

Supplemental Material: figures and tables

Table of contents

1 Assembly stats	2
2 Genome size estimate	3
3 Principal component analysis	4
3.1 Expanded sample set	4
3.2 IsoSeq reference	5
4 Population statistics	6
4.1 Estimates of <i>Fst</i>	6
4.2 Estimates of <i>pi</i>	7
4.3 Inferred from SNP data	8
5 Phylogenetics	9
5.1 Genetrees	9
6 Morphological identificaiton	13
7 Principal component analyses	14
7.1 Cluster C2	14
7.2 Cluster B1+B2	15
7.3 Temporal variation	16
7.4 K-means clustering subclustering	17
7.5 <i>Fst</i> values between subpopulations	18
7.6 <i>Fst</i> values between populations and regions	19

1 Assembly stats

Table S1 : Genome assembly statistics.

measurement	primary	alternate
total scaffolds	2,386	5,480
total contigs	3,646	9,440
scaffold total size, GB	3.33	2.69
scaffold N50, MB	10.4	4.6
scaffolds >10MB	96	31
scaffolds >1MB	449	625

Table S2 : Genome BUSCO statistics.

measurement	primary	alternate
BUSCO complete	89.7%	86.9%
single-copy	84.8%	80.5%
duplicated	4.9%	6.4%
fragmented	5.3%	5.6%
missing	5.0%	7.5%

2 Genome size estimate

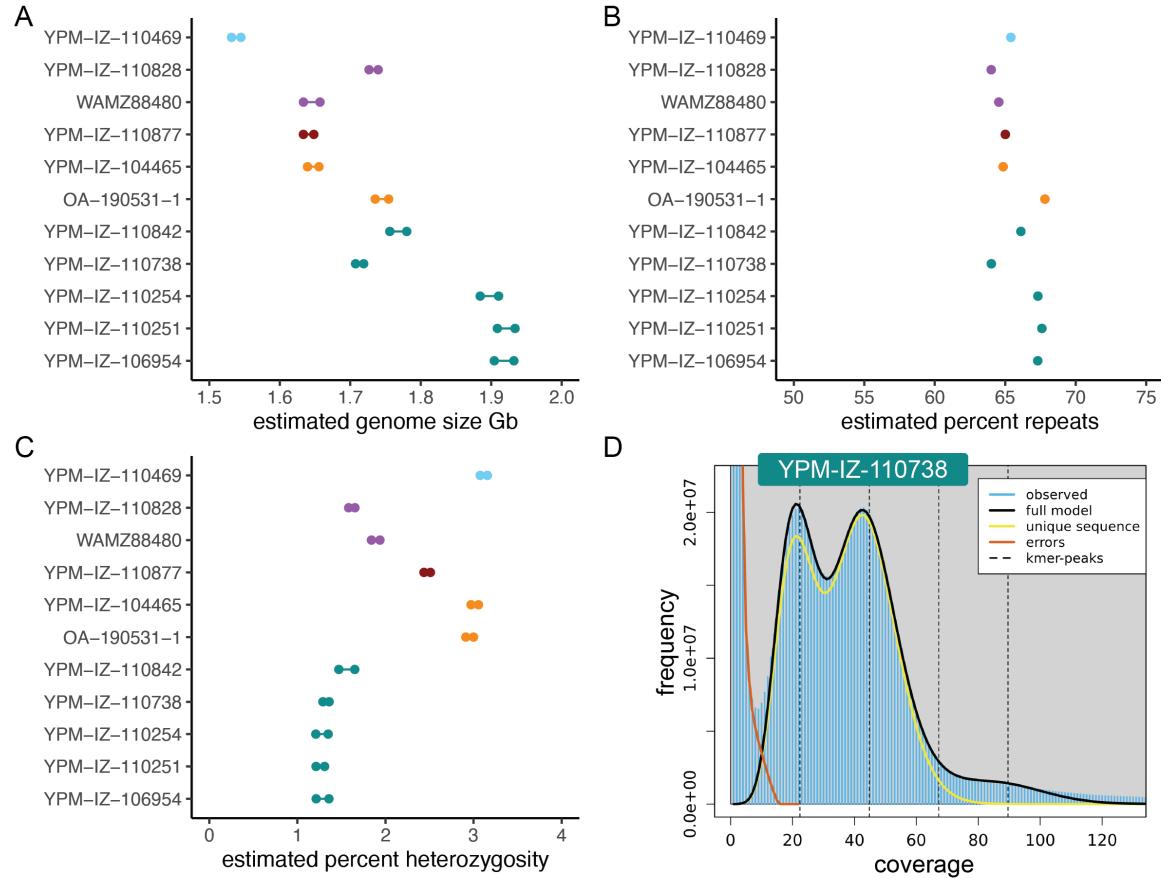


Figure S1 : Genome statistics for samples sequenced to a depth sufficient for a good model fit with a k-mer approach. A, genome size estimates from `GenomeScope`, in gigabases. Colors indicate clusters as in Fig. 2. B, estimated percent of the genome that is repeat sequences. C, estimated genome percent heterozygosity. D, example `GenomeScope` model fit for one Atlantic specimen.

3 Principal component analysis

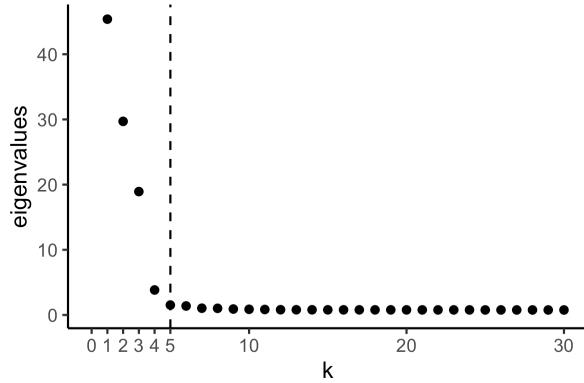


Figure S2 : Eigenvalues of the covariance matrix. The optimal number of components ($k=5$), as determined with PCANGSD, is shown with a dotted line.

3.1 Expanded sample set

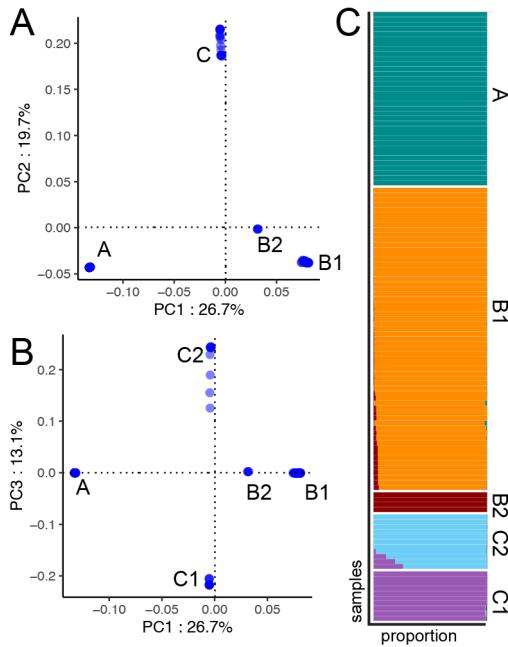


Figure S3 : Principal component analysis (PCA, A-B) and admixture analysis (C) of 121 samples, mapping reads to non-repeat regions of the reference genome.

