# Incomplete lineage sorting and ancient admixture, and speciation without morphological change in ghost-worm cryptic species

José Cerca[1,2,3], Angel G. Rivera-Colón[4], Mafalda S. Ferreira[5,6,7], Mark Ravinet[8,9], Michael D. Nowak[3], Julian M. Catchen[4] and Torsten H. Struck[3]

[1] Department of Environmental Science, Policy, and Management, University of California, University of California, Berkeley, Berkeley, CA, United States of America
[2] Department of Natural History, NTNU University Museum, Norwegian University of Science and Technology, Trondheim, Norway
[3] Natural History Museum, University of Oslo, Oslo, Norway
[4] Department of Evolution, Ecology, and Behavior, University of Illinois at Urbana-Champaign, Urbana Champaign, IL, United States of America
[5] Division of Biological Sciences, University of Montana, Missoula, MT, United States of America
[6] Departamento de Biologia, Universidade do Porto, Porto, Porto, Portugal
[7] CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO Laboratório Associado, Universidade do Porto, Porto, Porto, Portugal
[8] School of Life Sciences, University of Nottingham, Nottingham, United Kingdom
[9] Centre for Ecological and Evolutionary Synthesis, University of Oslo, Oslo, Norway

## ABSTRACT

Morphologically similar species, that is cryptic species, may be similar or quasi-similar owing to the deceleration of morphological evolution and stasis. While the factors underlying the deceleration of morphological evolution or stasis in cryptic species remain unknown, decades of research in the field of paleontology on punctuated equilibrium have originated clear hypotheses. Species are expected to remain morphologically identical in scenarios of shared genetic variation, such as hybridization and incomplete lineage sorting, or in scenarios where bottlenecks reduce genetic variation and constrain the evolution of morphology. Here, focusing on three morphologically similar *Stygocapitella* species, we employ a whole-genome amplification method (WGA) coupled with double-digestion restriction-site associated DNA sequencing (ddRAD) to reconstruct the evolutionary history of the species complex. We explore population structure, use population-level statistics to determine the degree of connectivity between populations and species, and determine the most likely demographic scenarios which generally reject for recent hybridization. We find that the combination of WGA and ddRAD allowed us to obtain genomic-level data from microscopic eukaryotes ($\sim$1 millimetre) opening up opportunities for those working with population genomics and phylogenomics in such taxa. The three species share genetic variance, likely from incomplete lineage sorting and ancient admixture. We speculate that the degree of shared variation might underlie morphological similarity in the Atlantic species complex.

## INTRODUCTION

The characterization and delimitation of species and populations using DNA sequencing and barcoding has led to the discovery of 'hidden species diversity' in previously established species (*Knowlton, 1993*; *Bickford et al., 2007*; *Pfenninger & Schwenk, 2007*; *Struck et al., 2018*). The initial interest in this hidden diversity, that is cryptic species, fuelled a debate on whether these lineages resulted from biases of a morphologically oriented classification of biodiversity or whether they resulted from underlying biological phenomena. On one side, proponents of the "artefact model" suggest that populations and species naturally accumulate morphological differences, and it is only the limitations associated with scientific methods that impede the discovery of those differences (*Korshunova et al., 2017*). On the one other side, the "evolutionary framework" suggests that the deceleration of morphological evolution is a plausible expectation, given the observation of stasis, niche conservatism and constraints in nature. While some of this diversity is potentially attributed to taxonomic artefacts (*Korshunova et al., 2017*), morphologically similar species—'true' cryptic species—have been discovered in various branches of the tree of life, thus representing an important part of biodiversity (*Pfenninger & Schwenk, 2007*; *Pérez-Ponce de León & Poulin, 2016*; *Cerca, Purschke & Struck, 2018*; *Fišer, Robinson & Malard, 2018*).

Following centuries of morphologically oriented taxonomy, the existence of "true" cryptic species entails a challenge to the delimitation, discovery and classification of species (*Bickford et al., 2007*; *Fišer, Robinson & Malard, 2018*; *Struck et al., 2018*). In the case of morphologically similar species, species delimitation relying on morphology alone will fail to capture the existing species diversity (*Pante et al., 2015*; *Fišer, Robinson & Malard, 2018*; *Chenuil et al., 2019*; *Struck & Cerca, 2019*), resulting in the lumping of different species into a single species complex. While much has been written on the consequences of cryptic species in terms of biological systematics, we have only recently begun to understand the impact of cryptic species in other fields of biology. When species are poorly delimited, determination of biogeographic breaks (*Weber, Stöhr & Chenuil, 2019*; *Cerca et al., 2020a*), inferences on the evolutionary history (*Wada, Kameda & Chiba, 2013*; *Swift, Daglio & Dawson, 2016*; *Struck et al., 2018*; *Dufresnes et al., 2019*), and the determination of ecological richness of an ecosystem (*Chenuil et al., 2019*) may be severely compromised. These problems extend outside fundamental fields of biology when species complexes are medically-relevant, such as the *Anopheles* cryptic species complex where not every morphologically-similar species is capable of transmitting malaria (*Erlank, Koekemoer & Coetzee, 2018*) or in parasite species (*De León & Nadler, 2010*; *Nadler & De Len, 2011*), but also in cases of conservation management (*Bickford et al., 2007*; *Bernardo, 2011*).

While the discovery of cryptic species complexes has increased in the last few years, the resulting debate has focused on whether these are taxonomic artefacts or biologically

relevant species. Consequently, the causes underlying morphological similarity remain mostly unexplored. Despite this hindrance, an important source of information may come from palaeontology where stasis has been studied for decades (*Eldredge & Gould, 1972*; *Gould, 2002*), and from the subsequent integration of this evidence with neontological data. A particularly insightful contribution is that of *Futuyma (2010)*, which suggests that stasis may result from certain ecological, genetic and developmental scenarios. Genetic scenarios include shared genetic variation, potentially resulting from hybridization or ILS, homogenizing morphological divergence; genetic constraints resulting from epistatic reactions or pleiotropy, or constrains from the lack of genetic variation due to repeated bottlenecks or founder effects; stabilizing selection on morphology (*Futuyma, 2010*). Some of these scenarios including stabilizing selection (*Lee & Frost, 2002*; *Novo et al., 2010*; *Novo et al., 2012*; *Lavoué et al., 2011*; *Smith et al., 2011*; *Santamaria et al., 2016*; *Zuccarello, West & Kamiya, 2018*), bottlenecks and founder effects (*Dornburg et al., 2016*; *Valtueña et al., 2016*) have been proposed to explain similarity on cryptic species. However, this remains untested since evidence for morphological similarity comes mostly from the interpretation of indirect methods, such as phylogenetic trees.

The *Stygocapitella* genus includes 11 described species with only four morphotypes and no significant quantitative morphological differences between some species (*Cerca et al., 2020a*; *Cerca et al., 2020b*). Morphologically identical species occur in sympatry and overlap in their distribution along the Northern European, Atlantic American, and Pacific American coastlines. In a previous study, we confirmed that three North Atlantic species— *Stygocapitella westheidei*, *S. subterranea*, and *S. josemariobrancoi*—are morphologically identical (*Cerca et al., 2020a*); nonetheless, we were not able to determine the causes underlying morphological similarity with certainty. Preliminary results from selected DNA markers indicated that morphological similarity potentially stems from niche conservatism and niche tracking, coupled with the fluctuating dynamics of their habitats and/or genetic constraints (*Cerca et al., 2020b*). Here, using genomic data, we extend these efforts by focusing on the causes linked to genetic variation underlying morphological similarity (see above). Following *Futuyma (2010)*, we hypothesize that (1) bottlenecks and founder effects reduce genetic variation, thus resulting in morphological similarity; (2) morphological similarity results from recent admixture; (3) shared genetic variation due to incomplete lineage sorting and ancient admixture underlies morphological similarity.

## METHODS AND MATERIALS

### Study system

*Stygocapitella* is part of the meiofauna, being generally found above the high-water line of sheltered gravel or sandy beaches. To collect individuals, we selected sampling areas based on old records or by assessing beaches using google maps (Table S1; Fig. 1). At each site, we drew a transect from the high-water line to the foot of the dune, digging a 1-meter deep hole every meter starting at the high-water line. In each hole, we collected sediment samples every 15 cm of depth with a volume of about 500 cm$^3$. Sediment samples were brought to the laboratory and interstitial invertebrates were extracted using the $MgCl_2$
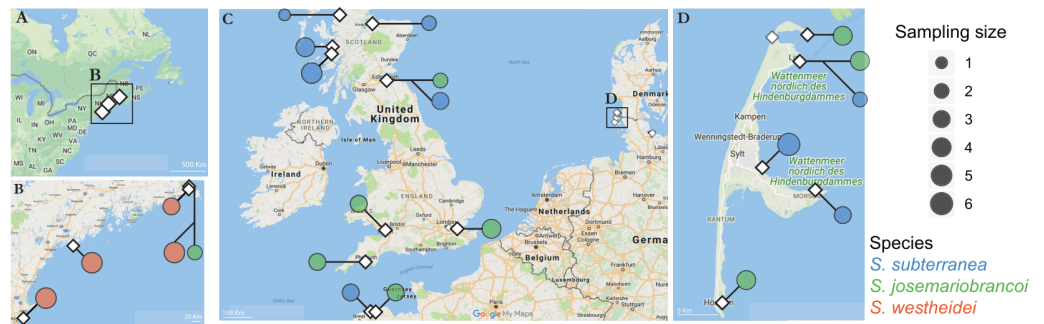
**Figure 1** **Sampling locations across the Northern Atlantic.** (A–B) North America, (C) United Kingdom, France and Germany, and the Island of Sylt in Germany. The three species are displayed in different colours: orange (*Stygocapitella westheidei*), green (*S. josemariobrancoi*) and blue (*S. subterranea*). Sampling locations with multiple circles denote populations in sympatry. Species are delimited using the COI, 16S, 18S and ITS1 barcodes.

Full-size ⌾ DOI: 10.7717/peerj.10896/fig-1

method, and isolated using a dissecting microscope (*Westheide & Purschke, 1988*). After identifying *Stygocapitella*, we collected and preserved these in a ∼70% ethanol solution for DNA extraction.

## DNA extraction and molecular species barcoding

Since *Stygocapitella westheidei*, *S. subterranea* and *S. josemariobrancoi* are morphologically indistinguishable (*Cerca et al., 2020a*), we barcoded individual individuals using 16S, 18S, ITS1 and COI as described in *Cerca et al. (2020b)* and *Cerca et al. (2020a)* (Table S2 includes NCBI reference-IDs). In brief, we extracted DNA from single individuals using either phenol-chloroform or the E.Z.N.A. Tissue DNA Kit (Omega Bio-Tek), and obtained COI (*Astrin & Stüben, 2008*), 18S (*Hillis & Dixon, 1991*), ITS1 (*Cerca et al., 2020a*), and 16S (*Palumbi et al., 1991*; *Zanol et al., 2010*) sequences using PCR. Amplified genetic markers were sequenced by Sanger-sequencing at Macrogen-Europe. For detailed information on amplification, primer sequences and extraction please see *Cerca et al. (2020b)* and *Cerca et al. (2020a)*.

## Library preparation and Illumina sequencing

We selected 50 *Stygocapitella josemariobrancoi*, 47 *S. subterranea* and 24 *S. westheidei* for library preparation (Table S2). Due to the reduced body size, DNA extractions of *Stygocapitella* yield low concentrations of DNA, therefore, to overcome this problem, we used a combination of whole genome amplification (WGA) (*Golombek et al., 2013*; *De Medeiros & Farrell, 2018*) followed by a double-digestion Restriction site-Associated DNA sequencing protocol (ddRAD) (*Baird et al., 2008*; *Peterson et al., 2012*). To complete the WGA reaction, DNA of a single individual is first denatured and mixed with random hexamer primers and the Phi29 DNA polymerase (Illustra Genomiphi HY DNA Amplification Kit; GE©Healthcare Life Science). Following the manufacturer's instructions, 2.5 µl of template DNA were mixed with 22.5 µl of sample buffer, and incubated at 95 °C for three minutes for denaturation. After this, we added 22.5 µl of reaction buffer and 2.5 µl of enzyme mix to the DNA-sample buffer solution, incubated

the solution at 30°C for four hours for DNA amplification and an enzyme heat-inactivation at 65 °C for ten minutes. DNA was purified using AMPure XP beads, and resuspended in ddH₂O. The concentration of the amplified DNA was determined with Qubit and the fragment size distribution with a fragment-analyzer.

