



iNaturalist

Received: 13 May 2020

Revised: 9 July 2020

Accepted: 10 July 2020

DOI: 10.1002/ece3.6727

ORIGINAL RESEARCH

Ecology and Evolution
Open Access

WILEY

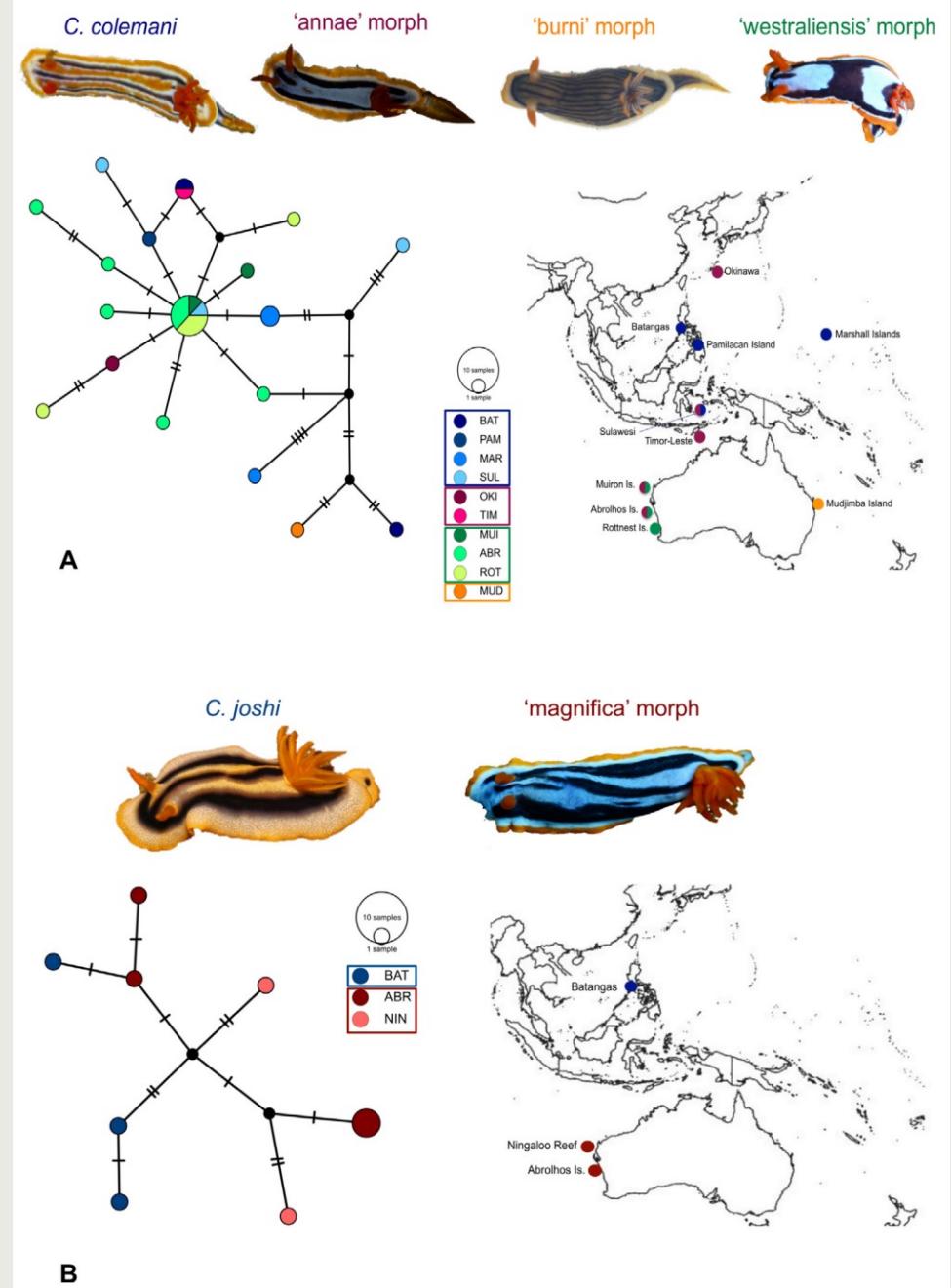
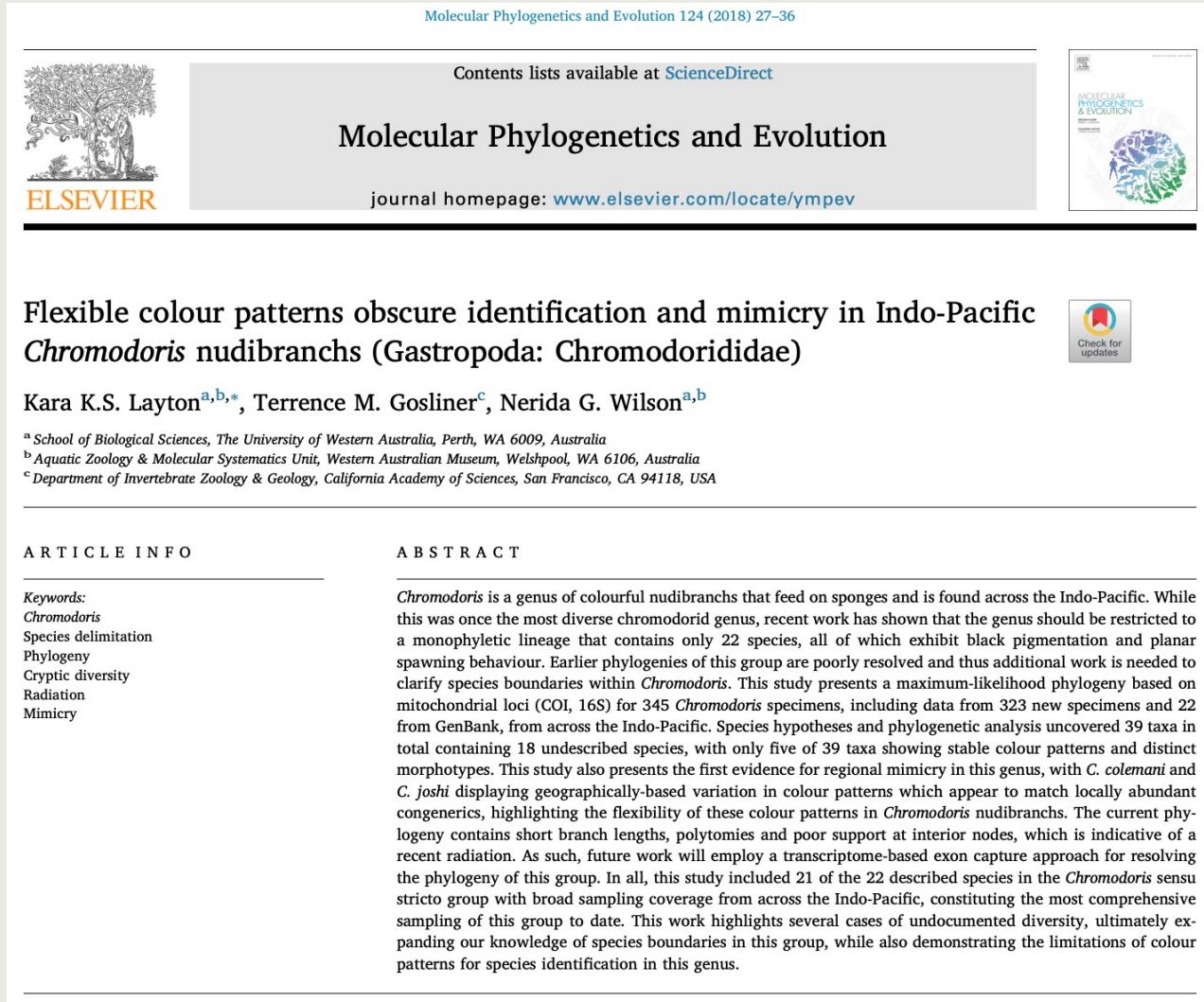
Mimicry and mitonuclear discordance in nudibranchs: New insights from exon capture phylogenomics