3.2 IsoSeq reference

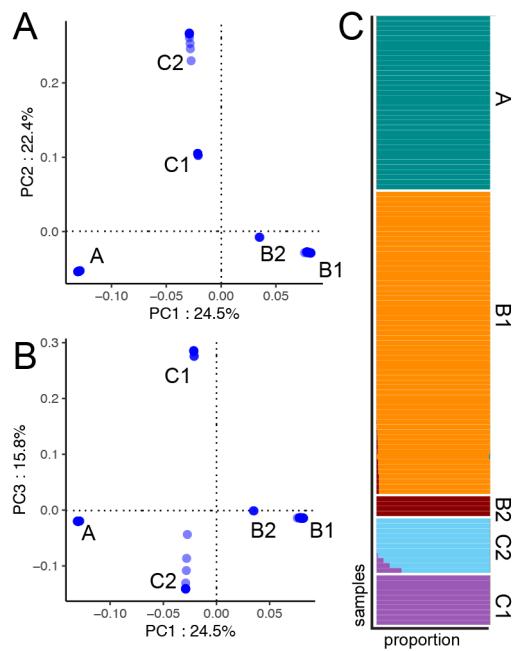


Figure S4 : PCA (A-B) and admixture analysis (C) of 121 samples, mapping reads to a reference transcriptome assembled from IsoSeq data.

4 Population statistics

4.1 Estimates of Fst

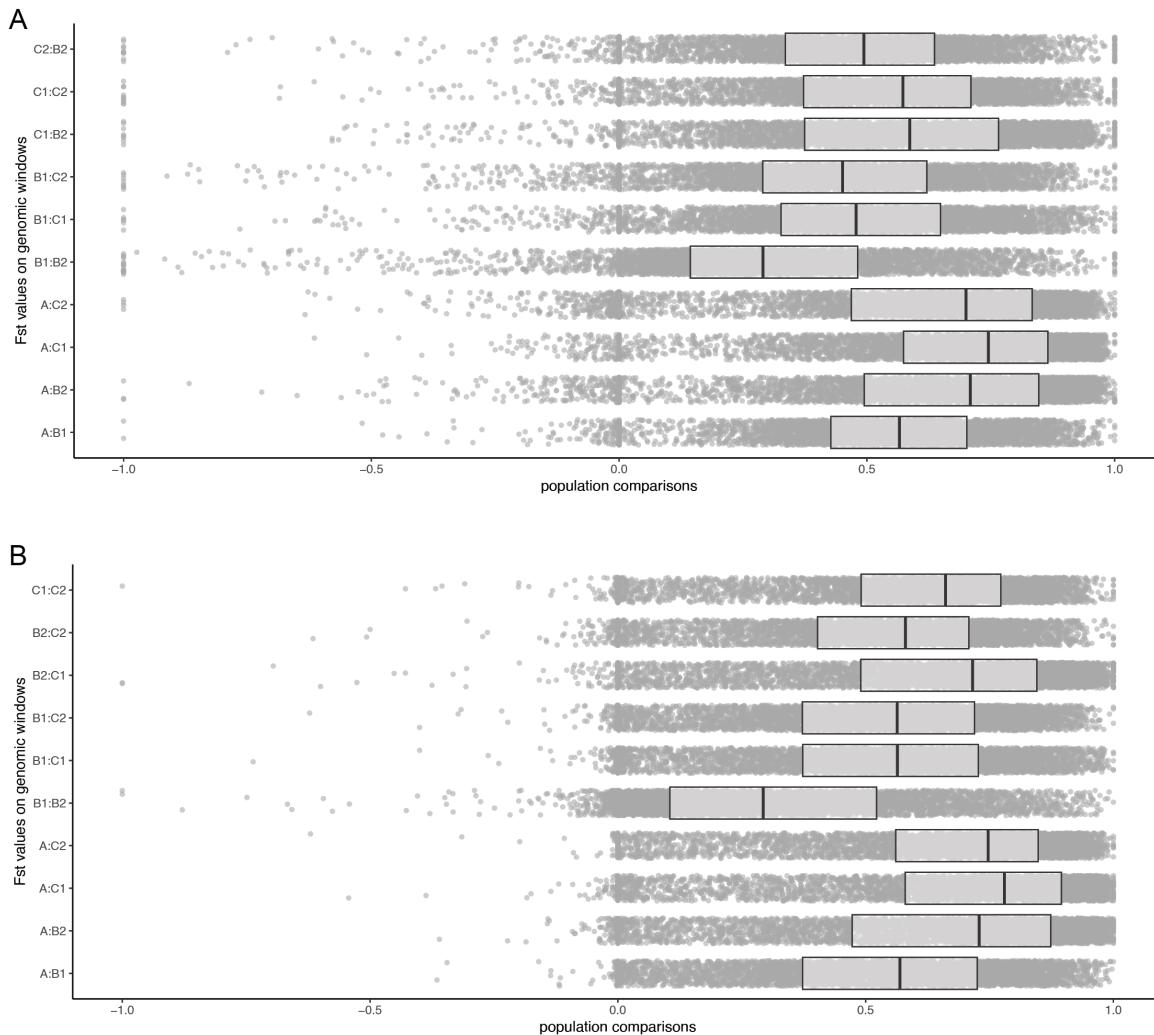


Figure S5 : Fst values between clusters as defined in Figs. 1-2, box-plots show mean and quartile values. A, Fst values across 5,000 randomly sampled non-repeat genomic windows. B, Fst values across 5,000 randomly sampled transcriptomic windows.