For each individual, 500 ng of amplified DNA was digested in 25 μl including 0.5 μl of each restriction enzyme (Pst-I HF and Mse-I, each 20 units/μl) and 2.5 μl Cut-smart buffer. The digestion reaction was carried out at 37 °C for two hours. Digested DNA was purified using Ampure-beads and resuspended in 22 μl ddH₂O, and Illumina adaptors with barcodes were ligated to the digested DNA in a 25 μl reaction including 20.5 μl sample DNA, 1 μl T4 DNA-ligase, 2.5 μl 10X T4 ligase buffer and 1 μl adapter P1/2-mix. This reaction was incubated for 30 min at 25 °C, and the enzyme inactivated for 10 min at 65 °C. The barcoded libraries were pooled, cleaned using AMPure XP beads, and eluted in 100 μl of ultra-purified water. We ran a size-selection step using Blue Pippin's 100–600 bp cassette (BDF2010) selecting for fragment-length between 300–600 bp followed by cleaning with AMPure XP beads to remove short fragments. The library was amplified in 200 μl including 100 μl Q5 HiFi MasterMix, 5 μl Primer mix and 20 μl DNA, in 18 PCR cycles (initial denaturation: 98 °C for 30s; 18 cycles of 98 °C for 10 s, 60 °C for 15 s, 72 °C for 15 s; and a final elongation of 72 °C for 2 min). Finally, two cleaning-quantification steps using AMPure XP beads were done, and the libraries were sent for Illumina Sequencing on an Illumina Hi-Seq 4000.

## Individual identification and multi-marker phylogeny

Raw 16S, 18S, ITS1 and COI sequences were assembled, and ends were automatically trimmed to remove primers and low-quality ends using Geneious v6.8.1, (Table S2). Each consensus sequence was queried against the default NCBI database (nr/nt) using BLAST (*Altschul et al., 1990*; *Camacho et al., 2009*) to exclude potential contamination. For each gene, sequences were aligned using mafft v7.310, using a maximum of 1,000 iterations, and the ends of the sequences trimmed until the first position without missing data. The accurate localpair algorithm was used for all genes (*Katoh & Standley, 2013*), with the exception of ITS1, which had a single peak, where the globalpair algorithm was applied as it is optimized for gappy sequences. The dataset was concatenated using FASconCAT v1.1 (*Kück & Meusemann, 2010*), and a partitioned phylogenetic tree was obtained using IQ-tree v1.6.10, by applying 1,000 fastbootstrap replications. Model determination was done automatically by IQ-tree, and it included TIM2+I for ITS1, JC+I for 18S, TN+R2 for 16S and HKY+I for COI. The congruence between these genes, and in the individuals used has previously been determined in *Cerca et al. (2020b)* and *Cerca et al. (2020a)*.

## De novo RAD assembly

All bioinformatic work is available in https://github.com/jcerca/Papers/tree/main/Stygocapitella_PeerJ. Since no reference genome is available for *Stygocapitella*, we used the *de novo* assembly approach implemented in Stacks v2.2 to identify RAD loci (*Rochette & Catchen, 2017*; *Rochette, Rivera-Colón & Catchen, 2019*). The first module of Stacks, 'process_radtags', was executed using flags tailored to improve data quality (-clean,

-quality, -rescue). To optimize Stacks' parameters, we ran the *de novo* pipeline repeatedly using different -M (mismatch between stacks within individuals) and -n (mismatches between stacks between individuals) values, as suggested by best practices (*Paris, Stevens & Catchen, 2017*). The total number of loci resulting from different -M -n values were plotted and analyzed, selecting -M 3 and -n 3 for the final dataset. Populations, as required for the population map, were defined based on the species and sampling site (total of 22 populations—Table S1).

Since we observed a considerably high level of missing data in the dataset (>90% missing data), we tested and implemented a new method to improve RADseq datasets. Missing data is especially problematic in RADseq as it can lead to erroneous inference of population-genetic parameters (*Arnold et al., 2013*; *Gautier et al., 2013*; *Hodel et al., 2017*). However, applying stringent filtering for missing data has been shown to prune parsimonious-informative loci, with best-practices suggesting non-conservative pruning of the data (*Huang & Lacey Knowles, 2016*; *Lee et al., 2018*; *Crotti et al., 2019*). To mitigate missing data while avoiding stringent filtering, we applied a novel procedure, which allowed us to retrieve more loci from our data (*Cerca et al., 2021*). In brief, we ran Stacks for every population present in the population map—22 times in total (Table S1)—thus lowering phylogenetic distance in the dataset. Since phylogenetic distance (biological origin) and artefacts in generating and processing data can lead to allelic dropout (*O'Leary et al., 2018*), lowering phylogenetic distance will isolate dropout caused by artefacts in library preparation (e.g., DNA size-selection, low DNA concentrations, poor digestion), and loss of information due to whole genome amplification (*De Medeiros & Farrell, 2018*). For each population, we identified samples with >45% missing loci and removed these from a final analysis (*hereafter* the *clean* dataset). To evaluate how the optimization impacted the final number of loci we compared the number of loci and missingness in the dataset before cleaning (hereafter the *uncleaned* dataset) and in the cleaned dataset using the values of -r 25 (a locus has to be present in at least 25% of individuals comprising a population to be considered) and -p 4 (a locus has to be present in at least 4 populations to be considered). Finally, we included a technical replicate in the dataset (individuals 222 01 and 222 01R), and checked whether it was coherently placed in all the analyses.

## Population genomics and phylogenomics

From the clean dataset, we extracted a single nucleotide polymorphism-only dataset (SNPs; hereafter *variant* dataset) and an all-sites dataset (containing non-variant and variant positions; hereafter *all-sites* dataset). Separating the data in these two datasets is necessary to meet the assumptions of some statistical tests which may require the presence of non-variant positions to calculate ratios of variant and non-variant sites. The variant dataset was pruned by selecting -r 50 (a locus has to be present in at least 50% of individuals comprising a population), and -p 8 (a locus has to be part in at least 8 populations) as loci cut-offs, using the 'populations' program included in Stacks, resulting in 4,737 RAD-loci. After this initial round of cleaning, we used vcftools v0.1.13 (*Danecek et al., 2011*) to further prune the dataset for 5% minimum allele frequency (–maf), and for mean loci coverage values between 10–100 (–min-meanDP 10 –max-meanDP 100) and

removed 12 individuals which had missingness above >90% (–missing-indv; Table S2). The combination of coverage filters together with the -M -m optimization procedure mentioned above optimizes the generation of RADseq loci by removing loci which may artificially come together (i.e., repetitive regions). Finally, to decrease the effect of physical linkage in the data, we used a custom BASH script which kept only one polymorphism (SNP) per RAD locus, resulting in a final dataset of 3,428 SNPs. Using this dataset, we assessed genetic variation by means of a principal component analysis (PCA), a multi-dimensional scaling (MDS) analysis and an ADMIXTURE analysis. PCA and MDS are model-free approaches to estimate population structure, being complementary as PCA assumes 'mean values' for missing data (i.e., dragging individuals with high missingness to the center) whereas MDS does not. PCA was computed using the R package Adegenet (*Jombart & Ahmed, 2011*) and MDS with plink v1.9 (*Chang et al., 2015*). ADMIXTURE is a model-oriented approach to determine population structure based on the presence/absence of heterozygotes (*Alexander, Novembre & Lange, 2009*). We ran ADMIXTURE assuming 1-6 clusters (K), running a total of 5 replicates for each K, and determined the best K by estimating the cross-validation error (Fig. S1). Considering the potential for admixture of individuals in sympatry (Fig. 1), we used $f3$ statistics, included as part of the Admixtools package, as a direct test for detecting hybridization (*Patterson et al., 2012*; *Peter, 2016*). These statistics consists of a 3 populations test where a focal population is derived from admixture between the other two populations. When this score is negative, it suggests that admixture likely has occurred. We estimated errors and confidence intervals on the $f3$ statistics by partitioning the dataset into blocks and applying a jackknife bootstrapping. Finally, we inferred species-level divergence by estimating Weir and Cockerham's $F_{ST}$ using vcftools (*Weir & Cockerham, 1984*; *Danecek et al., 2011*).

The all-sites dataset was obtained by extracting FASTA sequences from Stacks. To run the phylogenomic analysis, we wrote a Perl script to reorganize the data into loci, concatenated all loci in a supermatrix using FasConCat-G (*Kück & Longo, 2014*) and ran a partitioned-tree using IQ-tree v1.6.10 (*Nguyen et al., 2015*) specifying 1,000 fast bootstrap replications (*Chernomor, Von Haeseler & Minh, 2016*; *Hoang et al., 2017*), and locus-specific models which were determined as part of the run (*Kalyaanamoorthy et al., 2017*). To explore the effect of missing data on the tree topology, we ran BaCoCa (*Kück & Struck, 2014*), a pipeline which provides summary-level statistics on the concatenation matrix and tree, such as the % of positions with missing data shared by a pair of individuals. Pairwise positive overlap values were plotted to the tree topology using the R package ape (*Paradis & Schliep, 2018*). Additionally, we ran an Unweighted Pair Group Method with Arithmetic mean (UPGMA) tree using only the average % of pairwise shared data per individual (i.e., pairwise percentage of shared data between taxa which do not have an indel, ambiguous character state, or a missing character state). The UPGMA tree was run to understand whether taxa were grouped based on the overall pattern of missing data. Additionally, since RADseq loci (represented by allele 0 and allele 1) are not phased and since the labelling of 0 and 1 are arbitrary, we obtained a consensus sequence for each individual. This was done by running the consambig (consensus ambiguous) module included in the EMBOSS pipeline (*Rice, Longden & Bleasby, 2000*). We did a species network analysis using SPLITSTREE v4

(*Huson & Bryant, 2006*) to complement phylogenetic inference since the network does not enforce dichotomous branching at each node.

To gauge population-level patterns and diversity, we selected loci from the all-sites dataset without missing-data at the population-level and estimated summary statistics including nucleotide diversity ($\pi$), Waterson's estimator of genetic diversity (S) and Tajima's D using DNAsp v6 (*Rozas et al., 2017*). This dataset was not pruned for minimum allele frequency. The selection of sites without missing data is grounded on best-practices as missing data can lead to the under- or overestimation of some of these parameters (*Arnold et al., 2013*). Importantly, we selected only populations where 3 or more individuals (i.e., >5 'loci') had data available (Table S2).

Finally, we evaluated various demographic scenarios using fastsimcoal2, using the same dataset for the previous analysis which included running fastsimcoal2 (*Excoffier et al., 2013*). Fastsimcoal2 uses the site-frequency spectrum (SFS) and a coalescent-simulation framework based on an arbitrary user-defined scenario to infer population sizes, strength of gene flow and times of coalescence. To assess these models we calculated AIC and likelihood. Likelihood is calculated by running the 'best parameters' for each specified scenario multiple times and obtaining the distribution of likelihood estimates. AIC was calculated using a script available in https://speciationgenomics.github.io/fastsimcoal2/. To implement these simulations, we used the phylogeny obtained with *Stygocapitella subterranea* and *S. westheidei* as sister species, and *S. josemariobrancoi* as sister to the remaining two. We defined the following models: no gene flow, ancient gene flow (i.e., between *Stygocapitella josemariobrancoi* and the stem lineage of *S. subterranea* and *S. westheidei*), geographic gene flow (i.e., similar as the ancient gene flow, but also with modern gene flow between the sympatric *Stygocapitella subterranea* and *S. josemariobrancoi*; note we refer to 'modern' as opposed to 'ancient', that is, in existing lineages), modern gene flow (gene flow between all three modern species), all gene flow (gene flow between all three species and the two ancestral lineages), modern gene flow only between *S. josemariobrancoi* and *S. subterranea*, modern gene flow only between *S. josemariobrancoi* and *S. westheidei*, modern gene flow only between *S. westheidei* and *S. subterranea*, gene flow in ancestral times and between *S. subterranea* and *S. westheidei*, and ancient gene flow and modern gene flow between *S. josemariobrancoi* and *S. westheidei*. When included in the model, gene flow was moduled as asymmetric. Each model was run 10,000 times with an assumed mutation rate of 1.2e−8, and the best fitting scenario was evaluated using likelihood, by running it 100 times.

