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

José Cerca [Corresp., 1, 2, 3] , **Angel Rivera-Colon** [4] , **Mafalda Ferreira** [5, 6, 7] , **Mark Ravinet** [8, 9] , **Michael Nowak** [3] , **Julian Catchen** [4] , **Torsten Struck** [3]

[1] Department of Environmental Science, Policy, and Management, University of California, University of California, Berkeley, Berkeley, California, United States

[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, Illinois, United States

[5] Division of Biological Sciences, University of Montana, Missoula, Montana, United States

[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

Corresponding Author: José Cerca
Email address: jose.cerca@ntnu.no

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 (~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.

1 **Incomplete lineage sorting and ancient admixture, and speciation without morphological**

2 **change in ghost-worm cryptic species**

3 **José Cerca**[1,2,3], Angel Rivera-Colón[4], Mafalda S. Ferreira[5,6,7], Mark Ravinet[8,9], Michael D.

4 Nowak[1], Julian Catchen[4], Torsten H. Struck[1]

5 [1] Natural History Museum, University of Oslo, Oslo, Norway

6 [2] Department of Environmental Science, Policy, and Management, University of California,

7 Berkeley, CA, USA

8 [3] Department of Natural History, NTNU University Museum, Norwegian University of Science

9 and Technology, Trondheim, Norway

10 [4] Department of Evolution, Ecology, and Behavior, University of Illinois at Urbana-Champaign

11 [5] CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO Laboratório

12 Associado, Universidade do Porto, Vairão, Portugal

13 [6] Departamento de Biologia, Faculdade de Ciências da Universidade do Porto, Porto, Portugal

14 [7] Division of Biological Sciences, University of Montana, Missoula, Montana, USA

15 [8] Centre for Ecological and Evolutionary Synthesis, University of Oslo

16 [9] School of Life Sciences, University of Nottingham, UK

18 Corresponding Author:

19 José Cerca[1]

20 ORCID 0000-0001-7788-4367

21 Sars' Gate 1, Oslo, Norway

22 Email address: jose.cerca@ntnu.no

24 **Abstract**

25 Morphologically similar species, that is cryptic species, may be similar or quasi-similar owing to

26 the deceleration of morphological evolution and stasis. While the factors underlying the

27 deceleration of morphological evolution or stasis in cryptic species remain unknown, decades of

28 research in the field of paleontology on punctuated equilibrium have originated clear hypotheses.

29 Species are expected to remain morphologically identical in scenarios of shared genetic

30 variation, such as hybridization and incomplete lineage sorting, or in scenarios where bottlenecks

31 reduce genetic variation and constrain the evolution of morphology. Here, focusing on three

32 morphologically similar *Stygocapitella* species, we employ a whole-genome amplification

33 method (WGA) coupled with double-digestion restriction-site associated DNA sequencing

34 (ddRAD) to reconstruct the evolutionary history of the species complex. We explore population

35 structure, use population-level statistics to determine the degree of connectivity between

36 populations and species, and determine the most likely demographic scenarios which generally

37 reject for recent hybridization. We find that the combination of WGA and ddRAD allowed us to

38   obtain genomic-level data from microscopic eukaryotes (~1 millimetre) opening up opportunities
39   for those working with population genomics and phylogenomics in such taxa. The three species
40   share genetic variance, likely from incomplete lineage sorting and ancient admixture. We
41   speculate that the degree of shared variation might underlie morphological similarity in the
42   Atlantic species complex.
43
44   **Introduction**

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

60   Following centuries of morphologically oriented taxonomy, the existence of "true" cryptic
61   species entails a challenge to the delimitation, discovery and classification of species (Bickford et
62   al. 2007; Fišer et al. 2018; Struck et al. 2018). In the case of morphologically similar species,
63   species delimitation relying on morphology alone will fail to capture the existing species diversity
64   (Pante et al. 2015; Fišer et al. 2018; Chenuil et al. 2019; Struck and Cerca 2019), resulting in the
65   lumping of different species into a single species complex. While much has been written on the
66   consequences of cryptic species in terms of biological systematics, we have only recently begun
67   to understand the impact of cryptic species in other fields of biology. When species are poorly
68   delimited, determination of biogeographic breaks (Weber et al. 2019; Cerca et al. 2020a),
69   inferences on the evolutionary history (Wada et al. 2013; Swift et al. 2016; Struck et al. 2018;

70   Dufresnes et al. 2019), and the determination of ecological richness of an ecosystem (Chenuil et
71   al. 2019) may be severely compromised. These problems extend outside fundamental fields of
72   biology when species complexes are medically-relevant, such as the *Anopheles* cryptic species
73   complex where not every morphologically-similar species is capable of transmitting malaria
74   (Erlank et al. 2018) or in parasite species (De León and Nadler 2010; Nadler and De Len 2011),
75   but also in cases of conservation management (Bickford et al. 2007; Bernardo 2011).

76   While the discovery of cryptic species complexes has increased in the last few years, the
77   resulting debate has focused on whether these are taxonomic artefacts or biologically relevant
78   species. Consequently, the causes underlying morphological similarity remain mostly unexplored.
79   Despite this hindrance, an important source of information may come from palaeontology where
80   stasis has been studied for decades (Eldredge and Gould 1972; Gould 2002), and from the
81   subsequent integration of this evidence with neontological data. A particularly insightful
82   contribution is that of Futuyma (2010), which suggests that stasis may result from certain
83   ecological, genetic and developmental scenarios. Genetic scenarios include shared genetic
84   variation, potentially resulting from hybridization or ILS, homogenizing morphological
85   divergence; genetic constraints resulting from epistatic reactions or pleiotropy, or constrains from
86   the lack of genetic variation due to repeated bottlenecks or founder effects; stabilizing selection on
87   morphology (Futuyma 2010). Some of these scenarios including stabilizing selection (Lee and
88   Frost 2002; Novo et al. 2010, 2012; Lavoué et al. 2011; Smith et al. 2011; Santamaria et al. 2016;
89   Zuccarello et al. 2018), bottlenecks and founder effects (Dornburg et al. 2016; Valtueña et al.
90   2016) have been proposed to explain similarity on cryptic species. However, this remains untested
91   since evidence for morphological similarity comes mostly from the interpretation of indirect
92   methods, such as phylogenetic trees.

93   The *Stygocapitella* genus includes 11 described species with only four morphotypes and
94   no significant quantitative morphological differences between some species (Cerca et al. 2020a,
95   b). Morphologically identical species occur in sympatry and overlap in their distribution along the
96   Northern European, Atlantic American, and Pacific American coastlines. In a previous study, we
97   confirmed that three North Atlantic species - *Stygocapitella westheidei*, *S. subterranea*, and *S.*
98   *josemariobrancoi* – are morphologically identical (Cerca et al. 2020a), nonetheless, we were not
99   able to determine the causes underlying morphological similarity with certainty. Preliminary

100   results from selected DNA markers indicated that morphological similarity potentially stems from

101   niche conservatism and niche tracking, coupled with the fluctuating dynamics of their habitats

102   and/or genetic constraints (Cerca et al. 2020b). Here, using genomic data, we extend these efforts

103   by focusing on the causes linked to genetic variation underlying morphological similarity (see

104   above). Following Futuyma (2010), we hypothesize that 1) bottlenecks and founder effects reduce

105   genetic variation, thus resulting in morphological similarity; 2) morphological similarity results

106   from recent admixture; 3) shared genetic variation due to incomplete lineage sorting and ancient

107   admixture underlies morphological similarity.

Methods and materials

108

Study system

109

110   *Stygocapitella* is part of the meiofauna, being generally found above the high-water line of

111   sheltered gravel or sandy beaches. To collect individuals, we selected sampling areas based on old

112   records or by assessing beaches using google maps (Supplementary Table 1; Fig. 1). At each site,

113   we drew a transect from the high-water line to the foot of the dune, digging a 1-meter deep hole

114   every meter starting at the high-water line. In each hole, we collected sediment samples every 15

115   cm of depth with a volume of about 500 cm^3. Sediment samples were brought to the laboratory

116   and interstitial invertebrates were extracted using the $MgCl_2$ method, and isolated using a

117   dissecting microscope (Westheide and Purschke 1988). After identifying *Stygocapitella*, we

118   collected and preserved these in a ~70% ethanol solution for DNA extraction.

DNA extraction and molecular species barcoding

119

120   Since *Stygocapitella westheidei*, *S. subterranea* and *S. josemariobrancoi* are

121   morphologically indistinguishable (Cerca et al. 2020a), we barcoded individual individuals using

122   16S, 18S, ITS1 and COI as described in Cerca et al. (2020b, a) (Supplementary Table 2 includes

123   NCBI reference-ID). In brief, we extracted DNA from single individuals using either phenol-

124   chloroform or the E.Z.N.A. Tissue DNA Kit (Omega Bio-Tek), and obtained COI (Astrin and

125   Stüben 2008), 18S (Hillis and Dixon 1991), ITS1 (Cerca et al. 2020a), and 16S (Palumbi et al.

126   1991; Zanol et al. 2010) sequences using PCR. Amplified genetic markers were sequenced by

127   Sanger-sequencing at Macrogen-Europe. For detailed information on amplification, primer

128   sequences and extraction please see Cerca et al. (2020b, a).

Library preparation and Illumina sequencing

129

130   We selected 50 *Stygocapitella josemariobrancoi*, 47 *S. subterranea* and 24 *S. westheidei*

131 for library preparation (Supplementary Table 2). Due to the reduced body size, DNA extractions

132 of *Stygocapitella* yield low concentrations of DNA, therefore, to overcome this problem, we used

133 a combination of whole genome amplification (WGA) (Golombek et al. 2013; de Medeiros and

134 Farrell 2018) followed by a double-digestion Restriction site-Associated DNA sequencing

135 protocol (ddRAD) (Baird et al. 2008; Peterson et al. 2012). To complete the WGA reaction, DNA

136 of a single individual is first denatured and mixed with random hexamer primers and the Phi29

137 DNA polymerase (Illustra Genomiphi HY DNA Amplification Kit; GE© Healthcare Life

138 Science). Following the manufacturer's instructions, 2.5µl of template DNA were mixed with

139 22.5µl of sample buffer, and incubated at 95ºC for three minutes for denaturation. After this, we

140 added 22.5µl of reaction buffer and 2.5µl of enzyme mix to the DNA-sample buffer solution,

141 incubated the solution at 30ºC for four hours for DNA amplification and an enzyme heat-

142 inactivation at 65ºC for ten minutes. DNA was purified using AMPure XP beads, and resuspended

143 in ddH$_2$O. The concentration of the amplified DNA was determined with Qubit and the fragment

144 size distribution with a fragment-analyzer.