4.2 Estimates of π

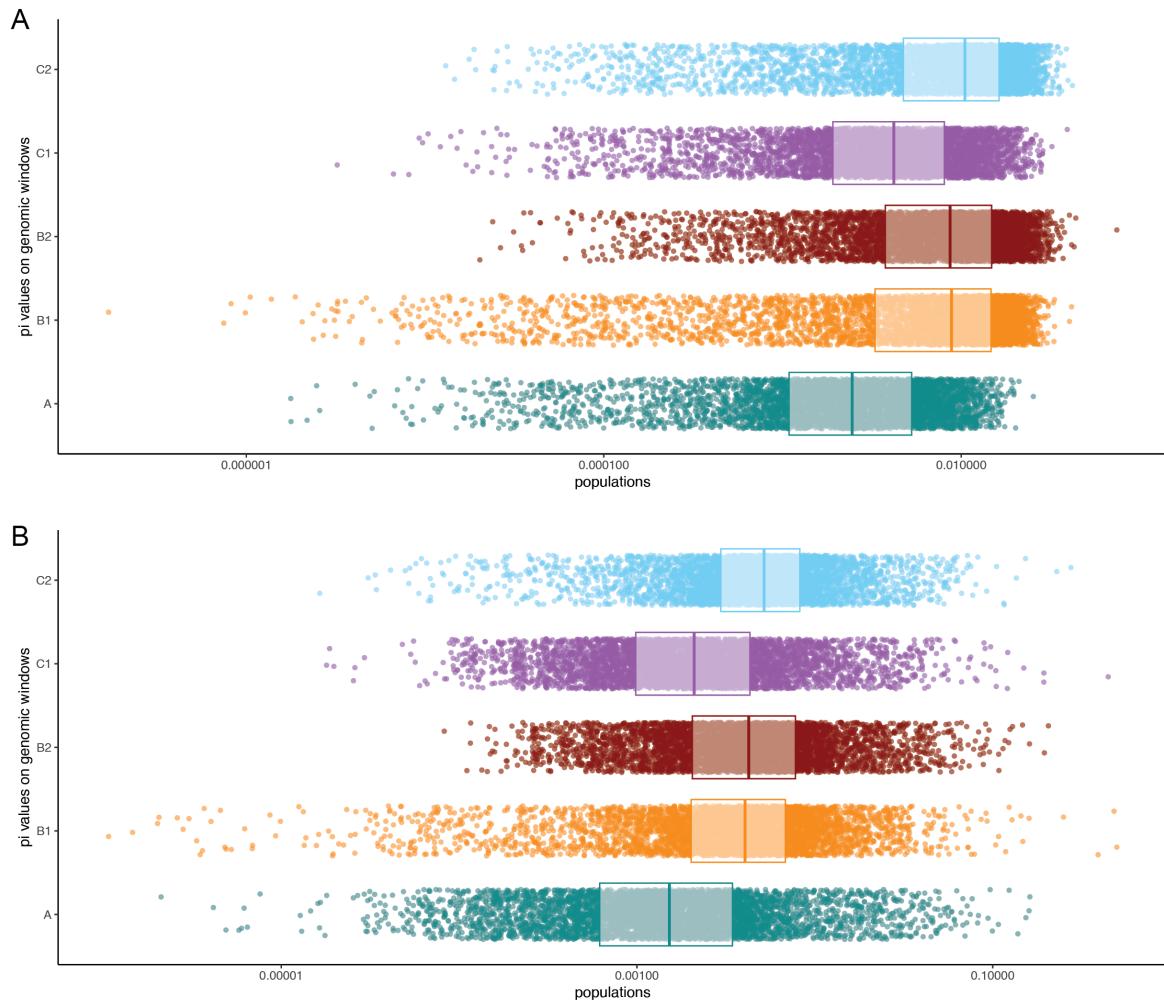


Figure S6 : π values between clusters, as defined in Figs. 1-2, box-plots show mean and quartile values. A, π values across 5,000 randomly sampled non-repeat genomic windows. B, π values across 5,000 randomly sampled transcriptomic windows.

4.3 Inferred from SNP data

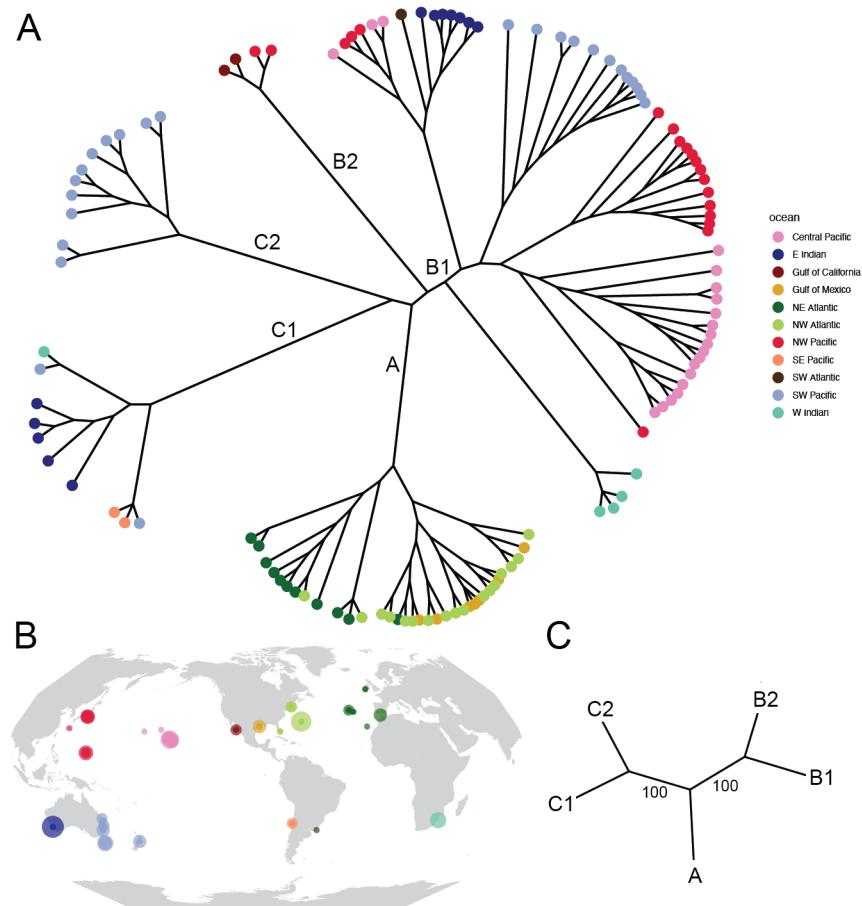


Figure S7 : A, unrooted phylogeny of specimens, inferred with SVDquartets on ~800k high-quality SNPs from non-repeat genomic regions. Internal branches are annotated with cluster labels, as defined in Figs. 1-2. Colors indicate region, as in (B). C, unrooted species phylogeny using the same SNP data. Coalescent bootstrap values are shown at internal branches.

