Species delimitation and mitogenome phylogenetics in the subterranean genus *Pseudoniphargus* (Crustacea: Amphipoda)

Morten Stokkan\(^a\), José A. Jurado-Rivera\(^b\), Pedro Oromí\(^c\), Carlos Juan\(^a,b\), Damià Jaume\(^a\), Joan Pons\(^a,⁎\)

\(^a\) IMEDEA (CSIC-UIB), Mediterranean Institute for Advanced Studies, C/ Miquel Marqués 21, Esportes, 07190 Balearic Islands, Spain

\(^b\) Dept. of Biology, Universitat de les Illes Balears, Ctra. Valldemossa km 7.5, Palma 07122, Balearic Islands, Spain

\(^c\) Dept. of Animal Biology, Edaphology and Geology, Universidad de La Laguna, Avda. Astrofísico Francisco Sánchez, s/n. Campus de Anchieta, Ap. correos 456, La Laguna 38206, Tenerife, Spain

**A B S T R A C T**

The amphi-Atlantic distributions exhibited by many thalassoid stygobiont (obligate subterranean) crustaceans have been explained by fragmentation by plate tectonics of ancestral shallow water marine populations. The amphipod stygobiont genus *Pseudoniphargus* is distributed across the Mediterranean region but also in the North Atlantic archipelagos of Bermuda, Azores, Madeira and the Canaries. We used species delimitation methods and mitogenome phylogenetic analyses to clarify the species diversity and evolutionary relationships within the genus and timing their diversification. Analyses included samples from the Iberian Peninsula, northern Morocco, the Balearic, Canarian, Azores and Madeira archipelagoes plus Bermuda. In most instances, morphological and molecular-based species delimitation analyses yielded consistent results. Notwithstanding, in a few cases either incipient speciation with no involvement of detectable morphological divergence or species crypticism were the most plausible explanations for the disagreement found between morphological and molecular species delimitations. Phylogenetic analyses based on a robust calibrated mitochondrial tree suggested that *Pseudoniphargus* lineages have a younger age than for other thalassoid amphipods displaying a disjunct distribution embracing both sides of the Atlantic Ocean. A major split within the family was estimated to occur at the Paleocene, when a lineage from Northern Iberian Peninsula diverged from the rest of pseudoniphargids. Species diversification in the peri-Mediterranean area was deduced to occur in early Miocene to Tortonian times, while in the Atlantic islands it started in the Pliocene. Our results show that the current distribution pattern of *Pseudoniphargus* resulted from a complex admix of relatively ancient vicariance events and several episodes of long-distance dispersal.

1. Introduction

Cryptic lineage complexes, defined as groups of genetically distinct but morphologically similar or indistinguishable species, have been reported in many taxonomic groups and environments (Pfenninger and Schwenk, 2007). They are common among amphipod crustaceans, one of the most diverse and ecologically important groups of invertebrates in continental waters (Witt and Hebert, 2000; Witt et al., 2006; Fišer et al., 2018; Copiłaş-Ciocianu et al., 2018). Here, the frequency of cryptic species is particularly high among obligate subterranean (stygobiont) forms, where the strong directional selection imposed by the extremely demanding habitat conditions favors the emergence of morphological convergence (Lefebure et al., 2006, 2007; Delić et al., 2017). In this regard, the shared loss or reduction of eyes and of body pigmentation, together with the elongation of appendages, are only the gross and more remarkable manifestations of this syndrome among cave amphipods (Barr and Holsinger, 1985).

*Pseudoniphargus* Chevreux, 1901 and the monotypic *Parapsuedoniphargus* Notenboom, 1988 are the only members of the amphipod family Pseudoniphargidae Karaman, 1993. Pseudoniphargids are exclusively stygobiont and occur in a broad array of subterranean habitats, spanning from brackish-water coastal wells and caves to the hyporheic medium of mountain streams located up to 1420 m above sea level (Stock, 1980; Notenboom, 1987a). Despite none of its members has ever been recorded in ordinary marine habitats, the family is considered to be of marine (thalassoid) origin. This is based on (i) the presence of several species in transitional habitats with the sea, such as brackish-water anchialine caves or intertidal coastal springs (Stock, 1980; Stock et al., 1986; Stock and Abreu, 1992; Bréhier and Jaume, 2009); (ii) the occurrence of limnic species on islands never connected to continental landmasses; and (iii) the limitation of continental records of the family to occur only on areas formerly occupied by epicontinental seas (Notenboom, 1991). *Pseudoniphargus* currently comprises 71 species (Horton et al., 2018), all showing an extremely localized distribution and often representing single-site endemics (Stock, 1980; Notenboom, 1986, 1987a, 1987b; Karaman and Ruflo, 1989; Jaume, 1991; Messouli et al., 2006; Bréhier and Jaume, 2009; Stokkan et al., 2018). The genus is present in...
North Africa (Morocco and Algeria), the Mediterranean region—including most of its islands—, the Iberian Peninsula plus the North Atlantic archipelagos of Canaries, Madeira and Azores, whereas two species occur on Bermuda, on the west side of the mid-Atlantic rift (Stock et al., 1986). *Pseudoniphargus* is however absent from the American continent, in contrast to the distribution shown by other thalassoid stygobiont crustaceans such as *Metacrangonyx* amphipods or *Tethysbaena* thermosbaenaceans, present on both shores of the Atlantic (Jaume and Christenson, 2001; Wagner, 1994).

The occupation of inland subterranean waters by *Pseudoniphargus* is assumed to result from the stranding of populations of a marine ancestor after the withdrawal of the sea during periods of marine regression (Stock, 1977, 1980). This process led to the isolation of populations and their eventual speciation on the newly emerged land, to the point that the branching pattern within a determinate *Pseudoniphargus* lineage might be synchronous with the sequence of regression events through which the stygobionts became isolated from their marine ancestor (Stock, 1980; Notenboom, 1988, 1991).

*Pseudoniphargus*, as most stygobionts, have extremely reduced long-range dispersal due to intrinsic biological and ecological features. On one side, the lack of vision and of body pigmentation increases considerably their vulnerability to predation, competition or to damaging radiations in epigean habitats. Furthermore, as in the rest of amphipods and other groups such as the thermosbaenaceans, females brood embryos in a marsupium until they complete their development. No eggs or free-swimming larvae are released to the water column to be transported passively by sea currents, which is one of the ordinary ways through which aquatic invertebrates attain a trans-oceanic distribution. In these subterranean forms, dispersal is so reduced that there is virtually no gene flow between neighboring anchialine cave populations (Bauzà-Ribot et al., 2011; Cánovas et al., 2016).