145   For each individual, 500 ng of amplified DNA was digested in 25 µl including 0.5 µl of

146 each restriction enzyme (Pst-I HF and Mse-I, each 20 units/µl) and 2.5 µl Cut-smart buffer. The

147 digestion reaction was carried out at 37º C for two hours. Digested DNA was purified using

148 Ampure-beads and resuspended in 22 µl ddH$_2$O, and Illumina adaptors with barcodes were ligated

149 to the digested DNA in a 25 µl reaction including 20.5 µl sample DNA, 1 µl T4 DNA-ligase, 2.5

150 µl 10X T4 ligase buffer and 1 µl adapter P1/2-mix. This reaction was incubated for 30 minutes at

151 25ºC, and the enzyme inactivated for 10 minutes at 65ºC. The barcoded libraries were pooled,

152 cleaned using AMPure XP beads, and eluted in 100 µl of ultra-purified water. We ran a size-

153 selection step using Blue Pippin's 100-600 bp cassette (BDF2010) selecting for fragment-length

154 between 300-600 bp followed by cleaning with AMPure XP beads to remove short fragments. The

155 library was amplified in 200 µl including 100 µl Q5 HiFi MasterMix, 5 µl Primer mix and 20 µl

156 DNA, in 18 PCR cycles (initial denaturation: 98ºC for 30s; 18 cycles of 98ºC for 10 seconds, 60ºC

157 for 15 seconds, 72ºC for 15 seconds; and a final elongation of 72ºC for 2 minutes). Finally, two

158 cleaning-quantification steps using AMPure XP beads were done, and the libraries were sent for

159 Illumina Sequencing on an Illumina Hi-Seq 4000.

Individual identification and multi-marker phylogeny

161 Raw 16S, 18S, ITS1 and COI sequences were assembled, and ends were automatically

162 trimmed to remove primers and low-quality ends using Geneious v6.8.1, (Supplementary Table

163 2). Each consensus sequence was queried against the default NCBI database (nr/nt) using BLAST

164 (Altschul et al. 1990; Camacho et al. 2009) to exclude potential contamination. For each gene,

165 sequences were aligned using mafft v7.310, using a maximum of 1,000 iterations, and the ends of

166 the sequences trimmed until the first position without missing data. The accurate localpair

167 algorithm was used for all genes (Katoh and Standley 2013), with the exception of ITS1, which

168 had a single peak, where the globalpair algorithm was applied as it is optimized for gappy

169 sequences. The dataset was concatenated using FASconCAT v1.1 (Kück and Meusemann 2010),

170 and a partitioned phylogenetic tree was obtained using IQ-tree v1.6.10, by applying 1,000

171 fastbootstrap replications. Model determination was done automatically by IQ-tree, and it included

172 TIM2+I for ITS1, JC+I for 18S, TN+R2 for 16S and HKY+I for COI. The congruence between

173 these genes, and in the individuals used has previously been determined in Cerca et al. (2020b, a).

De novo RAD assembly

175 All bioinformatic work is available in

176 https://github.com/jcerca/Papers/tree/main/Stygocapitella_PeerJ . Since no reference genome is

177 available for *Stygocapitella*, we used the *de novo* assembly approach implemented in Stacks v2.2

178 to identify RAD loci (Rochette and Catchen 2017; Rochette et al. 2019). The first module of

179 Stacks, 'process_radtags', was executed using flags tailored to improve data quality (--clean, --

180 quality, --rescue). To optimize Stacks' parameters, we ran the *de novo* pipeline repeatedly using

181 different -M (mismatch between stacks within individuals) and -n (mismatches between stacks

182 between individuals) values, as suggested by best practices (Paris et al. 2017). The total number

183 of loci resulting from different -M -n values were plotted and analyzed, selecting -M 3 and -n 3

184 for the final dataset. Populations, as required for the population map, were defined based on the

185 species and sampling site (total of 22 populations - Supplementary Table 1).

186   Since we observed a considerably high level of missing data in the dataset (>90% missing
187   data), we tested and implemented a new method to improve RADseq datasets. Missing data is
188   especially problematic in RADseq as it can lead to erroneous inference of population-genetic
189   parameters (Arnold et al. 2013; Gautier et al. 2013; Hodel et al. 2017). However, applying stringent
190   filtering for missing data has been shown to prune parsimonious-informative loci, with best-
191   practices suggesting non-conservative pruning of the data (Huang and Lacey Knowles 2016; Lee
192   et al. 2018; Crotti et al. 2019). To mitigate missing data while avoiding stringent filtering, we
193   applied a novel procedure, which allowed us to retrieve more loci from our data (Cerca et al
194   *submitted*). In brief, we ran Stacks for every population present in the population map – 22 times
195   in total (Supplementary Table 1) – thus lowering phylogenetic distance in the dataset. Since
196   phylogenetic distance (biological origin) and artefacts in generating and processing data can lead
197   to allelic dropout (O'Leary et al. 2018), lowering phylogenetic distance will isolate dropout caused
198   by artefacts in library preparation (e.g. DNA size-selection, low DNA concentrations, poor
199   digestion), and loss of information due to whole genome amplification (de Medeiros and Farrell
200   2018). For each population, we identified samples with >45% missing loci and removed these
201   from a final analysis (*hereafter* the *clean* dataset). To evaluate how the optimization impacted the
202   final number of loci we compared the number of loci and missingness in the dataset before cleaning
203   (hereafter the *uncleaned* dataset) and in the cleaned dataset using the values of -r 25 (a locus has
204   to be present in at least 25% of individuals comprising a population to be considered) and -p 4 (a
205   locus has to be present in at least 4 populations to be considered). Finally, we included a technical
206   replicate in the dataset (individuals 222 01 and 222 01R), and checked whether it was coherently
207   placed in all the analyses.

208 Population genomics and phylogenomics

209   From the clean dataset, we extracted a single nucleotide polymorphism-only dataset (SNPs;
210   hereafter *variant* dataset) and an all-sites dataset (containing non-variant and variant positions;
211   hereafter *all-sites* dataset). Separating the data in these two datasets is necessary to meet the
212   assumptions of some statistical tests which may require the presence of non-variant positions to
213   calculate ratios of variant and non-variant sites. The variant dataset was pruned by selecting -r 50
214   (a locus has to be present in at least 50% of individuals comprising a population), and -p 8 (a locus
215   has to be part in at least 8 populations) as loci cut-offs, using the 'populations' program included

216   in Stacks, resulting in 4,737 RAD-loci. After this initial round of cleaning, we used vcftools

217   v0.1.13 (Danecek et al. 2011) to further prune the dataset for 5% minimum allele frequency (--

218   maf), and for mean loci coverage values between 10-100 (--min-meanDP 10 --max-meanDP 100)

219   and removed 12 individuals which had missingness above >90% (--missing-indv; Supplementary

220   Table 2). The combination of coverage filters together with the -M -m optimization procedure

221   mentioned above optimizes the generation of RADseq loci by removing loci which may artificially

222   come together (i.e. repetitive regions). Finally, to decrease the effect of physical linkage in the

223   data, we used a custom BASH shell script which kept only one polymorphism (SNP) per RAD

224   locus, resulting in a final dataset of 3,428 SNPs. Using this dataset, we assessed genetic variation

225   by means of a principal component analysis (PCA), a multi-dimensional scaling (MDS) analysis

226   and an ADMIXTURE analysis. PCA and MDS are model-free approaches to estimate population

227   structure, being complementary as PCA assumes 'mean values' for missing data (i.e. dragging

228   individuals with high missingness to the center) whereas MDS does not. PCA was computed using

229   the R package Adegenet (Jombart and Ahmed 2011) and MDS with plink v1.9 (Chang et al. 2015).

230   ADMIXTURE is a model-oriented approach to determine population structure based on the

231   presence/absence of heterozygotes (Alexander et al. 2009). We ran ADMIXTURE assuming 1-6

232   clusters (K), running a total of 5 replicates for each K, and determined the best K by estimating

233   the cross-validation error (Supplementary Figure 1). Considering the potential for admixture of

234   individuals in sympatry (Figure 1), we used $f3$ statistics, included as part of the Admixtools

235   package, as a direct test for detecting hybridization (Patterson et al. 2012; Peter 2016). These

236   statistics consists of a 3 populations test where a focal population is derived from admixture

237   between the other two populations. When this score is negative, it suggests that admixture likely

238   has occurred. We estimated errors and confidence intervals on the $f3$ statistics by partitioning the

239   dataset into blocks and applying a jackknife bootstrapping. Finally, we inferred species-level

240   divergence by estimating Weir and Cockerham's F_{ST} using vcftools (Weir and Cockerham 1984;

241   Danecek et al. 2011).

242   The all-sites dataset was obtained by extracting FASTA sequences from Stacks. To run the

243   phylogenomic analysis, we wrote a Perl script to reorganize the data into loci, concatenated all loci

244   in a supermatrix using FasConCat-G (Kück and Longo 2014) and ran a partitioned-tree using IQ-

245   tree v1.6.10 (Nguyen et al. 2015) specifying 1,000 fast bootstrap replications (Chernomor et al.

246   2016; Hoang et al. 2017), and locus-specific models which were determined as part of the run

247 (Kalyaanamoorthy et al. 2017). To explore the effect of missing data on the tree topology, we ran
248 BaCoCa (Kück and Struck 2014), a pipeline which provides summary-level statistics on the
249 concatenation matrix and tree, such as the % of positions with missing data shared by a pair of
250 individuals. Pairwise positive overlap values were plotted to the tree topology using the R package
251 ape (Paradis and Schliep 2018). Additionally, we ran an Unweighted Pair Group Method with
252 Arithmetic mean (UPGMA) tree using only the average % of pairwise shared data per individual
253 (i.e. pairwise percentage of shared data between taxa which do not have an indel, ambiguous
254 character state, or a missing character state). The UPGMA tree was run to understand whether taxa
255 were grouped based on the overall pattern of missing data. Additionally, since RADseq loci
256 (represented by allele 0 and allele 1) are not phased and since the labelling of 0 and 1 are arbitrary,
257 we obtained a consensus sequence for each individual. This was done by running the consambig
258 (consensus ambiguous) module included in the EMBOSS pipeline (Rice et al. 2000). We did a
259 species network analysis using SPLITSTREE v4 (Huson and Bryant 2006) to complement
260 phylogenetic inference since the network does not enforce dichotomous branching at each node.

261 To gauge population-level patterns and diversity, we selected loci from the all-sites dataset
262 without missing-data at the population-level and estimated summary statistics including nucleotide
263 diversity (π), Waterson's estimator of genetic diversity (S) and Tajima's D using DNAsp v6 (Rozas
264 et al. 2017). This dataset was not pruned for minimum allele frequency. The selection of sites
265 without missing data is grounded on best-practices as missing data can lead to the under- or
266 overestimation of some of these parameters (Arnold et al. 2013). Importantly, we selected only
267 populations where 3 or more individuals (i.e. > 5 'chromosomes') had data available
268 (Supplementary Table 2).

