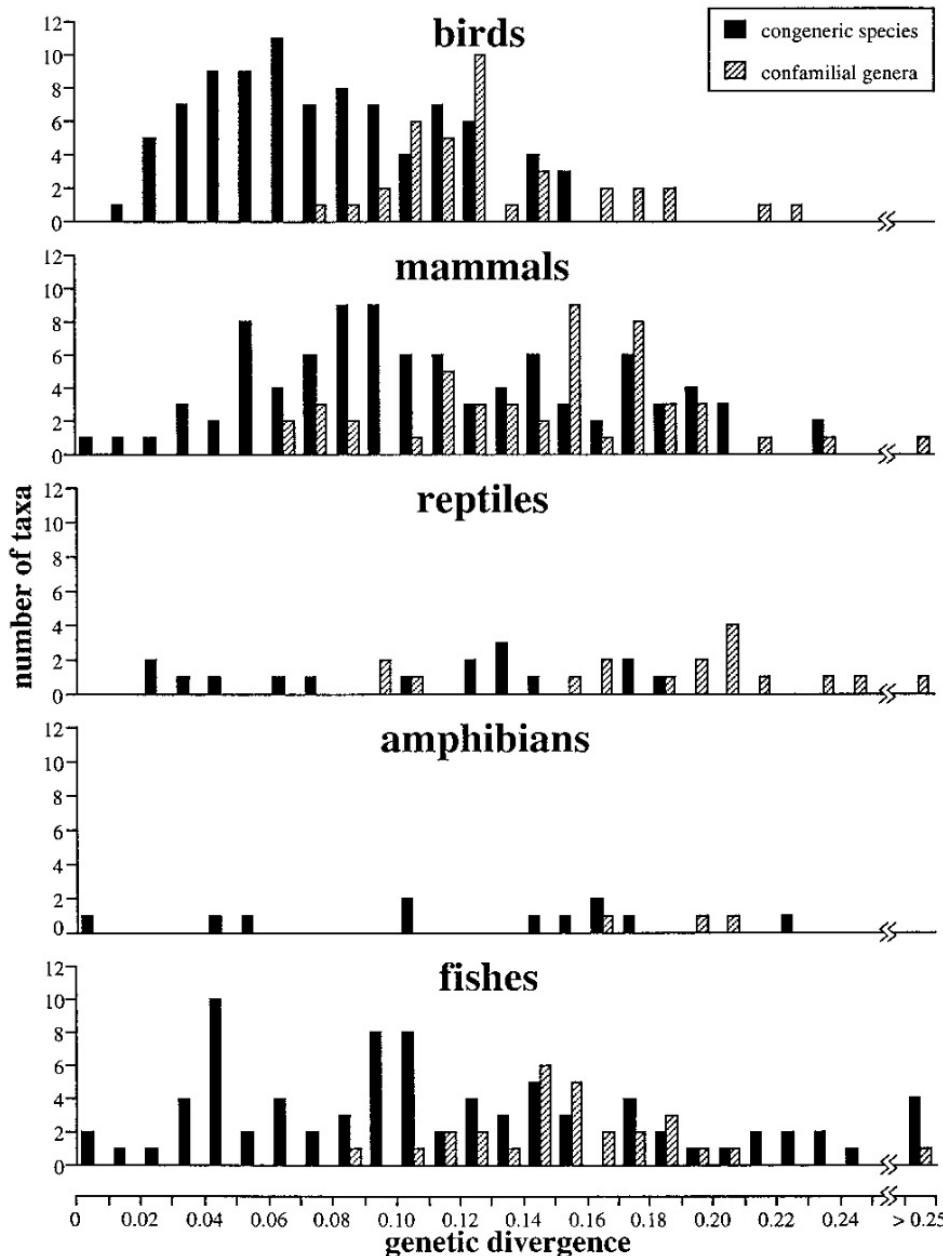
Should I use *de novo* or
referenced-based assembly?