5 Phylogenetics

5.1 Genetrees

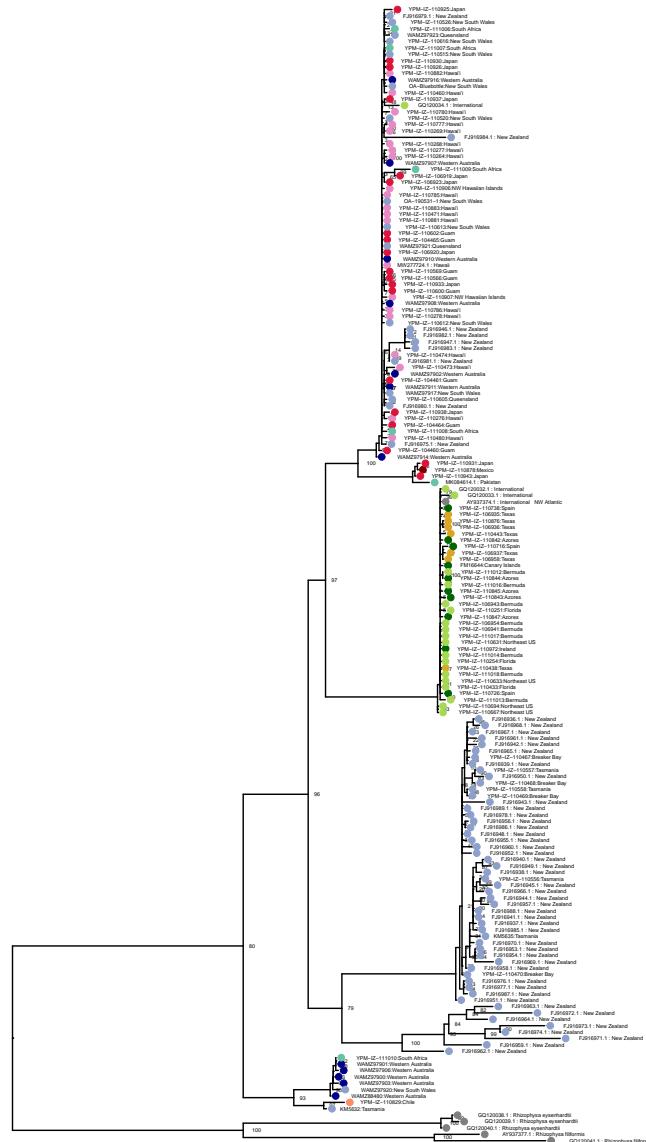


Figure S8 : Phylogeny of cytochrome oxidase 1 (CO1) sequences, assembled using *in silico* PCR from reads, and publicly available sequences from NCBI. Colors indicate region, as in Fig. XXXB. Bootstrap support values shown at nodes. Rooted using *Rhizophysa* as an outgroup.

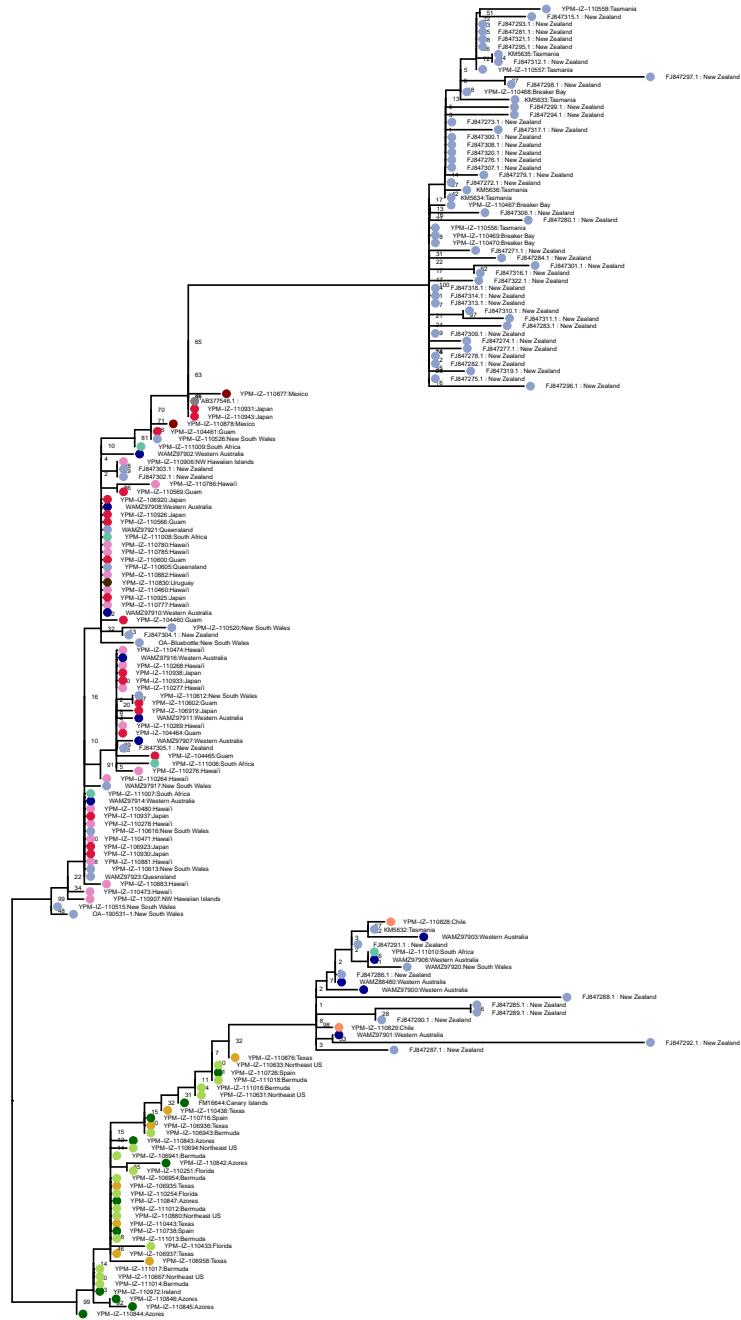


Figure S9 : Unrooted phylogeny of Internal Transcribed Spacer (ITS) sequences, assembled using *in silico* PCR from reads, and publicly available sequences from NCBI. Colors indicate region, as in Fig. XXXB. Bootstrap support values shown at nodes. Rooted using *Rhizophysa* as an outgroup.

