

Supplementary tables and figures

Table of contents

1 Genome assembly statistics	2
1.1 Genome size estimate	3
2 Population genomics	4
2.1 Expanded sample set	4
2.2 Iso-Seq reference	5
3 Population statistics	6
3.1 Estimates of <i>Fst</i>	6
3.2 Estimates of <i>pi</i>	7
4 Phylogenetics	8
4.1 Phylogeny inferred from SNP data	8
4.2 Genetrees	9
5 Morphological identificaiton	13
6 Subpopulation analyses	14
6.1 Cluster B1+B2	14
6.2 Temporal variation	15
6.3 K-means clustering subclustering	17
6.4 <i>Fst</i> values between subpopulations	18
6.5 <i>Fst</i> values between lineages and regions	19

1 Genome assembly statistics

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%

1.1 Genome size estimate

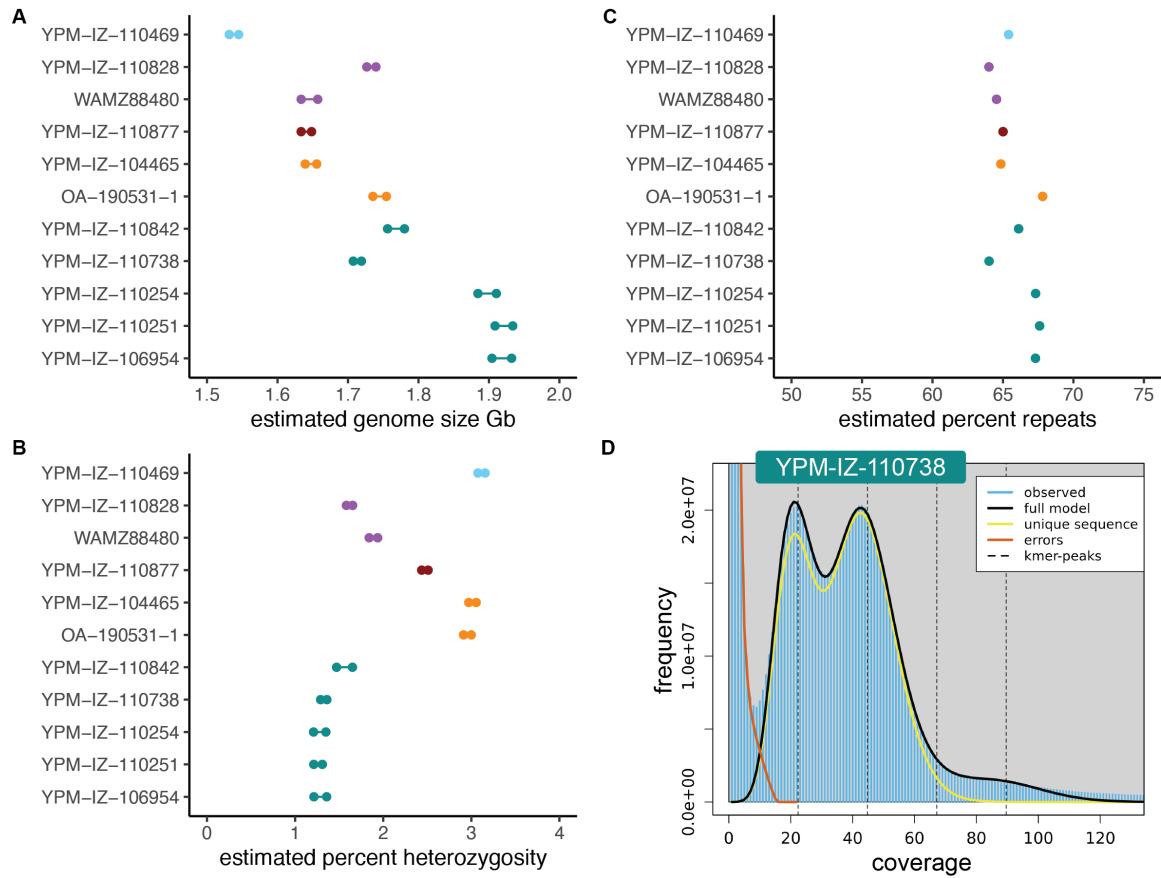


Figure S1 : Genome statistics for samples sequenced to a depth sufficient for a good model fit with a k-mer approach. A, genome size estimates from `GenomeScope`, in gigabases. Colors indicate clusters as in Fig. 2. B, estimated genome percent heterozygosity. C, estimated percent of the genome that is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

2 Population genomics

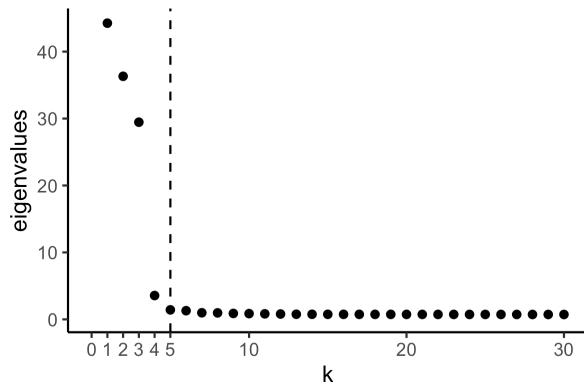


Figure S2 : Eigenvalues of the covariance matrix. The optimal number of components ($k=5$), as determined with PCANGSD, is shown with a dotted line.

2.1 Expanded sample set

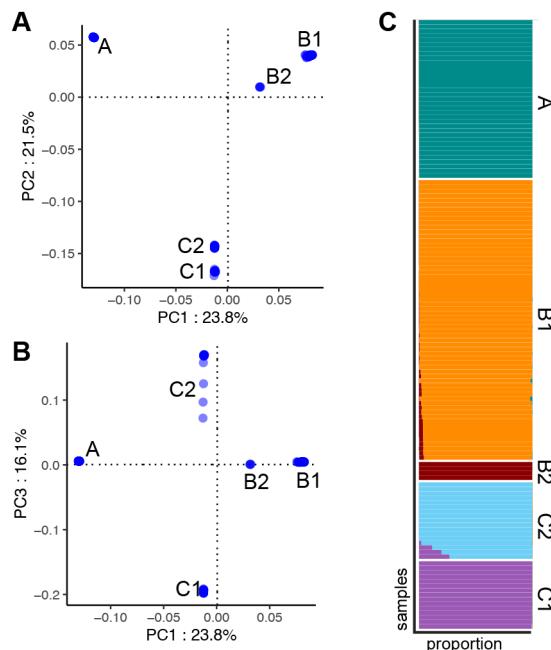


Figure S3 : Principal component analysis (PCA, A-B) and admixture analysis (C) of 133 high and moderate quality samples, mapping reads to non-repeat regions of the reference genome.