269 Finally, we evaluated various demographic scenarios using fastsimcoal2, using the same
270 dataset for the previous analysis which included running fastsimcoal2 (Excoffier et al. 2013).
271 Fastsimcoal2 uses the site-frequency spectrum (SFS) and a coalescent-simulation framework
272 based on an arbitrary user-defined scenario to infer population sizes, strength of gene flow and
273 times of coalescence. To assess these models we calculated AIC and likelihood. Likelihood is
274 calculated by running the 'best parameters' for each specified scenario multiple times and
275 obtaining the distribution of likelihood estimates. AIC was calculated using a script available in
276 https://speciationgenomics.github.io/fastsimcoal2/. To implement these simulations, we used the

277   phylogeny obtained with *Stygocapitella subterranea* and *S. westheidei* as sister species, and *S.*
278   *josemariobrancoi* as sister to the remaining two. We defined the following models: no gene flow,
279   ancient gene flow (i.e. between *Stygocapitella josemariobrancoi* and the stem lineage of *S.*
280   *subterranea* and *S. westheidei*), geographic gene flow (i.e. similar as the ancient gene flow, but
281   also with modern gene flow between the sympatric *Stygocapitella subterranea* and *S.*
282   *josemariobrancoi*; note we refer to 'modern' as opposed to 'ancient', that is, in existing lineages),
283   modern gene flow (gene flow between all three modern species), all gene flow (gene flow between
284   all three species and the two ancestral lineages), modern gene flow only between *S.*
285   *josemariobrancoi* and *S. subterranea*, modern gene flow only between *S. josemariobrancoi* and *S.*
286   *westheidei*, modern gene flow only between *S. westheidei* and *S. subterranea*, gene flow in
287   ancestral times and between *S. subterranea* and *S. westheidei*, and ancient gene flow and modern
288   gene flow between *S. josemariobrancoi* and *S. westheidei*. When included in the model, gene flow
289   was moduled as asymmetric. Each model was run 10,000 times with an assumed mutation rate of
290   1.2e-8, and the best fitting scenario was evaluated using likelihood, by running it 100 times.

291   **Results**

292   **Tree of selected molecular markers**

293   We compiled a dataset comprising 4,147 bp (the COI fragment consisted of 629 bp, 16S of
294   548 bp, ITS1 of 1,150 and 18S of 1,817 bp), from which 716 sites were phylogenetic-informative
295   sites. From a total of 69 individuals, we obtained 67 16S sequences, 61 COI sequences, 28 18S
296   sequences and 31 ITS1 sequences. Every species was recovered as monophyletic (Figure 2 A) with
297   bootstrap support values of 100 for *S. josemariobrancoi*, of 86 for *S. subterranea* and of 100 for
298   *S. westheidei* (Figure 2 A). The retrieved tree topology includes *S. westheidei* and *S. subterranea*
299   as sister species. *Stygocapitella josemariobrancoi* as sister to the clade comprising *S. subterranea*
300   and *S. westheidei*. Single gene trees show concordance between markers (Supplementary Figures
301   2-5)

302   **Genomic dataset**

303   We obtained a total of 1,277,919,764☐ sequencing reads from two Illumina HiSeq4000
304   lanes. After demultiplexing and cleaning the data with process_radtags, we retained 899,112,800☐
305   reads (107,830,588 reads were discarded for having ambiguous barcodes, 270,174,154☐ for

306 ambitious RADtags, and 802,222 for low quality). When comparing the clean and unclean
307 dataset, the approach to lower allelic dropout yielded a substantial increase of the number of loci.
308 In detail, after running Stacks for each population individually, we removed 17 *S. subterranea* out
309 of a total of 47, 7 *S. westheidei* out of 24, and 16 out of 50 for *S. josemariobrancoi* (roughly ~33%
310 of the dataset, Supplementary Table 2). The uncleaned dataset yielded 179,742 loci (55,037,190
311 sites including 628,031 variant sites), whereas the cleaned data yielded 368,696 loci (112,725,106
312 sites including 1,100,431 variant sites). When pruned with common denominators included in the
313 module Populations (-r 50 -p4), the unclean dataset yielded 109,369 loci, and the cleaned yielded
314 272,134 loci. Individual-level missing data was reduced from 84.89% in the uncleaned dataset to
315 80.79% in the cleaned dataset. We validated this approach by comparing PCA, MDS and
316 phylogenomic trees using both cleaned and the uncleaned dataset (see
317 https://ecoevorxiv.org/47tka). A comprehensive investigation of this strategy including additional
318 datasets will be published in a separate paper (Cerca et al. *submitted*).

319 **Genomic trees and networks**

320 The phylogenomic tree (Figure 2C) shows a slightly different topology from the tree
321 obtained with selected molecular markers. The branches representing *Stygocapitella subterranea*,
322 *S. westheidei*, and *S. josemariobrancoi* have a bootstrap support of 93, 99 and 100, respectively.
323 The tree topology is broadly similar to the selected marker phylogenetic tree, with *S.*
324 *josemariobrancoi* being sister to the clade comprising *S. subterranea* and *S. westheidei*. However,
325 strictly speaking, none of the species is recovered as monophyletic, since three individuals
326 identified as *S. josemariobrancoi* are not placed with *S. josemariobrancoi*. Specifically, 422 04
327 from Bristol Channel nests within *S. subterranea*, 422 05 from Bristol Channel nests within *S.*
328 *westheidei* and 401 03 from St. Efflam is positioned as sister to *S. subterranea* (Figure 2C,
329 individuals denoted by arrows). Importantly, mapping of shared pairwise data in the tree topology
330 suggests that these trends are not driven by missing data, since the branches representing the three
331 aforementioned individuals do not exhibit elevated levels of missing data (Figure 2C). The
332 UPGMA tree, which is solely built on a pairwise matrix of missing-data, shows that *S. subterranea*
333 and *S. josemariobrancoi* are generally separated, interlaced by individuals from *S. westheidei*
334 (Supplementary Figure 6). While this suggests that the three species have different amounts of
335 missing data, the fact that individuals are generally mixed suggests that missing data is not driving

336   phylogenetic reconstruction. For example, the three individuals resulting in a paraphyletic

337   reconstruction are not placed closely to their sister taxa in the tree, therefore indicating that missing

338   data has no influence in the paraphyletic position of these individuals (Supplementary Figure 6).

339   Finally, the generated phylogenomic consensus tree shows a similar topology to the that in Figure

340   2 (Supplementary Figure 7). The three samples causing paraphyly of the lineages in the

341   phylogenomic tree are placed within *S. subterranea* (Bristol Channel 422 04), as sister to the

342   lineage *S. josemariobrancoi* and *S. westheidei* (St. Efflam 401 03), and as the first branch of *S.*

343   *josemariobrancoi* (Bristol Channel 422 05).

344   In the phylogenetic network, *Stygocapitella westheidei* is separated from the remaining two

345   species, occupying a separate and relatively compact area of the network. *Stygocapitella*

346   *subterranea* is mostly confined to one small section of the network, however, three individuals are

347   very close to the center of the network (398 04, 398 08 and 398 09 from Keitum; Figure 3). In line

348   with the results from the phylogenomic tree, Bristol Channel 422 04, which is identified as part of

349   *S. josemariobrancoi*, is nested within *S. subterranea* in the network. *Stygocapitella*

350   *josemariobrancoi*, on the other hand, is clearly stretched and set apart in the network, occupying

351   a large area (Figure 3). While most individuals are nested within a condensed and remote portion

352   of the network, the individuals Bristol Channel 422 05, St Efflam 401 03 and 401 04 lie in an

353   intermediate position between the center of the network and the majority of individuals from *S.*

354   *josemariobrancoi* (Figure 3). This is broadly in agreement with the phylogenomic tree, which

355   shows Bristol Channel 422 05 nested with *S. westheidei* (Figure 2C) and St. Efflam 401 03 sister

356   to *S. subterranea* (Figure 2C). The distance between most individuals belonging to *S.*

357   *josemariobrancoi* and the rest of the network (centre of the network, and the two remaining

358   species) suggests there is a greater degree of differentiation in this species.

359   **Population structure, differentiation and summary statistics**

360   The PCA separates the three species across the first two principal components, (which

361   together explain 30.4% of the variance; Figure 4A). Three individuals stand out, including Bristol

362   Channel 422 04 (labelled as *S. josemariobrancoi*) which is placed closely with *S. subterranea*

363   individuals, Bristol Channel 422 05 which occupies an intermediate position between *S. westheidei*

364   and *S. josemariobrancoi*, and Lubec 428 02 which is relatively distant from the remaining *S.*

365   *westheidei* individuals. The multi-dimensional scaling plot, which is less affected by missing data,

366   separates the species into three distinct clusters (MDS; Figure 4B). In coherence with the previous
367   analyses, we detect several taxa with intermediate positions: Bristol Channel 422 04 (labelled as
368   *S. josemariobrancoi*) is closer to the *S. subterranea* cluster than to the *S. josemariobrancoi*;
369   Hoernum 169 09 (labelled as *S. josemariobrancoi*) in an intermediate position between these two
370   species; Bristol Channel 422 05, St. Efflam 401 04, 401 03, 401 05 (all labelled as *S.*
371   *josemariobrancoi*) are found in an intermediate position between *S. josemariobrancoi* and *S.*
372   *westheidei*; Lubec 429 02 (labelled as *S. westheidei*) is also distant from the *S. westheidei* cluster,
373   being relatively close to Bristol Channel 422 05 (Figure 4B).

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

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

393   Summary statistics suggest that populations of *S. josemariobrancoi* have a higher degree
394   of genetic variation. Waterson's estimate, S, ranges between 2.11-2.39 in *S. subterranea* and
395   between 1.81-2.35 in *S. westheidei,* and between 1.89-6.83 in *S. josemariobrancoi,* with Bristol

396   Channel (S = 4.92), and Gravesend (S = 6.93; Table 3). This is similar when analysing nucleotide

397   polymorphisms (π), with *S. subterranea* ranging between 0.002-0.0037, *S. westheidei* between

398   0.0021-0.0034, but *S. josemariobrancoi* ranging between 0.0026-0.0099 again with Gravesend and

399   Bristol Channel as outliers (Gravesend π = 0.0099, Bristol Channel π = 0.0086; Table 3).

400   Interestingly, the sympatric sites do not reveal any signal of higher polymorphism, as it would be

401   expected in scenarios of on-going hybridization. For instance, π = 0.0033 and S = 2.13 for *S.*

402   *subterranea* in Musselburgh, and π = 0.0045 and S = 2.49 for *S. josemariobrancoi* in Musselburgh.

403   Hausstrand, for which we were only able to obtain data for *S. josemariobrancoi* retrieved a π =

404   0.0038 and a S = 2.55, while the population Lubec of *S. westheidei* shows π = 0.0029 and S = 2.15.