Under these assumptions, the amphi-Atlantic distribution displayed by *Pseudoniphargus* and other thalassoid stygobionts is better explained in terms of vicariance rather than of dispersal. It has been proposed that their current distribution would be the outcome of the fragmentation by plate tectonics of ancient, ancestral shallow-water marine populations.
pelagoes, plus Bermuda (Fig. 1; Supplementary Table S1). The dataset
northern Morocco, the Balearic, Canarian, Azores and Madeira archi-
(mostly wells and caves) distributed along the Iberian Peninsula,

2. Material and methods
dellah distance dispersal in the biogeographic history of
pothesis and an approximate time frame for species diversi-
coding mitochondrial genes (mitogenomics). This phylogenetic hy-
ephalon and an approximate time frame for species diversifica-

Here we present a phylogenetic survey of Pseudoniphargus including
representatives from all major geographical areas where the genus has
been recorded. We use a large mitochondrial cytochrome oxidase sub-
unit I (coxI) sequence dataset to explore species diversity under several
species delimitation criteria and establish how widespread is crypticism
among Pseudoniphargus. We select a set of key representative taxa
among those resulting from the species delimitation analyses to build a
robust phylogeny of the genus based on the complete set of protein-
coding mitochondrial genes (mitogenomes). This phylogenetic hy-
thesis and an approximate time frame for species diversification are
used to assess the potential roles played by both vicariance and long-
distance dispersal in the biogeographic history of Pseudoniphargus.

2. Material and methods

2.1. Taxon sampling

We collected representatives of Pseudoniphargus from 62 localities
(mostly wells and caves) distributed along the Iberian Peninsula,
northern Morocco, the Balearic, Canarian, Azores and Madeira archi-
pelagoes, plus Bermuda (Fig. 1; Supplementary Table S1). The dataset
includes 30 out of the 71 species currently recognized in the genus, plus
17 presumed new species not yet formally described. Specimens were
preserved in the field in 96% ethanol immediately after collection.
Voucher specimens and DNA aliquots were deposited at IMEDEA
(Spain). Most species were collected at single localities except for a few
cases where the same species was recorded in multiple sites, viz. P.
brevipedunculatus (ten localities on the Azores Islands), P. elongatus (five
localities in N Spain), P. mercadali (three localities on the Balearic
Islands) and P. ruffoi (three localities in Morocco). Two species appeared
syntopically at seven sampling sites, viz. P. triasi and P. davisi on the
same well at Cabrera (Balearic Islands), P. carpalis and P. grandimana
in Red Bay cave (Bermuda), whereas P. longipes and two cryptic species (P.
sp6-Morocco A and B) were found sharing the same well at Sidi Ab-
dellah (Taza; Morocco).

2.2. PCR amplification and sequencing of cox1 sequences

Genomic DNA was individually purified from whole specimens
using the Qiagen DNeasy Blood & Tissue kit (Qiagen, Hilden, Germany)
following the manufacturer’s protocol. A total of 410 individuals were
sequenced from 30 described species plus 17 recognized herein as po-
tentially new (Supplementary Table S1). We aimed to sequence a
minimum of five specimens per site whenever possible, but the actual
number varied largely between one and 25 depending on sampling
success (Supplementary Table S1). Taxa were identified based on their
original descriptions (Bréhier and Jaume, 2009, and references therein;
Messouli et al., 2006; Stokkan et al., 2018). PCR amplification of
mitochondrial cytochrome c oxidase subunit I (cox1) was performed
mainly using the primer-pair LCO/HCO (Folmer et al., 1994), but in
some cases we used an internal degenerated primer set especially de-
signed for this genus (amplicon of 463 bp), based on the sequences
amplified with the previous primers (PSNi_cox1_F_A39 5'-GCTATGCT-
TTTGGATTTAATTTATGTT-3' and PSNi_cox1_R_A71 5'-CAAAA-
CAGATGGTATAAAGAATTGGRCTCNCC-3'; Stokkan et al., 2016). PCR
and Sanger sequencing followed protocols described elsewhere
(Bauzá-Ribot et al., 2011). DNA sequences were edited in CodonCode
Aligner v5.1.5 (CodonCode Corporation, Denham, MA, USA) and
aligned using MAFFT (Katoh, 2013).

2.3. Single locus species delimitation analyses

Two tree-based and one distance species delimitation methods were
explored. First, we applied the Generalized Mixed Yule Coalescent
model (GMYC; Pons et al., 2006; Fujisawa and Barralough, 2013). This
method estimates species boundaries implementing a likelihood ratio
test to compare the null model assuming a single coalescent branching
rate across a clock-like tree with a complex model including both
coalescent and Yule branching models. The R library SPLITS v1.0-19
(Ezard and Fujisawa, 2009) was used in this analysis specifying a single
threshold as recommended by Fujisawa and Barralough (2013) on an
ultrametric tree estimated with BEAST v1.8.2 (Drummond and
Rambaut, 2007). The GMYC support value of each node representing a
transition from species to population was calculated as the sum of
Akaikie weights of candidate delimitation models where the node is
included (Fujisawa and Barralough, 2013). This allows assessing
whether the branch leading to a node contains a coalescence-to-
speciation threshold under different coalescent models (Fujisawa and
Barralough, 2013). A node support value of 1 means that all coalescent
models tested support the existence of a speciation event on that
branch, while lower values indicate that fewer coalescent models sup-
port a particular speciation event. We set an arbitrary minimum value
of 0.9 to consider a GMYC entity as supported by most coalescent
models. To perform the analysis, 1st plus 2nd codon positions and 3rd
positions were considered as separated partitions due to their marked
differences in nucleotide composition and substitution rates (Stokkan
et al., 2016). Models of nucleotide substitution were evaluated with
MrAIC.pl v1.4.6 (Nylander, 2004) yielding TRN + G and HKY + G as
best-fitting models for 1st + 2nd and 3rd codons positions, respec-
tively. The analysis implemented a coalescent constant population size
as tree model and a single uncorrelated relaxed clock with a log-normal
distribution. Substitution rate was arbitrarily set with a mean of the
standard 0.0115 nucleotide substitutions per site and lineage per mil-

The few cases in which this hypothesis has been tested using time-
calibrated molecular phylogenetic methods have shown a relative
congruence between genetic divergence and vicariance by plate tec-
tonics. Thus, in metacrangonyctid amphipods, the major diversification
episodes established in a dated robust phylogeny were compatible with
a plate tectonics-driven, vicariant scenario (Bauzá-Ribot et al., 2012).
But similar studies carried out on Typhlatya cave shrimps concluded
that not only ancient vicariance, but also ancient long-distance dis-
persal could have contributed to their current distribution pattern
(Botello et al., 2013; Jurado-Rivera et al., 2017).


990
model, implemented in this case with a lower cut-off value (X = 1.0) since the default slope threshold (X = 1.5) was not significant and produced a single species. These values were used to estimate the inflection point between two distributions of ranked frequencies of pairwise distances using ten recursive steps, i.e. the minimum optimal gap between intra- and interspecific distances.