Kara K. S. Layton^{1,2,3} | Jose I. Carvajal² | Nerida G. Wilson^{1,2}

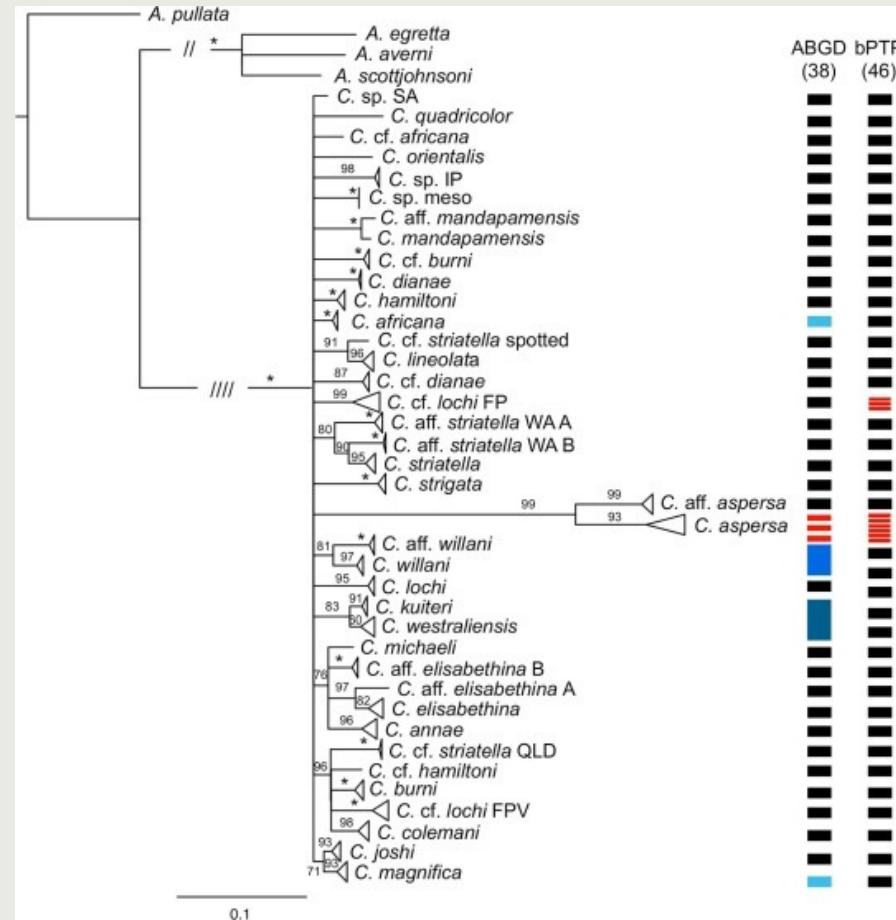
CG2 October 10th 2024

Lina Marie Raubold

Introduction



Introduction

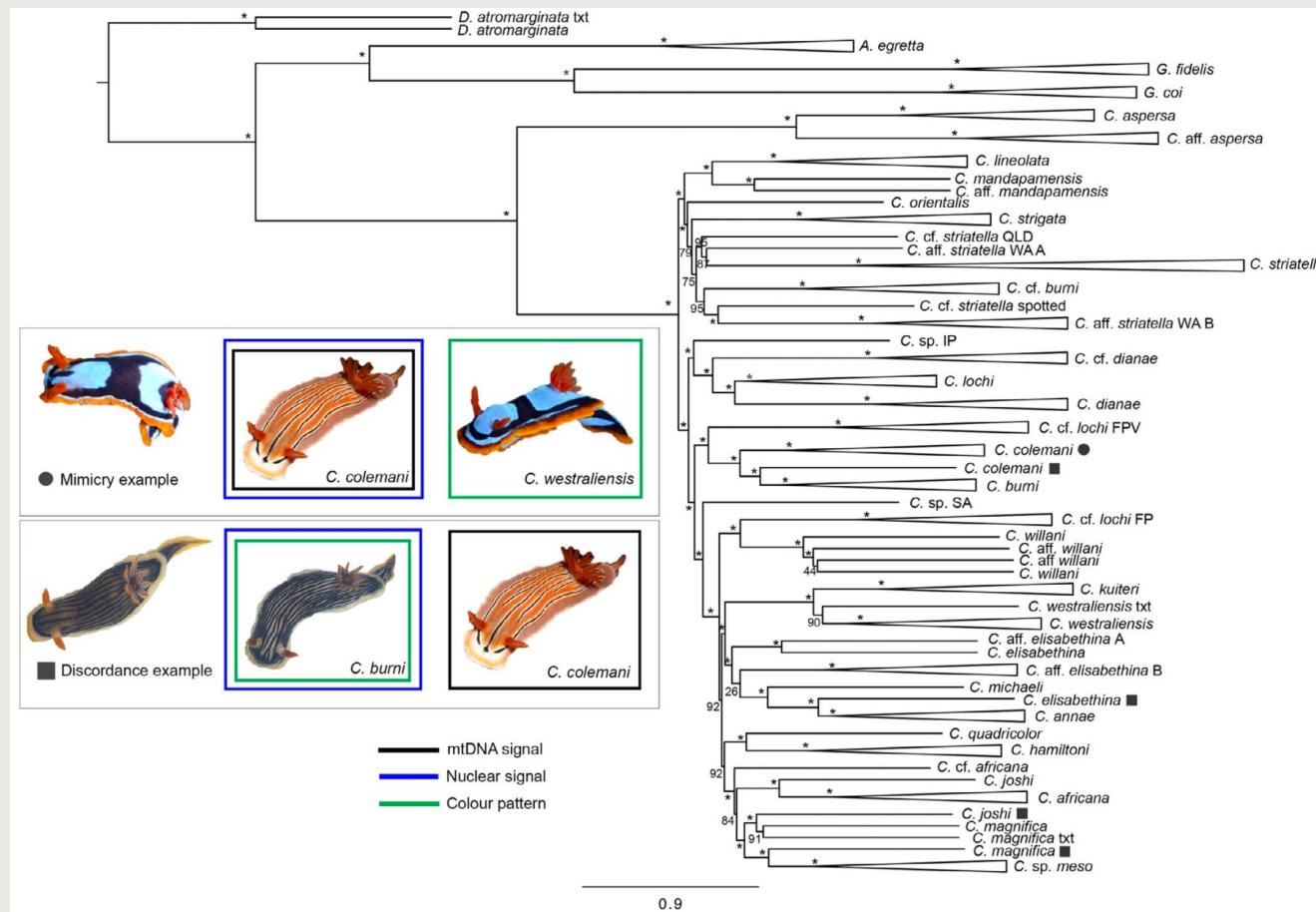


- Recent radiation
- Geographical mimicry
- COI and color pattern not very helpful

Main findings

Mimicry and mitonuclear discordance in nudibranchs: New insights from exon capture phylogenomics

Kara K. S. Layton^{1,2,3}  | Jose I. Carvajal² | Nerida G. Wilson^{1,2} 



- Transcriptome-based exon capture approach