405   Tajima's D excludes the possibility for bottlenecks, as none of the populations exhibits significant

406   Tajima's D (i.e. values below -2 or above +2; Table 3).

407   Simulation of demographic scenarios suggests admixture may have happened in ancestral

408   branches (Figure 6). All the best supported scenarios are provided in

409   https://github.com/jcerca/Papers/blob/main/Stygocapitella_PeerJ/11_fscScenarios/   and   as

410   supplementary. Three of the top four most supported scenarios using likelihood suggest ancient

411   admixture: geographic gene flow (i.e. gene flow between the two ancestral branches and

412   afterwards between *S. subterranea* and *S. josemariobrancoi*), ancient gene flow (i.e. gene flow

413   between the two ancestral branches), ancient gene flow and between western lineages (i.e. gene

414   flow between the two ancestral branches and afterwards between *S. westheidei* and *S.*

415   *josemariobrancoi*). The remaining scenario suggests no gene flow. The most supported of these

416   scenarios was the geographic gene flow, which includes coalescent times of 411,402 generations

417   for the first coalescent event (presumably, *Stygocapitella* has a generation time of a single year),

418   and 21,680,033 for the second coalescent event. The ancient gene flow scenario, which was the

419   second most supported, includes estimates of 585,534 generations for the first coalescent event

420   and 16,749,615 generations for the second coalescent event. The no gene flow scenario is the 3rd

421   most supported scenario, but the first and second coalescent events are suggested to have occurred

422   451 and 12,834 years ago, being clearly at odds with previous evidence suggesting these lineages

423   diverged millions of years ago (Struck et al. 2017; Cerca et al. 2020a). The fourth most supported

424   scenario, ancient gene flow and gene flow between western lineages, has coalescent times of

425   3,566,294 and 18,353,555 (Figure 6). Finally, while the AIC assessment provide slightly different

426 results (Supplementary Figure 8), the second and third most supported scenarios are the geographic

427 gene flow and ancient gene flow; being thus in agreement with the likelihood results.

428 **Discussion**

429   Morphological similarity through extended periods of times, or stasis, has been

430 hypothesized to occur under three possible scenarios underlied by genetics: homogenous genetic

431 variation (due to e.g. ILS, hybridization), genetic constraints (e.g. pleiotropy), and lack of genetic

432 variation (e.g. bottlenecks and founder effects) (Futuyma 2010). While our sampling design does

433 not account for genetic constraints, we study the evolutionary history of *Stygocapitella* species

434 seeking to determine signals of loss of genetic variation or shared genetic variance. We find that

435 the three morphologically similar *Stygocapitella* species herein studied share genetic variation and

436 exclude the possibility of recent bottlenecks or recent admixture. Demographic and admixture

437 analysis reveal signatures of ancestral admixture and incomplete lineage sorting during the

438 divergence of these three species. We discuss the possible implications of these processes to the

439 genomic underpinnings of indistinguishable morphology among cryptic species.

440   **Whole genome amplification and the generation of RADseq data**

441   We show that WGA combined with RAD sequencing may become an important tool for

442 microscopic eukaryote genomics. RADseq library preparation typically requires 200-500 ng of

443 DNA per individual, yet DNA extraction of a single *Stygocapitella* individual typically yields 10-

444 100 ng of DNA, thus representing a challenge to obtain genome-level data. One potential solution

445 is to pool individuals. However, pooling may not be ideal when dealing with morphologically

446 similar species, especially when they occur in sympatry, as observed in *Stygocapitella*, since the

447 identification of individuals based solely on morphology may be impossible. Arguably, one of the

448 major advantages of RADseq is to open up population genomics and phylogenomics as approaches

449 for non-model systems at an affordable cost. In such systems, experimental designs may benefit

450 from the inclusion of the largest number of individuals possible, which encompass the whole

451 spectrum of populations or species to determine species boundaries, phylogeography, population

452 structure and the phylogeny. Pooling individuals from different species or populations together

453 may lead to an incorrect inference of the phylogeny when species boundaries are not known, but

454 also skew allelic variation. Consequently, pooling of individuals may result in difficulties during

455   data processing and interpretation, and may require extra efforts such as barcoding of individuals
456   before pooling. In the view of these challenges, we optimized and applied a whole-genome
457   amplification protocol to obtain genome level data, thus confirming its potential for population
458   genomic-inference and phylogenetics (de Medeiros and Farrell 2018, 2020).

459   **Bottlenecks, meiofaunal dispersal and morphological similarity**

460   The few phylogeographic studies available for meiofaunal organisms have generally
461   detected founder effects and bottlenecks, and have discussed how colonization dynamics may be
462   determined by a series of bottlenecks (Casu and Curini-Galletti 2006; Derycke et al. 2007; Andrade
463   et al. 2011), however, we find no evidence for bottlenecks in this dataset when using summary
464   statistics and Tajima's D. Evidence for the prevalence of bottlenecks in meiofauna has been further
465   supported by experimental evidence which showed that colonization of new areas may be
466   characterized by founder effects and bottlenecks, which, in turn, are expected to shape the genetic
467   differentiation of meiofaunal populations (Derycke et al. 2007, 2013). Broadly, this follows
468   hypotheses on marine-invertebrate biogeography which suggests that repeated extinction and
469   recolonization dynamics may be involved in shaping genetic differentiation in populations and
470   species (Andrade et al. 2011; Derycke et al. 2013). We find no indication of bottlenecks in
471   *Stygocapitella,* as suggested by non-significant Tajima's D values (Table 3), even though that part
472   of the sampling distribution of *Stygocapitella* included areas which were glaciated only 10,000
473   years ago (Wares and Cunningham 2001). Two possibilities may explain this scenario and support
474   our findings. First, the hypothesis that meiofauna disperse through a series of bottlenecks may
475   require more evidence. Indeed, when scoring a total of 752 papers, we could only detect 48 studies
476   focusing on biogeography and 7 on evolutionary biology (including population genetics) (Cerca
477   et al. 2018), thus suggesting that meiofaunal biogeography is in its early days and more studies are
478   needed for more solid conclusions. This hypothesis may also be at odds with evidence that
479   meiofauna may be, indeed, good dispersers (reviewed in Cerca et al. 2018). Even if dispersal is
480   carried by only a limited group of individuals, concomitant with the idea of a founder effect and
481   bottleneck colonization, the dispersal of more organisms from 'source populations' (multiple
482   waves of dispersal) through time would eventually homogenize genetic variation in newly
483   colonized areas. Second, to the best of our knowledge, this is the first work to focus on population-
484   genomic level data in meiofauna. Typically, works have focused on sequencing a limited

485   combination of mitochondrial and nuclear genes (e.g. (Derycke et al. 2005, 2007, 2008; Kieneke

486   et al. 2012; Leasi and Norenburg 2014). While non-recombining data, such as mitochondrial

487   markers, may provide an ideal indication for the occurrence of bottlenecks, these effects should be

488   confirmed on complementary genomic regions. For instance, using the 16S mitochondrial marker,

489   we previously detected single haplotypes in populations of *S. subterreanea* or *S. josemariobrancoi*.

490   These populations had a statistically significant Tajima's D (*S. josemariobrancoi* individuals from

491   Bristol Channel; *S. subterranea* individuals from Glenancross) (Cerca et al. 2020b). While this

492   pattern was in conflict with that of the nuclear ITS1 (Cerca et al. 2020a), it is further rejected when

493   using genome-level data, which provides a more comprehensive, and independent, assessment of

494   genomic variance. This suggests that the dispersal-by-bottlenecks idea in meiofauna warrants more

495   data, and that biogeography of meiofauna will benefit from more genomic studies.

496   Overall, we suggest that morphological similarity in *Stygocapitella* is unlikely to result

497   from the lack of standing genetic variation due to re-occurring bottlenecks. Under this hypothesis,

498   it is expected that bottlenecks reduce genetic variation, which will in turn limit morphological

499   evolution, thus leading to stasis (Futuyma 2010). Given the lack of evidence for recent bottlenecks

500   in *Stygocapitella*, this does not seem plausible. In addition to the evidence for the lack of recent

501   bottlenecks, the fact that the remaining 8 *Stygocapitella* species live in similar habitats and are

502   distributed throughout the world (Cerca et al. 2020a) indirectly suggests that bottlenecks may not

503   be typical in the evolutionary history of the group.

504   **Evidence for Incomplete Lineage Sorting and Admixture**

505   We find clear evidence for shared genetic variation in *Stygocapitella*. The most

506   conspicuous evidence for this comes from the admixture analysis, which clearly demonstrates

507   admixed populations in the three species (Figure 5). This evidence is further supported by

508   individuals with intermediate positions in the MDS – a test which is robust to missing data (Figure

509   4). However, several evidences do not support a preponderant role of recent admixture. First, we

510   obtained no evidence for admixture when using F-statistics, since we find only positive F-values

511   (Table 1). Second, contrary to the expectation of ongoing gene flow, we do not observe higher

512   levels of heterozygosity in sympatric populations (Lubec in the USA, Musselburgh in Scotland,

513   Hausstrand in Germany; Table 3) where individuals of different species are found in the same

514   sediment sample in close proximity (volume ranging from 50-500 cm^3). Third, admixture often

515   generates incongruence between mitochondrial and nuclear markers (Melo-Ferreira et al. 2012;
516   Sloan et al. 2017), which is not seen in single-marker trees (Supplementary Figures 2-5). Fourth,
517   models with exclusive recent admixture are generally poorly supported by the demographic
518   analysis (Figure 6). In contrast, three out of the four most supported demographic scenarios suggest
519   ancient admixture, and one supported no gene flow at all (Figure 6). The scenario with no gene
520   flow inferred coalescent times of 451 and 13,834 generations or years (1 generation is expected to
521   be 1 year, Günter Purschke pers. comm)] which are not compatible with estimates of the splitting
522   age of the three *Stygocapitella* species (~5-30 million years ago; Cerca et al. 2020b). Given that
523   reduced times of coalescence are a typical signature of simulations that do not account for gene
524   flow, when it has occurred in empirical data (Leaché et al. 2019), it is likely that incomplete lineage
525   sorting alone cannot explain the patterns of shared variation among *Stygocapitella* species. In other
526   words, the demographic analysis supports a scenario that includes ancient admixture. The three
527   scenarios with ancient admixture vary in the presence or absence of admixture after the second
528   coalescence event (*S. subterranea* and *S. josemariobrancoi*): in one scenario, admixture is
529   exclusive to the ancestral branch; in the remaining two, gene flow between *S. josemariobrancoi*
530   and either *S. westheidei* or *S. subterranea* occur. Given the lack of support for on-going gene flow
531   between species by the F_{ST}, summary statistics, and F-statistics (Tables 1-3, Supplementary
532   Figures 1-5), admixture may have occurred immediately before or after the speciation event of *S.*
533   *westheidei* or *S. subterranea*, but not in recent times (i.e. the last generations). Furthermore, the
534   occurrence of ancient admixture can affect the inference of recent admixture when not take the
535   phylogeny into account (Malinsky et al. 2018; Ferreira et al. 2020), this may explain the
536   incongruence between some of our analysis. Therefore, while the demographic analysis suggests
537   the occurrence of admixture among *S. josemariobrancoi* and the other species, future studies are
538   necessary to confidently dissect and determine the role of recent gene flow in the system with
539   independent analyses. For example, these studies will benefit from using whole-genome data to
540   determine whether interspecific divergence in regions of the genome show gene-species tree
541   discordance, thereby dissecting ILS and recent hybridization (Joly et al. 2009; Giska et al. 2019).
542   Also, the demographic analysis favouring a preponderant role of ancient admixture does not
543   exclude the occurrence of ILS, and the beforementioned approach would also allow to clarify the
544   relative contribution of ILS and gene flow to shared patterns of variation among species. In sum,

545   to the extent that we can speculate, our data suggests that shared genetic variance is more likely
546   explained by an evolutionary history including incomplete lineage sorting and ancient geneflow.