### 2.4. Mitochondrial genome sequences

The complete or almost complete DNA sequences for mitogenomes of 28 *Pseudoniphargus* putative species out of a total of 47 were obtained using any of these three different methods: (i) long PCR mitochondrial fragments sequenced in a Roche GS FLX or GS Junior platform (eight species) as in Bauzà-Ribot et al. (2012); (ii) sequencing the whole genome at low coverage in Illumina platforms (18 species) (see methods in Pons et al., 2014; and Stokkan et al., 2016); and (iii) in cases of low genomic DNA yield, the mitogenome was amplified by multiple displacement amplification using specifically designed primers followed by sequencing in an Illumina platform (four species) (Blanco et al., 1989) (see Supplementary Methods). Indexed libraries were quantified by qPCR, pooled in equimolar concentration, and paired-end sequenced in a MiSeq v2 (2 × 150) or one lane using HiSeq2500 (2 × 150 or 2 × 160 reads). Fastq files were generated and demultiplexed with bcl2fastq software (Illumina) and adapters and low-quality bases removed from reads using Trimmmomatic v0.33 (Bolger et al., 2014). Contigs were assembled de novo with Trinity v2014-04-13p1 (Grabherr et al., 2011), and Ray v2.3.1 (Boisvert et al., 2010) using collapsed and truncated paired-end reads only. Mitogenome contigs were annotated in MITOS webserver (Bertani et al., 2013). Start and stop codon regions of protein-coding genes (PCGs) and the 5' and 3' ends of ribosomal genes were refined manually by comparing them to sequences available on databases. Secondary structures of tRNAs were corroborated with tRNAscan-SE (Schattner et al., 2005).

### 2.5. Mitochondrial phylogenetic analyses

We included the mitochondrial PCG sequences of 19 outgroup taxa taken from the GenBank and metAMiGA databases to root the *Pseudoniphargus* tree (Feijão et al., 2006): one myid, eight isopods and ten amphipods (see Supplementary Table S2). The newly obtained *Pseudoniphargus* mitochondrial genome sequences were first translated to protein, aligned with Muscle v3.8.31 (Edgar, 2004) with default parameters and back-translated to nucleotides conserving the integrity of codon triplets but excluding terminal stop codons. The mitogenomes of three species reported elsewhere (*Pseudoniphargus sorbensianus* LN871175; *P. gorbeanus* LN871176 and *P. davidi* FR872383) were included in the analyses (Stokkan et al., 2016). Regions aligned ambiguously, in particular at 5' and 3' sequence ends due to uneven sequence lengths were removed as codons with Gblocks v0.91b (Castresana, 2000) under relaxed conditions. After trimming, most of the 13 mitochondrial PCGs retained more than 94% of the original positions of the alignment except the genes nad4l (91.26%), nad2 (90.57%), nad6 (85.41%), and the highly variable atp8 (74.55%).

Phylogenetic trees were built under a maximum likelihood framework using both DNA and protein datasets in IQ-TREE (Nguyen et al., 2015), with the best partition scheme and models estimated in PartitionFinder (Lanfear et al., 2012) and branch lengths calculated independently for each partition (i.e. unlinked branch lengths). Tree topologies were also assessed under parsimony criterion in PAUP v4.0 beta 10 with 1000 random addition replicates, saving 50 trees per replicate, and heuristic searches with tree-bisection-reconnection branch-swapping (Swofford, 2002). Node support was assessed by 1000 bootstrap replicates in ML and parsimony. Tree searches were also performed under the Bayesian criterion in PhyloBayes mpi v1.5a implementing the GTR + CAT model, as this model has been shown to be less prone to phylogenetic errors caused by compositional bias or saturation (Lartillot and Philippe, 2004; Lartillot et al., 2009). Two independent runs were sampled every 1000 generations until convergence was reached. To assess nucleotide substitution saturation, the Xia saturation test was implemented using the whole dataset and for first, second and third codon sites separately as implemented in DAMBE v6.0.0 (Xia and Lemey, 2009; Xia et al., 2003). Saturation was also explored plotting corrected and uncorrected branch lengths estimated under a fixed topology in IQ-TREE. Correlation between patristic distances was estimated with the software Patristic (Fourment and Gibbs, 2006).

### 2.6. Calibration of the molecular clock and dating

There are no fossils of *Pseudoniphargus* known, and the amphipod fossil record is not older than Eocene in age (niphargids and crangonyctids preserved in Baltic amber 54–40 million years in age; Jażdżewski et al., 2014; Starr et al., 2016). For that reason, we relied on palaeogeographical evidence to constraint the age of a set of nodes in our *Pseudoniphargus* tree. Specifically, the geological age of oceanic islands or the age of emergence from the sea of continental areas where sister species occur were assumed to precede the divergence of sister taxa on those areas. The ages of the three geological events related to the diversification of particular *Pseudoniphargus* lineages were considered as soft constraints with a parametric log-normal distribution that assigns non-zero probabilities to all possible age values. The midpoint of the extent of each geological event was arbitrarily considered as the age with the highest probability in the distribution. Congruence among calibration points was explored performing a cross-validation analysis, i.e. comparing three additional BEAST runs, each calibrated with a single geological event. The geological events used for this purpose in our analysis were: (i) The emergence of the Basque Country (western edge of the Pyrenees; northern Spain) at the Early Oligocene (33.7 Ma; Rögl, 1998). There, a triplet of sister species occurs (*P. gorbeanus*/ *P. unisexualis*/ *P. sp1-Basque*). Assuming that the divergence of these three species from a common marine ancestor was driven by past episodes of sea level regression (Stock, 1980; Notenboom, 1991; Boutin and Coineau, 1990), it could not have happened before the first subaerial exposure of that area. We constrained this node with a strong prior by implementing a log-normal distribution with a mean (M) of 11.5 Ma and standard deviation (SD) of 0.1 (95% confidence interval 9.6–13.4). (ii) The establishment of the so-called Southern Riffian Corridor across northern Morocco between 8 and 6.1 Ma (Achalhi et al., 2016). This gateway enabled the temporary direct connection between the Mediterranean and the Atlantic Ocean up to the closing of the Strait of Taza during the Messinian. The sister species *P. ruffoi*/ *P. longipes* are found one at each side of this vanished sea corridor. As in the preceding case, we consider that the origin of these two species cannot be older than that age, assuming they derive from a common marine ancestor. We defined a normal distribution as a soft constraint for this node with M = 11.5 Ma and SD of 0.1 (5.8–8.5). (iii) The emergence of La Palma (Canary Islands), which harbors the sister species *P. cupicola*/ *P. multidens*. The age of the split between these two taxa cannot be older than the first subaerial exposure of the island (ca. 2.0 Ma; Carracedo et al., 2001). We defined a log-normally distributed function as a soft constraint for this node with M = 2.0 Ma and SD of 0.1 (1.6–2.4).