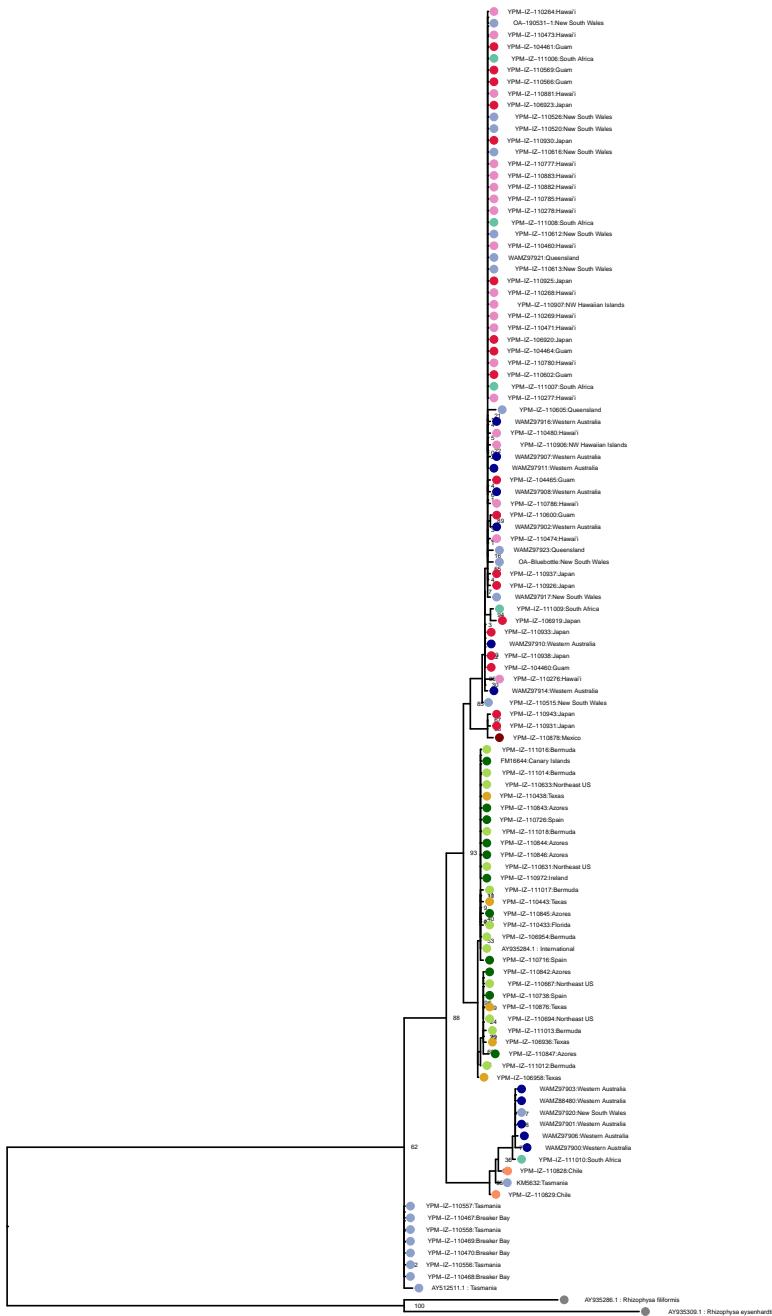


Figure S10 : Phylogeny of 16S ribosomal RNA sequences, assembled using *in silico* PCR from reads, and publicly available sequences from NCBI. Colors indicate region, as in Fig. XXXB. Bootstrap support values shown at nodes. Rooted using *Rhizophysa* as an outgroup.

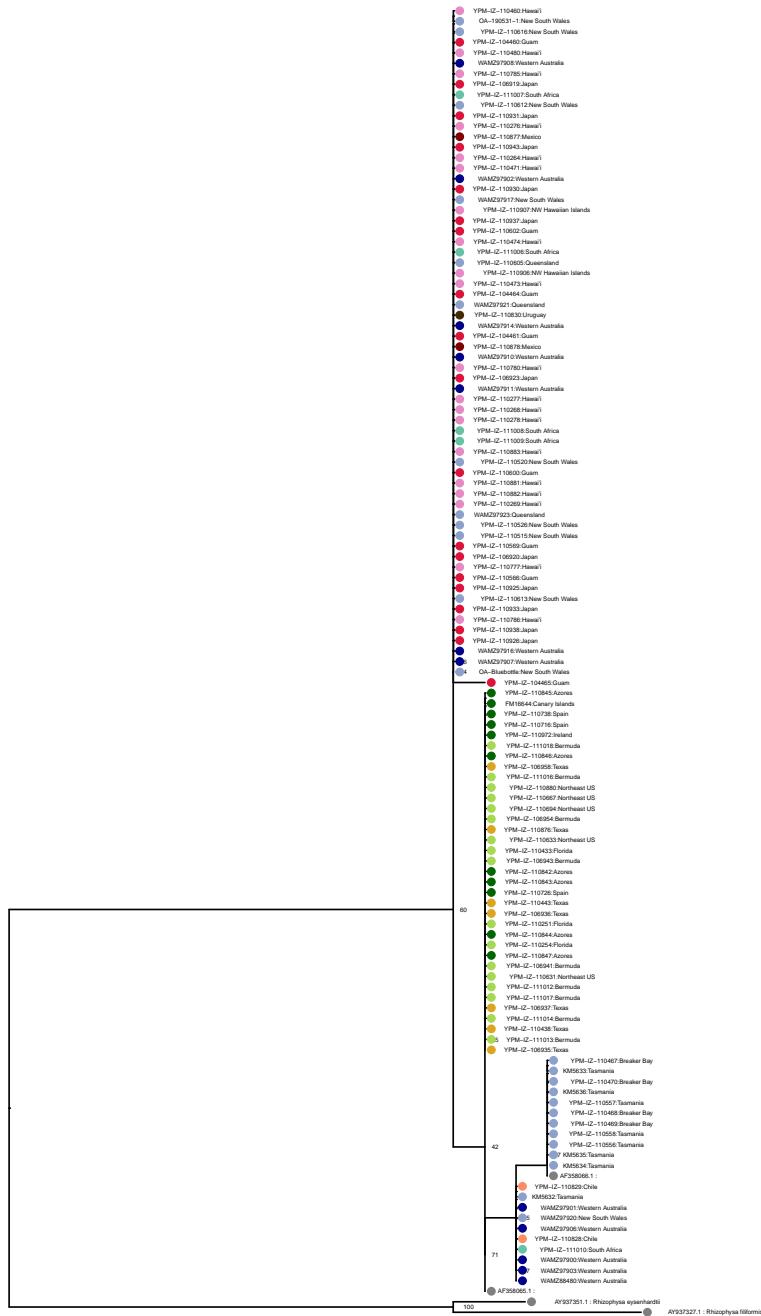


Figure S11 : Phylogeny of 18S ribosomal RNA sequences, assembled using *in silico* PCR from reads, and publicly available sequences from NCBI. Colors indicate region, as in Fig. XXXB. Bootstrap support values shown at nodes. Rooted using *Rhizophysa* as an outgroup.

6 Morphological identificaiton

The diagram illustrates morphological features for four species: *P. physalis*, *P. utriculus*, *P. megalista*, and *P. sp C2*. The features are:

- crest:float height**: As tall as float (green), greater than float (red).
- float:crest length**: Anterior projection less than 1/4 crest length (green), anterior projection greater than 3/4 crest length (red).
- crest color**: Reddish (green), yellow-tipped (red).
- float color**: Yellow-tipped (green), reddish (red).
- feeding body color**: Reddish (green), yellow-tipped (red).
- num. major tentacles**: Greater than 2 (green), 1-2 (green), 2 (red).
- gap between growth zones**: Present (green), absent (red).