547   Evidence for ancient admixture or incomplete lineage sorting is further seen in the
548   phylogenomic analysis. Phylogenetic approaches, which seek to reconstruct the evolutionary
549   history of lineages, often fail to resolve the evolutionary history and the 'true tree-topology' when
550   the taxa in question have high rates of incomplete lineage sorting or admixture (Kubatko and
551   Degnan 2007; Degnan and Rosenberg 2009), but incongruence may also result from tree-building
552   errors, paralogy or horizontal gene transfer (Scornavacca and Galtier 2017). We discard tree-
553   building errors based on the following evidence. First, individuals with intermediate positions in
554   the PCA and MDS correspond to those causing paraphyly in the trees. Second, when exploring the
555   effects of missing data through a) labelling the tree with % of missing data (Figure 2); and b)
556   constructing a cladogram based only on the shear % of missing data (i.e. UPGMA tree,
557   Supplementary Figure 6); we find the placement of intermediate individuals is not guided by
558   missing data. Should missing data determine their placement, we would expect these specimens to
559   nest in close proximity in the UPGMA tree. Finally, in the phylogenetic network, the individuals
560   which are also far removed from the remaining *S. josemariobrancoi* individuals, occupying central
561   positions or being paraphyletic in the phylogenetic network correspond to those paraphyletic in
562   the tree and in intermediate positions in the PCA and MDS.

563   In sum, incomplete lineage sorting and ancient hybridization are known to contribute to
564   levels of shared variation among species (Pease et al. 2016; Malinsky et al. 2018; Edelman et al.
565   2019), even at deep evolutionary levels (Song et al. 2015; Suh et al. 2015). The development of
566   tools which employ the substantial amount of modern genomic data has allowed separating cases
567   of ILS and ancient admixture, showing that ancient hybridization can have a strong impact in the
568   levels of shared variation among species complexes (Malinsky et al. 2018; Li et al. 2019; Taylor
569   and Larson 2019; Ferreira et al. 2020), even after several million years of divergence (Barth et al.
570   2020; Suvorov et al. 2020). Future work should employ multispecies-network methods or
571   coalescent simulations (Joly et al. 2009) to determine the relative role of ILS and ancient
572   admixture. However, it does not seem unlikely that both processes might have thus contributed to
573   levels of shared variation across 5-30 millions of the divergence of *Stygocapitella* (Cerca 2020).

574   **Incomplete lineage sorting and morphological similarity**

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

593   **Conclusions**

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

602   **Acknowledgements**

603   JC is grateful to Tim Worsfold, Andy Mackie, Henning Reiss, Lis Jørgensen for laboratory

604   space in the UK and Norway. We thank Audun Schrøder-Nielsen and Lisbeth Thorbek for

605   assistance in laboratory work, and Inês Modesta for fieldwork support. We are grateful to Diego

606   Fontaneto, Gerardo Perez-Ponce de Leon and an anonymous reviewer for their comments,

607   suggestions and critiques, which have led to the substantial improvement of this manuscript. We

608   acknowledge the use of Norwegian national infrastructure for high-performance computing and

609   storage via the projects NN9408K and NS9408K, respectively. Fieldwork was partly supported by

610   a Den Grevelige Hjelmstjerne-Rosencroneske Stiftelse ved UiO (JC), and THS was partly

611   supported by the EU Assemble program. A UiO:LifeScience internationalization support and a

612   Godfrey Hewitt mobility award allowed JC to visit Julian Catchen at the UIUC. JC dedicates this

613   paper to

Author contributions

615   JCe, MDM, THS designed this study. JCe and THS collected organisms in the field. JCe extracted

616   DNA and prepared RADseq libraries. JCe did data analysis with support of MR (population

617   genomics), JCa (RADseq), AR-C (RADseq), MSF (ILS), and THS (phylogenetics). JCe drafted

618   the manuscript, which was read, commented on and approved by all the remaining authors.

References

620   Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in

621   unrelated individuals. Genome Res 19:1655–1664. https://doi.org/10.1101/gr.094052.109

622   Altschul SF, Gish W, Miller W, et al (1990) Basic local alignment search tool. J Mol Biol

623   215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

624   Andrade SCS, Norenburg JL, Solferini VN (2011) Worms without borders: Genetic diversity

625   patterns in four Brazilian *Ototyphlonemertes* species (Nemertea, Hoplonemertea). Mar Biol

626   158:2109–2124. https://doi.org/10.1007/s00227-011-1718-3

627   Arnold B, Corbett-Detig RB, Hartl D, Bomblies K (2013) RADseq underestimates diversity and

628   introduces genealogical biases due to nonrandom haplotype sampling. Mol Ecol 22:3179–

629   3190. https://doi.org/10.1111/mec.12276

630   Astrin JJ, Stüben PE (2008) Phylogeny in cryptic weevils: Molecules, morphology and new

631   genera of western Palaearctic Cryptorhynchinae (Coleoptera: Curculionidae). Invertebr Syst

632   22:503–522. https://doi.org/10.1071/IS07057

633   Baird NA, Etter PD, Atwood TS, et al (2008) Rapid SNP discovery and genetic mapping using

634   sequenced RAD markers. PLoS One 3:e3376. https://doi.org/10.1371/journal.pone.0003376

635   Barth JMI, Gubili C, Matschiner M, et al (2020) Stable species boundaries despite ten million

636   years of hybridization in tropical eels. Nat Commun 11:1–13.

637   https://doi.org/10.1038/s41467-020-15099-x

638   Bernardo J (2011) A critical appraisal of the meaning and diagnosability of cryptic evolutionary

639   diversity, and its implications for conservation in the face of climate change. Clim Chang

640   Ecol Syst 380–438. https://doi.org/10.1017/CBO9780511974540.019

641   Bickford D, Lohman DJ, Sodhi NS, et al (2007) Cryptic species as a window on diversity and

642   conservation. Trends Ecol Evol 22:148–155. https://doi.org/10.1016/j.tree.2006.11.004

643   Camacho C, Coulouris G, Avagyan V, et al (2009) BLAST+: Architecture and applications.

644   BMC Bioinformatics 10:1–9. https://doi.org/10.1186/1471-2105-10-421

645   Casu M, Curini-Galletti M (2006) Genetic evidence for the existence of cryptic species in the

646   mesopsammic flatworm *Pseudomonocelis ophiocephala* (Rhabditophora: Proseriata). Biol

647   J Linn Soc 87:553–576. https://doi.org/10.1111/j.1095-8312.2006.00588.x

648   Cerca J, Meyer C, Purschke G, Struck TH (2020a) Delimitation of cryptic species drastically

649   reduces the geographical ranges of marine interstitial ghost-worms (Stygocapitella;

650   Annelida, Sedentaria). Mol Phylogenet Evol 143:106663.

651   https://doi.org/10.1016/j.ympev.2019.106663

652   Cerca J, Meyer C, Stateczny D, et al (2020b) Deceleration of morphological evolution in a

653   cryptic species complex and its link to paleontological stasis. Evolution (N Y) n/a:1–16.

654   https://doi.org/10.1111/evo.13884

655   Cerca J, Purschke G, Struck TH (2018) Marine connectivity dynamics: clarifying cosmopolitan

656   distributions of marine interstitial invertebrates and the meiofauna paradox. Mar Biol

657   165:123. https://doi.org/10.1007/s00227-018-3383-2

658   Chang CC, Chow CC, Tellier LCAM, et al (2015) Second-generation PLINK: Rising to the
659   challenge of larger and richer datasets. Gigascience 4:1–16. https://doi.org/10.1186/s13742-
660   015-0047-8

661   Chenuil A, Cahill AE, Délémontey N, et al (2019) Problems and questions posed by cryptic
662   species. A framework to guide future studies. pp 77–106

663   Chernomor O, Von Haeseler A, Minh BQ (2016) Terrace aware data structure for phylogenomic
664   inference from supermatrices. Syst Biol 65:997–1008.
665   https://doi.org/10.1093/sysbio/syw037

666   Crotti M, Barratt CD, Loader SP, et al (2019) Causes and analytical impacts of missing data in
667   RADseq phylogenetics: Insights from an African frog (Afrixalus). Zool Scr 48:157–167.
668   https://doi.org/10.1111/zsc.12335

669   Danecek P, Auton A, Abecasis G, et al (2011) The variant call format and VCFtools.
670   Bioinformatics 27:2156–2158. https://doi.org/10.1093/bioinformatics/btr330

671   De León GPP, Nadler SA (2010) What we don't recognize can hurt us: A plea for awareness
672   about cryptic species. J Parasitol 96:453–464. https://doi.org/10.1645/GE-2260.1

673   de Medeiros BAS, Farrell BD (2018) Whole genome amplification in double-digest RAD-seq
674   results in adequate libraries but fewer sequenced loci. PeerJ.
675   https://doi.org/10.7717/peerj.5089

676   de Medeiros BAS, Farrell BD (2020) Evaluating insect-host interactions as a driver of species
677   divergence in palm flower weevils. Commun Biol. https://doi.org/10.1038/s42003-020-
678   01482-3

679   Degnan JH, Rosenberg NA (2009) Gene tree discordance, phylogenetic inference and the
680   multispecies coalescent. Trends Ecol Evol 24:332–340.
681   https://doi.org/10.1016/j.tree.2009.01.009

682   Derycke S, Backeljau T, Moens T (2013) Dispersal and gene flow in free-living marine

683   nematodes. Front Zool 10:1. https://doi.org/10.1186/1742-9994-10-1

684   Derycke S, Remerie T, Backeljau T, et al (2008) Phylogeography of the *Rhabditis* (Pellioditis)

685   marina species complex: Evidence for long-distance dispersal, and for range expansions and

686   restricted gene flow in the northeast Atlantic. Mol Ecol 17:3306–3322.