For the calibration analyses we excluded outgroups and rooted the trees according to the topologies obtained in the analyses including outgroups. Tree topologies, model parameter values and node ages were co-estimated and optimized in Beast v1.8.2 implementing the best partition scheme and models described above. Bayesian analyses were run for 150 million generations. Convergence of the run was assessed in Tracer v1.6 (Rambaut et al., 2015) ensuring parameter values to have ESS values > 200. Mean values and confidence intervals of the parameters and ages were estimated in TreeAnnotator after a burnin of the first 15 million generations. We implemented default priors for all parameters except for clock and Yule rate priors since the ones by
default did not converged during path sampling. The substitution rate prior (sctl.mean) for 1st codon sites was set as a log normal distribution in real space with $M = 0.005$ and $SD = 1.5$ (95% confidence interval $8.58 \times 10^{-05}$–$3.07 \times 10^{-02}$); $M = 0.1$ and $SD = 1.5$ ($1.72 \times 10^{-04}$–$6.14 \times 10^{-02}$) for 2nd codon positions; and $M = 0.5$ and $SD = 1.0$ ($4.27 \times 10^{-03}$–$0.21$) for 3rd codon sites and as global rate. The tree prior for Yule birth rate parameter was also set to a log-normal distribution in real space with $M = 0.5$ and $SD = 3.0$ (95% confidence interval $5.5 \times 10^{-05}$–$1.98$). We estimated node ages under a strict clock, and two types of relaxed clocks where rates on descendant branches are independent of the rate at the parent branch: uncorrelated log-normal clock (UCLN) and random local clocks (RLC). Runs implementing RLC were run for 500 million generations since it took about 280 million generations until convergence, which were discarded as burnin. Different partition schemes and clocks were compared based on Bayes Factors estimated by marginal likelihoods using path sampling and the stepping stone model as implemented in Beast. We performed 100 steps of 5 million generations each using a path scheme with a betaQuantile 0.33 (Baele et al., 2012), discarding 25% of the run as burnin.

3. Results

3.1. Phylogeny and species delimitation based on cox 1 sequences

The mitochondrial cox1 sequences obtained from 410 Pseudoniphargus individuals showed no indels, stop codons or singleton non-synonymous substitutions, suggesting the absence of nuclear mitochondrial DNA segments in the dataset (accession numbers LS482437-LS482846). Invariant and parsimony-informative nucleotide positions accounted for 53.3% and 43.7%, respectively. Only two localities placed in close proximity on the island of Cabrera (Balearic Islands), where P. davidi occur, and two wells harboring P. brevipedunculatus on Pico Island (Azores) showed shared haplotypes, while all other collecting sites rendered locality-specific mitochondrial sequences. After collapsing into identical sequences, a total of 182 haplotypes were retained. Bayesian and ML phylogenetic trees showed four major clades, with weakly supported relationships among them (Fig. 2):

(i) Clade G1, including exclusively species from northern Spain, from the upper reaches of the Ebro river basin and valleys and massifs of the Basque Country; (ii) Clade G2, comprising a single species from the Cantabrian mountains in northern Spain; (iii) Clade G3, a mostly Mediterranean clade embracing species from S and SE Spain, Portugal, the Balearic Islands and northern Morocco, but also including a species from Gran Canaria (P. sp2-Canaries); and (iv) clade G4, included most of the species from Atlantic Islands (Canaries, Madeira, Azores, Bermuda) but also several taxa from the Balearics, Portugal and SE Spain.

The GMYC algorithm applied to the ultrametric tree favored a two-parameter model including both coalescent and Yule branching over the null model with a single coalescent parameter ($P < 1e^{-15}$). The single threshold approach retrieved 63 entities defined herein as Molecular Operational Taxonomic Units (MOTUs), with a confidence interval ranging from 59 to 67 (Table 1; Fig. 2). Twenty-one of these MOTUs were represented each by single specimens, whereas the remaining 42 MOTUs embraced multiple individuals (Fig. 2). Thirty of the GMYC groupings were consistent with known species, whereas other 17 were previously recognized based on the morphological analyses herein undertaken (considered here as undescribed morphotypes). Most of the MOTUs had a high GMYC support (i.e. most of the coalescent models tested validate them), with values ranging between 0.9 and 1.0, with the exception of the groupings identified within P. brevipedunculatus, P. elongatus, P. gorbeanus, P. longipes and P. sp1-Andalusia and P. sp3-Andalusia (see Fig. 2 and Supplementary Fig. S1 for precise values).

The PTP method discriminated among 55 MOTUs, 44 of which were consistent with those recognized by GMYC (Table 1; Fig. 2). This method in turn rendered a slightly better correspondence between morphologically defined species and MOTUs than the GMYC. Thus, PTP and morphology confirmed the same species for P. sp1-Andaluica, P. sp3-Andalusia, P. brevipedunculatus and P. cupicola. On the contrary, PTP split P. gorbeanus in one more group than GMYC but merged the sequences of P. guernicae and P. sp1-Baseque, two species recognized both morphologically and by GMYC, in a single entity.

The distance-based method ABGD estimated the maximal intraspecific distance to be $P = 4.64\%$, retrieving a total of 49 MOTUs
Table 1

Comparison of the taxonomically recognized and newly proposed species of *Pseudoniphargus* based on morphology with the molecular operational taxonomic units (MOTUs) estimated herein by three species delimitation methods. Numbers in brackets indicate number of MOTUs recognized when considering GMYC and PTP support values (nodes supported by more than 90% of coalescent models in GMYC or tree-topologies in PTP; see Material and Methods and Fig. 2). A 0.5 value indicates that two sister species were merged together in a single unit by a particular method. Asterisks denote cases where delimitation was based on a single specimen or single unit by a particular method. Haplotype sequence.