	<i>P. physalis</i>	<i>P. utriculus</i>	<i>P. megalista</i>	<i>P. sp C2</i>
feeding body color	reddish	yellow-tipped		
float color		clear, glassy		
crest color				
crest:float height	as tall as float			
float:crest length	anterior projection <1/4 crest length	anterior projection >3/4 crest length		
gap between growth zones		present		
num. major tentacles	>2			
		1-2	>2	

Legend:

- Strong indicator, sufficient to confirm (green box)
- Weak indicator, must all be present to confirm (light green box)
- Counterindicator, sufficient to exclude (red box)

Figure S12 : Rules-based analysis for positive identification of inaturalist images based on scored traits.

7 Principal component analyses

7.1 Cluster C2

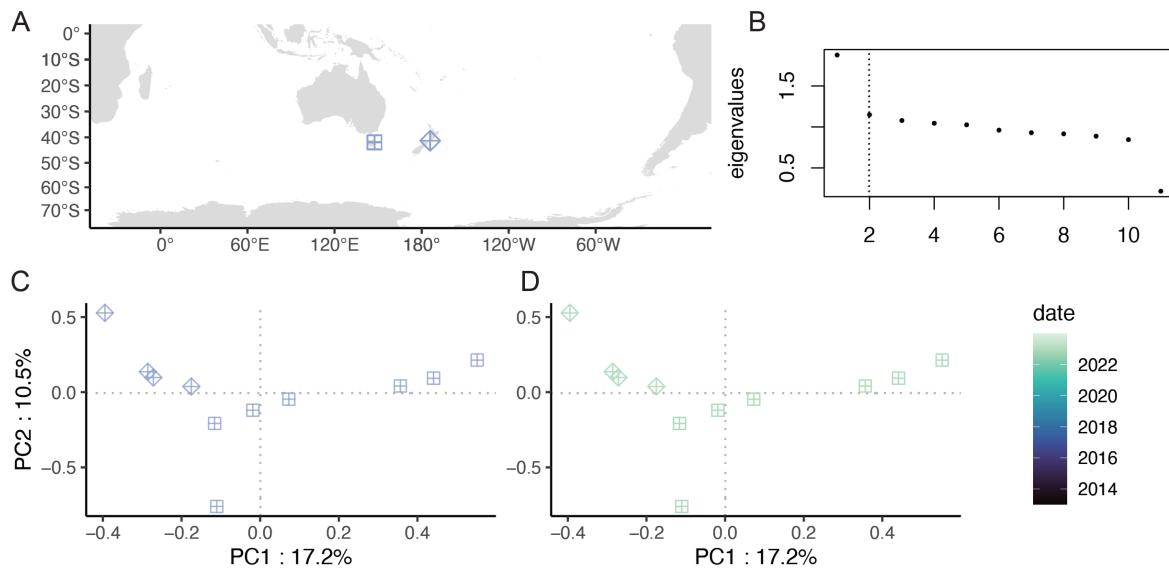


Figure S13 : PCA of samples in cluster C2. A, map of samples. B, eigenvalues of the covariance matrix. C, points arranged on the first two PCs, colored by region. D, colored by collection date.

7.2 Cluster B1+B2

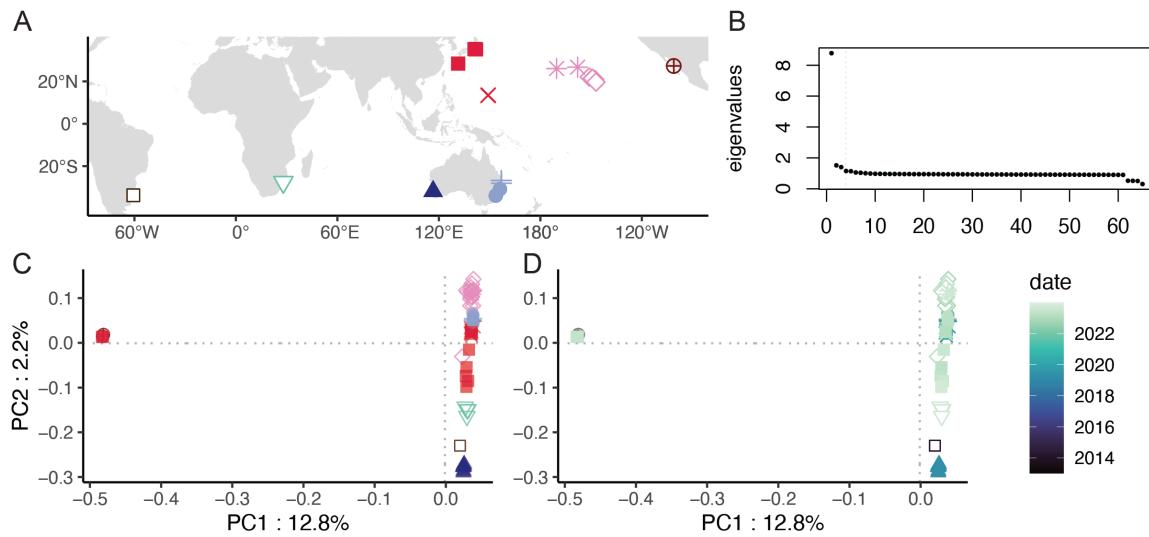


Figure S14 : PCA of samples in clusters B1 and B2 together. A, map of samples. B, eigenvalues of the covariance matrix. C, points arranged on the first two PCs, colored by region. D, colored by collection date.