687   https://doi.org/10.1111/j.1365-294X.2008.03846.x

688   Derycke S, Remerie T, Vierstraete A, et al (2005) Mitochondrial DNA variation and cryptic

689   speciation within the free-living marine nematode *Pellioditis marina*. Mar Ecol Prog Ser

690   300:91–103. https://doi.org/10.3354/meps300091

691   Derycke S, Vynckt R Van, Vanoverbeke J, et al (2007) Colonization patterns of Nematoda on

692   decomposing algae in the estuarine environment: Community assembly and genetic

693   structure of the dominant species *Pellioditis marina*. Limnol Oceanogr 52:992–1001.

694   https://doi.org/10.4319/lo.2007.52.3.0992

695   Dornburg A, Federman S, Eytan RI, Near TJ (2016) Cryptic species diversity in sub-Antarctic

696   islands: A case study of Lepidonotothen. Mol Phylogenet Evol 104:32–43.

697   https://doi.org/10.1016/j.ympev.2016.07.013

698   Dufresnes C, Strachinis I, Suriadna N, et al (2019) Phylogeography of a cryptic speciation

699   continuum in Eurasian spadefoot toads (Pelobates). Mol Ecol 3257–3270.

700   https://doi.org/10.1111/mec.15133

701   Edelman NB, Frandsen PB, Miyagi M, et al (2019) Genomic architecture and introgression

702   shape a butterfly radiation. Science (80-) 599:594–599

703   Eldredge N, Gould SJ (1972) Punctuated Equilibria: An alternative to phylogetic gradualism. In:

704   Schopf TJM (ed) Models in paleobiology. Freeman, Cooper and Co., San Francisco., pp

705   82–115

706   Erlank E, Koekemoer LL, Coetzee M (2018) The importance of morphological identification of

707   African anopheline mosquitoes (Diptera: Culicidae) for malaria control programmes. Malar

708   J 17:1–7. https://doi.org/10.1186/s12936-018-2189-5

709   Estes S, Arnold SJ (2007) Resolving the paradox of stasis: models with stabilizing selection

710   explain evolutionary divergence on all timescales. Am Nat 169:227–244.

711   https://doi.org/10.1086/510633

712   Ferreira MS, Jones MR, Callahan CM, et al (2020) The legacy of recurrent introgression during

713   the radiation of hares. Syst Biol

714   Fišer C, Robinson CT, Malard F (2018) Cryptic species as a window into the paradigm shift of

715   the species concept. Mol Ecol 27:613–635. https://doi.org/10.1111/mec.14486

716   Futuyma DJ (2010) Evolutionary constraint and ecological consequences. Evolution (N Y)

717   64:1865–1884. https://doi.org/10.1111/j.1558-5646.2010.00960.x

718   Gautier M, Gharbi K, Cezard T, et al (2013) The effect of RAD allele dropout on the estimation

719   of genetic variation within and between populations. Mol Ecol 22:3165–3178.

720   https://doi.org/10.1111/mec.12089

721   Giska I, Farelo L, Pimenta J, et al (2019) Introgression drives repeated evolution of winter coat

722   color polymorphism in hares. Proc Natl Acad Sci U S A 116:24150–24156.

723   https://doi.org/10.1073/pnas.1910471116

724   Golombek A, Tobergte S, Nesnidal MP, et al (2013) Mitochondrial genomes to the rescue –

725   Diurodrilidae in the myzostomid trap. Mol Phylogenet Evol 68:312–326

726   Gould SJ (2002) The structure of evolutionary theory

727   Hansen TF, Houle D (2004) Evolvability, stabilizing selection, and the problem of stasis. In:

728   Pigliucci M, Preston K (eds) Phenotypic integration: studying the ecology and evolution of

729   complex phenotypes. Oxford University Press, New York, pp 130–154

730   Hillis DM, Dixon MT (1991) Ribosomal DNA: Molecular evolution and phylogenetic inference.

731   Q Rev Biol 66:411–453

732   Hoang DT, Chernomor O, von Haeseler A, et al (2017) Ufboot2: Improving the ultrafast

733   bootstrap approximation. 35:518–522. https://doi.org/10.5281/zenodo.854445

734   Hodel RGJ, Chen S, Payton AC, et al (2017) Adding loci improves phylogeographic resolution

735   in red mangroves despite increased missing data: Comparing microsatellites and RAD-Seq

736   and investigating loci filtering. Sci Rep 7:1–14. https://doi.org/10.1038/s41598-017-16810-
737   7

738   Huang H, Lacey Knowles L (2016) Unforeseen consequences of excluding missing data from
739   next-generation sequences: Simulation study of rad sequences. Syst Biol 65:357–365.
740   https://doi.org/10.1093/sysbio/syu046

741   Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. Mol
742   Biol Evol 23:254–267. https://doi.org/10.1093/molbev/msj030

743   Joly S, McLenachan PA, Lockhart PJ (2009) A statistical approach for distinguishing
744   hybridization and incomplete lineage sorting. Am Nat 174:. https://doi.org/10.1086/600082

745   Jombart T, Ahmed I (2011) adegenet 1.3-1: New tools for the analysis of genome-wide SNP
746   data. Bioinformatics 27:3070–3071. https://doi.org/10.1093/bioinformatics/btr521

747   Kalyaanamoorthy S, Minh BQ, Wong TKF, et al (2017) ModelFinder: Fast model selection for
748   accurate phylogenetic estimates. Nat Methods 14:587–589.
749   https://doi.org/10.1038/nmeth.4285

750   Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7:
751   Improvements in performance and usability. Mol Biol Evol 30:772–780.
752   https://doi.org/10.1093/molbev/mst010

753   Kieneke A, Martínez Arbizu PM, Fontaneto D (2012) Spatially structured populations with a low
754   level of cryptic diversity in European marine Gastrotricha. Mol Ecol 21:1239–54.
755   https://doi.org/10.1111/j.1365-294X.2011.05421.x

756   Knowlton N (1993) Sibling species in the sea. Annu Rev Ecol Syst 24:189–216

757   Korshunova T, Martynov A, Bakken T, Picton B (2017) External diversity is restrained by
758   internal conservatism: New nudibranch mollusc contributes to the cryptic species problem.
759   Zool Scr 46:683–692. https://doi.org/10.1111/zsc.12253

760   Kubatko LS, Degnan JH (2007) Inconsistency of phylogenetic estimates from concatenated data
761   under coalescence. Syst Biol 56:17–24. https://doi.org/10.1080/10635150601146041

762   Kück P, Longo GC (2014) FASconCAT-G: Extensive functions for multiple sequence alignment
763   preparations concerning phylogenetic studies. Front Zool 11:1–8.
764   https://doi.org/10.1186/s12983-014-0081-x

765   Kück P, Meusemann K (2010) FASconCAT: Convenient handling of data matrices. Mol
766   Phylogenet Evol 56:1115–1118. https://doi.org/10.1016/j.ympev.2010.04.024

767   Kück P, Struck TH (2014) BaCoCa - A heuristic software tool for the parallel assessment of
768   sequence biases in hundreds of gene and taxon partitions. Mol Phylogenet Evol 70:94–98.
769   https://doi.org/10.1016/j.ympev.2013.09.011

770   Lavoué S, Miya M, Arnegard ME, et al (2011) Remarkable morphological stasis in an extant
771   vertebrate despite tens of millions of years of divergence. Proc Biol Sci / R Soc 278:1003–
772   1008. https://doi.org/10.1098/rspb.2010.1639

773   Leaché AD, Zhu T, Rannala B, Yang Z (2019) The Spectre of Too Many Species. Syst Biol
774   68:168–181. https://doi.org/10.1093/sysbio/syy051

775   Leasi F, Norenburg JL (2014) The necessity of DNA taxonomy to reveal cryptic diversity and
776   spatial distribution of meiofauna, with a focus on Nemertea. PLoS One 9:.
777   https://doi.org/10.1371/journal.pone.0104385

778   Lee CE, Frost BW (2002) Morphological stasis in the Eurytemora affinis species complex
779   (Copepoda: Temoridae). In: Hydrobiologia. pp 111–128

780   Lee KM, Kivela SM, Ivanov V, et al (2018) Information Dropout Patterns in Restriction Site
781   Associated DNA Phylogenomics and a Comparison with Multilocus Sanger Data in a
782   Species-Rich Moth Genus. Syst Biol 1–15. https://doi.org/10.1093/sysbio/syy029

783   Li G, Figueiró H V., Eizirik E, et al (2019) Recombination-Aware Phylogenomics Reveals the
784   Structured Genomic Landscape of Hybridizing Cat Species. Mol Biol Evol 36:2111–2126.
785   https://doi.org/10.1093/molbev/msz139

786   Malinsky M, Svardal H, Tyers AM, et al (2018) Whole-genome sequences of Malawi cichlids
787   reveal multiple radiations interconnected by gene flow. Nat Ecol Evol 2:1940–1955.
788   https://doi.org/10.1038/s41559-018-0717-x

789   Melo-Ferreira J, Boursot P, Carneiro M, et al (2012) Recurrent introgression of mitochondrial
790   DNA among hares (lepus spp.) revealed by species-tree inference and coalescent
791   simulations. Syst Biol 61:367–381. https://doi.org/10.1093/sysbio/syr114

792   Nadler SA, De Len GPP (2011) Integrating molecular and morphological approaches for
793   characterizing parasite cryptic species: Implications for parasitology. Parasitology
794   138:1688–1709. https://doi.org/10.1017/S003118201000168X

795   Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ (2015) IQ-TREE: A fast and effective
796   stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol
797   32:268–274. https://doi.org/10.1093/molbev/msu300

798   Novo M, Almodóvar A, Fernández R, et al (2010) Cryptic speciation of hormogastrid
799   earthworms revealed by mitochondrial and nuclear data. Mol Phylogenet Evol 56:507–512.
800   https://doi.org/10.1016/j.ympev.2010.04.010

801   Novo M, Almodóvar A, Fernández R, et al (2012) Appearances can be deceptive: Different
802   diversification patterns within a group of mediterranean earthworms (Oligochaeta,
803   Hormogastridae). Mol Ecol 21:3776–3793. https://doi.org/10.1111/j.1365-
804   294X.2012.05648.x

805   O'Leary SJ, Puritz JB, Willis SC, et al (2018) These aren't the loci you'e looking for: Principles
806   of effective SNP filtering for molecular ecologists. Mol Ecol 27:3193–3206.
807   https://doi.org/10.1111/mec.14792

808   Palumbi S, Martin A, Romano S, et al (1991) The simple fool's guide to PCR, version 2.

809   Pante E, Puillandre N, Viricel A, et al (2015) Species are hypotheses: avoid connectivity
810   assessments based on pillars of sand. Mol Ecol 24:525–544.
811   https://doi.org/10.1111/mec.13048

812   Paradis E, Schliep K (2018) Ape 5.0: An environment for modern phylogenetics and
813   evolutionary analyses in R. Bioinformatics 35:526–528.
814   https://doi.org/10.1093/bioinformatics/bty633

815   Paris JR, Stevens JR, Catchen JM (2017) Lost in parameter space: a road map for stacks.

816   Methods Ecol Evol 8:1360–1373. https://doi.org/10.1111/2041-210X.12775

817   Patterson N, Moorjani P, Luo Y, et al (2012) Ancient admixture in human history. Genetics
818   192:1065–1093. https://doi.org/10.1534/genetics.112.145037