<table>
<thead>
<tr>
<th>Species</th>
<th># ind</th>
<th>p-dist</th>
<th>GMYC</th>
<th>PTP</th>
<th>ABGD</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>P. burgensis</em></td>
<td>2</td>
<td>0.000</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><em>P. elongatus</em></td>
<td>10</td>
<td>0.046</td>
<td>5(4)</td>
<td>3(2)</td>
<td>2</td>
</tr>
<tr>
<td><em>P. gorbeanus</em></td>
<td>22</td>
<td>0.013</td>
<td>2(2)</td>
<td>3(1)</td>
<td>1</td>
</tr>
<tr>
<td><em>P. guernicae</em></td>
<td>14</td>
<td>0.008</td>
<td>1</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td><em>P. incatatus</em></td>
<td>2</td>
<td>0.000</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><em>P. jereanu</em></td>
<td>6</td>
<td>0.001</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><em>P. uniaosalis</em></td>
<td>5</td>
<td>0.001</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><em>P. sp1-Andalusia</em></td>
<td>4</td>
<td>0.001</td>
<td>1</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td><em>P. pedri</em></td>
<td>7</td>
<td>0.001</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><em>P. grandis</em></td>
<td>4</td>
<td>0.003</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><em>P. latipes</em></td>
<td>8</td>
<td>0.002</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><em>P. morosoi</em></td>
<td>6</td>
<td>0.001</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><em>P. sorosanensis</em></td>
<td>11</td>
<td>0.000</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><em>P. stocks</em></td>
<td>3</td>
<td>0.000</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><em>P. sp1-Andalusia</em></td>
<td>11</td>
<td>0.013</td>
<td>2(1)</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><em>P. sp2-Andalusia</em></td>
<td>5</td>
<td>0.001</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><em>P. sp3-Andalusia</em></td>
<td>5</td>
<td>0.010</td>
<td>2(1)</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><em>P. sp1-Murcia</em></td>
<td>3</td>
<td>0.003</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><em>P. merca</em></td>
<td>2</td>
<td>n/a</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><em>P. sp1-Portugal</em></td>
<td>1</td>
<td>n/a</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><em>P. sp2-Portugal</em></td>
<td>4</td>
<td>0.000</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><em>P. davini</em></td>
<td>19</td>
<td>0.001</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><em>P. mercadali</em></td>
<td>18</td>
<td>0.073</td>
<td>4(2)</td>
<td>4(2)</td>
<td>3</td>
</tr>
<tr>
<td><em>P. pipe</em></td>
<td>12</td>
<td>0.007</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><em>P. priya</em></td>
<td>5</td>
<td>0.009</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><em>P. triasi</em></td>
<td>2</td>
<td>0.000</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><em>P. sp1-Balea</em></td>
<td>8</td>
<td>0.003</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><em>P. asociate</em></td>
<td>7</td>
<td>0.001</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><em>P. caprocal</em></td>
<td>26</td>
<td>0.013</td>
<td>2(2)</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><em>P. grandfas</em></td>
<td>25</td>
<td>0.001</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><em>P. mulitdias</em></td>
<td>22</td>
<td>0.004</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><em>P. salinas</em></td>
<td>7</td>
<td>0.001</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><em>P. sp1-Canaries</em></td>
<td>2</td>
<td>0.000</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><em>P. sp2-Canaries</em></td>
<td>2</td>
<td>0.000</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><em>P. brei</em></td>
<td>60</td>
<td>0.018</td>
<td>4(3)</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><em>P. sp1-Azores</em></td>
<td>7</td>
<td>0.014</td>
<td>2(1)</td>
<td>2(1)</td>
<td>1</td>
</tr>
<tr>
<td><em>P. portosan</em></td>
<td>11</td>
<td>0.002</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><em>P. carpas</em></td>
<td>4</td>
<td>0.000</td>
<td>1</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td><em>P. grandfus</em></td>
<td>3</td>
<td>0.009</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><em>P. lopites</em></td>
<td>14</td>
<td>0.004</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><em>P. ruff</em></td>
<td>5</td>
<td>0.003</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><em>P. sp1-Morocco</em></td>
<td>3</td>
<td>0.007</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><em>P. sp2-Morocco</em></td>
<td>3</td>
<td>0.000</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><em>P. sp3-Morocco</em></td>
<td>1</td>
<td>n/a</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><em>P. sp4-Morocco</em></td>
<td>1</td>
<td>n/a</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><em>P. sp5-Morocco</em></td>
<td>1</td>
<td>n/a</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><em>P. sp6-Morocco</em></td>
<td>6</td>
<td>0.038</td>
<td>2(2)</td>
<td>2(2)</td>
<td>2</td>
</tr>
</tbody>
</table>

(Table 1). Forty out of the 47 morphologically recognized species corresponded to a single entity in this analysis, except for *P. grandfus* and *P. carpas*, that were merged into one, as it was the case for *P. guernicae* and *P. sp1-Balea*. *Pseudoniphargus mercadali* was split into three entities using ABGD, whereas *P. sp6-Morocco* and *P. elongatus* were split each into two.

In summary, species coincidently detected by diagnostic morphological features and three molecular species delimitation methods accounted for ca. 80% of the cases (38 species) (Table 1). In six other cases (13%), the discrepancy found between the morphological classification and one of the three delimitation methods might suggest we are in front of transition events from populations to species (incipient speciation) but where the morphologies of descendants have not diverged yet. Finally, in three assignments (7%), genetic lineages appear to be relatively old phylogenetically, but have proven to be undistinguishable morphologically and should be considered as true cryptic species (between two and four phyletic lineages present in *P. mercadali* and *P. elongatus*, and two in *P. sp6-Morocco* (Table 1).

3.2. Mitochondrial genomes

Specimens representative of 28 MOTUs were selected for sequencing their mitochondrial genome and study of their phylogenetic relationships, assuring an even sampling of all *Pseudoniphargus* distributional areas. Three additional MOTUs represented by the mitogenomes of other *Pseudoniphargus* species obtained previously were also included in the analyses (Stokkan et al., 2016). The mitogenome sequence coverage obtained was dependent of the amplification and sequencing methods used but was generally higher than 50x (Table 2; Supplementary Methods). In two species, *Pseudoniphargus* sp2-Portugal and *P. salinus*, the PCG sequences were incomplete as only two non-overlapping fragments were obtained probably due to technical difficulties in sequencing long polyA/T runs. In addition, the analysis of the mitogenome sequences obtained from pooled individuals of *P. sp6-Morocco* showed the presence of two divergent mitogenomes (divergence > 10%), confirming the presence of two syntopic cryptic species in their sampling locality.

The 29 newly obtained mitogenomes showed the 37 genes characteristic of metazoans: 13 PCGs, 22 rRNA genes, plus two rRNA genes with a length ranging 14–15 kb. Gene arrangement was not conserved across species as described previously (Stokkan et al., 2016). In addition to the three gene orders already described in the latter paper, two of them involving shifts of protein-coding genes, another two gene arrangements affecting the position of several tRNA genes were found in the sampled species (not shown).

3.3. Mitochondrial genome phylogeny

The individual alignments of the 13 PCGs of the *Pseudoniphargus* species plus outgroups after excluding ambiguous positions were concatenated in a single dataset of 10,833 bp (corresponding to 3611 amino acids). Sequences showed a high A + T nucleotide bias, especially at third codon positions, as is usually the case for arthropod mitochondrial DNA (Hassanin, 2006; Stokkan et al., 2016). Xia’s test showed relatively low levels of nucleotide substitution saturation under the assumption of a symmetric tree even using outgroup sequences (p < 0.001) and all codon positions. Low to moderate saturation was also found at 1st and 2nd codon positions (0.3 and 0.4, compared to a critical value of 0.8). A higher, but still non-critical value was recovered for 3rd codon positions (0.79 vs. 0.83). This was confirmed by plotting branch lengths estimated under the simple F81 + G substitution model versus the complex GTR + G (Supplementary Fig. S2). After refining the results obtained in PartitionFinder using Bayesian Information Scores (BIC), the best partition consisted of subdividing the nucleotide dataset in four partitions: 1st codon positions; 2nd codon positions; 3rd codon positions coded on the + strand and 3rd codon positions coded on the – strand (the genes nad1, nad4, nad4L, nad5) (Table 3), while the best substitution model for each partition consisted on GTR + G.