- Identified 2925 exons & 1630 genes
- Mitonuclear discordance in previously identified mimics
- Likely deriving from introgression/mitochondrial capture

Methods - Overview

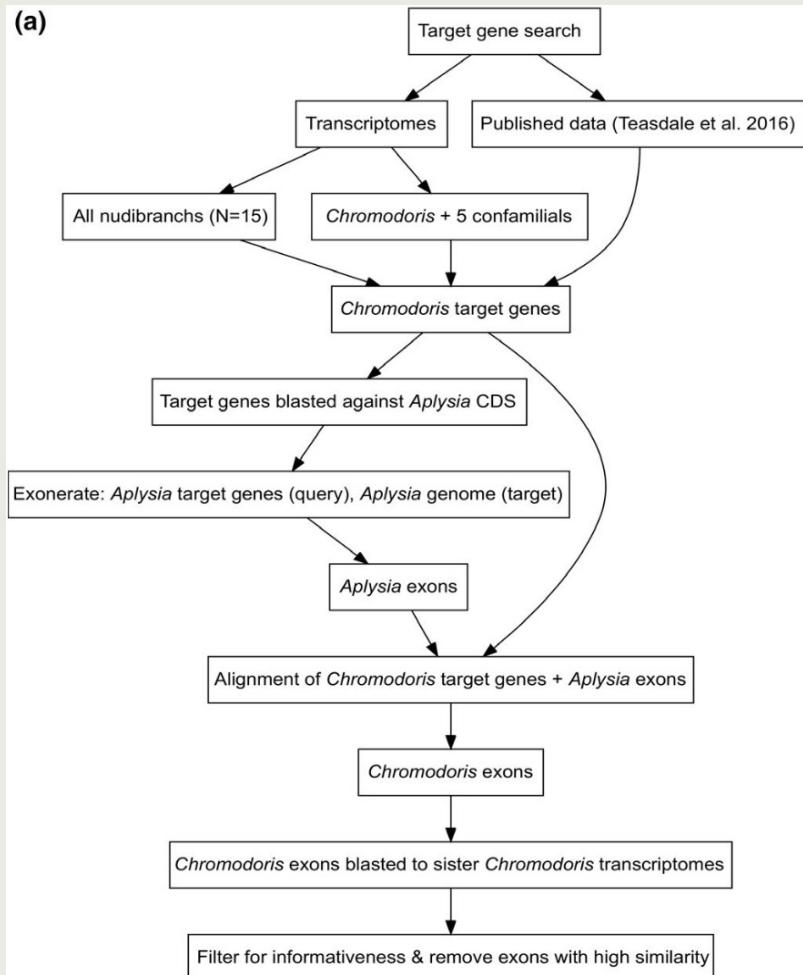


1) Bait design



2) Exon capture phylogenomics





- Transcriptome sequencing
- Transcriptomes >20 M reads: random subset of 20M paired end reads
- *de novo* assembly
Agalma (transcriptome pipeline)
- Removal of transcripts with <0.003 fragments/Kb Million
- 15 transcriptomes for ML phylogeny
IQtree v1.6.8