## RESULTS

### Tree of selected molecular markers

We compiled a dataset comprising 4,147 bp (the COI fragment consisted of 629 bp, 16S of 548 bp, ITS1 of 1,150 and 18S of 1,817 bp), from which 716 sites were phylogenetic-informative sites. From a total of 69 individuals, we obtained 67 16S sequences, 61 COI sequences, 28 18S sequences and 31 ITS1 sequences. Every species was recovered as monophyletic (Fig. 2A) with bootstrap support values of 100 for *S. josemariobrancoi*, of 86 for *S. subterranea* and of 100 for *S. westheidei* (Fig. 2A). The retrieved tree topology
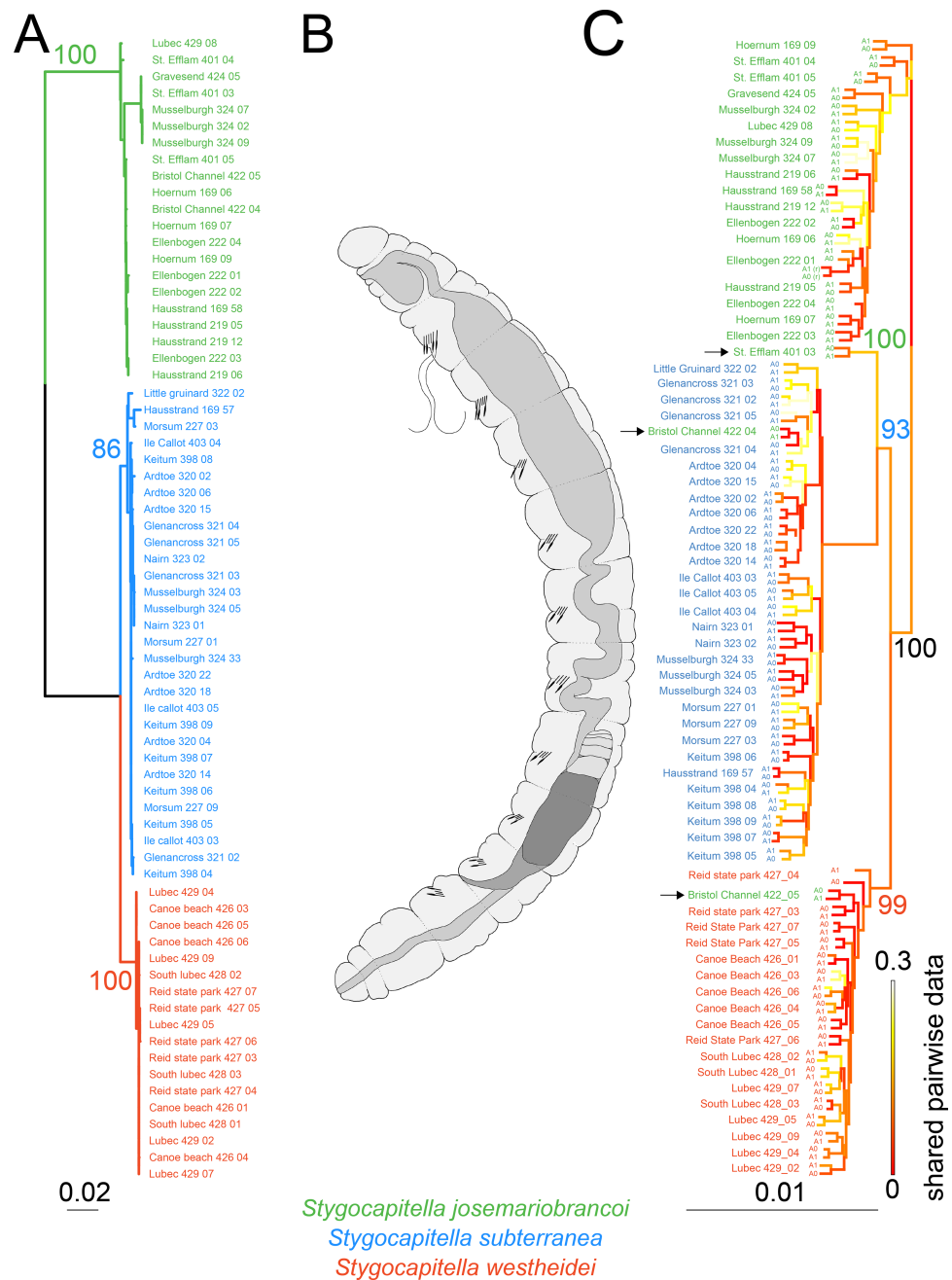
**Figure 2  Phylogenetic reconstruction and Scanning Electron Microscopy images of *Stygocapitella*.**
(A) Maximum likelihood phylogeny of a concatenated, partitioned dataset (COI, 16S, 18S and ITS1), with scale provided on the bottom. Coloration follows species, with blue representing *Stygocapitella subterranea*, green representing *S. josemariobrancoi*, and orange *S. westheidei*. Bootstrap support for the branches representing species are provided on top of the branches. Every species is retrieved as monophyletic. (B) Drawing of the *Stygocapitella westheidei, S. subterranea, S. josemariobrancoi* morphotype. For more information on the classification and distinction of the morphotypes see *Cerca et al. (2020a)* and *Cerca et al. (2020b)*. (C) Phylogenomic tree based on 4,737 RADseq loci. Bootstrap support is provided for the main branches.(continued on next page...)

Full-size 🖼 DOI: 10.7717/peerj.10896/fig-2

**Figure 2 (…continued)**
Coloration follows species with blue representing *Stygocapitella subterranea*, green representing *S. josemariobrancoi*, and orange *S. westheidei*. Three specimens, denoted by arrows, are identified as 'paraphyletic', including Bristol Channel 422 04 (identified as *S. josemariobrancoi*, nested within *S. subterranea*), Bristol Channel 422 05 (identified as *S. josemariobrancoi*, nested within *S. westheidei*) and St. Efflam 401 03 (identified as *S. josemariobrancoi*, nested sister to *S. subterranea*). The tree topology is coloured with shared pairwise data as estimated by BaCoCa. Allele 0 and allele 1 are displayed for all specimens. Specimen 222_01 is a technical replicate and is therefore represented twice. Shared pairwise data was calculated by integrating BaCoCa's information on pairwise missing data.

includes *S. westheidei* and *S. subterranea* as sister species. *Stygocapitella josemariobrancoi* as sister to the clade comprising *S. subterranea* and *S. westheidei*. Single gene trees show concordance between markers (Figs. S2–S5)

## Genomic dataset

We obtained a total of 1,277,919,764 sequencing reads from two Illumina HiSeq4000 lanes. After demultiplexing and cleaning the data with process_radtags, we retained 899,112,800 reads (107,830,588 reads were discarded for having ambiguous barcodes, 270,174,154 for ambitious RADtags, and 802,222 for low quality). When comparing the clean and unclean dataset, the approach to lower allelic dropout yielded a substantial increase of the number of loci. In detail, after running Stacks for each population individually, we removed 17 *S. subterranea* out of a total of 47, 7 *S. westheidei* out of 24, and 16 out of 50 for *S. josemariobrancoi* (roughly ∼33% of the dataset, Table S2). The uncleaned dataset yielded 179,742 loci (55,037,190 sites including 628,031 variant sites), whereas the cleaned data yielded 368,696 loci (112,725,106 sites including 1,100,431 variant sites). When pruned with common denominators included in the module Populations (-r 50 -p4), the unclean dataset yielded 109,369 loci, and the cleaned yielded 272,134 loci. Individual-level missing data was reduced from 84.89% in the uncleaned dataset to 80.79% in the cleaned dataset. We validated this approach by comparing PCA, MDS and phylogenomic trees using both cleaned and the uncleaned dataset (see *Cerca et al. (2021)*). A comprehensive investigation of this strategy including additional datasets will be published in a separate paper (*Cerca et al., 2021*).

## Genomic trees and networks

The phylogenomic tree (Fig. 2C) shows a slightly different topology from the tree obtained with selected molecular markers. The branches representing *Stygocapitella subterranea*, *S. westheidei*, and *S. josemariobrancoi* have a bootstrap support of 93, 99 and 100, respectively. The tree topology is broadly similar to the selected marker phylogenetic tree, with *S. josemariobrancoi* being sister to the clade comprising *S. subterranea* and *S. westheidei*. However, strictly speaking, none of the species is recovered as monophyletic, since three individuals identified as *S. josemariobrancoi* are not placed with *S. josemariobrancoi*. Specifically, 422 04 from Bristol Channel nests within *S. subterranea*, 422 05 from Bristol Channel nests within *S. westheidei* and 401 03 from St. Efflam is positioned as sister to *S. subterranea* (Fig. 2C, individuals denoted by arrows). Importantly, mapping of shared pairwise data in the tree topology suggests that these trends are not driven by missing

data, since the branches representing the three aforementioned individuals do not exhibit elevated levels of missing data (Fig. 2C). The UPGMA tree, which is solely built on a pairwise matrix of missing-data, shows that *S. subterranea* and *S. josemariobrancoi* are generally separated, interlaced by individuals from *S. westheidei* (Fig. S6). While this suggests that the three species have different amounts of missing data, the fact that individuals are generally mixed suggests that missing data is not driving phylogenetic reconstruction. For example, the three individuals resulting in a paraphyletic reconstruction are not placed closely to their sister taxa in the tree, therefore indicating that missing data has no influence in the paraphyletic position of these individuals (Fig. S6). Finally, the generated phylogenomic consensus tree shows a similar topology to the that in Fig. 2 (Fig. S7). The three samples causing paraphyly of the lineages in the phylogenomic tree are placed within *S. subterranea* (Bristol Channel 422 04), as sister to the lineage *S. josemariobrancoi* and *S. westheidei* (St. Efflam 401 03), and as the first branch of *S. josemariobrancoi* (Bristol Channel 422 05).

In the phylogenetic network, *Stygocapitella westheidei* is separated from the remaining two species, occupying a separate and relatively compact area of the network. *Stygocapitella subterranea* is mostly confined to one small section of the network, however, three individuals are very close to the center of the network (398 04, 398 08 and 398 09 from Keitum; Fig. 3). In line with the results from the phylogenomic tree, Bristol Channel 422 04, which is identified as part of *S. josemariobrancoi*, is nested within *S. subterranea* in the network. *Stygocapitella josemariobrancoi*, on the other hand, is clearly stretched and set apart in the network, occupying a large area (Fig. 3). While most individuals are nested within a condensed and remote portion of the network, the individuals Bristol Channel 422 05, St Efflam 401 03 and 401 04 lie in an intermediate position between the center of the network and the majority of individuals from *S. josemariobrancoi* (Fig. 3). This is broadly in agreement with the phylogenomic tree, which shows Bristol Channel 422 05 nested with *S. westheidei* (Fig. 2C) and St. Efflam 401 03 sister to *S. subterranea* (Fig. 2C). The distance between most individuals belonging to *S. josemariobrancoi* and the rest of the network (centre of the network, and the two remaining species) suggests there is a greater degree of differentiation in this species.