2.2 Iso-Seq reference

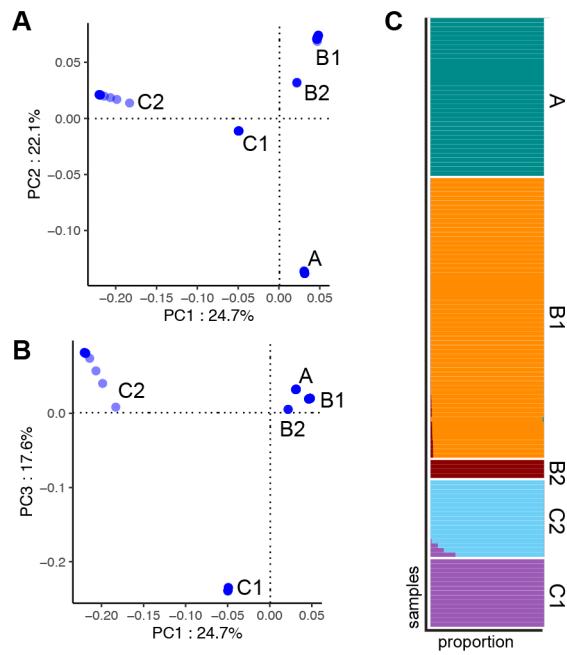


Figure S4 : PCA (A-B) and admixture analysis (C) of 133 samples, mapping reads to a reference transcriptome assembled from Iso-Seq data.

3 Population statistics

3.1 Estimates of Fst

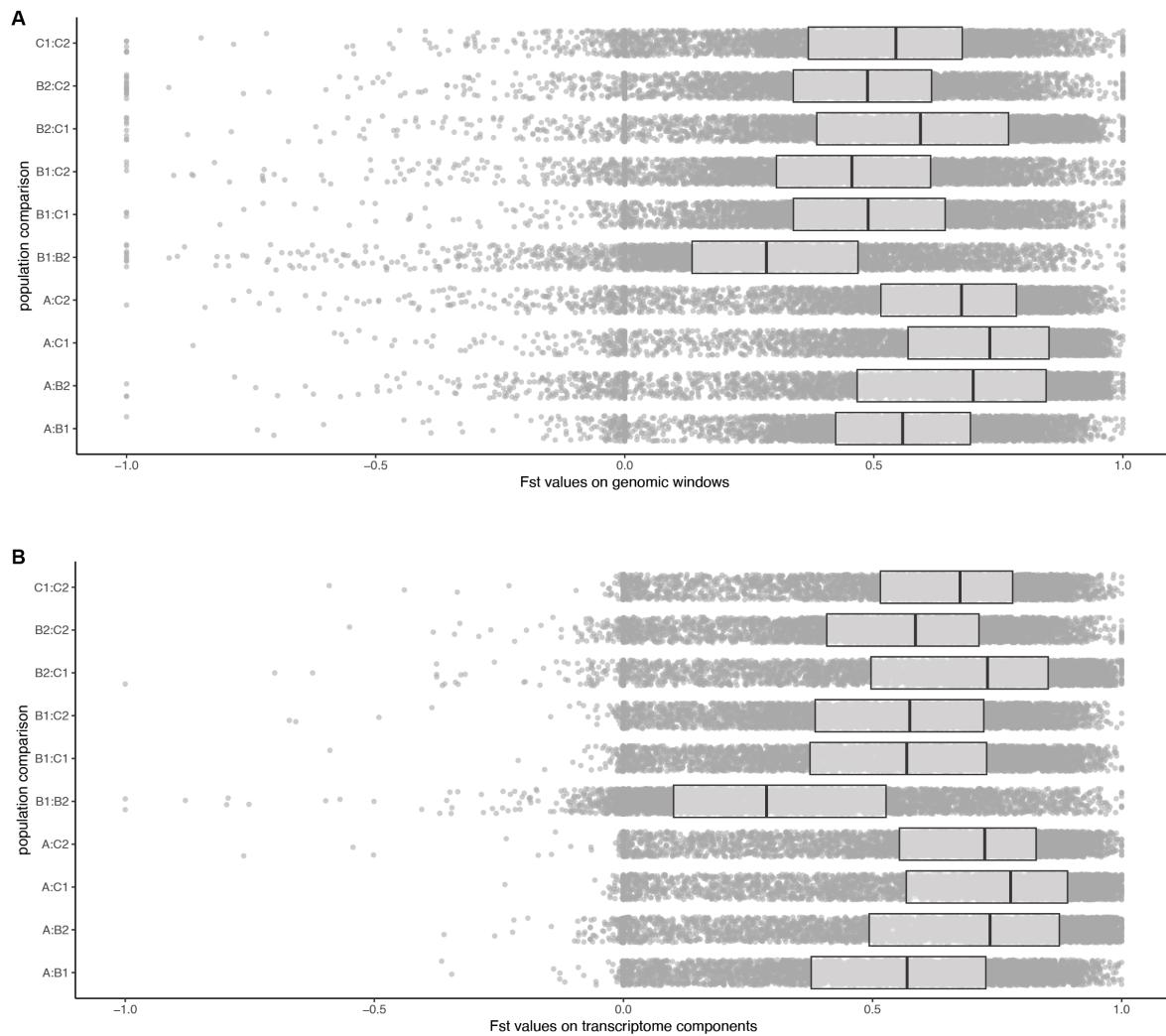


Figure S5 : Fst values between clusters as defined in Figs. 1-2, box-plots show mean and quartile values. A, Fst values across 5,000 randomly sampled non-repeat genomic windows. B, Fst values across 5,000 randomly sampled transcripts.