The ML phylogenetic tree including outgroups retrieved two main *Pseudoniphargus* mitogenome lineages, of which one appears subdivided in turn into another three clades (Fig. 3), in accordance with the coxl phylogeny (Fig. 2). The major clade, coincident with coxl species-group G1, comprises exclusively the three taxa occurring around the Gulf of Biscay in northern Spain (*Pseudoniphargus gorbeanus*, *P. uniaosalis* and *P. sp1-Balea*). The second major clade includes all the remaining sampled species of the genus (Fig. 3). This clade can be further subdivided into clade G2, including only the species *P. elongatus*, which is distributed throughout Santander and the north of Burgos provinces in...
northern Spain, and clade G3, composed of a group of species from Morocco, the Balearic Islands, SW and SE Iberian Peninsula plus the new species *P. sp2*—Canaries from Gran Canaria (Canary Islands). Finally, clade G4 shows a surprisingly large geographic distribution including Madeira (*P. portosancti*), Mallorca in the Balearics (*P. mercadali*), the south of the Iberian Peninsula (4 spp.), and also Bermuda (*P. carpalis*), Azores (2 spp.), and the Canary Islands of La Gomera (*P. gomerae*), La Palma (2 spp.) and El Hierro (*P. salinus*).

### Table 2

List of the *Pseudoniphargus* species whose mitogenome sequences were obtained in the present study. GenBank accession numbers and sampling locality are indicated, as are details on the sequencing method used (SM), sequence length, number of reads (x10^3), coverage depth and average length of mapped reads in bp. Sequencing methods are as follows: 1) Long PCR Fragments and GS FLX or GS JUNIOR sequencing, 2) Whole genome sequencing at low coverage in Illumina platforms with individually tagged libraries (Miseq 2x150 bp, HiSeq2500 2x150 bp or 2x160 bp, and 3) Multiple displacement amplification and Miseq 2x150 bp without tagging.

Details of the mitogenomes of three individually tagged libraries (Miseq 2x150 bp, Hiseq2500 2x150 bp or 2x160 bp, and 3) Multiple displacement amplification methods are as follows: 1) Long PCR Fragments and GS FLX or GS JUNIOR sequencing, 2) Whole genome sequencing at low coverage in Illumina platforms with one composed of included third codon positions of BIC). The best model for all partitions was GTR + G with 4 categories to estimate the alpha parameter. # The best partition scheme found by Partition Finder.

### Table 3

Comparison of different partition schemes based on Bayesian information Criterion (BIC = −2Ln + Kn(n), where Ln is the tree Likelihood, K the number of free parameters of the models including the number of branches per partition, and n the number of sites (10833 pb per nucleotide sequences, and 3611 characters per amino-acid and codon models). Tree length (TL) and increment of BIC values relative to the best partition scheme found in Partition Finder are also indicated (% Δ BIC). The best model for all partitions was GTR + G with 4 categories to estimate the alpha parameter. # The best partition scheme found by Partition Finder included third codon positions of atp8 gene within first codon sites partitions. * The best partition scheme for amino-acid sequences in Partition Finder was partition one composed of atp6, cox1, cox2, cox3, nad1 genes and the model mtART + F + I + G and partition 2 with *P. coob*, *P. mercadali*, *P. carpalis*, *Azores* (2 spp.), and the Canary Islands of *La Gomera* (*P. gomerae*), *La Palma* (2 spp.) and El Hierro (*P. salinus*).
The other partitioning strategies analyzed using the nucleotide sequence dataset resulted in identical or nearly identical topologies showing high support values for most nodes except for the relationships among some of the taxa within clade G3 (Fig. 3). Similar results were obtained using parsimony, treating data as 62 different codon categories (GY98) and implementing the CAT model (Fig. 3) or using protein sequences analyzed by different methods and substitution models (Supplementary Fig. S3).

3.4. Calibration of molecular clock and dating

The best fitting clock model resulted from setting three independent uncorrelated log-normal clocks (1st codon, 2nd codon, 3rd codon), since third codon positions were considered a single partition in despite of the coding strand, and a Yule diversification model (Table 4). The chronogram of Pseudoniphargus diversification obtained after using the age of three geological events as node constraints to calibrate the tree (see Material and Methods) is shown in Fig. 4. The root of the tree, corresponding to the initial diversification of the sampled Pseudoniphargus taxa was estimated to fall in the Eocene, about 52.8 Ma (HPD95% = 46.5–59.5), while subsequent major splits took place at the early Oligocene (ca. 32 Ma; HPD95% = 27.5–36.4) and during the Miocene (24.6 and 22.5 Ma for clades G3 and G4, respectively). Splits leading to the oceanic island species present on the Canaries, Azores and Bermuda occurred in the transition from the Miocene to the Pliocene (ca. 5 Ma, Fig. 4). Other partition schemes and the use of a single biogeographic constrain rendered overlapping age posterior distributions except when only the age of La Palma was used as age constraint, which rendered significantly younger ages (not shown).

#### Table 4

List of Bayesian analyses constrained under different molecular clocks, partition schemes, and diversification models ranked by Bayes Factors. Marginal likelihoods were estimated using a path-sampling method which values were used to calculate Bayes Factors. Data was either partitioned as four partitions (1st codon, 2nd codon, 3rd codon plus strand, 3rd codon minus strand), three partitions (1st, 2nd, 3rd), or two partitions (1st + 2nd, 3rd). Abbreviations: UCLN (uncorrelated log-normal clock), RLC (random local clock), and strict (strict clock).

<table>
<thead>
<tr>
<th>Scheme and model</th>
<th>Marginal Ln</th>
<th>Bayes factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 partitions 3 UCLN Yule</td>
<td>-162220.017</td>
<td>0.000</td>
</tr>
<tr>
<td>4 partitions 3 UCLN Birth-Death</td>
<td>-162235.267</td>
<td>-30.499</td>
</tr>
<tr>
<td>4 partitions 4 UCLN Birth-Death</td>
<td>-162237.696</td>
<td>-35.357</td>
</tr>
<tr>
<td>4 partitions 4 UCLN Yule</td>
<td>-162243.669</td>
<td>-47.304</td>
</tr>
<tr>
<td>4 partitions 3 RLC Yule</td>
<td>-162302.005</td>
<td>-163.976</td>
</tr>
<tr>
<td>4 partitions 4 RLC Yule</td>
<td>-162331.782</td>
<td>-223.529</td>
</tr>
<tr>
<td>4 partitions 2 UCLN Yule</td>
<td>-162416.241</td>
<td>-392.448</td>
</tr>
<tr>
<td>4 partitions 2 UCLN Birth-Death</td>
<td>-162418.614</td>
<td>-394.448</td>
</tr>
<tr>
<td>4 partitions 4 strict Yule</td>
<td>-162441.814</td>
<td>-443.593</td>
</tr>
<tr>
<td>4 partitions 3 strict Yule</td>
<td>-162450.010</td>
<td>-459.986</td>
</tr>
<tr>
<td>3 partitions 3 UCLN Yule</td>
<td>-163814.156</td>
<td>-3188.278</td>
</tr>
<tr>
<td>3 partitions 3 UCLN Birth-Death</td>
<td>-164341.140</td>
<td>-4242.246</td>
</tr>
<tr>
<td>3 partitions 1 UCLN Yule</td>
<td>-164447.132</td>
<td>-4454.230</td>
</tr>
<tr>
<td>2 partitions 2 UCLN Yule</td>
<td>-164447.132</td>
<td>-4454.230</td>
</tr>
<tr>
<td>Excluding 3rd codon sites</td>
<td>-63545.708</td>
<td>0.000</td>
</tr>
<tr>
<td>2 partitions 2 UCLN Birth-Death</td>
<td>-63605.540</td>
<td>-119.663</td>
</tr>
<tr>
<td>2 partitions 2 UCLN Yule</td>
<td>-63637.950</td>
<td>-184.484</td>
</tr>
<tr>
<td>2 partitions 1 RLC Yule</td>
<td>-63833.142</td>
<td>-574.87</td>
</tr>
</tbody>
</table>