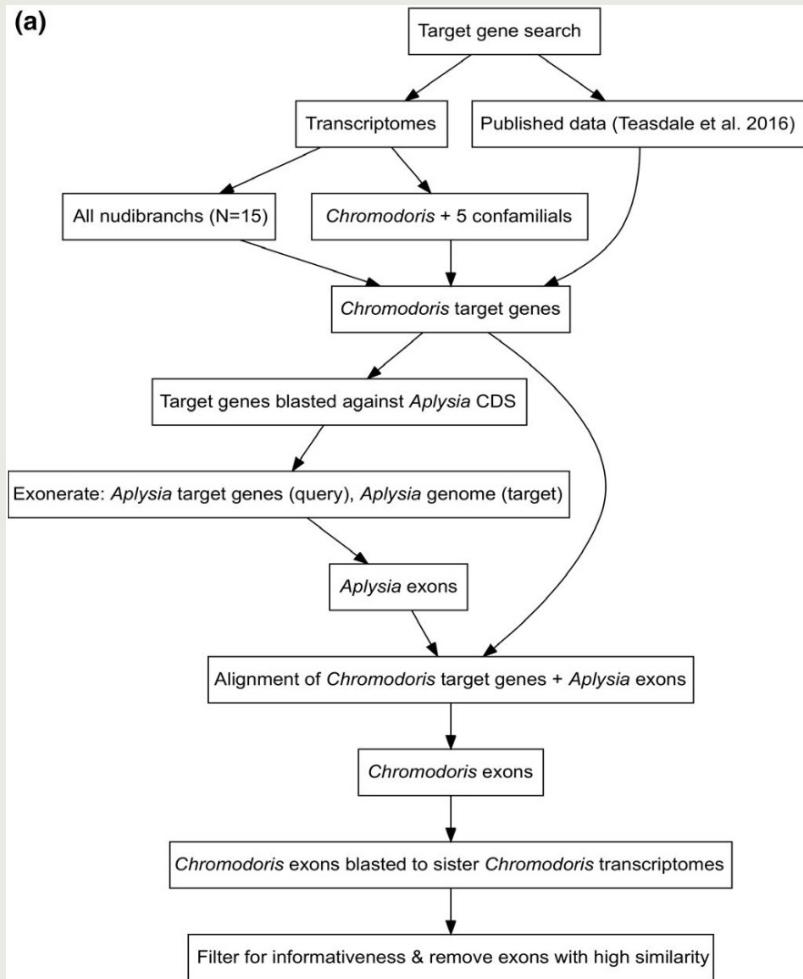
RNA extraction

Transcriptome sequencing & *de novo* assembly

ID single copy homologous genes

ID exon/intron boundaries

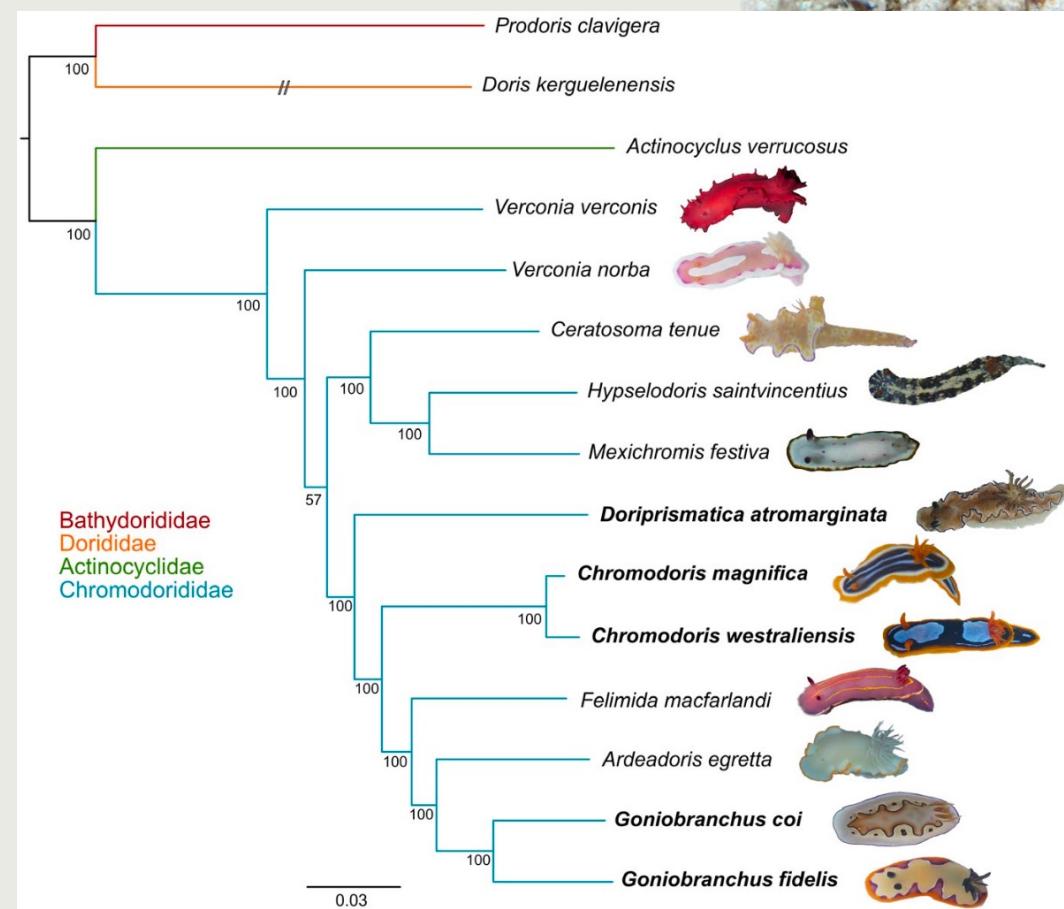
(a)