3.2 Estimates of π

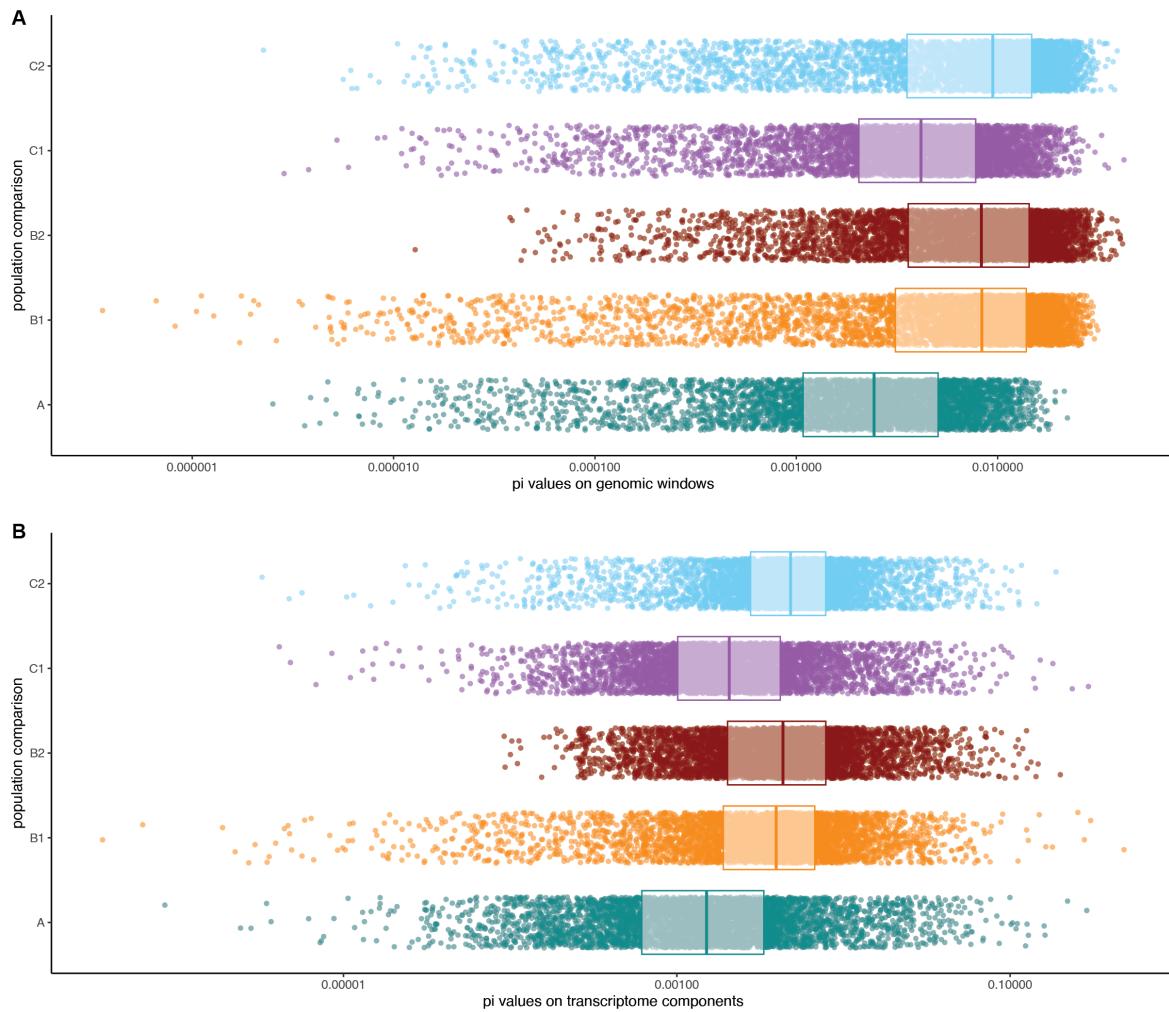


Figure S6 : π values between clusters, as defined in Figs. 1-2, box-plots show mean and quartile values. A, π values across 5,000 randomly sampled non-repeat genomic windows. B, π values across 5,000 randomly sampled transcripts.

4 Phylogenetics

4.1 Phylogeny inferred from SNP data

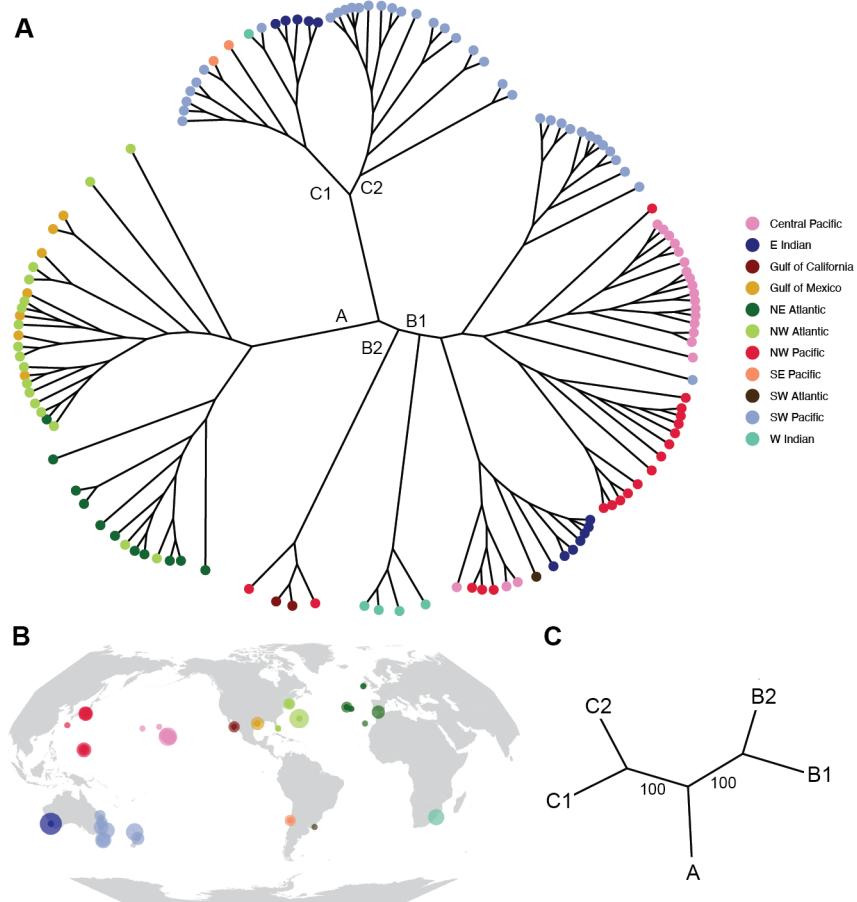


Figure S7 : A, unrooted phylogeny of specimens, inferred with `SVDQuartets` on ~800k high-quality SNPs from non-repeat genomic regions. Internal branches are annotated with cluster labels, as defined in Figs. 1-2. Colors indicate region, as in (B). C, unrooted species phylogeny using the same SNP data.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