819   Pease JB, Haak DC, Hahn MW, Moyle LC (2016) Phylogenomics Reveals Three Sources of
820   Adaptive Variation during a Rapid Radiation. PLoS Biol 14:1–24.
821   https://doi.org/10.1371/journal.pbio.1002379

822   Pérez-Ponce de León G, Poulin R (2016) Taxonomic distribution of cryptic diversity among
823   metazoans: not so homogeneous after all. Biol Lett 12:20160371.
824   https://doi.org/10.1098/rsbl.2016.0371

825   Peter BM (2016) Admixture, population structure, and f-statistics. Genetics 202:1485–1501.
826   https://doi.org/10.1534/genetics.115.183913

827   Peterson BK, Weber JN, Kay EH, et al (2012) Double digest RADseq: An inexpensive method
828   for de novo SNP discovery and genotyping in model and non-model species. PLoS One 7:.
829   https://doi.org/10.1371/journal.pone.0037135

830   Pfenninger M, Schwenk K (2007) Cryptic animal species are homogeneously distributed among
831   taxa and biogeographical regions. BMC Evol Biol 7:121. https://doi.org/10.1186/1471-
832   2148-7-121

833   Rice P, Longden L, Bleasby A (2000) EMBOSS: The European Molecular Biology Open
834   Software Suite. Trends Genet 16:276–277. https://doi.org/10.1016/S0168-9525(00)02024-2

835   Rochette N, Rivera-Colón A, Catchen JM (2019) Stacks2: Analytical methods for paired-end
836   sequencing improve RADseq-based population genomics. Mol Ecol.
837   https://doi.org/10.1111/mec.15253

838   Rochette NC, Catchen JM (2017) Deriving genotypes from RAD-seq short-read data using
839   Stacks. Nat Protoc 12:2640–2659. https://doi.org/10.1038/nprot.2017.123

840   Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, et al (2017) DnaSP 6: DNA sequence
841   polymorphism analysis of large data sets. Mol Biol Evol 34:3299–3302.

842   https://doi.org/10.1093/molbev/msx248

843 Santamaria CA, Mateos M, Dewitt TJ, Hurtado LA (2016) Constrained body shape among
844   highly genetically divergent allopatric lineages of the supralittoral isopod Ligia occidentalis
845   (Oniscidea). Ecol Evol 6:1537–1554. https://doi.org/10.1002/ece3.1984

846 Scornavacca C, Galtier N (2017) Incomplete lineage sorting in mammalian phylogenomics. Syst
847   Biol 66:112–120. https://doi.org/10.1093/sysbio/syw082

848 Sloan DB, Havird JC, Sharbrough J (2017) The on-again, off-again relationship between
849   mitochondrial genomes and species boundaries. Mol Ecol 26:2212–2236.
850   https://doi.org/10.1111/mec.13959

851 Smith MA, Eveleigh ES, McCann KS, et al (2011) Barcoding a quantified food web: Crypsis,
852   concepts, ecology and hypotheses. PLoS One 6:.
853   https://doi.org/10.1371/journal.pone.0014424

854 Song S, Liu L, Edwards S V, Wu S (2015) Erratum: Resolving conflict in eutherian mammal
855   phylogeny using phylogenomics and the multispecies coalescent model (Proceedings of the
856   National Academy of Sciences of the United States of America (2012) 109 (14942-14947)
857   (DOI:10.1073/pnas.1211733109)). Proc Natl Acad Sci U S A 112:E6079.
858   https://doi.org/10.1073/pnas.1518753112

859 Struck TH, Cerca J (2019) Cryptic species and their Evolutionary significance. eLS 1–9.
860   https://doi.org/10.1002/9780470015902.a0028292

861 Struck TH, Feder JL, Bendiksby M, et al (2018) Finding evolutionary processes hidden in
862   cryptic species. Trends Ecol Evol 1–11. https://doi.org/10.1016/j.tree.2017.11.007

863 Suh A, Smeds L, Ellegren H (2015) The dynamics of incomplete lineage sorting across the
864   ancient adaptive radiation of neoavian birds. PLoS Biol 13:1–18.
865   https://doi.org/10.1371/journal.pbio.1002224

866 Suvorov A, Kim BY, Wang J, et al (2020) Widespread introgression across a phylogeny of 155
867   Drosophila genomes. bioRxiv 2020.12.14.422758

868 Swift HF, Daglio LG, Dawson MN (2016) Three routes to crypsis: stasis, convergence, and
869   parallelism in the *Mastigias* species complex (Scyphozoa, Rhizostomeae). Mol Phylogenet
870   Evol 99:103–115. https://doi.org/10.1016/j.ympev.2016.02.013

871 Taylor SA, Larson EL (2019) Insights from genomes into the evolutionary importance and
872   prevalence of hybridization in nature. Nat Ecol Evol 3:170–177.
873   https://doi.org/10.1038/s41559-018-0777-y

874 Valtueña FJ, López J, Álvarez J, et al (2016) Scrophularia arguta, a widespread annual plant in
875   the Canary Islands: a single recent colonization event or a more complex phylogeographic
876   pattern? Ecol Evol 6:4258–4273. https://doi.org/10.1002/ece3.2109

877 Wada S, Kameda Y, Chiba S (2013) Long-term stasis and short-term divergence in the
878   phenotypes of microsnails on oceanic islands. Mol Ecol 22:4801–10.
879   https://doi.org/10.1111/mec.12427

880 Wares JP, Cunningham CW (2001) Phylogeography and historical ecology of the North Atlantic
881   intertidal. Evolution 55:2455–2469. https://doi.org/10.1111/j.0014-3820.2001.tb00760.x

882 Weber AA-T, Stöhr S, Chenuil A (2019) Species delimitation in the presence of strong
883   incomplete lineage sorting and hybridization. Mol Phylogenet Evol 131:240218.
884   https://doi.org/10.1101/240218

885 Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure.
886   Evolution (N Y) 38:1358–1370. https://doi.org/10.2307/2408641

887 Westheide W, Purschke G (1988) Organism processing. In: Higgins RP, Thiel H (eds)
888   Introduction to the study of meiofauna. Smithsonian Institution Press, Washington, pp 146–
889   160

890 Zanol J, Halanych KM, Struck TH, Fauchald K (2010) Phylogeny of the bristle worm family
891   Eunicidae (Eunicida, Annelida) and the phylogenetic utility of noncongruent 16S, COI and
892   18S in combined analyses. Mol Phylogenet Evol 55:660–676.
893   https://doi.org/10.1016/j.ympev.2009.12.024

894 Zuccarello GC, West JA, Kamiya M (2018) Non-monophyly of Bostrychia simpliciuscula

895   (Ceramiales, Rhodophyta): Multiple species with very similar morphologies, a revised

896   taxonomy of cryptic species. Phycol Res 66:100–107. https://doi.org/10.1111/pre.12207

897

898

899

Figure 1

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.

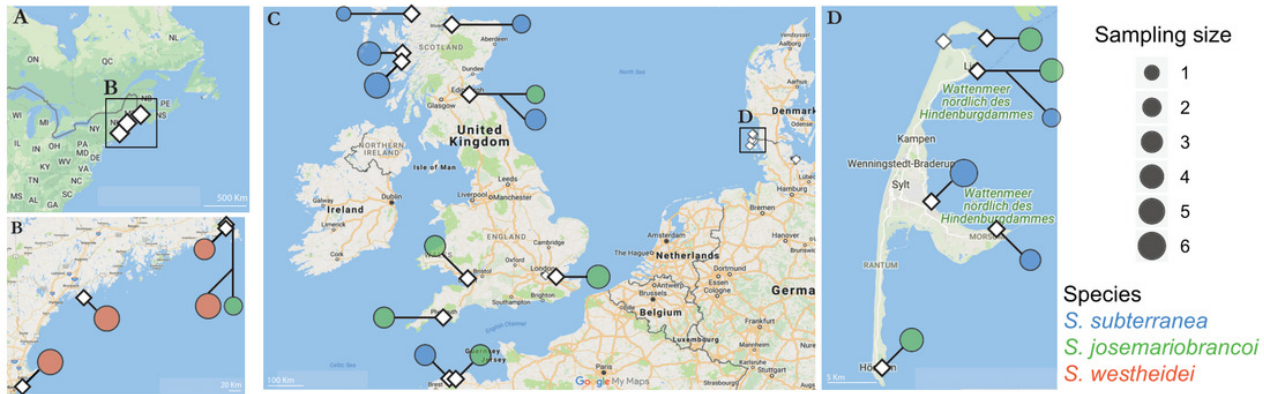



Figure 2

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 (2020). C) Phylogenomic tree based on 4,737 RADseq loci. Bootstrap support is provided for the main branches. 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.

Stygocapitella josemariobrancoi
Stygocapitella subterranea
Stygocapitella westheidei