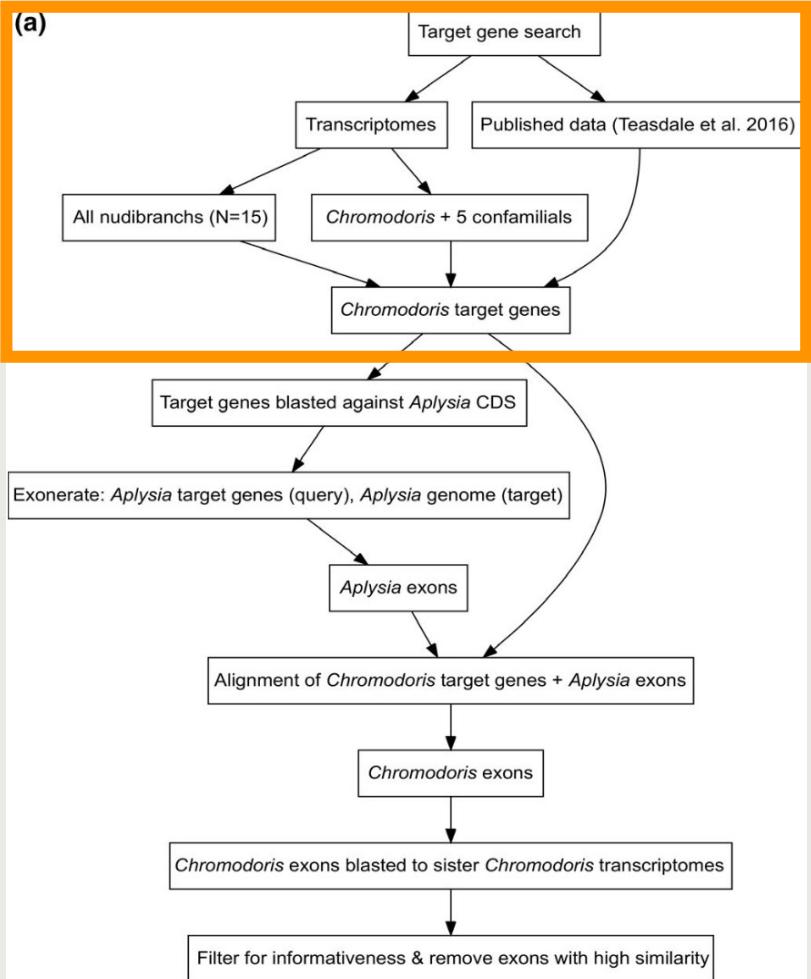


15 transcriptomes
for ML phylogeny
IQtree v1.6.8

Single partition
GTR20+F
substitution model

Framework for
interpreting gene
and taxon selection
downstream

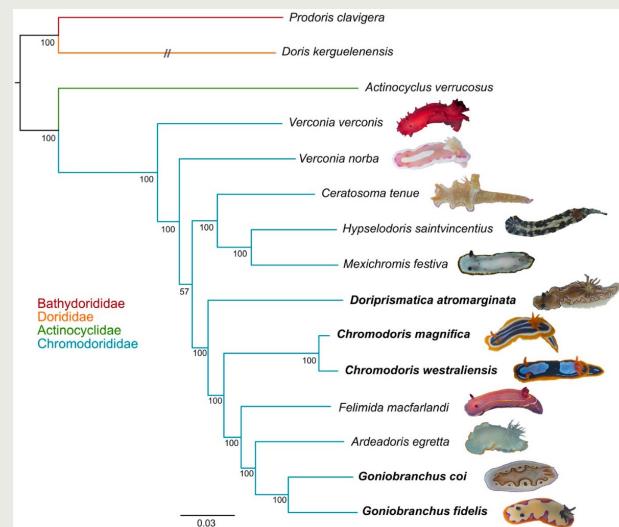


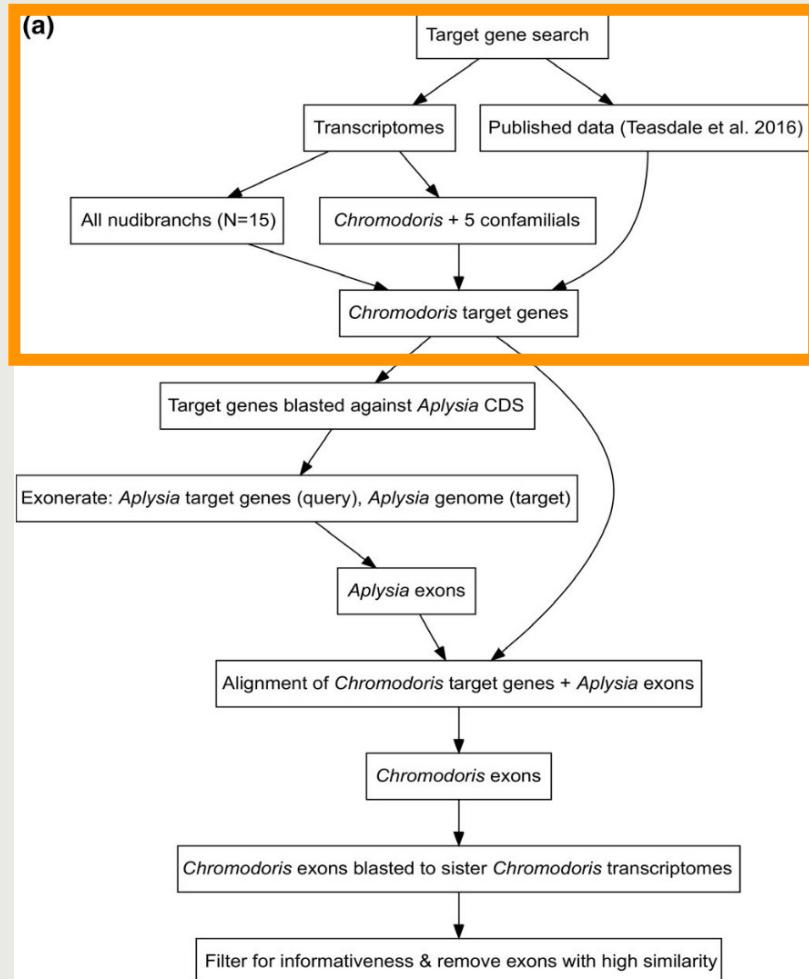


- Comparisons across 3 taxonomic levels:
- Fast 5, all nudibranchs ($n=15$) and Eupulmonata (Teasdale *et al.*, 2016)
- ID of single copy homologous genes
Agalma phylogeny pipeline
Fast 5: nucleotide search
All nudibranchs: amino acid search
- DNA extraction from output
Agalma matrix2genes function