The mean substitution rate estimated for the combined 13 mitochondrial PCGs was 2.23 × 10⁻² per site, million years and lineage, with a 95% confidence interval with a minimum of 1.99 × 10⁻² and a maximum of 2.65 × 10⁻². This value corresponds to a pairwise
divergence of 4.46%, nearly doubling the commonly implemented value of 2.3% (Brower, 1994). The mean substitution rate for first, second and third codon sites were estimated as $0.72 \times 10^{-2}$ ($0.64–0.81$), $0.33 \times 10^{-2}$ ($0.29–0.37$) and $5.65 \times 10^{-2}$ ($4.99–6.35$), respectively. As expected, rate in third codon sites was about 10-fold higher than those estimated in first and second codon sites and also showed larger confidence intervals.

4. Discussion

4.1. Species boundaries and species delimitation methods

A preliminary morphology-based assessment of the species diversity present in our Pseudoniphargus samples enabled the assignment of specimens to a minimum of 30 already known described species plus other 17 presumably new. Results of three different methods of species delineation based on single-locus (cox1) data performed on the same set of specimens (two of them based on gene trees and one on distance matrices) showed a general agreement with the morphological analysis (Table 1; Fig. 2). Two of the presumed new species revealed by molecular delimitation methods (Pseudoniphargus sp1-Basque and P. sp1-Balearics) were confirmed morphologically afterwards, constituting an instance of “inadequate taxonomy revealed through the application of molecular methods” (Astrin et al., 2012). In other cases of presumably new species, morphological divergence was not recognizable or character states were too variable as to enable a formal taxonomic description based solely on morphological features. These cases may represent either evolutionarily young lineages in which incipient speciation is currently ongoing in absence of detectable morphological differentiation, or cryptic and reproductively isolated old lineages that deserve an appropriate taxonomic treatment (Delić et al., 2017).

Four possible cryptic species complexes were revealed in our study. The densely sampled P. gorbeanus appears split into two to three independently evolving lineages occurring in sympatry at Alava (North Spain). However, when using the more stringent criterion of only accepting GMYC or PTP highly-supported MOTUs, only GMYC favors the presence of two entities while the other methods concur in considering P. gorbeanus as a single species (Table 1; Fig. 2). Seemingly, Pseudoniphargus elongatus was divided into a minimum of two and a maximum of four highly supported cryptic allopatric lineages depending on the method used (Table 1; Fig. 2). Since a high number of individuals were sequenced in both species, these presumed cryptic lineages could be just a byproduct of sampling bias (Fujisawa and Barraclough, 2013; Zhang et al., 2013). Nevertheless, other species that were also densely sampled in our study—such as P. multidens and P. portosancti—were not split into different entities by the same analyses. Crypticism was also disclosed for the Menorcan and Mallorcan lineages of P. mercadali (Balearic Islands), showing at least two highly supported entities. The same pattern was found in P. brevipedunculatus on three Azores Islands (São Miguel, Faial and Pico), suggesting the occurrence of incipient allopatric speciation in absence of morphological divergence. Similar results have been reported in other insular stygobiont amphipods and thermosbaenaceans (Bauzà-Ribot et al., 2011; Cánovas et al., 2016). In contrast with the former cases of species crypticism, we found two cases...
of sympatric species where members of each pair were clearly differentiated by diagnostic morphological features (i.e. *P. guernicae* and *P. spl1*-Basque in the Iberian northern clade G1, and the Bermudian *P. carpalis* and *P. grandimanus* in clade G4) but they were not distinctly supported as molecular MOTUs, likely due to very low inter- and intraspecific genetic divergences. GMYC has been shown to produce more entities than other methods, resulting in an overestimation of species numbers. This is usually due to low genetic divergence between lineages with the overlap of inter- and intraspecific divergences, or lack of reciprocal monophyly between sister clades (Hendrich et al., 2010; Talavera et al., 2013; Kekkonen et al., 2015). Nevertheless, the GMYC method seems to be better suited for detecting early speciation events, where both intra- and interspecific method appears to be better suited for detecting early speciation events, where both intra- and interspecific distances are low, as it is able to detect entities with lower intra-MOTU distances than PTP or other methods. The limitations associated to coxI polymorphisms and a single population-to-species threshold suppose that eventual occurrence of cases of split and lump of species are to be expected, so failures in identification are likely to occur if species are delimited based on molecular delimitation approaches only (Talavera et al., 2013). For that reason, GMYC and other molecular delimitation methods should be supplemented with additional information (i.e. morphological features) and should also take into account support values for defined MOTUs (i.e. assessing both uncertainties in tree topology and the support obtained by different coalescent models in GMYC) (Talavera et al., 2013; Fujisawa and Barrachlough, 2013).

4.2. Mitogenome phylogenetics and biogeographic patterns

Our phylogenetic analysis shows the family Pseudoniphargidae to be at least Early Tertiary in age (Fig. 4). The family experienced a first major split during the Paleocene, when a branch (Clade G1) leading to the current Basque Country species diverged from the rest of pseudoniphargids. The occupation of continental waters by members of this Basque clade, assuming it was coeval to species diversification (Stock, 1980), could not take place before the end of the Eocene (about 40 Ma), when the current Basque area started to emerge from the sea (Plaziat, 1981; Rögl, 1998).

The remaining *Pseudoniphargus* experienced three further major splits, of which the more basal (Clade G2) led to *P. elongatus*, a species from the northern Spanish provinces of Santander and Burgos, adjacent to the Basque Country. This split was virtually synchronous with the first occupation of continental waters by the Basque lineage and might have resulted from the same episodes of marine regression affecting the southern coastline of the Gulf of Biscay at the late Eocene or Early Oligocene (Plaziat, 1981; Rögl, 1998).