## Population structure, differentiation and summary statistics

The PCA separates the three species across the first two principal components, (which together explain 30.4% of the variance; Fig. 4A). Three individuals stand out, including Bristol Channel 422 04 (labelled as *S. josemariobrancoi*) which is placed closely with *S. subterranea* individuals, Bristol Channel 422 05 which occupies an intermediate position between *S. westheidei* and *S. josemariobrancoi*, and Lubec 428 02 which is relatively distant from the remaining *S. westheidei* individuals. The multi-dimensional scaling plot, which is less affected by missing data, separates the species into three distinct clusters (MDS; Fig. 4B). In coherence with the previous analyses, we detect several taxa with intermediate positions: Bristol Channel 422 04 (labelled as *S. josemariobrancoi*) is closer to the *S. subterranea* cluster than to the *S. josemariobrancoi*; Hoernum 169 09 (labelled as *S. josemariobrancoi*) in an intermediate position between these two species; Bristol Channel 422 05, St. Efflam 401 04,
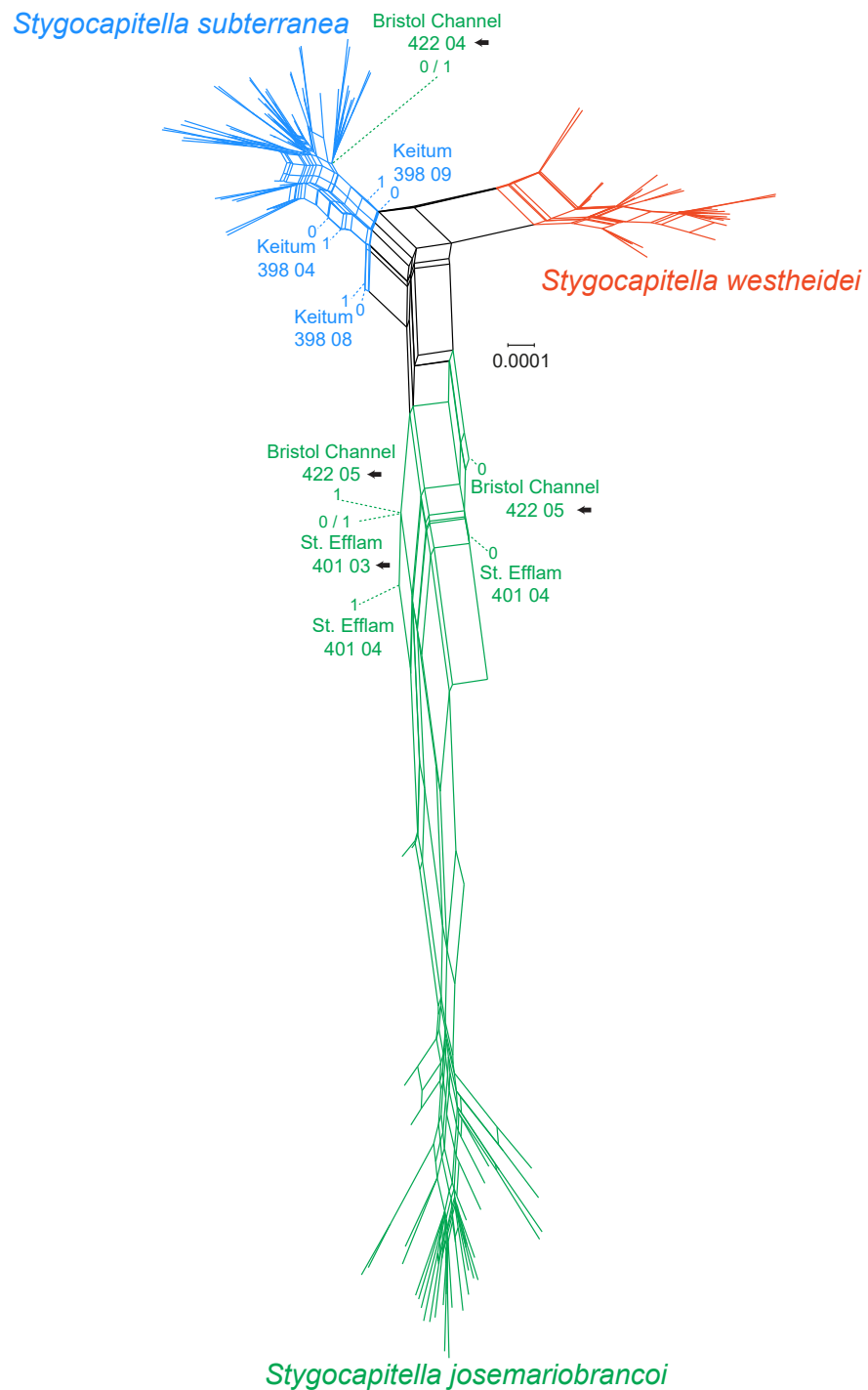
**Figure 3** **Phylogenetic network based on 4,737 RADseq loci.** Coloration follows species with blue representing *Stygocapitella subterranea*, green representing *S. josemariobrancoi*, and orange representing *S. westheidei*. Specimens with arrows represent specimens which were as paraphyletic in the phylogenomic tree (Fig. 2C). *S. josemariobrancoi* is clearly stretched, indicating a greater differentiation from the remaining two species. In congruence with the phylogenomic analysis, Bristol Channel 422 04 is nested within *S. subterranea*.

Full-size ☑ DOI: 10.7717/peerj.10896/fig-3
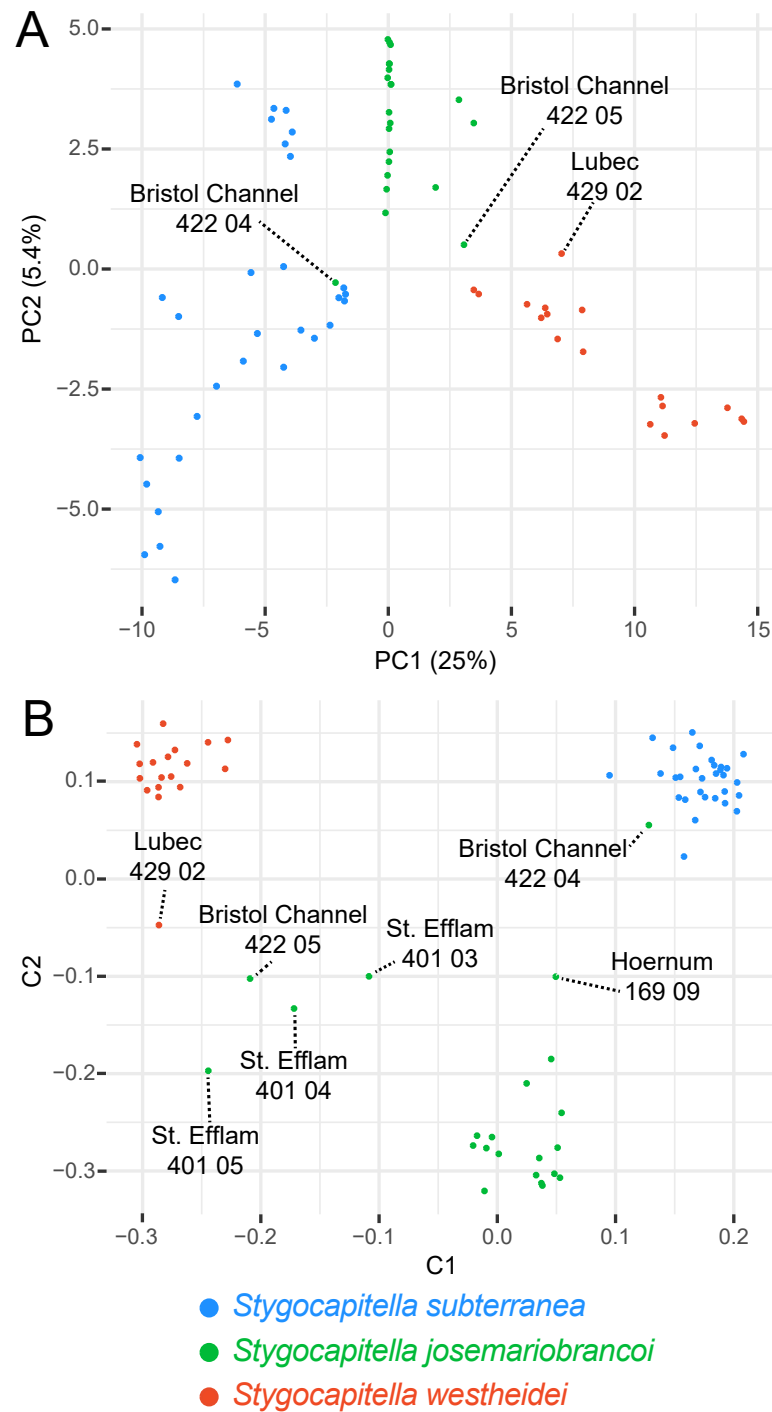
**Figure 4** **Principal Component Analysis (A) and Multi-dimensional scaling (B) of 3,428 SNPs.** The percentage of explained variation is displayed along with the axis for each PC. Lineage (species) are given in different colours. Specimens with 'intermediate positions' are highlighted both analyses, indicating potential shared generated variation between specimens.

Full-size 🖼 DOI: 10.7717/peerj.10896/fig-4
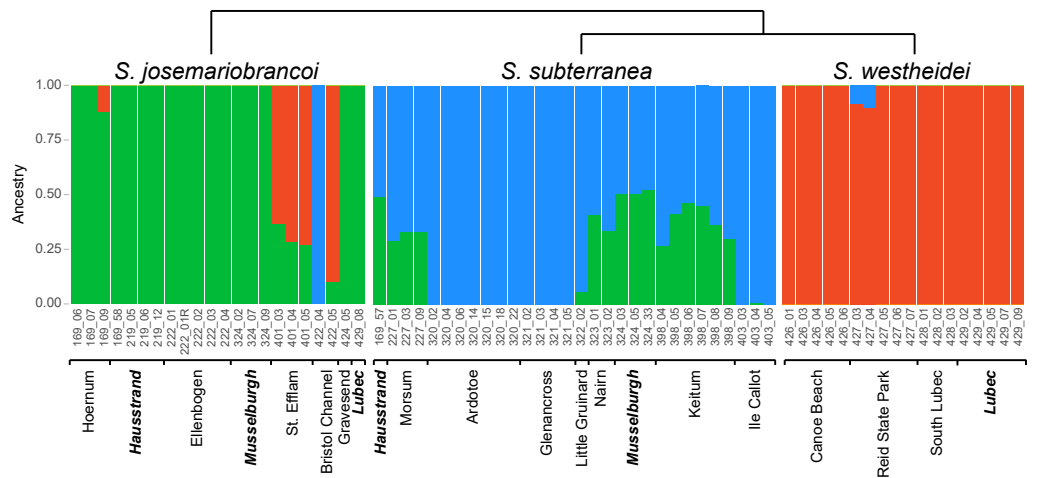
**Figure 5** **ADMIXTURE analysis of 3,428 SNPs shows shared genetic variation.** *Stygocapitella josemariobrancoi*, *S. subterranea* and *S. westheidei* are plotted consecutively from left to right. Populations and specimen-ids are denoted at the bottom, with sympatric-populations in bold and italics. The cladogram follows the tree topology retrieved in Figs. 2A and 2B.

Full-size 🔍 DOI: 10.7717/peerj.10896/fig-5

401 03, 401 05 (all labelled as *S. josemariobrancoi*) are found in an intermediate position between *S. josemariobrancoi* and *S. westheidei*; Lubec 429 02 (labelled as *S. westheidei*) is also distant from the *S. westheidei* cluster, being relatively close to Bristol Channel 422 05 (Fig. 4B).

The ADMIXTURE analysis confirms shared genetic signal among species. The most supported cluster size was $K = 3$ (Fig. S1) and is plotted in Fig. 5. In agreement with the phylogenetic network, the MDS and the PCA, *S. westheidei* is the species with the least amount of admixture, with only 2 individuals sharing a relatively low degree of ancestry with *S. subterranea*. A majority of *S. subterranea* individuals (17 out of 30) share genetic variation with *S. josemariobrancoi*. *S. josemariobrancoi* has 5 individuals which are admixed from *S. westheidei*, and an individual (Bristol 422 04) identified as having a *S. subterranea* ancestry. Notably, individuals from sympatric areas and belonging to *S. josemariobrancoi* and *S. westheidei* (Hausstrand, Musselburgh, Lubec) show no signal of shared ancestry. However, five *S. subterranea* individuals with shared ancestry belong to two sympatric sites (Hausstrand and Musselburgh). $f3$ statistics were positive, thus suggesting that the observed patterns of admixture are unlikely to be due to recent admixture (Table 1). Notably, two out of three scenarios retrieved Z scores >3 (threshold used for significance), including *S. subterranea* and *S. josemariobrancoi* as source and *S. westheidei* as target, and *S. josemariobrancoi* and *S. westheidei* as sources and *S. subterranea* as target.

Notably, $F_{ST}$ estimates among species are high, thus indicating isolation. Pairwise $F_{ST}$ comparisons were lower between *S. josemariobrancoi* and either of the remaining species: 0.53 for *S. josemariobrancoi* vs *S. subterranea*, 0.492 *S. josemariobrancoi* vs *S. westheidei* and 0.664 for S. *subterranea vs S. westheidei* (Table 2).

Summary statistics suggest that populations of *S. josemariobrancoi* have a higher degree of genetic variation. Waterson's estimate, S, ranges between 2.11–2.39 in *S. subterranea*

**Table 1** **$f3$ statistics testing for hybridization between lineages.** Each row represents a scenario where two species are the source for admixture, and the third species is the target of hybridization. A $f3$ statistic, the standard error (SE) and a Z-score value calculated with jackknife is provided for each scenario based.

| Source 1 | Source 2 | Target | $f3$ | SE | Z |
|---|---|---|---|---|---|
| S. subterranea | S. josemariobrancoi | S. westheidei | 0.91 | 0.08 | 11.424 |
| S. josemariobrancoi | S. westheidei | S. subterranea | 1.09 | 0.24 | 4.575 |
| S. westheidei | S. subterranea | S. josemariobrancoi | 0.25 | 0.10 | 2.605 |

**Table 2** **Weir-Cockherham Fst-estimate.** Estimates are provided in the lower part of the table, and the number of individuals included in the pairwise estimation is provided in the upper part of the table.

| Species | *S. subterranea* | *S. josemariobrancoi* | *S. westheidei* |
|---|---|---|---|
| S. subterranea | – | 52 | 48 |
| S. josemariobrancoi | 0.530 | – | 40 |
| S. westheidei | 0.664 | 0.492 | – |

and between 1.81–2.35 in *S. westheidei*, and between 1.89–6.83 in *S. josemariobrancoi*, with Bristol Channel ($S = 4.92$), and Gravesend ($S = 6.93$; Table 3). This is similar when analysing nucleotide polymorphisms ($\pi$), with *S. subterranea* ranging between 0.002–0.0037, *S. westheidei* between 0.0021–0.0034, but *S. josemariobrancoi* ranging between 0.0026–0.0099 again with Gravesend and Bristol Channel as outliers (Gravesend $\pi = 0.0099$, Bristol Channel $\pi = 0.0086$; Table 3). Interestingly, the sympatric sites do not reveal any signal of higher polymorphism, as it would be expected in scenarios of on-going hybridization. For instance, $\pi = 0.0033$ and $S = 2.13$ for *S. subterranea* in Musselburgh, and $\pi = 0.0045$ and $S = 2.49$ for *S. josemariobrancoi* in Musselburgh. Hausstrand, for which we were only able to obtain data for *S. josemariobrancoi* retrieved a $\pi = 0.0038$ and a $S = 2.55$, while the population Lubec of *S. westheidei* shows $\pi = 0.0029$ and $S = 2.15$. Tajima's D excludes the possibility for bottlenecks, as none of the populations exhibits significant Tajima's D (i.e., values below -2 or above +2; Table 3).