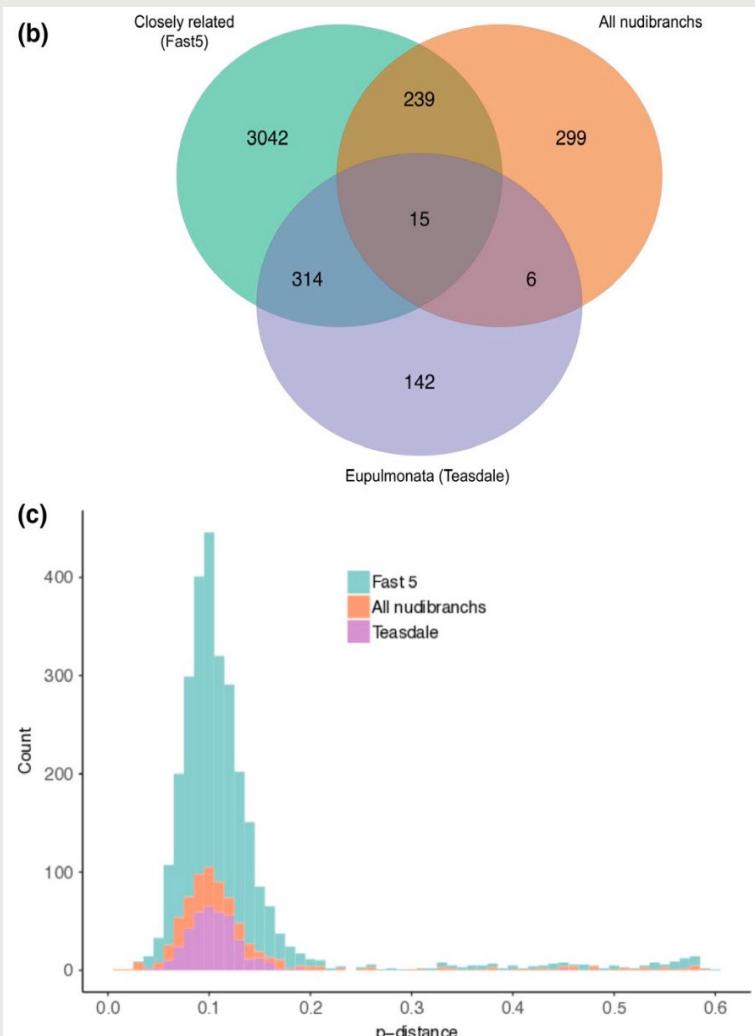
Wiki





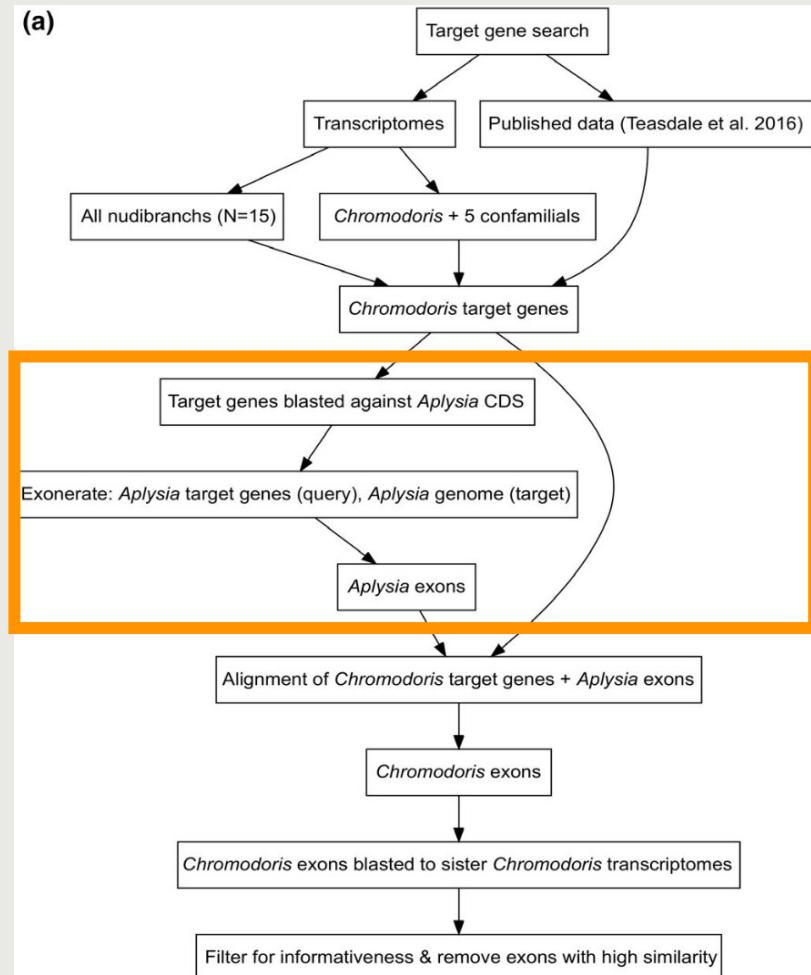
- Pairwise distance calculation between 2 most complete transcriptomes **Geneious v9.0**
ID distribution of evolutionary rates between results from both searches (Fast 5 vs. All nudibranchs)
- Addition of previously published target gene set ($n=500$) from study on Eupulmonata (Teasdale *et al.*, 2016)





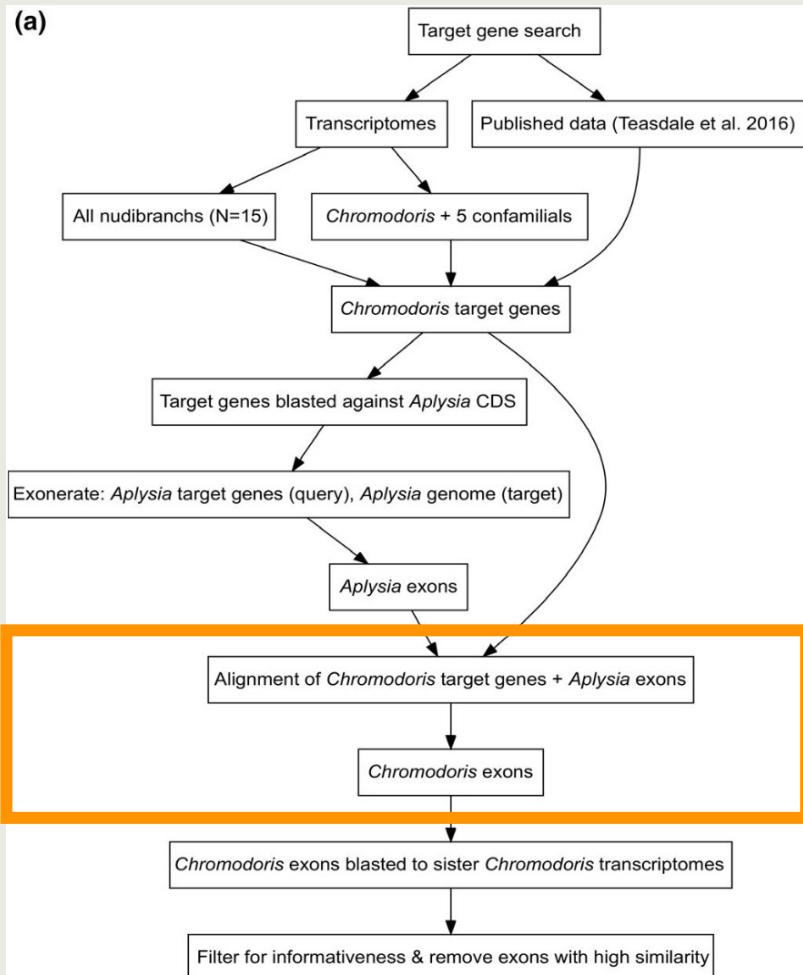
- Pairwise distance calculation between 2 most complete transcriptomes Geneious v9.0
- ID distribution of evolutionary rates between results from both searches (Fast 5 vs. All nudibranchs)
- Addition of previously published target gene set ($n=500$) from study on Eupulmonata (Teasdale *et al.*, 2016)