4.2 Genetrees

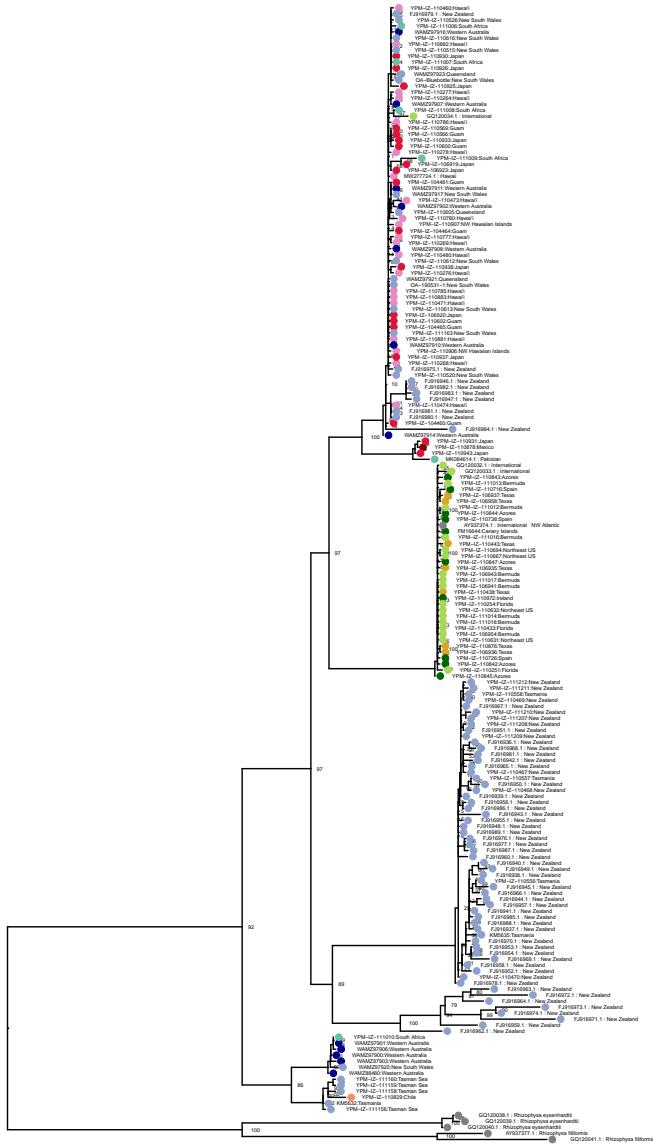


Figure S8 : Phylogeny of cytochrome oxidase 1 (CO1) sequences, assembled using *in silico* PCR from reads, and publicly available sequences from NCBI. Colors indicate region, as in Fig. 7B. Bootstrap support values shown at nodes. Rooted using *Rhizophysa* as an outgroup.

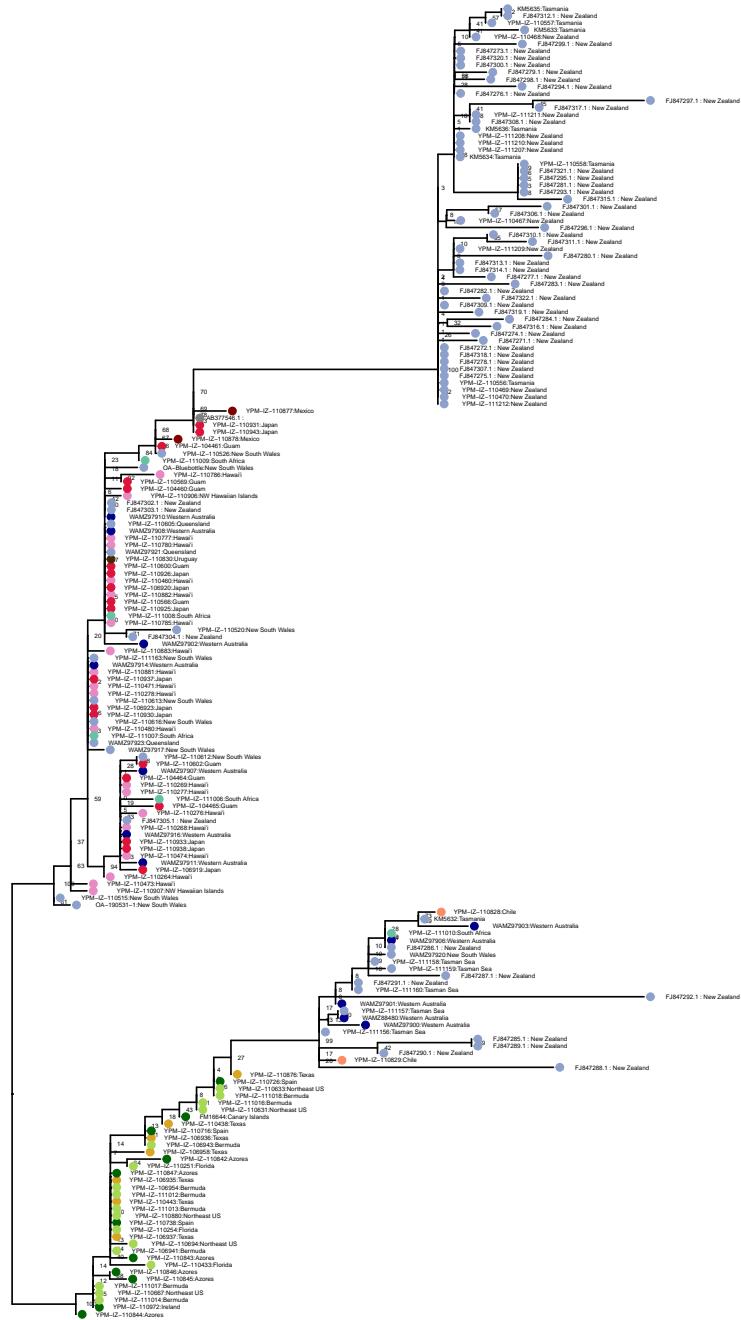


Figure S9 : Unrooted phylogeny of Internal Transcribed Spacer (ITS) sequences, assembled using *in silico* PCR from reads, and publicly available sequences from NCBI. Colors indicate region, as in Fig. 7B. Bootstrap support values shown at nodes.