Simulation of demographic scenarios suggests admixture may have happened in ancestral branches (Fig. 6). All the best supported scenarios are provided in https://github.com/jcerca/Papers/blob/main/Stygocapitella_PeerJ/11_fscScenarios/ and as supplementary. Three of the top four most supported scenarios using likelihood suggest ancient admixture: geographic gene flow (i.e., gene flow between the two ancestral branches and afterwards between *S. subterranea* and *S. josemariobrancoi*), ancient gene flow (i.e., gene flow between the two ancestral branches), ancient gene flow and between western lineages (i.e., gene flow between the two ancestral branches and afterwards between *S. westheidei* and *S. josemariobrancoi*). The remaining scenario suggests no gene flow. The most supported of these scenarios was the geographic gene flow, which includes coalescent times of 411,402 generations for the first coalescent event (presumably, *Stygocapitella* has a generation time of a single year), and 21,680,033 for the second coalescent event. The ancient gene flow scenario, which was the second most supported, includes estimates of 585,534 generations for the first coalescent event and 16,749,615 generations for the second coalescent event. The no gene flow scenario is the 3rd most supported scenario, but the

**Table 3  Summary statistics for the various analysed populations.** For each site we provide the number of specimens and chromossomes, number of loci considered, S (Waterson's estimate), the averaged $\pi$, and the averaged Tajima's D. Analysed populations include those with >5 chromossomes, with the exception of Plymouth which had no samples with no missing data.

| Species | Site | Number of specimens (chromossomes) | Number of loci analyzed (without missing data) | S (Waterson's estimate) | Averaged $\pi$ | Averaged Tajima's D |
|---|---|---|---|---|---|---|
| *Stygocapitella subterranea* | Ardtoe | 7 (14) | 122 | 2.1066 | 0.0020 | −0.1843 |
| | Glenancross | 4 (8) | 658 | 2.1763 | 0.0029 | 0.1123 |
| | Hausstrand | 1 (not analyzed) | – | | – | |
| | Ile Callot | 3 (6) | 1301 | 2.3912 | 0.0037 | 0.1958 |
| | Keitum | 6 (12) | 40 | 2.525 | 0.0030 | −0.1858 |
| | Little Gruinard | 1 (not analyzed) | – | – | – | – |
| | Morsum | 3 (6) | 1923 | 2.1482 | 0.0032 | 0.1326 |
| | Musselburgh | 3 (6) | 1166 | 2.1329 | 0.0033 | 0.1571 |
| | Nairn | 2 (not analyzed) | – | – | – | – |
| *Stygocapitella josemariobrancoi* | Bristol Channel | 2 (not analyzed) | – | – | – | – |
| | Ellenbogen | 4 (8) | 403 | 1.6377 | 0.0024 | −0.1993 |
| | Gravesend | 2 (not analyzed) | – | – | – | – |
| | Hausstrand | 4 (8) | 154 | 2.5519 | 0.0038 | −0.3862 |
| | Hoernum | 3 (6) | 261 | 2.6858 | 0.0046 | −0.0281 |
| | Lubec | 1 (not analyzed) | – | – | – | – |
| | Plymouth | 0 (not analyzed) | – | – | – | – |
| | Musselburgh | 3 (6) | 414 | 2.5918 | 0.0045 | −0.0356 |
| | Saint Efflam | 3 (6) | 572 | 3.3969 | 0.0059 | 0.1702 |
| *Stygocapitella westheidei* | Canoe Beach | 5 (10) | 577 | 1.8128 | 0.0021 | −0.3654 |
| | Lubec | 5(10) | 89 | 2.1461 | 0.0029 | −0.1044 |
| | Reid State Park | 5 (10) | 519 | 2.3468 | 0.0024 | −0.5266 |
| | South Lubec | 3 (6) | 524 | 2.0115 | 0.0034 | 0.1723 |

first and second coalescent events are suggested to have occurred 451 and 12,834 years ago, being clearly at odds with previous evidence suggesting these lineages diverged millions of years ago (*Struck et al., 2017*; *Cerca et al., 2020a*). The fourth most supported scenario, ancient gene flow and gene flow between western lineages, has coalescent times of 3,566,294 and 18,353,555 (Fig. 6). Finally, while the AIC assessment provide slightly different results (Fig. S8), the second and third most supported scenarios are the geographic gene flow and ancient gene flow; being thus in agreement with the likelihood results.

## DISCUSSION

Morphological similarity through extended periods of times, or stasis, has been hypothesized to occur under three possible scenarios underlied by genetics: homogenous genetic variation (due to e.g., ILS, hybridization), genetic constraints (e.g., pleiotropy), and lack of genetic variation (e.g., bottlenecks and founder effects) (*Futuyma, 2010*). While our sampling design does not account for genetic constraints, we study the evolutionary history of *Stygocapitella* species seeking to determine signals of loss of genetic variation or shared
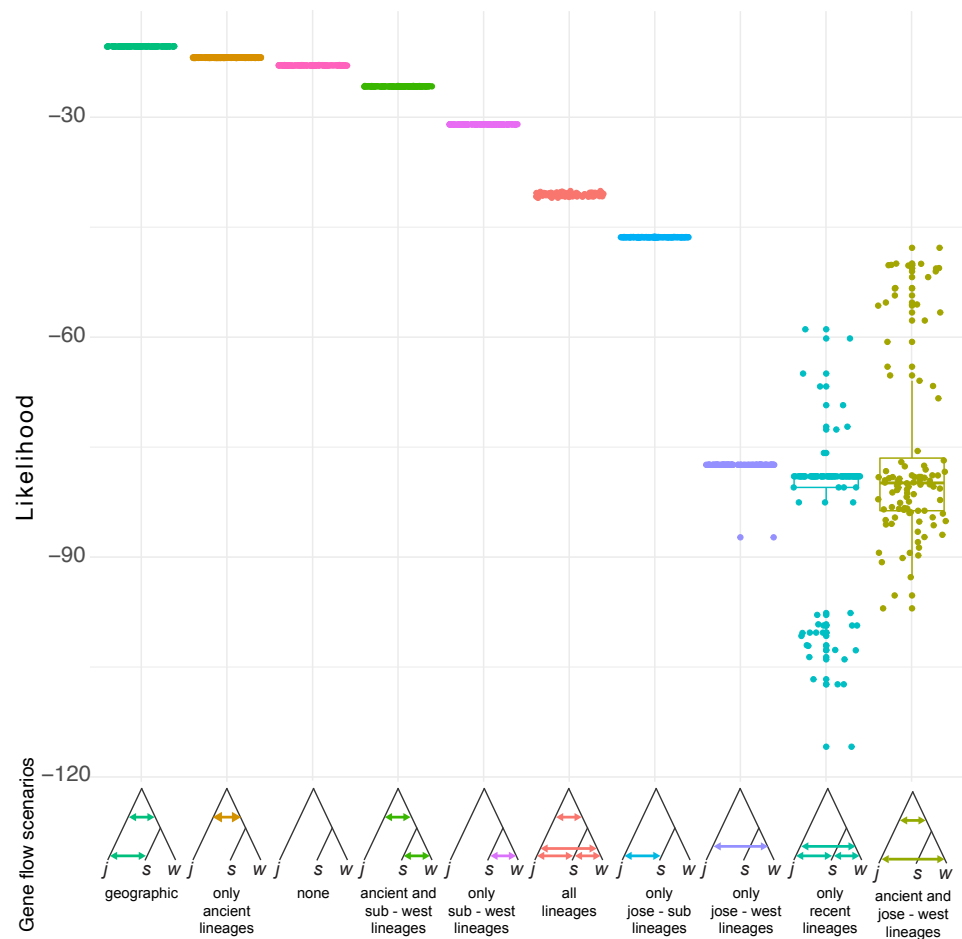
**Figure 6  Demographic scenarios considered.** The likelihood of different demographic scenarios is displayed on the Y axis. Based on the estimated phylogeny (Fig. 2), we modelled scenarios for (from left to right): (1) geographic gene flow (gene flow between *S. josemariobrancoi* and the ancient lineage, and *S. josemariobrancoi* and *S. subterranea*); (2) ancient gene flow (gene flow between *S. josemariobrancoi* and the lineage before the *S. subterranea* and *S. westheidei* split); (3) no gene flow at all; (4) ancient gene flow and gene flow between *S. subterranea* and *S. westheidei*; (5) gene flow only between *S. westheidei* and *S. subterranea*; (6) gene flow in every possible branch; (7) gene flow in sympatric, European linages; (8) gene flow between *S. josemariobrancoi* and *S. westheidei*; (9) gene flow between currently existing lineages; (10) ancient gene flow with gene flow between *S. josemariobrancoi* and *S. westheidei*.

Full-size 🖼 DOI: 10.7717/peerj.10896/fig-6

genetic variance. We find that the three morphologically similar *Stygocapitella* species herein studied share genetic variation and exclude the possibility of recent bottlenecks or recent admixture. Demographic and admixture analysis reveal signatures of ancestral admixture and incomplete lineage sorting during the divergence of these three species. We discuss the possible implications of these processes to the genomic underpinnings of indistinguishable morphology among cryptic species.

## Whole genome amplification and the generation of RADseq data

We show that WGA combined with RAD sequencing may become an important tool for microscopic eukaryote genomics. RADseq library preparation typically requires 200–500 ng of DNA per individual, yet DNA extraction of a single *Stygocapitella* individual typically yields 10–100 ng of DNA, thus representing a challenge to obtain genome-level data. One potential solution is to pool individuals. However, pooling may not be ideal when dealing with morphologically similar species, especially when they occur in sympatry, as observed in *Stygocapitella*, since the identification of individuals based solely on morphology may be impossible. Arguably, one of the major advantages of RADseq is to open up population genomics and phylogenomics as approaches for non-model systems at an affordable cost. In such systems, experimental designs may benefit from the inclusion of the largest number of individuals possible, which encompass the whole spectrum of populations or species to determine species boundaries, phylogeography, population structure and the phylogeny. Pooling individuals from different species or populations together may lead to an incorrect inference of the phylogeny when species boundaries are not known, but also skew allelic variation. Consequently, pooling of individuals may result in difficulties during data processing and interpretation, and may require extra efforts such as barcoding of individuals before pooling. In the view of these challenges, we optimized and applied a whole-genome amplification protocol to obtain genome level data, thus confirming its potential for population genomic-inference and phylogenetics (*De Medeiros & Farrell, 2018*; *De Medeiros & Farrell, 2020*).