During the Oligocene, the rest of *Pseudoniphargus* diverged in two clades (Fig. 4). One (Clade G3) comprises 13 species from southern Spain (Andalusia), Morocco and the Balearic Islands, plus one species each from Portugal and Gran Canaria (Canaries). The other one (Clade G4) includes eight species from the Canaries, Madeira, Azores and Bermuda, plus two from Portugal and one each from Andalusia, Murcia (SE Spain) and the Balearics. The most basal species in both clades correspond to Atlantic taxa – from Portugal in Clade G3; from Porto Santo Island (Madeira) in Clade G4 – which, combined with the Atlantic condition of Clades G1 & G2, suggesting that the origin of the family Pseudoniphargidae should be located on that ocean. The occupation of the Mediterranean by pseudoniphargids seems thus to be a relatively recent episode in the evolutionary history of the group.

Only the three species from the Basque Country and the four species from Morocco – aside the monospecific Clade G2– conform monophyly with a clearly delimited geographical projection, denoting occupation of these areas was performed only once and by members of a single lineage (Fig. 3; Fig. 4). In the rest of territories, the *Pseudoniphargus* assemblage is polyphyletic, indicating these areas accumulate remnants of several colonization waves. Thus, the Balearic Islands harbor representatives of four independent lineages, whereas the Atlantic islands, Portugal and southern Spain include three each.

Comparison between the reported known ages of emergence of the oceanic islands harboring *Pseudoniphargus* and the estimated age of the corresponding nodes in our dated cladogram of species relationships shows a close correspondence in the cases of the species from Porto Santo (Madeira), Gran Canaria (Canaries) and Sao Miguel (Azores), suggesting colonization of inland waters and speciation on these islands took place as soon as emerged land was available. Island and node ages were found to be compatible also in the case of Bermuda and La Gomera, though in both cases the islands seem to have been colonized long afterwards they were exposed subaerially (but see below). Only the species from Faial Island (Azores) and El Hierro (Canaries) show an estimated age that is noticeably older than the islands harboring them, and only in the former case is this difference noticeable. Interestingly, the species from El Hierro inhabits only brackish-water coastal wells and has never been reported from fresh, inland groundwater, denoting it is still in the first steps of the colonization process of emerged land.

The presence of *Pseudoniphargus* on Porto Santo Island (Madeira) and Gran Canaria (Canaries) can be explained without recourse to long-range transoceanic dispersal. Assuming their ancestors were shallow-water marine species associated to submerged island primordials, the area currently comprised between both islands and the Iberian Peninsula has harbored seamounts and now drowned archipelagos since the early Tertiary (Fernández-Palacios et al., 2011). These structures might have acted as stepping-stones in the past, enabling the movement of species from the Iberian coasts.

On the contrary, the estimated age and genealogical affiliation of the rest of *Pseudoniphargus* species in this oceanic clade, which conform a well-established monophylum sister to a Portuguese species, suggest recent episodes of long-range transoceanic dispersal have played a major role in the establishment of their distribution pattern. Thus, Bermuda conforms a mid-plate rise in an area devoid of adjacent seamounts or ridges that might have acted as stepping-stones to enable its colonization. The lavas that formed the original Bermuda shield volcano are dated at 47–40 Ma, whereas at 40–36 Ma the Bermuda platform had already risen to sea level (Vogt and Jung, 2007). Notwithstanding, that the island has probably been submerged repeatedly since then, for the last time during the Pleistocene (Olson and Hearty, 2003). Hence the occupation of the freshwater habitats of the island by *Pseudoniphargus* is surely much more recent that the age attributed to the Bermudian lineage, and in accord with this is the fact (already remarked in the case of the species from El Hierro) that the Bermudian species is euryhaline, denoting its recent origin (Stock et al., 1986). The primordial Bermuda was already placed far away from the East Atlantic coast when it started its existence, and despite this, the estimated age of the *Pseudoniphargus* lineage present on the island is of only 5 Ma. Furthermore, this lineage occupies a basal position within its insular Atlantic clade (Fig. 3; Fig. 4) where the estimated species age diminishes in a Bermuda-Azores-Canaries progression. This could be explained by an episode of “reverse migration” back to the east Atlantic coast once Bermuda had been reached by the ancestor of this lineage. Considering this set of circumstances, it is hardly tenable that the opening of the Atlantic had played any role in the establishment of the current amph-Atlantic distribution of *Pseudoniphargus*.

Whereas in the foregoing lineage from the Atlantic islands species diversification started in the Pliocene, most diversification in the peri-Mediterranean area took place in early Miocene to Tortonian times. During that period, the territory currently corresponding to southern Iberia, northern Morocco and the Balearics has evolved from a broad seaway between the Atlantic and the Mediterranean to a mosaic of islands and interior and marginal basins (Rögl, 1998; Martin et al., 2014; Achali et al., 2016). This scenario might have enabled species diversification associated to recurrent episodes of marine transgression and regression.
4.3. Conclusion

The data presented herein show the power of combining an extensive sampling across the distribution range of a lineage coupled with species delimitation methods and a phylogeny built from mitochondrial genomes to shed light on the species diversity, phylogenetic relationships and biogeographic history of a lineage of subterranean amphipods. Four possible cryptic species complexes hidden each within the formally recognized species Pseudoniphargus gorbeanus, P. elongatus, P. mercadali and P. brevipedunculatus have been identified in our study, while morphological characters and molecular delimitation methods delineated most of remaining species concomitantly. The obtained mitogenome-based phylogeny supports an origin of the genus Pseudoniphargus in the Early Tertiary, and the existence of four major clades, of which two show a clear limited geographical projection in the N Iberian Peninsula. The rest of the area of distribution was colonized multiple times in a complex pattern. Both vicariance and dispersal have contributed to the present distribution of Pseudoniphargus, but the opening of the Atlantic Ocean in the Mesozoic has not played any discernible role in the diversification and present biogeography of the genus.

Acknowledgements

This work was supported by the Spanish MINECO grants CGL2009-08256, CGL2012-33597 and CGL2016-76144-P, financed by the Agencia Española de Investigación, (AEI) and the European Regional Development Fund (FEDER). We thank the following colleagues for supplying Pseudoniphargus samples: Ana Sofía Rebollo (Cádiz); Manuel Naranjo and Sonia Martín (Gran Canaria); Toni Pérez Fernández (Villacarrillo); Manuel Baena (Córdoba); Ana I. Camacho (Madrid); Andrés Ros (Murcia); Kenneth Meland (Bergen); and Mohammed Yacoubi (Marrakech). In the Azores, special thanks are extended to Isabel Amorim and Fernando Pereira (Terceira); Paulino Costa (Pico) and Teofilo Braga, Emanuel Pacheco, Manuel Soares, Cassilda Medeiros, Rita Pinheiros, José P. Medeiros and SINAGA S.A. (São Miguel).

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version, at https://doi.org/10.1016/j.ympev.2018.07.002.

References