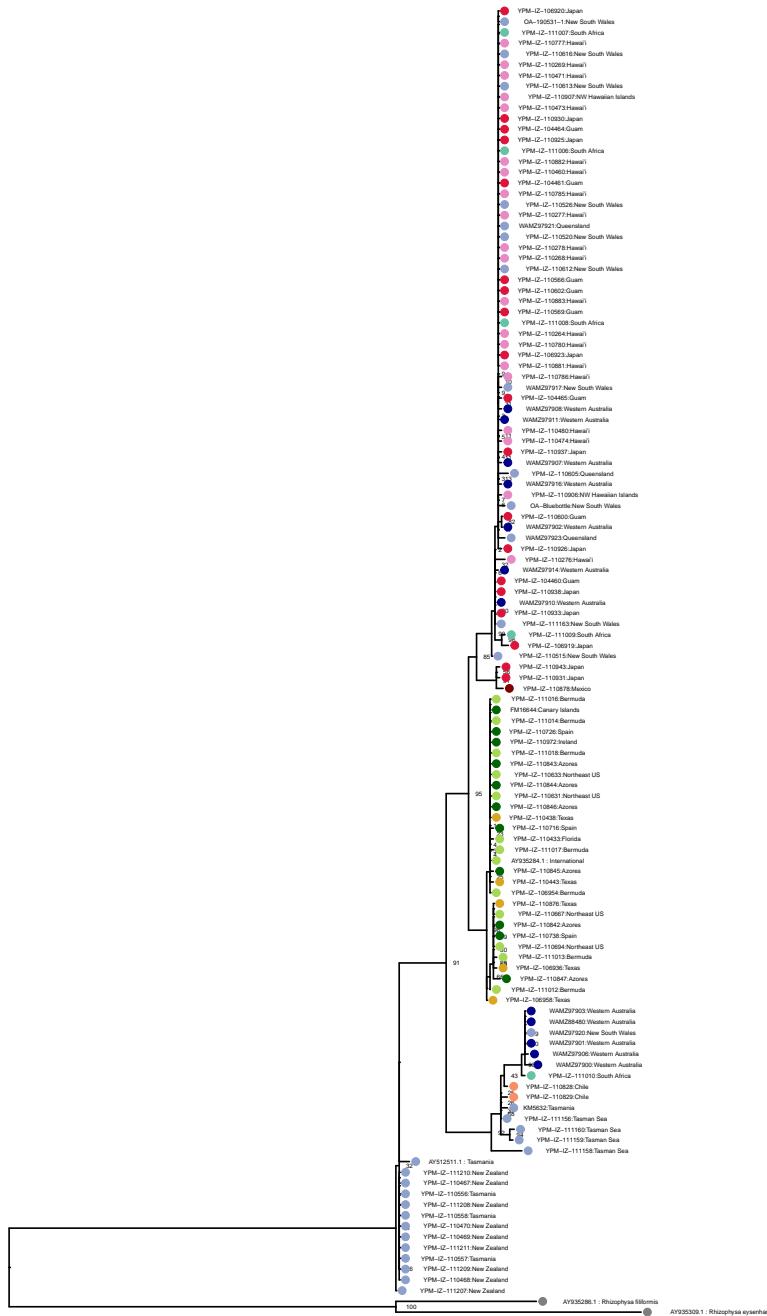


Figure S10 : Phylogeny of 16S ribosomal RNA sequences, assembled using *in silico* PCR from reads, and publicly available sequences from NCBI. Colors indicate region, as in Fig. 7B. Bootstrap support values shown at nodes. Rooted using *Rhizophysa* as an outgroup.

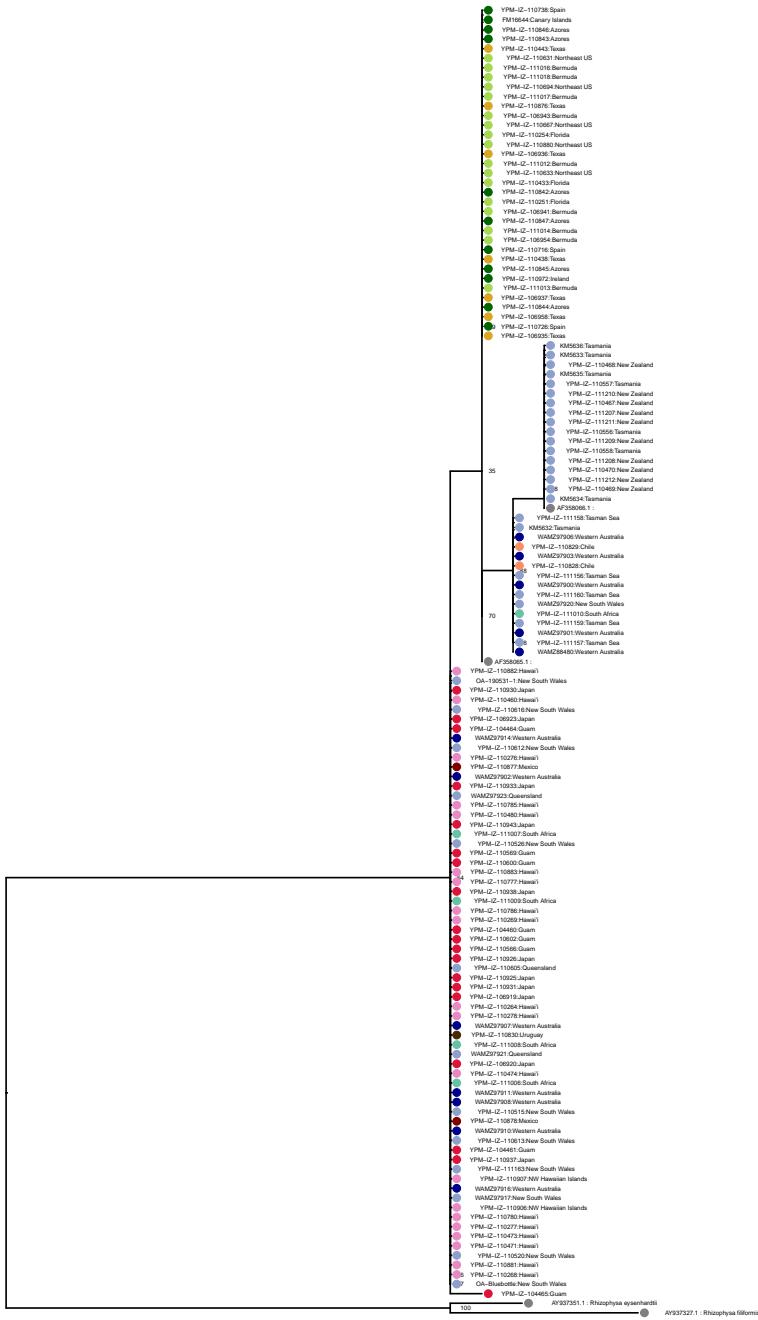


Figure S11 : Phylogeny of 18S ribosomal RNA sequences, assembled using *in silico* PCR from reads, and publicly available sequences from NCBI. Colors indicate region, as in Fig. 7B. Bootstrap support values shown at nodes. Rooted using *Rhizophysa* as an outgroup.