## Bottlenecks, meiofaunal dispersal and morphological similarity

The few phylogeographic studies available for meiofaunal organisms have generally detected founder effects and bottlenecks, and have discussed how colonization dynamics may be determined by a series of bottlenecks (*Casu & Curini-Galletti, 2006*; *Derycke et al., 2007*; *Andrade, Norenburg & Solferini, 2011*), however, we find no evidence for bottlenecks in this dataset when using summary statistics and Tajima's D. Evidence for the prevalence of bottlenecks in meiofauna has been further supported by experimental evidence which showed that colonization of new areas may be characterized by founder effects and bottlenecks, which, in turn, are expected to shape the genetic differentiation of meiofaunal populations (*Derycke et al., 2007*; *Derycke, Backeljau & Moens, 2013*). Broadly, this follows hypotheses on marine-invertebrate biogeography which suggests that repeated extinction and recolonization dynamics may be involved in shaping genetic differentiation in populations and species (*Andrade, Norenburg & Solferini, 2011*; *Derycke, Backeljau & Moens, 2013*). We find no indication of bottlenecks in *Stygocapitella,* as suggested by non-significant Tajima's D values (Table 3), even though that part of the sampling distribution of *Stygocapitella* included areas which were glaciated only 10,000 years ago (*Wares & Cunningham, 2001*). Two possibilities may explain this scenario and support our findings. First, the hypothesis that meiofauna disperse through a series of bottlenecks may require more evidence. Indeed, when scoring a total of 752 papers, we could only detect 48 studies focusing on biogeography and 7 on evolutionary biology (including population genetics) (*Cerca, Purschke & Struck, 2018*), thus suggesting that meiofaunal

biogeography is in its early days and more studies are needed for more solid conclusions. This hypothesis may also be at odds with evidence that meiofauna may be, indeed, good dispersers (reviewed in *Cerca, Purschke & Struck, 2018*). Even if dispersal is carried by only a limited group of individuals, concomitant with the idea of a founder effect and bottleneck colonization, the dispersal of more organisms from 'source populations' (multiple waves of dispersal) through time would eventually homogenize genetic variation in newly colonized areas. Second, to the best of our knowledge, this is the first work to focus on population-genomic level data in meiofauna. Typically, works have focused on sequencing a limited combination of mitochondrial and nuclear genes (e.g., *Derycke et al., 2005*; *Derycke et al., 2007*; *Derycke et al., 2008*; *Kieneke, Martínez Arbizu & Fontaneto, 2012*; *Leasi & Norenburg, 2014*). While non-recombining data, such as mitochondrial markers, may provide an ideal indication for the occurrence of bottlenecks, these effects should be confirmed on complementary genomic regions. For instance, using the 16S mitochondrial marker, we previously detected single haplotypes in populations of *S. subterreanea* or *S. josemariobrancoi*. These populations had a statistically significant Tajima's D (*S. josemariobrancoi* individuals from Bristol Channel; *S. subterranea* individuals from Glenancross) (*Cerca et al., 2020b*). While this pattern was in conflict with that of the nuclear ITS1 (*Cerca et al., 2020a*), it is further rejected when using genome-level data, which provides a more comprehensive, and independent, assessment of genomic variance. This suggests that the dispersal-by-bottlenecks idea in meiofauna warrants more data, and that biogeography of meiofauna will benefit from more genomic studies.

Overall, we suggest that morphological similarity in *Stygocapitella* is unlikely to result from the lack of standing genetic variation due to re-occurring bottlenecks. Under this hypothesis, it is expected that bottlenecks reduce genetic variation, which will in turn limit morphological evolution, thus leading to stasis (*Futuyma, 2010*). Given the lack of evidence for recent bottlenecks in *Stygocapitella*, this does not seem plausible. In addition to the evidence for the lack of recent bottlenecks, the fact that the remaining 8 *Stygocapitella* species live in similar habitats and are distributed throughout the world (*Cerca et al., 2020a*) indirectly suggests that bottlenecks may not be typical in the evolutionary history of the group.

## Evidence for incomplete lineage sorting and admixture

We find clear evidence for shared genetic variation in *Stygocapitella*. The most conspicuous evidence for this comes from the admixture analysis, which clearly demonstrates admixed populations in the three species (Fig. 5). This evidence is further supported by individuals with intermediate positions in the MDS—a test which is robust to missing data (Fig. 4). However, several evidences do not support a preponderant role of recent admixture. First, we obtained no evidence for admixture when using F-statistics, since we find only positive F-values (Table 1). Second, contrary to the expectation of ongoing gene flow, we do not observe higher levels of heterozygosity in sympatric populations (Lubec in the USA, Musselburgh in Scotland, Hausstrand in Germany; Table 3) where individuals of different species are found in the same sediment sample in close proximity (volume ranging from 50–500 $cm^3$). Third, admixture often generates incongruence between mitochondrial

and nuclear markers (*Melo-Ferreira et al., 2012*; *Sloan, Havird & Sharbrough, 2017*), which is not seen in single-marker trees (Figs. S2–S5). Fourth, models with exclusive recent admixture are generally poorly supported by the demographic analysis (Fig. 6). In contrast, three out of the four most supported demographic scenarios suggest ancient admixture, and one supported no gene flow at all (Fig. 6). The scenario with no gene flow inferred coalescent times of 451 and 13,834 generations or years (1 generation is expected to be 1 year, Günter Purschke pers. comm)] which are not compatible with estimates of the splitting age of the three *Stygocapitella* species (~5–30 million years ago; *Cerca et al., 2020b*). Given that reduced times of coalescence are a typical signature of simulations that do not account for gene flow, when it has occurred in empirical data (*Leaché et al., 2019*), it is likely that incomplete lineage sorting alone cannot explain the patterns of shared variation among *Stygocapitella* species. In other words, the demographic analysis supports a scenario that includes ancient admixture. The three scenarios with ancient admixture vary in the presence or absence of admixture after the second coalescence event (*S. subterranea* and *S. josemariobrancoi*): in one scenario, admixture is exclusive to the ancestral branch; in the remaining two, gene flow between *S. josemariobrancoi* and either *S. westheidei* or *S. subterranea* occur. Given the lack of support for on-going gene flow between species by the $F_{ST}$, summary statistics, and F-statistics (Tables 1–3, Figs. S1–S5), admixture may have occurred immediately before or after the speciation event of *S. westheidei* or *S. subterranea*, but not in recent times (i.e., the last generations). Furthermore, the occurrence of ancient admixture can affect the inference of recent admixture when not take the phylogeny into account (*Malinsky et al., 2018*; *Ferreira et al., 2020*), this may explain the incongruence between some of our analysis. Therefore, while the demographic analysis suggests the occurrence of admixture among *S. josemariobrancoi* and the other species, future studies are necessary to confidently dissect and determine the role of recent gene flow in the system with independent analyses. For example, these studies will benefit from using whole-genome data to determine whether interspecific divergence in regions of the genome show gene-species tree discordance, thereby dissecting ILS and recent hybridization (*Joly, McLenachan & Lockhart, 2009*; *Giska et al., 2019*). Also, the demographic analysis favouring a preponderant role of ancient admixture does not exclude the occurrence of ILS, and the beforementioned approach would also allow to clarify the relative contribution of ILS and gene flow to shared patterns of variation among species. In sum, to the extent that we can speculate, our data suggests that shared genetic variance is more likely explained by an evolutionary history including incomplete lineage sorting and ancient geneflow.

Evidence for ancient admixture or incomplete lineage sorting is further seen in the phylogenomic analysis. Phylogenetic approaches, which seek to reconstruct the evolutionary history of lineages, often fail to resolve the evolutionary history and the 'true tree-topology' when the taxa in question have high rates of incomplete lineage sorting or admixture (*Kubatko & Degnan, 2007*; *Degnan & Rosenberg, 2009*), but incongruence may also result from tree-building errors, paralogy or horizontal gene transfer (*Scornavacca & Galtier, 2017*). We discard tree-building errors based on the following evidence. First, individuals with intermediate positions in the PCA and MDS correspond to those causing

paraphyly in the trees. Second, when exploring the effects of missing data through (a) labelling the tree with % of missing data (Fig. 2); and (b) constructing a cladogram based only on the shear % of missing data (i.e., UPGMA tree, Fig. S6); we find the placement of intermediate individuals is not guided by missing data. Should missing data determine their placement, we would expect these specimens to nest in close proximity in the UPGMA tree. Finally, in the phylogenetic network, the individuals which are also far removed from the remaining *S. josemariobrancoi* individuals, occupying central positions or being paraphyletic in the phylogenetic network correspond to those paraphyletic in the tree and in intermediate positions in the PCA and MDS.

In sum, incomplete lineage sorting and ancient hybridization are known to contribute to levels of shared variation among species (*Pease et al., 2016*; *Malinsky et al., 2018*; *Edelman et al., 2019*), even at deep evolutionary levels (*Song et al., 2015*; *Suh, Smeds & Ellegren, 2015*). The development of tools which employ the substantial amount of modern genomic data has allowed separating cases of ILS and ancient admixture, showing that ancient hybridization can have a strong impact in the levels of shared variation among species complexes (*Malinsky et al., 2018*; *Li et al., 2019*; *Taylor & Larson, 2019*; *Ferreira et al., 2020*), even after several million years of divergence (*Barth et al., 2020*; *Suvorov et al., 2020*). Future work should employ multispecies-network methods or coalescent simulations (*Joly, McLenachan & Lockhart, 2009*) to determine the relative role of ILS and ancient admixture. However, it does not seem unlikely that both processes might have thus contributed to levels of shared variation across 5–30 millions of the divergence of *Stygocapitella* (*Cerca et al., 2020a*; *Cerca et al., 2020b*).

## Incomplete lineage sorting and morphological similarity

The debate on morphological similarity is slowly shifting from 'are cryptic species an artefact of systematics?' to 'what are the causes underlying morphological similarity?', following the evidence that speciation is not necessarily accompanied by morphological divergence (*Wada, Kameda & Chiba, 2013*; *Swift, Daglio & Dawson, 2016*; *Cerca et al., 2020b*). We have previously argued that the study of morphological similarity will benefit from predictions, models, and evidence from paleontological stasis (*Cerca et al., 2020a*), which positions that stasis may result from constraints, selective pressures on physiology and/or behaviour, stabilizing selection, niche conservatism (*Hansen & Houle, 2004*; *Estes & Arnold, 2007*; *Futuyma, 2010*). While similarity in different cryptic species complexes may stem from different causes, morphological similarity in the three studied *Stygocapitella* species complex is likely associated with homogeneous genetic variation caused by incomplete lineage sorting and ancestral admixture that occurred during the divergence of the complex (*Futuyma, 2010*). In such a scenario, it is expected that patterns of genetic variation remain similar for the species, thus resulting in the retention of symplesiomorphic morphological states (*Futuyma, 2010*) and in the deceleration of morphological evolution (*Cerca et al., 2020b*). In any case, future works using whole-genome data are necessary to, for example, detect if regions affected by incomplete lineage sorting and gene flow are disproportionally enriched for genes that usually contribute for morphological divergence in closely related

taxa. These works should also employ more variant-level data to confirm the patterns herein obtained.

## CONCLUSIONS

The increasing discovery of cryptic species has led to heated debates in systematics, mostly lacking an integration in an evolutionary framework. Here, we tested the hypotheses that morphological similarity may own to reduced genetic variation (bottlenecks, founder effects), recent admixture (shared genetic variation), or incomplete lineage sorting and ancient admixture. We found that morphological similarity in the three morphologically similar *Stygocapitella* species may own to incomplete lineage sorting or ancient admixture underlying shared genetic variation. Future works should focus on understanding whether reduced genetic variation or shared genetic variation underlies morphological similarity in other systems.

## ACKNOWLEDGEMENTS

## ADDITIONAL INFORMATION AND DECLARATIONS

### Funding

### Grant Disclosures

## Competing Interests

The authors declare there are no competing interests.

## Author Contributions

- José Cerca conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Angel Rivera-Colon analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.
- Mafalda Ferreira, Mark Ravinet and Julian Catchen analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.
- Michael Nowak conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.
- Torsten Struck conceived and designed the experiments, performed the experiments, analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.

## Data Availability

The following information was supplied regarding data availability:

Code is available at Github:

https://www.ebi.ac.uk/ena/browser/view/PRJEB40223.

The data is available in the ENA: PRJEB40223.

## Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.10896#supplemental-information.

## REFERENCES

**Alexander DH, Novembre J, Lange K. 2009.** Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* **19**:1655–1664 DOI 10.1101/gr.094052.109.

**Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990.** Basic local alignment search tool. *Journal of Molecular Biology* **215**:403–410 DOI 10.1016/S0022-2836(05)80360-2.

**Andrade SCS, Norenburg JL, Solferini VN. 2011.** Worms without borders: genetic diversity patterns in four Brazilian *Ototyphlonemertes* species (Nemertea, Hoplonemertea). *Marine Biology* **158**:2109–2124 DOI 10.1007/s00227-011-1718-3.

**Arnold B, Corbett-Detig RB, Hartl D, Bomblies K. 2013.** RADseq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. *Molecular Ecology* **22**:3179–3190 DOI 10.1111/mec.12276.

**Astrin JJ, Stüben PE. 2008.** Phylogeny in cryptic weevils: molecules, morphology and new genera of western Palaearctic Cryptorhynchinae (Coleoptera: Curculionidae). *Invertebrate Systematics* **22**:503–522 DOI 10.1071/IS07057.

**Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA. 2008.** Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLOS ONE* **3**:e3376 DOI 10.1371/journal.pone.0003376.