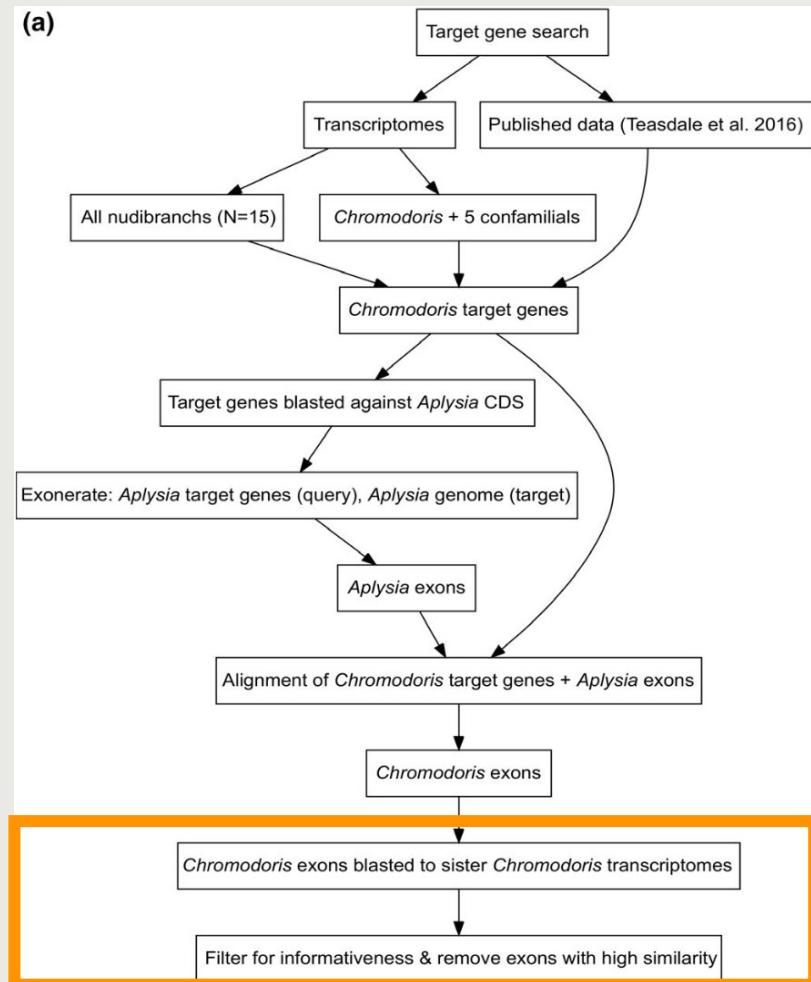


- Concat all 3 sets of target genes (Fast 5, All nudibranchia and Eupulmonata)
- ID corresponding genes in *Aplysia californica* (best assembled genome in Heterobranchia) **BLAST**
- ID introns **Exonerate v2.2.0**
map *Aplysia* genes against *Aplysia* genome
- ID *Aplysia* exons:
Parse output into individual exons
only save exons >200bp **custom Python script**





- ID *Chromodoris* exons **Exonrate v2.2.0** map *Aplysia* exons against *Chromodoris magnifica* and *Chromodoris westraliensis* transcriptomes
- ID best *Chromodoris* exons:
Parse output into individual exons
only save exons >115bp, >65% similar to *Aplysia* query exons, <than original query exons in length
select single best exons from both species based on highest scores **custom Python script**



- Select variable (=informative) genes reciprocal **BLAST** of best *Chromodoris* exons against both transcriptomes
- Retention of hits that were 92-99% similar + only exons that did not blast against any other exon in the target set
- Send off *Chromodoris* targets to Arbor Biosciences for bait design and synthesis



- Extracted DNA -> Arbor Biosciences for Library prep and target capture sequencing (Illumina HiSeq 2500 platform with 100 bp paired end reads)
- Read cleanup **Trimmomatic v0.36**
removal of adapter sequences + reads with a score below 15 in a 4bp sliding window + removal of reads <26bp
- Assemble cleaned reads into contigs of target regions **HybPiper v1.3.1**
reads mapped against reference file of concatenated bait sequences (Burrows-Wheeler-Algorithm)
- Assemble mapped reads de novo by gene into contigs **SPAdes v3.13.0**
- Trim contigs to just include exons **Exonrate v2.2.0**



- Trim contigs to just include exons **Exonrate v2.2.0**
 - Produce one fasta file for each gene + summary statistics **HybPiper**
 - Remove genes that show no/poor enrichment (contigs <50% of reference gene)
 - Align gene files **MAFFT v7**
 - Trim aligned genes **trimAl v1.2**
 - ID # of parsimony informative sites in final alignment **Geneious v9.0**
-
- 1630 gene alignments available for phylogenetic analyses

DNA extraction
& Target sequencing

Bioinformatic processing

Phylogenetic analysis

SNP calling

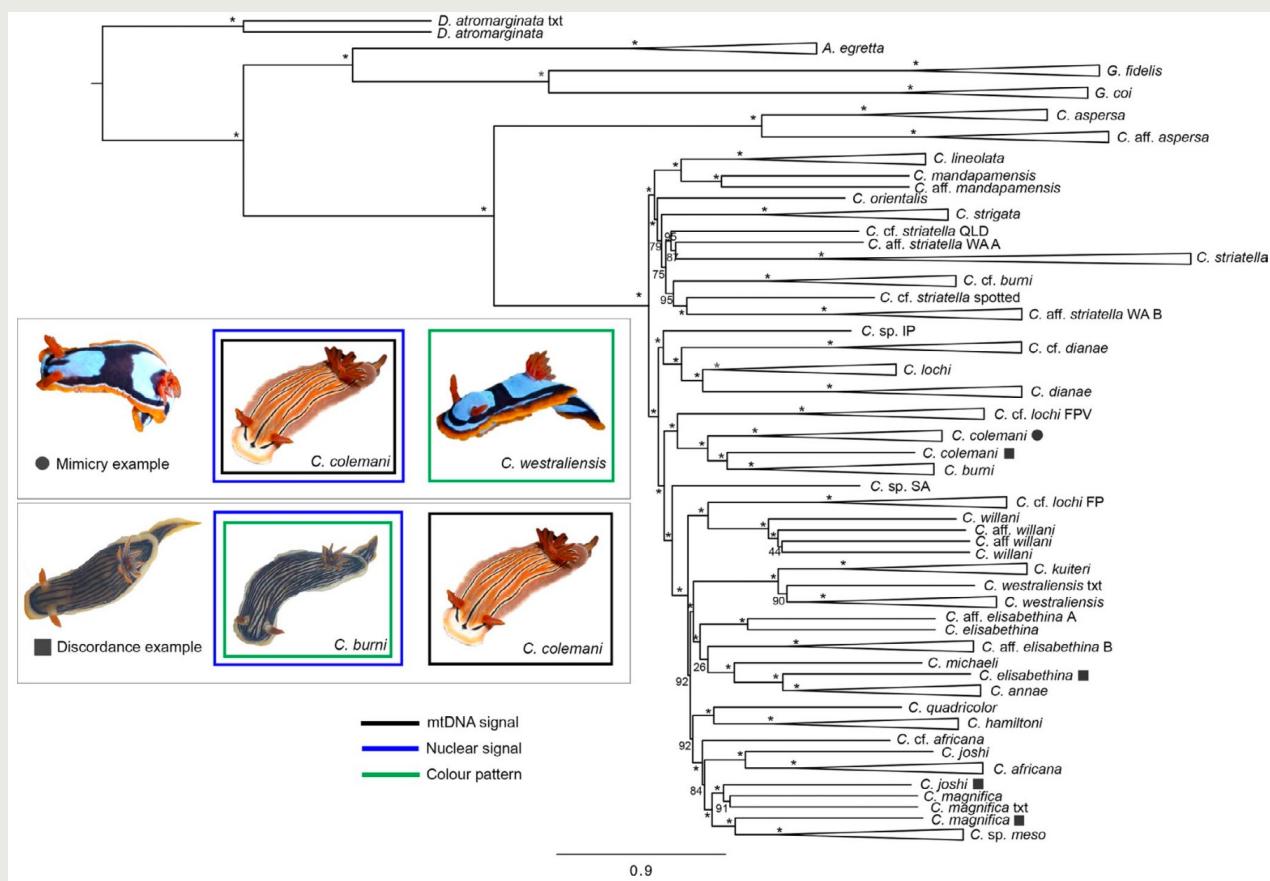


Data:

- 2 datasets (All nudibranchs + Eupulmonata & only Eupulmonata)
- Mitochondrial dataset (COI + 16S; Layton *et al.*, 2018) for tree landscape comparisons
Treespace package in R

Analysis:

- Concat gene alignments for model testing and ML analysis IQtree v1.6.8
- Individual ML trees for each gene for summary coalescence analysis ASTRAL II

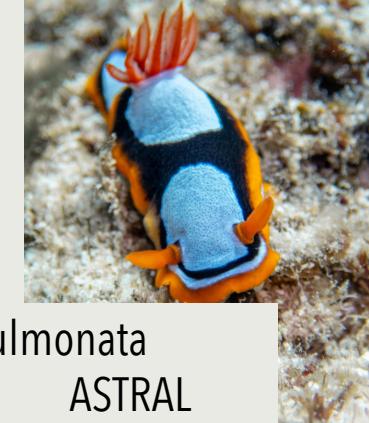


DNA extraction
& Target sequencing

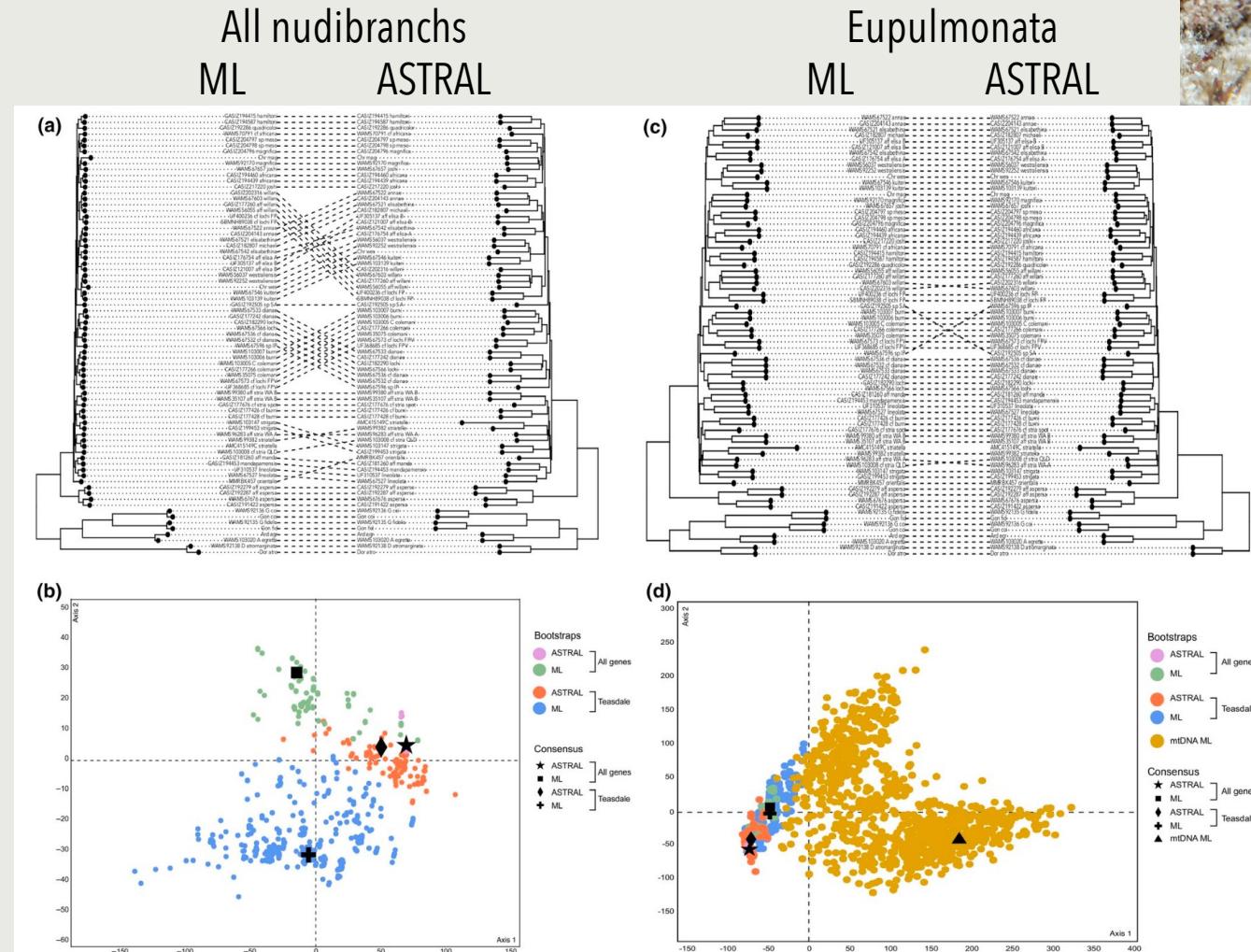
Bioinformatic processing

Phylogenetic analysis

SNP calling



- test for congruence in multiple sets of trees **cophylo function in phytools (R)**
- Exploration of topological variabilities for all trees:
all genes vs. just Eupulmonata
tree building: ML vs. ASTRAL
Treespace package in R (Kendall Colijn metric)





- Remove PCR duplicates from binary alignment maps (BAM) from HypPiper runs + sort alignments **Genome Analysis Tool Kit (GATK) v4**
- SNP variant calling **GATK - haplotype caller in gVCF mode**
- Variant selection after hard filtering step **GATK - selectvariants tool**
- Prune for linkage disequilibrium **PLINK 2.0**
remove SNP that were correlated to any other SNP in a 50 bp sliding window
- ID % of heterozygous fixed alleles **custom Python script**
search for alleles that were fixed differently in putative parental lineages for individuals where hybridization was suspected

DNA extraction
& Target sequencing

Bioinformatic processing

Phylogenetic analysis

SNP calling

Test for introgression among closely related species

- Hybridization detection using Phylogenetic Invariants HyDe
use concatenated exon alignment of all nudibranchs for full hybridization
detection analysis `run_hyde.py` script
- 80 significant results from HyDe: widespread introgression among closely
related species; only 2/4 indiv. with mitonuclear discordance retrieved as
possible hybrids -> both introgression and mitochondrial capture seem to
play a role





iNaturalist

Questions? 😊