5 Morphological identification

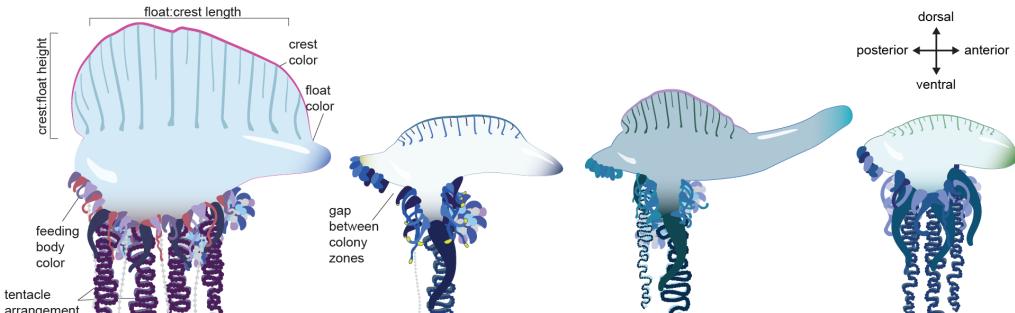


Diagram illustrating morphological features for identification:

- float:crest length:** The ratio of the height of the float to the length of the crest.
- crest color:** The color of the upper, pointed part of the body.
- float color:** The color of the lower, bulbous part of the body.
- feeding body color:** The color of the feeding apparatus.
- tentacle arrangement:** The position and number of tentacles relative to the float.
- gap between colony zones:** A visible gap between distinct growth zones on the body.
- dorsal:** The top side of the organism.
- posterior:** The rear side.
- anterior:** The front side.
- ventral:** The bottom side.

	<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		
arrangement principle tentacles	>2 central			
			reddish OR yellow-tipped	
		clear, glassy		
		purple		
				pink, purple
			no visible height	
			anterior projection >3/4 crest length	
			anterior projection <1/4 crest length	
		present		
		1-2 central		
			>2 central	
				>2 central

strong indicator, sufficient to confirm

weak indicator, must all be present to confirm

counterindicator, sufficient to exclude

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

6 Subpopulation analyses

6.1 Cluster B1+B2

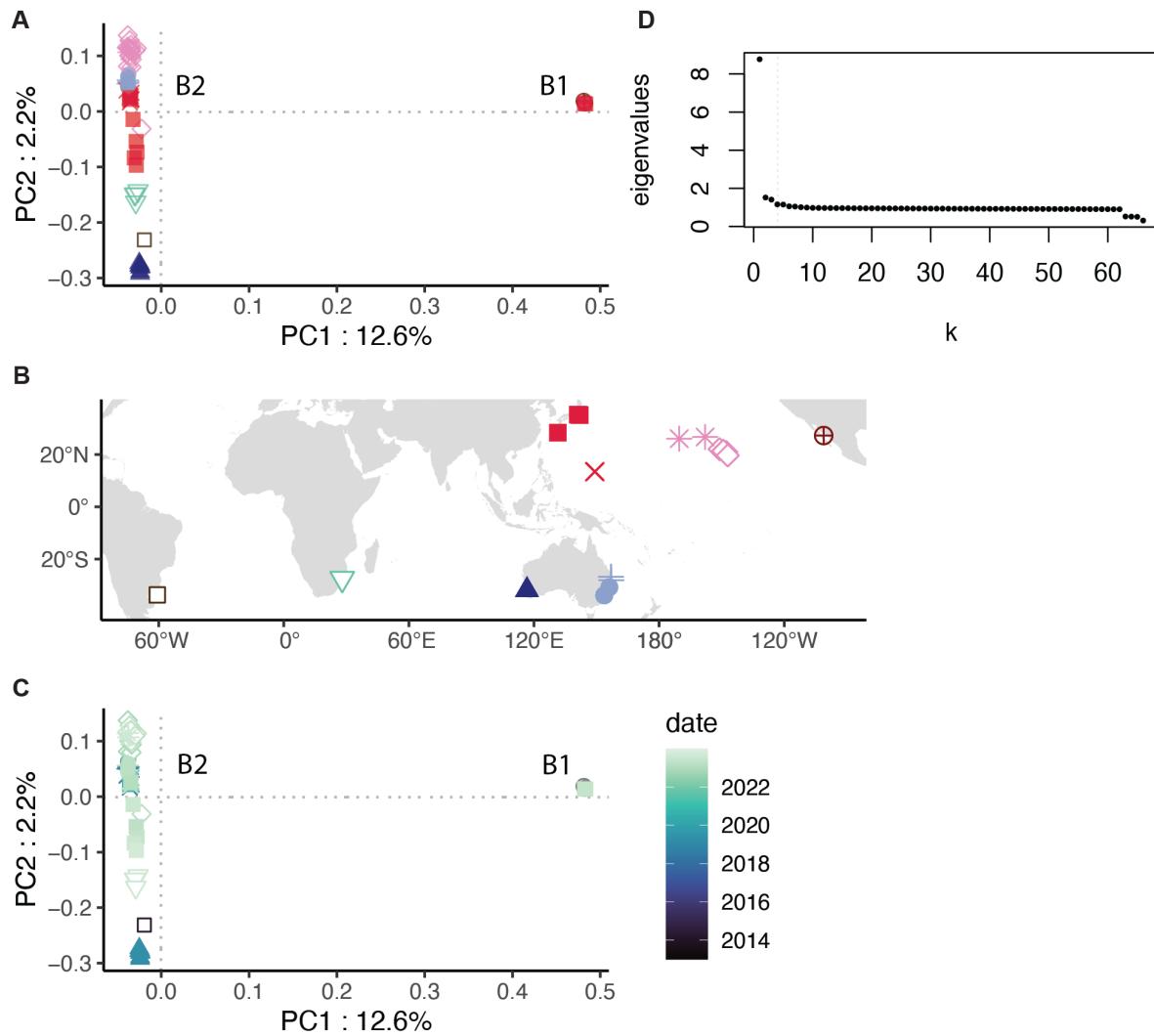


Figure S13 : A, PCA of samples in clusters B1 and B2 together. B, map of samples. C, PCA, colored by collection date. D, eigenvalues of the covariance matrix.