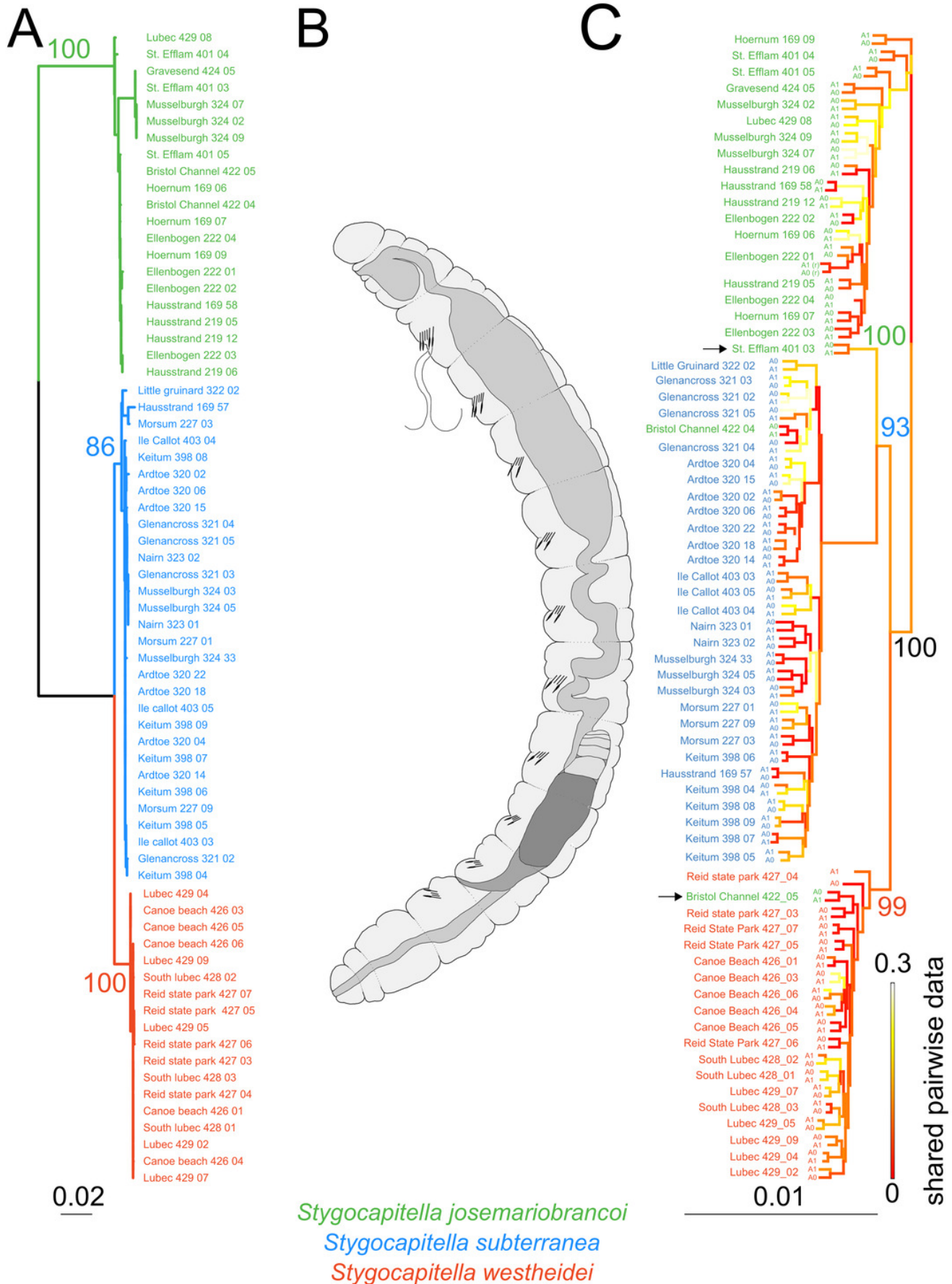



Figure 3

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.*

49152

Figure 4

Principal Component Analysis (A) and Multi-dimensional scaling (B) of 3,428 SNPs.

The percentage of explained variation is displayed with the axis for the PCA. Lineage (species) are given in different shapes, with colours represent populations. Specimens with 'intermediate positions' are highlighted both analyses, indicating potential shared generated variation between specimens.

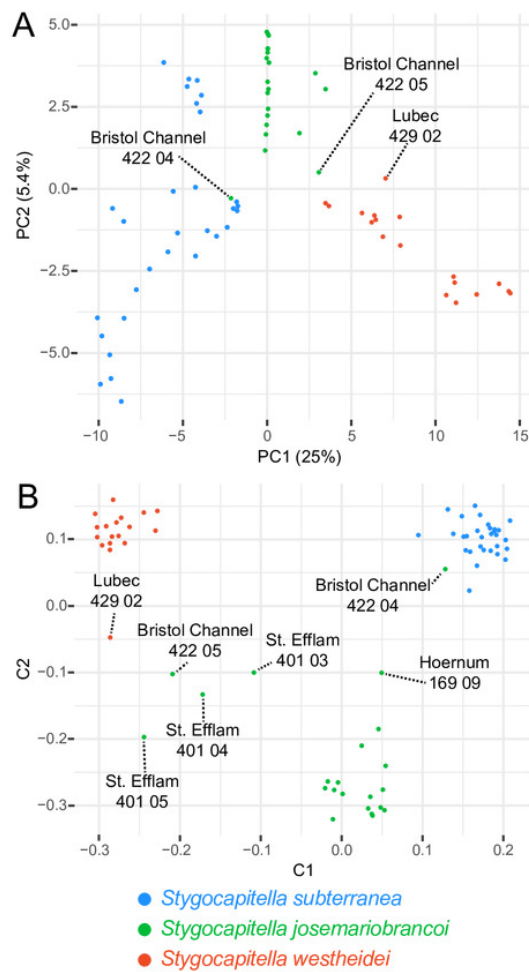



Figure 5

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 Figure 2 A and 2 B.

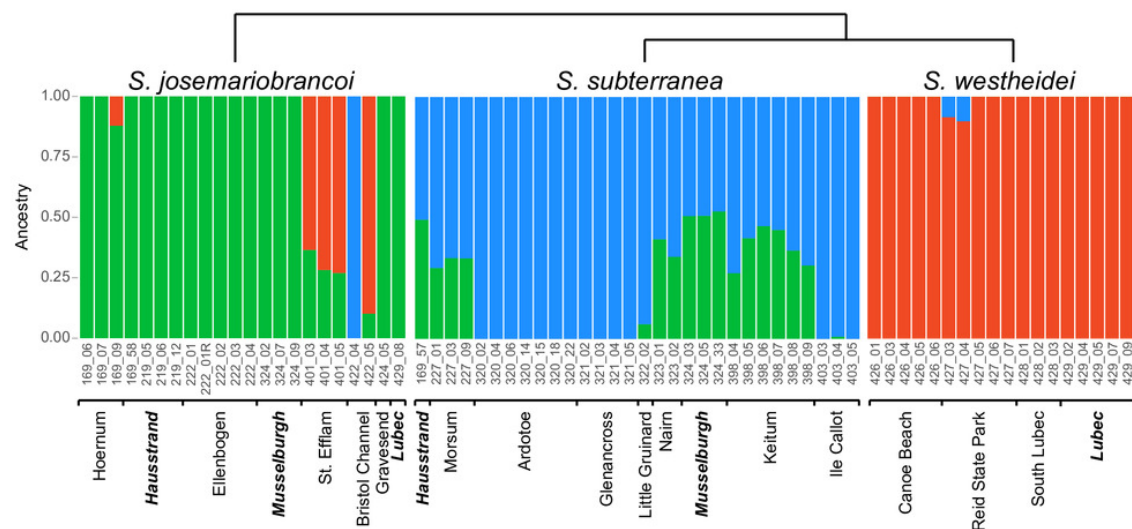
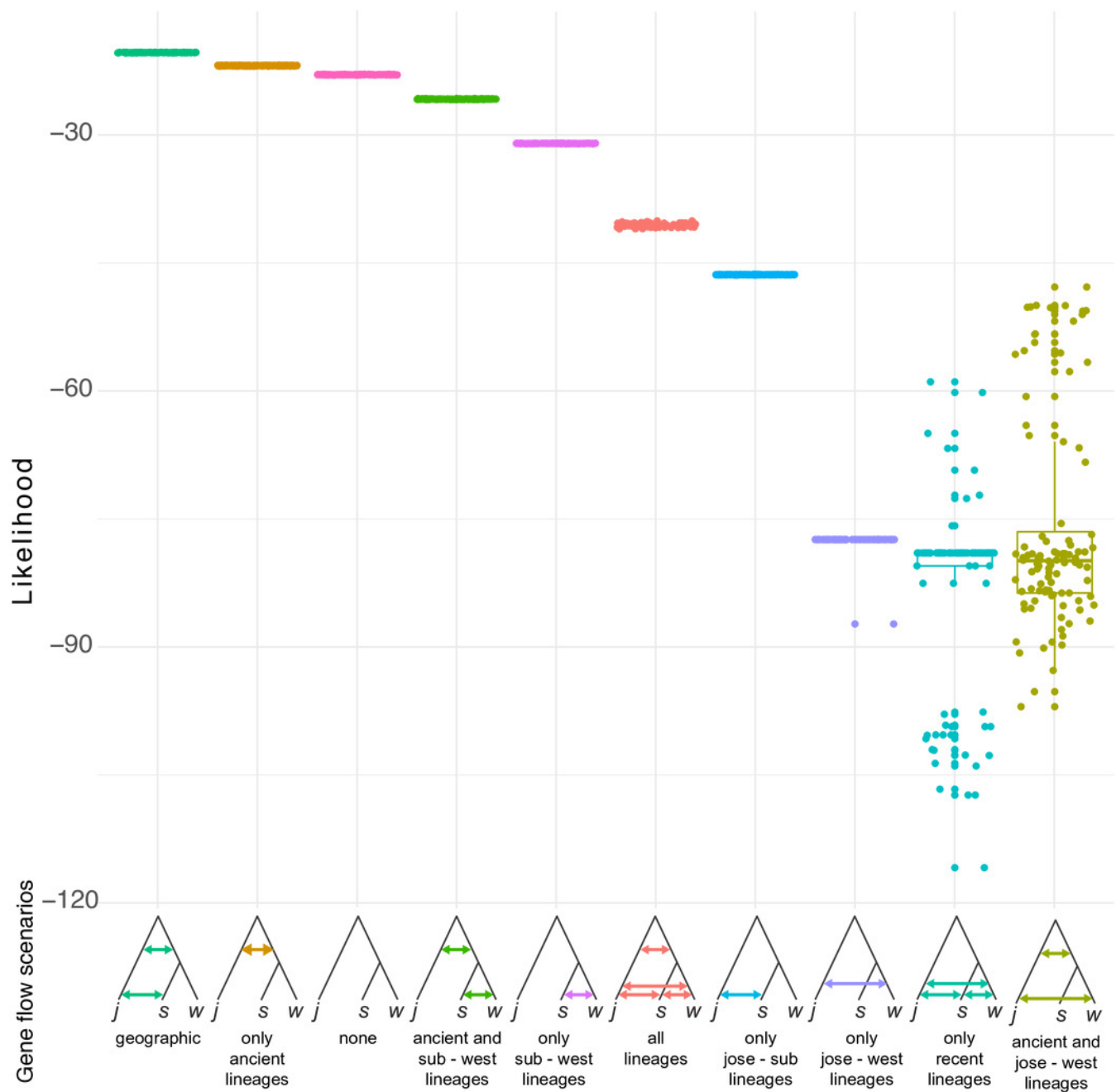



Figure 6

Figure 6. Demographic scenarios considered.

The likelihood of different demographic scenarios is displayed on the Y axis. Based on the estimated phylogeny (Figure 2), we modelled scenarios for (from left to right): 1) geographic gene flow (gene flowbetween *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*.

ƒ3 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 ƒ3 statistic, the standard error (SE) and a Z-score value calculated with jackknife is provided for each scenario based.

1 **Table 1 ƒ3 statistics testing for hybridization between lineages.** Each row represents a scenario where
2 two species are the source for admixture, and the third species is the target of hybridization. A ƒ3 statistic,
3 the standard error (SE) and a Z-score value calculated with jackknife is provided for each scenario based.

Source 1	Source 2	Target	ƒ3	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

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.

1   **Table 2** Weir-Cockherham Fst-estimate is provided in the lower part of the table, and the number of

2   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	-

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 π, 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.

1 **Table 3 Summary statistics** for the various analysed populations. For each site we provide the number of
2 specimens and chromossomes, number of loci considered, S (Waterson's estimate), the averaged π, and
3 the averaged Tajima's D. Analysed populations include those with >5 chromossomes, with the exception
4 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 π	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