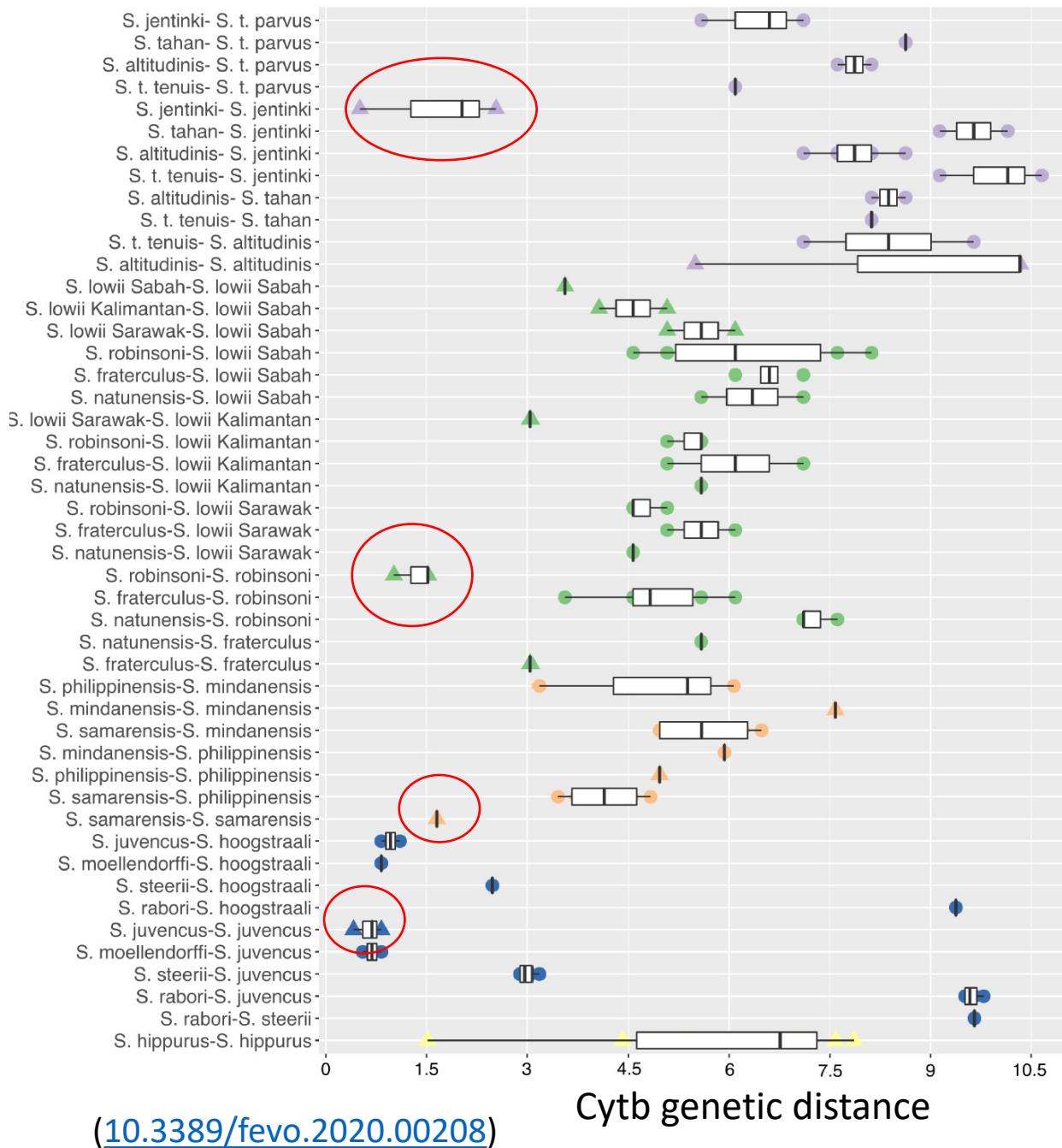
Better reference-based if you
have a good reference.

What is a good reference?

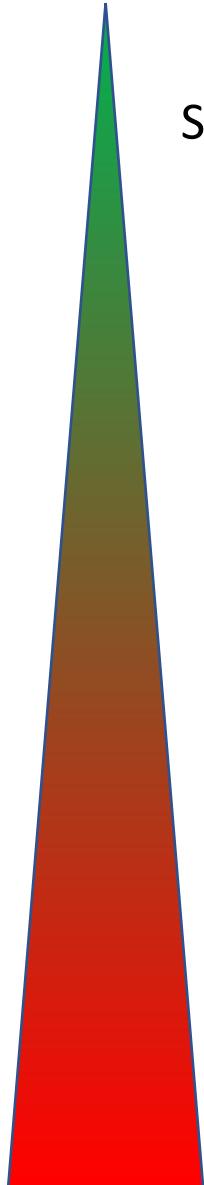
Preferably, same species
But same genus is also fine
(some problems in the control
region).

Cytochrome b variation





cytb divergence



Species

0-8 %

Mapping to a reference from the same species or genus.
(software: BWA, Geneious mapper)

Genus

5-18 %

Iterative mapping against a reference or use
conserved region as seed
(software: MITObim, NOVOPlasty, Geneious
iterative mapper)

Family

10-20 %

Use conserved region as seed or de novo
assembly
(software: MitoZ, NOVOPlasty)

Let's practice