6.2 Temporal variation

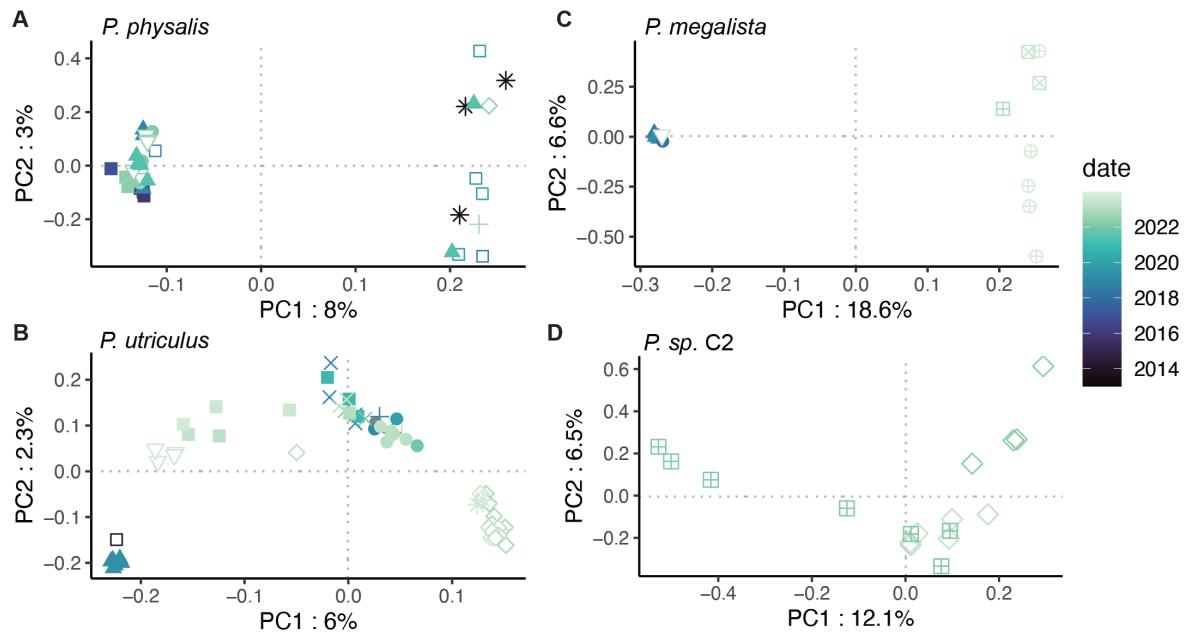


Figure S14 : A-D, Sample collection date, visualized on the first two PCs of genomic variation for each species. Shapes correspond to geographic region, as in Fig 4.

6.3 K-means clustering subclustering

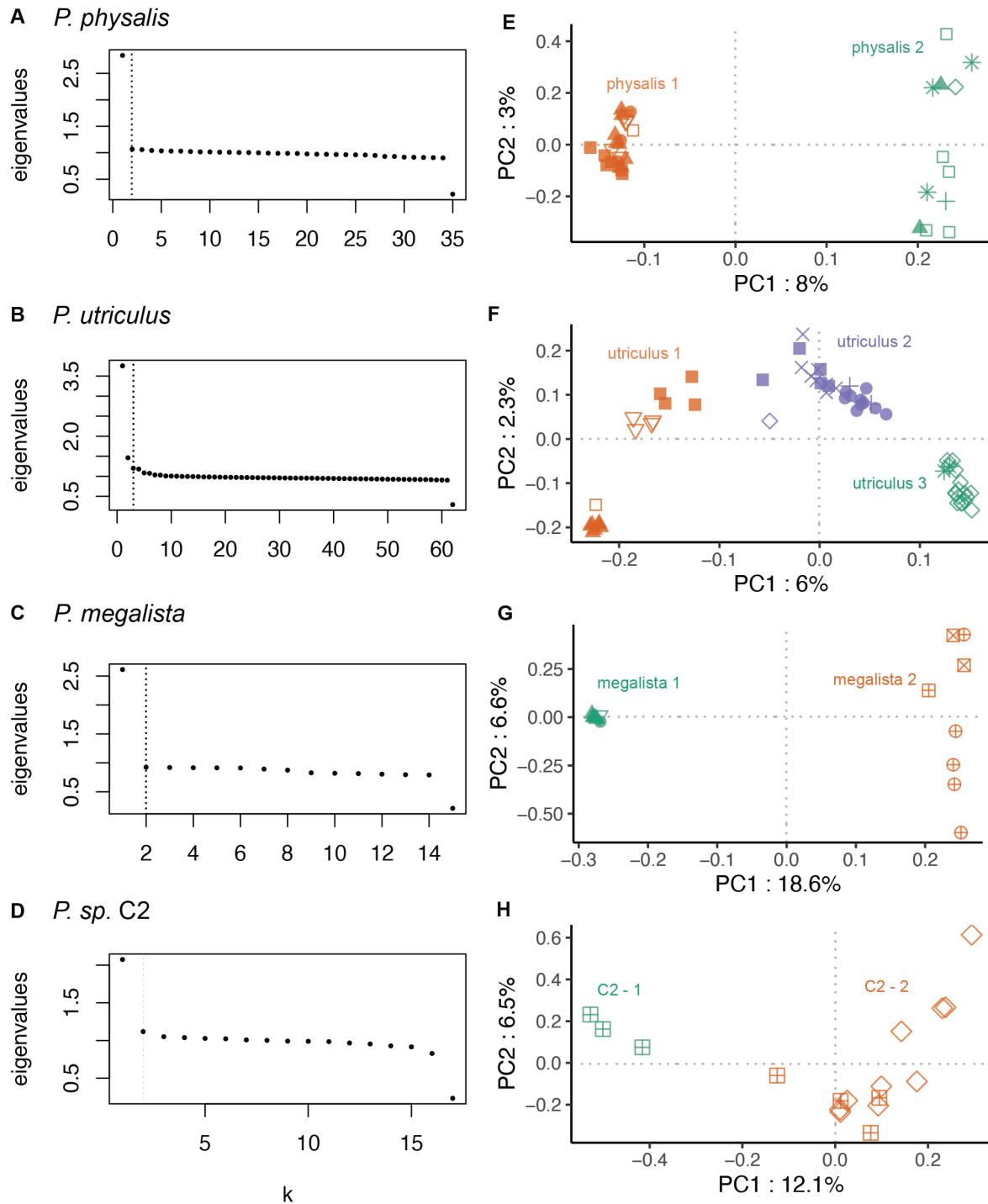


Figure S15 : A-D, eigenvalues of the covariance matrix, the optimal number of clusters marked by a dashed line. E-H, Result of k-means clustering within each of the four species. Shapes correspond to geographic region, as in Fig. 4