7.3 Temporal variation

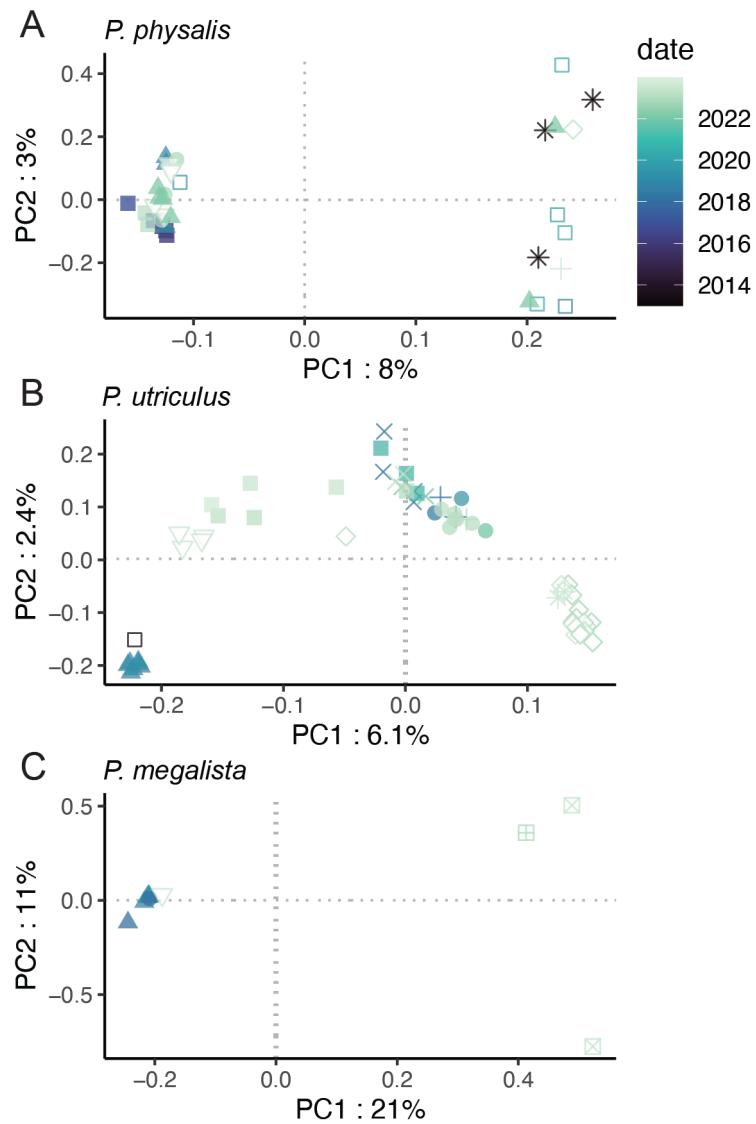


Figure S15 : Sample collection date, visualized on the first two PCs of genomic variation for *P. physalis* (A), *P. utriculus* (B), and *P. megalista* (C). Shapes correspond to region, as in Fig 4.

7.4 K-means clustering subclustering

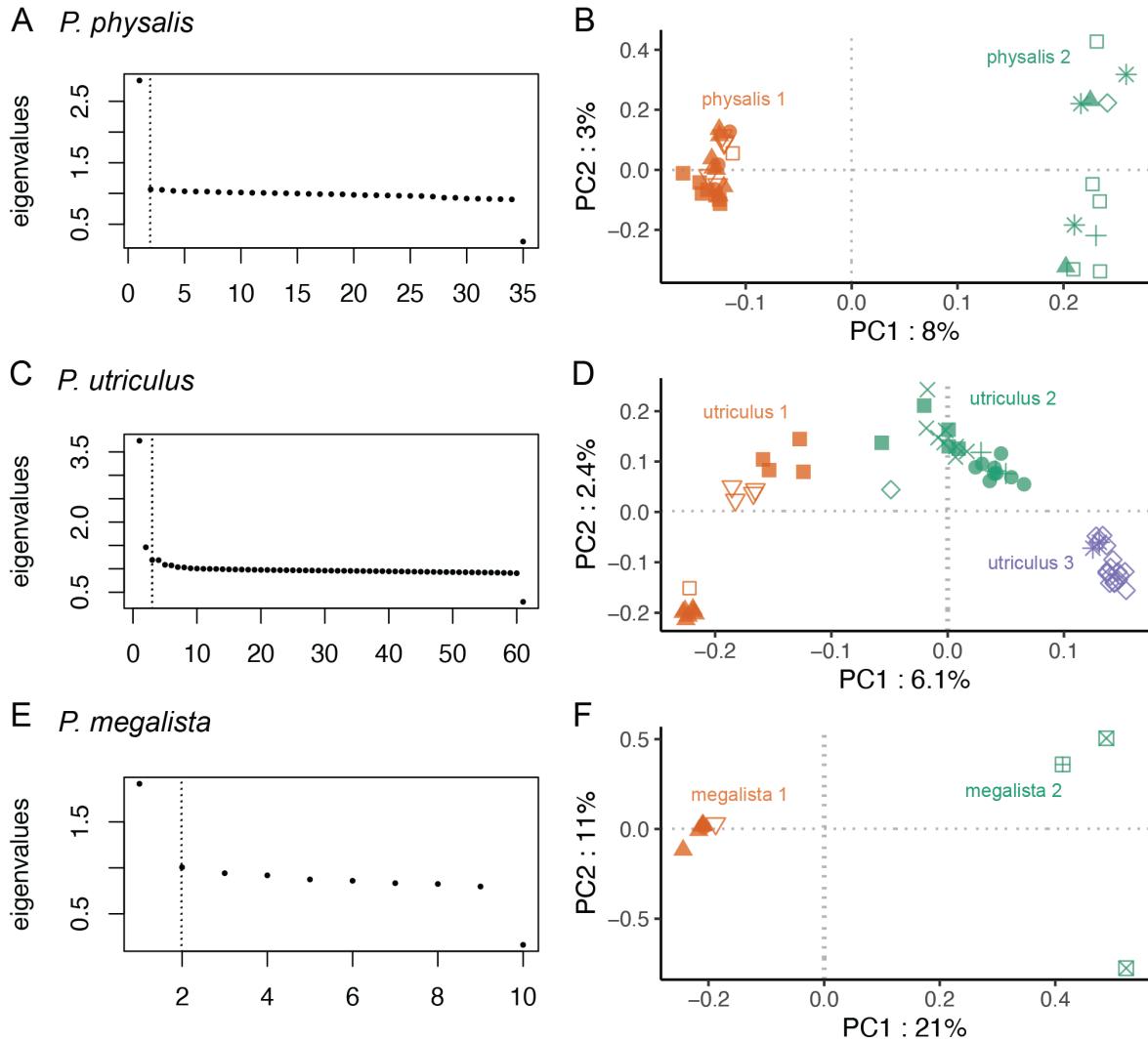


Figure S16 : Result of k-means clustering within each of the three species, *P. physalis* (A-B), *P. utriculus* (C-D), and *P. megalista* (D-E). A, C, E, eigenvalues of the covariance matrix, the optimal number of clusters marked by a dashed line. B, D, F, results of k-means clustering, visualized on the first two PCs. Shapes correspond to regions, shown in Fig. 4