**Barth JMI, Gubili C, Matschiner M, Tørresen OK, Watanabe S, Egger B, Han Y-S, Feunteun E, Sommaruga R, Jehle R, Schabetsberger R. 2020.** Stable species boundaries despite ten million years of hybridization in tropical eels. *Nature Communications* **11**:1–13 DOI 10.1038/s41467-020-15099-x.

**Bernardo J. 2011.** A critical appraisal of the meaning and diagnosability of cryptic evolutionary diversity, and its implications for conservation in the face of climate change. *Climatic Change Ecology and Systematics* **18**:380–438 DOI 10.1017/CBO9780511974540.019.

**Bickford D, Lohman DJ, Sodhi NS, Ng PKL, Meier R, Winker K, Ingram KK, Das I. 2007.** Cryptic species as a window on diversity and conservation. *Trends in Ecology & Evolution* **22**:148–155 DOI 10.1016/j.tree.2006.11.004.

**Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009.** BLAST+: architecture and applications. *BMC Bioinformatics* **10**:1–9 DOI 10.1186/1471-2105-10-421.

**Casu M, Curini-Galletti M. 2006.** Genetic evidence for the existence of cryptic species in the mesopsammic flatworm *Pseudomonocelis ophiocephala* (Rhabditophora: Proseriata). *Biological Journal of the Linnean Society* **87**:553–576 DOI 10.1111/j.1095-8312.2006.00588.x.

**Cerca J, Maurstad MF, Rochette N, Rivera-Colón A, Rayamajhi N, Catchen J, Struck TH. 2021.** Removing the bad apples: a simple bioinformatic method to improve loci-recovery in de novo RADseq data for non-model organisms. *Methods in Ecology and Evolution* in press.

**Cerca J, Meyer C, Purschke G, Struck TH. 2020a.** Delimitation of cryptic species drastically reduces the geographical ranges of marine interstitial ghost-worms (Stygocapitella; Annelida, Sedentaria). *Molecular Phylogenetics and Evolution* **143**:106663 DOI 10.1016/j.ympev.2019.106663.

**Cerca J, Meyer C, Stateczny D, Siemon D, Wegbrod J, Purschke G, Dimitrov D, Struck TH. 2020b.** Deceleration of morphological evolution in a cryptic species complex and its link to paleontological stasis. *Evolution* **n/a**:1–16 DOI 10.1111/evo.13884.

**Cerca J, Purschke G, Struck TH. 2018.** Marine connectivity dynamics: clarifying cosmopolitan distributions of marine interstitial invertebrates and the meiofauna paradox. *Marine Biology* **165**:123 DOI 10.1007/s00227-018-3383-2.

**Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ. 2015.** Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**:1–16 DOI 10.1186/s13742-015-0047-8.

**Chenuil A, Cahill AE, Délémontey N, Du Salliant du Luc E, Fanton H. 2019.** Problems and questions posed by cryptic species. A framework to guide future studies. 77–106.

**Chernomor O, Von Haeseler A, Minh BQ. 2016.** Terrace aware data structure for phylogenomic inference from supermatrices. *Systematic Biology* **65**:997–1008 DOI 10.1093/sysbio/syw037.

**Crotti M, Barratt CD, Loader SP, Gower DJ, Streicher JW. 2019.** Causes and analytical impacts of missing data in RADseq phylogenetics: insights from an African frog (Afrixalus). *Zoologica Scripta* **48**:157–167 DOI 10.1111/zsc.12335.

**Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R. 1000 Genomes Project Analysis Group. 2011.** The variant call format and VCFtools. *Bioinformatics* **27**:2156–2158 DOI 10.1093/bioinformatics/btr330.

**De León GPP, Nadler SA. 2010.** What we don't recognize can hurt us: a plea for awareness about cryptic species. *Journal of Parasitology* **96**:453–464 DOI 10.1645/GE-2260.1.

**De Medeiros BAS, Farrell BD. 2018.** Whole genome amplification in double-digest RAD-seq results in adequate libraries but fewer sequenced loci. *PeerJ* **6**:e5089 DOI 10.7717/peerj.5089.

**De Medeiros BAS, Farrell BD. 2020.** Evaluating insect-host interactions as a driver of species divergence in palm flower weevils. *Nature Communications* **3(1)**:749 DOI 10.1038/s42003-020-01482-3.

**Degnan JH, Rosenberg NA. 2009.** Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology & Evolution* **24**:332–340 DOI 10.1016/j.tree.2009.01.009.

**Derycke S, Backeljau T, Moens T. 2013.** Dispersal and gene flow in free-living marine nematodes. *Frontiers in Zoology* **10**:1 DOI 10.1186/1742-9994-10-1.

**Derycke S, Remerie T, Backeljau T, Vierstraete A, Vanfleteren J, Vincx M, Moens T. 2008.** Phylogeography of the *Rhabditis* (Pellioditis) marina species complex: evidence for long-distance dispersal, and for range expansions and restricted gene flow in the northeast Atlantic. *Molecular Ecology* **17**:3306–3322 DOI 10.1111/j.1365-294X.2008.03846.x.

**Derycke S, Remerie T, Vierstraete A, Backeljau T, Vanfleteren J, Vincx M, Moens T. 2005.** Mitochondrial DNA variation and cryptic speciation within the free-living marine nematode *Pellioditis marina*. *Marine Ecology Progress Series* **300**:91–103 DOI 10.3354/meps300091.

**Derycke S, Van Vynckt R, Vanoverbeke J, Vincx M, Moens T. 2007.** Colonization patterns of Nematoda on decomposing algae in the estuarine environment: community assembly and genetic structure of the dominant species *Pellioditis marina*. *Limnology and Oceanography* **52**:992–1001 DOI 10.4319/lo.2007.52.3.0992.

**Dornburg A, Federman S, Eytan RI, Near TJ. 2016.** Cryptic species diversity in sub-Antarctic islands: a case study of Lepidonotothen. *Molecular Phylogenetics and Evolution* **104**:32–43 DOI 10.1016/j.ympev.2016.07.013.

**Dufresnes C, Strachinis I, Suriadna N, Mykytynets G, Cogălniceanu D, Székely P, Vukov T, Arntzen JW, Wielstra B, Lymberakis P, Geffen E, Gafny S, Kumlutaş Y, Ilgaz Ç, Candan K, Mizsei E, Szabolcs M, Kolenda K, Smirnov N, Géniez P, Lukanov S, Crochet P-A, Dubey S, Perrin N, Litvinchuk SN, Denoël M. 2019.** Phylogeography of a cryptic speciation continuum in Eurasian spadefoot toads (Pelobates). *Molecular Ecology* **28**:3257–3270 DOI 10.1111/mec.15133.

**Edelman NB, Frandsen PB, Miyagi M, Clavijo B, Davey J, Dikow R, Garca-Accinelli G, Van Belleghem SM, Patterson N, Neafsey DE, Challis R, Kumar S, Moreira GRP, Salazar C, Chouteau M, Counterman BA, Papa R, Blaxter M, Reed RD, Dasmahapatra KK, Kronforst M, Joron M, Jiggins CD, McMillan WO, Di Palma F, Blumberg AJ, Wakeley J, Jaffe D, Mallet J. 2019.** Genomic architecture and introgression shape a butterfly radiation. *Science* **366**:594–599 DOI 10.1126/science.aaw2090.

**Eldredge N, Gould SJ. 1972.** Punctuated equilibria: an alternative to phylogetic gradualism. In: Schopf TJM, ed. *Models in paleobiology*. San Francisco: Freeman, Cooper and Co., 82–115.

**Erlank E, Koekemoer LL, Coetzee M. 2018.** The importance of morphological identification of African anopheline mosquitoes (Diptera: Culicidae) for malaria control programmes. *Malaria Journal* **17**:1–7 DOI 10.1186/s12936-018-2189-5.

**Estes S, Arnold SJ. 2007.** Resolving the paradox of stasis: models with stabilizing selection explain evolutionary divergence on all timescales. *The American Naturalist* **169**:227–244 DOI 10.1086/510633.

**Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M. 2013.** Robust Demographic Inference from Genomic and SNP Data. *PLoS Genetics* **9**.

**Ferreira MS, Jones MR, Callahan CM, Farelo L, Tolesa Z, Suchentrunk F, Boursot P, Mills LS, Alves PC, Good JM, Mills J. 2020.** The legacy of recurrent introgression during the radiation of hares. *Systematic Biology* Epub ahead of print 2020 02 December DOI 10.1093/sysbio/syaa088.

**Fišer C, Robinson CT, Malard F. 2018.** Cryptic species as a window into the paradigm shift of the species concept. *Molecular Ecology* **27**:613–635 DOI 10.1111/mec.14486.

**Futuyma DJ. 2010.** Evolutionary constraint and ecological consequences. *Evolution* **64**:1865–1884 DOI 10.1111/j.1558-5646.2010.00960.x.

**Gautier M, Gharbi K, Cezard T, Foucaud J, Kerdelhué C, Pudlo P, Cornuet J-M, Estoup A. 2013.** The effect of RAD allele dropout on the estimation of genetic variation within and between populations. *Molecular Ecology* **22**:3165–3178 DOI 10.1111/mec.12089.

**Giska I, Farelo L, Pimenta J, Seixas FA, Ferreira MS, Marques JP, Miranda I, Letty J, Jenny H, Hackländer K, Magnussen E, Melo-Ferreira J. 2019.** Introgression drives repeated evolution of winter coat color polymorphism in hares. *Proceedings of the National Academy of Sciences of the United States of America* **116**:24150–24156 DOI 10.1073/pnas.1910471116.

**Golombek A, Tobergte S, Nesnidal MP, Purschke G, Struck TH. 2013.** Mitochondrial genomes to the rescue—diurodrilidae in the myzostomid trap. *Molecular Phylogenetics and Evolution* **68**:312–326 DOI 10.1016/j.ympev.2013.03.026.

**Gould SJ. 2002.** *The structure of evolutionary theory.* Cambridge: Harvard University Press.

**Hansen TF, Houle D. 2004.** Evolvability, stabilizing selection, and the problem of stasis. In: Pigliucci M, Preston K, eds. *Phenotypic integration: studying the ecology and evolution of complex phenotypes.* New York: Oxford University Press, 130–154.

**Hillis DM, Dixon MT. 1991.** Ribosomal DNA: molecular evolution and phylogenetic inference. *Quarterly Review of Biology* **66**:411–453 DOI 10.1086/417338.

**Hoang DT, Chernomor O, Haeseler Avon, Minh BQ, Vinh LS. 2017.** Ufboot2: improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution* **35**:518–522 DOI 10.5281/zenodo.854445.

**Hodel RGJ, Chen S, Payton AC, McDaniel SF, Soltis P, Soltis DE. 2017.** Adding loci improves phylogeographic resolution in red mangroves despite increased missing data: comparing microsatellites and RAD-Seq and investigating loci filtering. *Scientific Reports* **7**:1–14 DOI 10.1038/s41598-017-16810-7.

**Huang H, Lacey Knowles L. 2016.** Unforeseen consequences of excluding missing data from next-generation sequences: simulation study of rad sequences. *Systematic Biology* **65**:357–365 DOI 10.1093/sysbio/syu046.

**Huson DH, Bryant D. 2006.** Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* **23**:254–267 DOI 10.1093/molbev/msj030.

**Joly S, McLenachan PA, Lockhart PJ. 2009.** A statistical approach for distinguishing hybridization and incomplete lineage sorting. *The American Naturalist* **174(2)**:E54–70 DOI 10.1086/600082.

**Jombart T, Ahmed I. 2011.** Adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* **27**:3070–3071 DOI 10.1093/bioinformatics/btr521.

**Kalyaanamoorthy S, Minh BQ, Wong TKF, Von Haeseler A, Jermiin LS. 2017.** ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* **14**:587–589 DOI 10.1038/nmeth.4285.

**Katoh K, Standley DM. 2013.** MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* **30**:772–780 DOI 10.1093/molbev/mst010.

**Kieneke A, Martínez Arbizu PM, Fontaneto D. 2012.** Spatially structured populations with a low level of cryptic diversity in European marine Gastrotricha. *Molecular Ecology* **21**:1239–1254 DOI 10.1111/j.1365-294X.2011.05421.x.

**Knowlton N. 1993.** Sibling species in the sea. *Annual Review of Ecology and Systematics* **24**:189–216 DOI 10.1146/annurev.es.24.110193.001201.

**Korshunova T, Martynov A, Bakken T, Picton B. 2017.** External diversity is restrained by internal conservatism: new nudibranch mollusc contributes to the cryptic species problem. *Zoologica Scripta* **46**:683–692 DOI 10.1111/zsc.12253.