6.4 Fst values between subpopulations

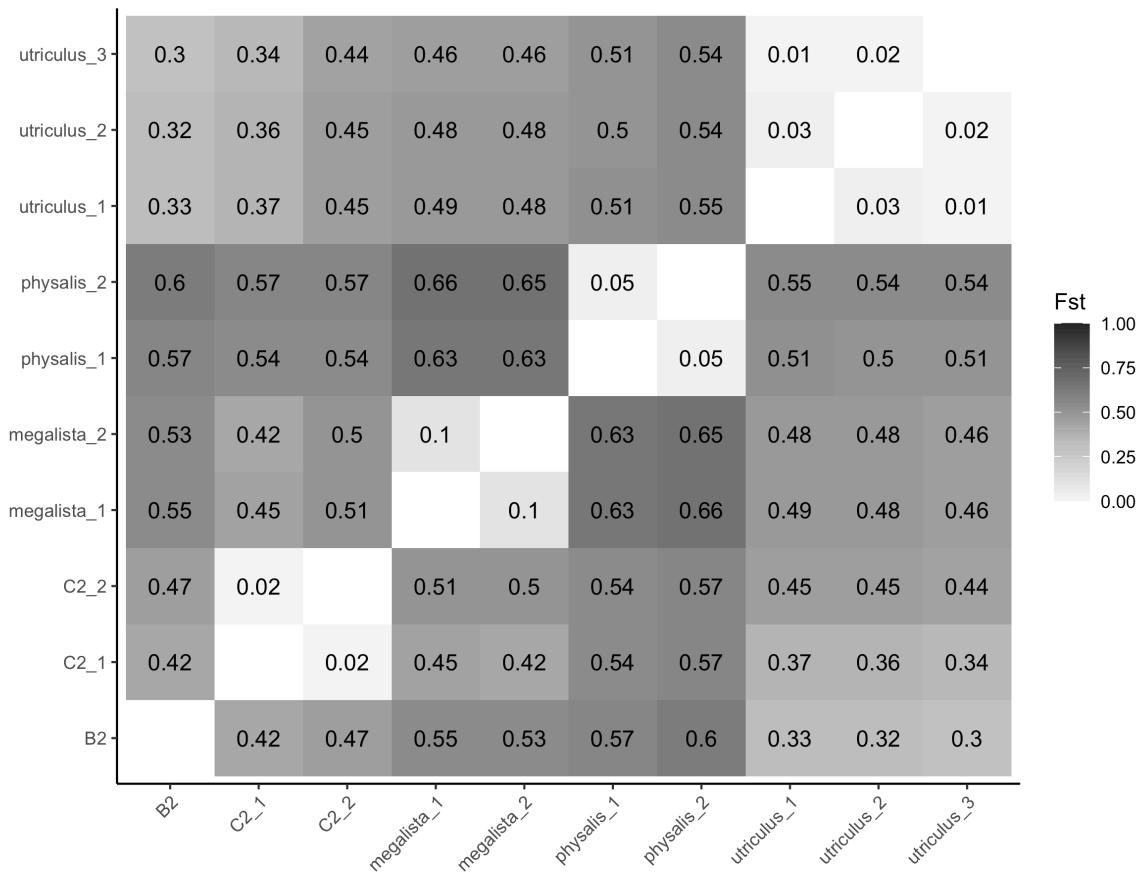


Figure S16 : Reciprocal average Fst values between subpopulations, defined by k-means clustering of the covariance matrix within populations. Lineage B2 was treated as one population.

6.5 Fst values between lineages and regions

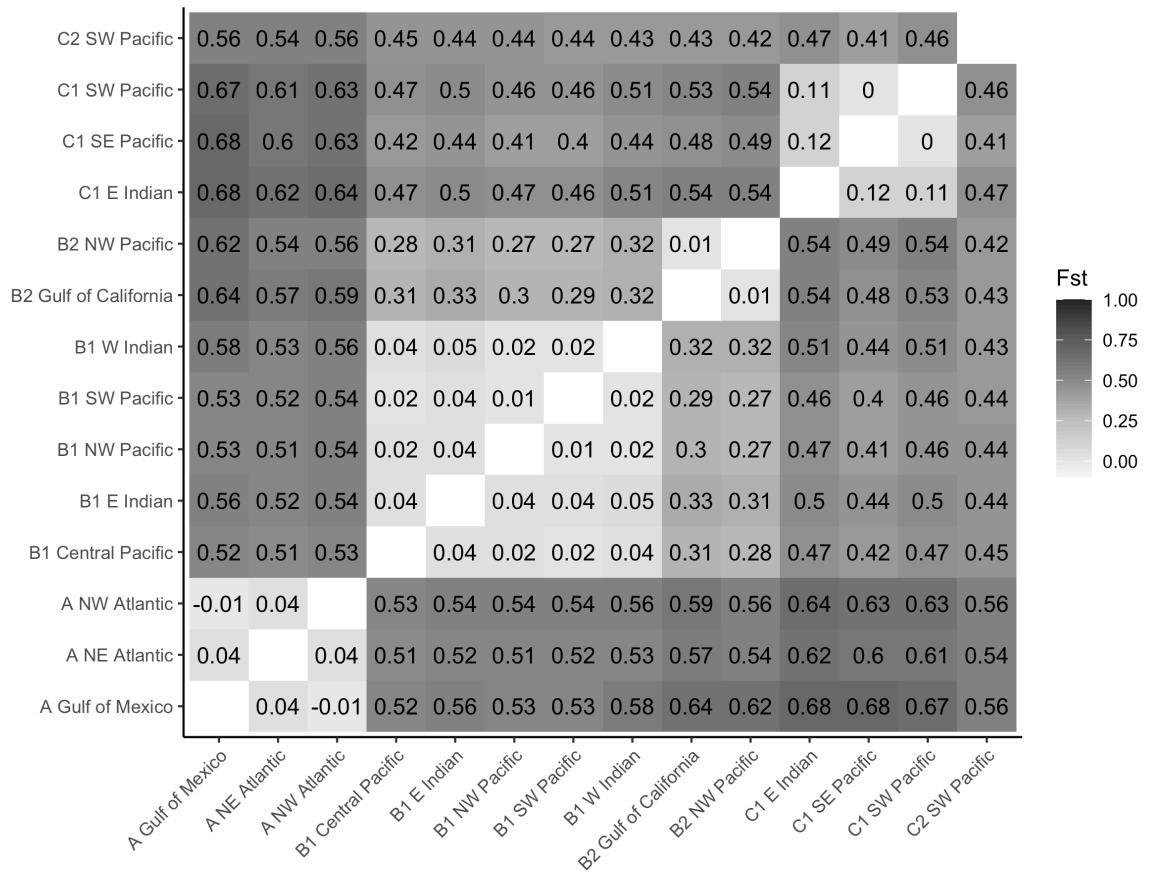


Figure S17 : Reciprocal average Fst values between subpopulations, defined by lineage and oceanic region, and excluding groupings with only one representative sample.