7.5 Fst values between subpopulations

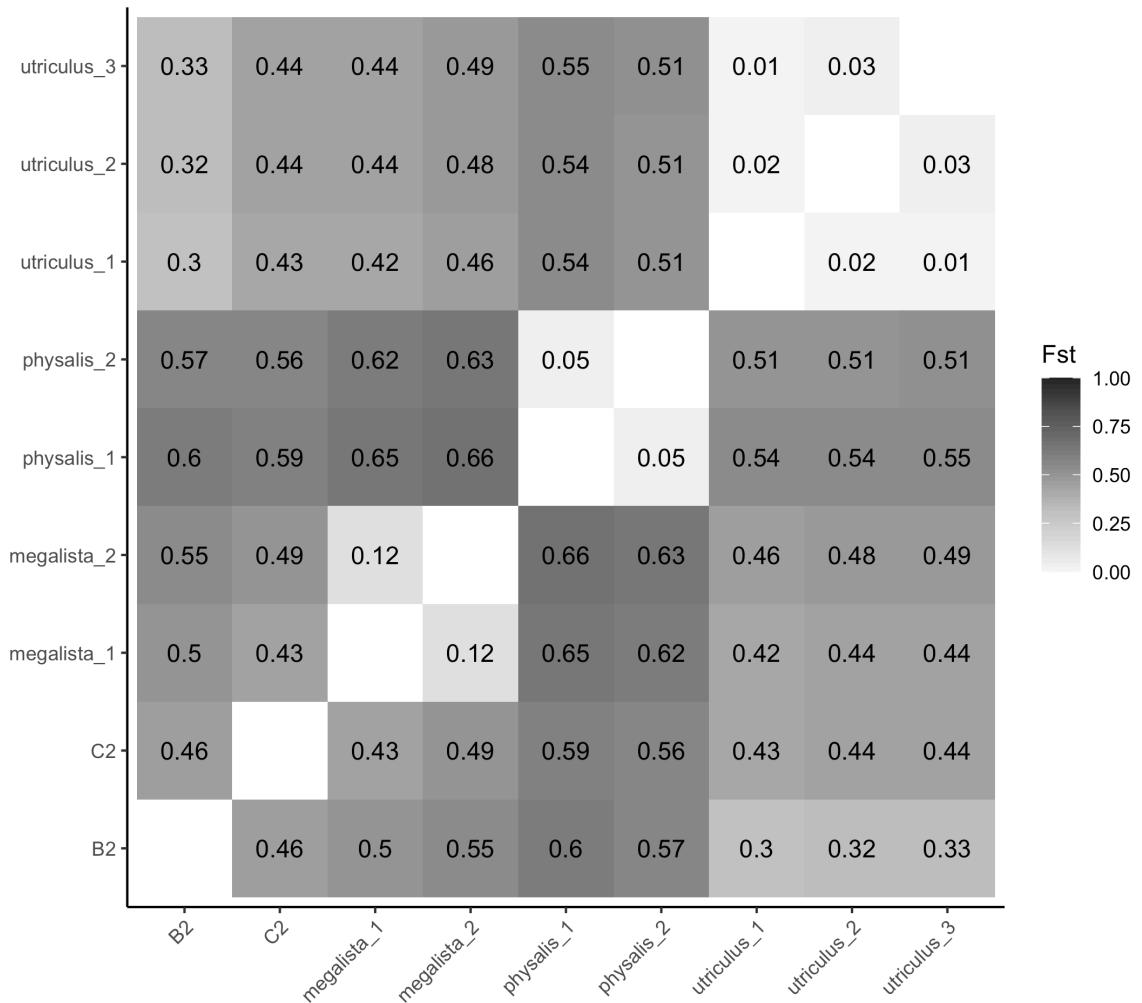


Figure S17 : Reciprocal average Fst values between subpopulations, defined by k-means clustering of the covariance matrix within populations. Populations B2 and C2 were each treated as one cluster given limited sampling.

7.6 Fst values between populations and regions

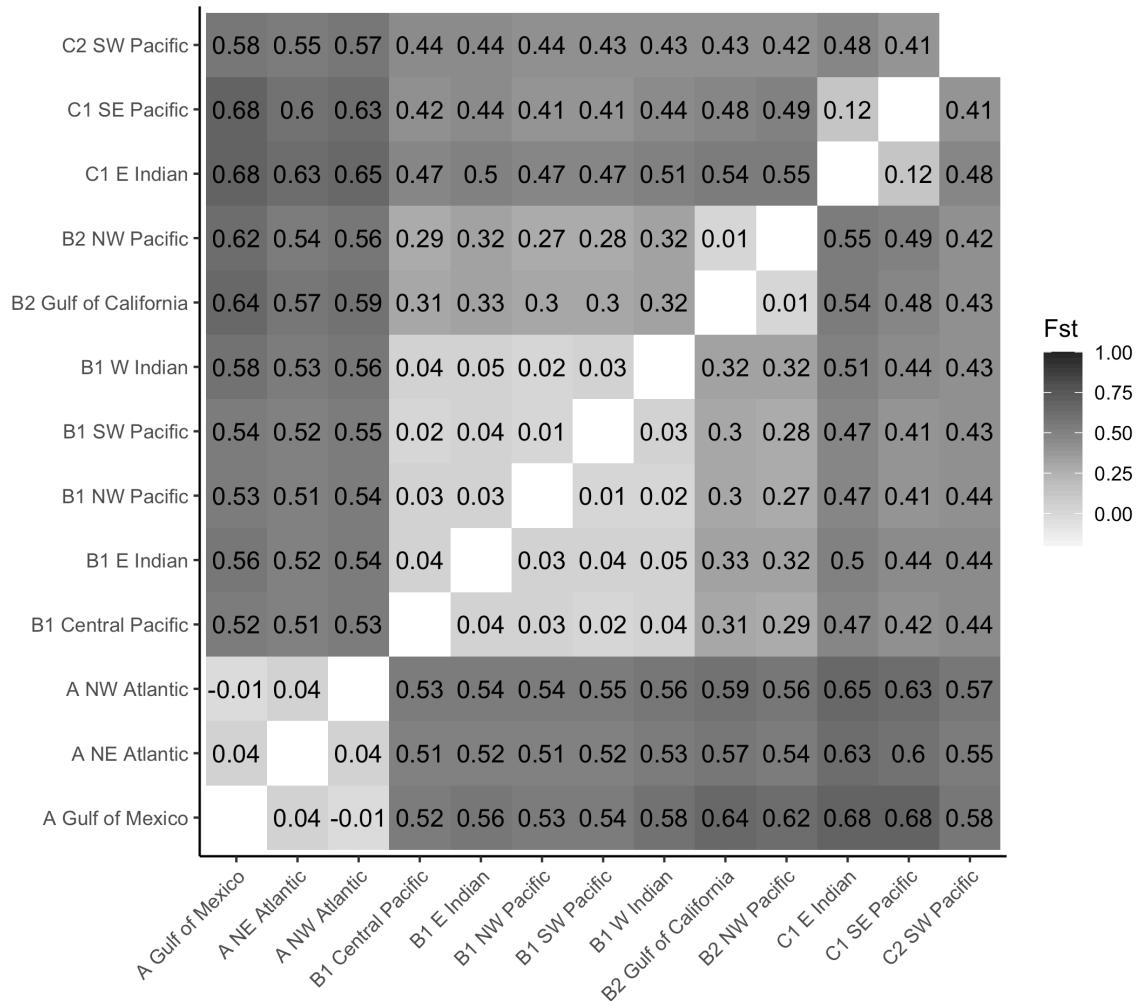


Figure S18 : Reciprocal average Fst values between subpopulations, defined by population and oceanic region, and excluding subpopulations with only one representative sample.

⋮