**Kubatko LS, Degnan JH. 2007.** Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Systematic Biology* **56**:17–24 DOI 10.1080/10635150601146041.

**Kück P, Longo GC. 2014.** FASconCAT-G: extensive functions for multiple sequence alignment preparations concerning phylogenetic studies. *Frontiers in Zoology* **11**:1–8 DOI 10.1186/s12983-014-0081-x.

**Kück P, Meusemann K. 2010.** FASconCAT: convenient handling of data matrices. *Molecular Phylogenetics and Evolution* **56**:1115–1118 DOI 10.1016/j.ympev.2010.04.024.

**Kück P, Struck TH. 2014.** BaCoCa—a heuristic software tool for the parallel assessment of sequence biases in hundreds of gene and taxon partitions. *Molecular Phylogenetics and Evolution* **70**:94–98 DOI 10.1016/j.ympev.2013.09.011.

**Lavoué S, Miya M, Arnegard ME, McIntyre PB, Mamonekene V, Nishida M. 2011.** Remarkable morphological stasis in an extant vertebrate despite tens of millions of years of divergence. *Proceedings of the Royal Society B: Biological Sciences* **278**:1003–1008 DOI 10.1098/rspb.2010.1639.

**Leaché AD, Zhu T, Rannala B, Yang Z. 2019.** The spectre of too many species. *Systematic Biology* **68**:168–181 DOI 10.1093/sysbio/syy051.

**Leasi F, Norenburg JL. 2014.** The necessity of DNA taxonomy to reveal cryptic diversity and spatial distribution of meiofauna, with a focus on Nemertea. *PLOS ONE* **9**:e104385 DOI 10.1371/journal.pone.0104385.

**Lee CE, Frost BW. 2002.** Morphological stasis in the Eurytemora affinis species complex (Copepoda: Temoridae). *Hydrobiologia* **480**:111–128 DOI 10.1023/A:1021293203512.

**Lee KM, Kivela SM, Ivanov V, Hausmann A, Kaila L, Wahlberg N, Mutanen M. 2018.** Information dropout patterns in restriction site associated dna phylogenomics and a comparison with multilocus sanger data in a species-rich moth genus. *Systematic Biology* **67**(**6**):925–939 DOI 10.1093/sysbio/syy029.

**Li G, Figueiró HV, Eizirik E, Murphy WJ. 2019.** Recombination-aware phylogenomics reveals the structured genomic landscape of hybridizing cat species. *Molecular Biology and Evolution* **36**:2111–2126 DOI 10.1093/molbev/msz139.

**Malinsky M, Svardal H, Tyers AM, Miska EA, Genner MJ, Turner GF, Durbin R. 2018.** Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. *Nature Ecology and Evolution* **2**:1940–1955 DOI 10.1038/s41559-018-0717-x.

**Melo-Ferreira J, Boursot P, Carneiro M, Esteves PJ, Farelo L, Alves PC. 2012.** Recurrent introgression of mitochondrial DNA among hares (lepus spp.) revealed by species-tree inference and coalescent simulations. *Systematic Biology* **61**:367–381 DOI 10.1093/sysbio/syr114.

**Nadler SA, De Len GPP. 2011.** Integrating molecular and morphological approaches for characterizing parasite cryptic species: implications for parasitology. *Parasitology* **138**:1688–1709 DOI 10.1017/S003118201000168X.

**Nguyen LT, Schmidt HA, Haeseler AVon, Minh BQ. 2015.** IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* **32**:268–274 DOI 10.1093/molbev/msu300.

**Novo M, Almodóvar A, Fernández R, Trigo D, Díaz Cosín DJ. 2010.** Cryptic speciation of hormogastrid earthworms revealed by mitochondrial and nuclear data. *Molecular Phylogenetics and Evolution* **56**:507–512 DOI 10.1016/j.ympev.2010.04.010.

**Novo M, Almodóvar A, Fernández R, Trigo D, Díaz-Cosín DJ, Giribet G. 2012.** Appearances can be deceptive: different diversification patterns within a group of mediterranean earthworms (Oligochaeta, Hormogastridae). *Molecular Ecology* **21**:3776–3793 DOI 10.1111/j.1365-294X.2012.05648.x.

**O'Leary SJ, Puritz JB, Willis SC, Hollenbeck CM, Portnoy DS. 2018.** These aren't the loci you'e looking for: principles of effective SNP filtering for molecular ecologists. *Molecular Ecology* **27**:3193–3206 DOI 10.1111/mec.14792.

**Palumbi S, Martin A, Romano S, Owen McMillan W, Stice L, Grabowski G. 1991.** The simple fool's guide to PCR. version 2.

**Pante E, Puillandre N, Viricel A, Arnaud-Haond S, Aurelle D, Castelin M, Chenuil A, Destombe C, Forcioli D, Valero M, Viard F, Samadi S. 2015.** Species are hypotheses: avoid connectivity assessments based on pillars of sand. *Molecular Ecology* **24**:525–544 DOI 10.1111/mec.13048.

**Paradis E, Schliep K. 2018.** Ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**:526–528 DOI 10.1093/bioinformatics/bty633.

**Paris JR, Stevens JR, Catchen JM. 2017.** Lost in parameter space: a road map for stacks. *Methods in Ecology and Evolution* **8**:1360–1373 DOI 10.1111/2041-210X.12775.

**Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T, Webster T, Reich D. 2012.** Ancient admixture in human history. *Genetics* **192**:1065–1093 DOI 10.1534/genetics.112.145037.

**Pease JB., Haak DC., Hahn MW., Moyle LC.. 2016.** Phylogenomics reveals three sources of adaptive variation during a rapid radiation. *PLOS Biology* **14**:1–24 DOI 10.1371/journal.pbio.1002379.

**Pérez-Ponce de León G, Poulin R. 2016.** Taxonomic distribution of cryptic diversity among metazoans: not so homogeneous after all. *Biology Letters* **12**:20160371 DOI 10.1098/rsbl.2016.0371.

**Peter BM. 2016.** Admixture, population structure, and f-statistics. *Genetics* **202**:1485–1501 DOI 10.1534/genetics.115.183913.

**Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. 2012.** Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLOS ONE* **7**:e37135 DOI 10.1371/journal.pone.0037135.

**Pfenninger M, Schwenk K. 2007.** Cryptic animal species are homogeneously distributed among taxa and biogeographical regions. *BMC Evolutionary Biology* **7**:121 DOI 10.1186/1471-2148-7-121.

**Rice P, Longden L, Bleasby A. 2000.** EMBOSS: the european molecular biology open software suite. *Trends in Genetics* **16**:276–277 DOI 10.1016/S0168-9525(00)02024-2.

**Rochette NC, Catchen JM. 2017.** Deriving genotypes from RAD-seq short-read data using Stacks. *Nature Protocols* **12**:2640–2659 DOI 10.1038/nprot.2017.123.

**Rochette N, Rivera-Colón A, Catchen JM. 2019.** Stacks2: analytical methods for paired-end sequencing improve RADseq-based population genomics. *Molecular Ecology* **28(21)**:4737–4754 DOI 10.1111/mec.15253.

**Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A. 2017.** DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular Biology and Evolution* **34**:3299–3302 DOI 10.1093/molbev/msx248.

**Santamaria CA, Mateos M, Dewitt TJ, Hurtado LA. 2016.** Constrained body shape among highly genetically divergent allopatric lineages of the supralittoral isopod Ligia occidentalis (Oniscidea). *Ecology and Evolution* **6**:1537–1554 DOI 10.1002/ece3.1984.

**Scornavacca C, Galtier N. 2017.** Incomplete lineage sorting in mammalian phylogenomics. *Systematic Biology* **66**:112–120 DOI 10.1093/sysbio/syw082.

**Sloan DB, Havird JC, Sharbrough J. 2017.** The on-again, off-again relationship between mitochondrial genomes and species boundaries. *Molecular Ecology* **26**:2212–2236 DOI 10.1111/mec.13959.

**Smith MA, Eveleigh ES, McCann KS, Merilo MT, McCarthy PC, Van Rooyen KI. 2011.** Barcoding a quantified food web: crypsis, concepts, ecology and hypotheses. *PLOS ONE* **6**:e14424 DOI 10.1371/journal.pone.0014424.

**Song S, Liu L, Edwards SV, Wu S. 2015.** Erratum: resolving conflict in eutherian mammal phylogeny using phylogenomics and the multispecies coalescent model (Proceedings of the National Academy of Sciences of the United States of America (2012) 109 (14942-14947) (DOI:10.1073/pnas.1211733109)). *Proceedings of the National Academy of Sciences of the United States of America* **112**:E6079 DOI 10.1073/pnas.1518753112.

**Struck TH, Cerca J. 2019.** Cryptic species and their evolutionary significance. *eLS* Epub ahead of print 2019 16 January DOI 10.1002/9780470015902.a0028292.

**Struck TH, Feder JL, Bendiksby M, Birkeland S, Cerca J, Gusarov VI, Kistenich S, Larsson K-H, Hsiang Liow L, Nowak MD, Stedje B, Bachmann L, Dimitrov D. 2018.** Finding evolutionary processes hidden in cryptic species. *Trends in Ecology & Evolution* **33(3)**:153–163 DOI 10.1016/j.tree.2017.11.007.

**Struck TH, Koczula J, Stateczny D, Meyer C, Purschke G. 2017.** Two new species in the annelid genus Stygocapitella (Orbiniida, Parergodrilidae) with comments on their biogeography. *Zootaxa* **4286**:301–332.

**Suh A, Smeds L, Ellegren H. 2015.** The dynamics of incomplete lineage sorting across the ancient adaptive radiation of neoavian birds. *PLOS Biology* **13**:1–18 DOI 10.1371/journal.pbio.1002224.

**Suvorov A, Kim BY, Wang J, Armstrong EE, Peede D, DAgostino ERR, Price DK, Wadell P, Lang M, Courtier-Orgogozo V, David JR, Petrov D, Matute DR, Schrider DR, Comeault AA. 2020.** Widespread introgression across a phylogeny of 155 Drosophila genomes. *bioRxiv* DOI 10.1101/2020.12.14.422758.

**Swift HF, Daglio LG, Dawson MN. 2016.** Three routes to crypsis: stasis, convergence, and parallelism in the *Mastigias* species complex (Scyphozoa, Rhizostomeae). *Molecular Phylogenetics and Evolution* **99**:103–115 DOI 10.1016/j.ympev.2016.02.013.

**Taylor SA, Larson EL. 2019.** Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. *Nature Ecology and Evolution* **3**:170–177 DOI 10.1038/s41559-018-0777-y.

**Valtueña FJ, López J, Álvarez J, RodrguezRiaño T, OrtegaOlivencia A. 2016.** Scrophularia arguta, a widespread annual plant in the Canary Islands: a single recent colonization event or a more complex phylogeographic pattern? *Ecology and Evolution* **6**:4258–4273 DOI 10.1002/ece3.2109.

**Wada S, Kameda Y, Chiba S. 2013.** Long-term stasis and short-term divergence in the phenotypes of microsnails on oceanic islands. *Molecular Ecology* **22**:4801–4810 DOI 10.1111/mec.12427.

**Wares JP, Cunningham CW. 2001.** Phylogeography and historical ecology of the North Atlantic intertidal. *Evolution* **55**:2455–2469 DOI 10.1111/j.0014-3820.2001.tb00760.x.

**Weber AA-T, Stöhr S, Chenuil A. 2019.** Species delimitation in the presence of strong incomplete lineage sorting and hybridization. *Molecular Phylogenetics and Evolution* **131**:240218 DOI 10.1101/240218.

**Weir BS, Cockerham CC. 1984.** Estimating F-statistics for the analysis of population structure. *Evolution* **38**:1358–1370 DOI 10.2307/2408641.

**Westheide W, Purschke G. 1988.** Organism processing. In: Higgins RP, Thiel H, eds. *Introduction to the study of meiofauna.* Washington, D.C.: Smithsonian Institution Press, 146–160.

**Zanol J, Halanych KM, Struck TH, Fauchald K. 2010.** Phylogeny of the bristle worm family Eunicidae (Eunicida, Annelida) and the phylogenetic utility of noncongruent 16S, COI and 18S in combined analyses. *Molecular Phylogenetics and Evolution* **55**:660–676 DOI 10.1016/j.ympev.2009.12.024.

**Zuccarello GC, West JA, Kamiya M. 2018.** Non-monophyly of Bostrychia simpliciuscula (Ceramiales, Rhodophyta): Multiple species with very similar morphologies, a revised taxonomy of cryptic species. *Phycological Research* **66**:100–107 DOI 10.1111/pre.12207.