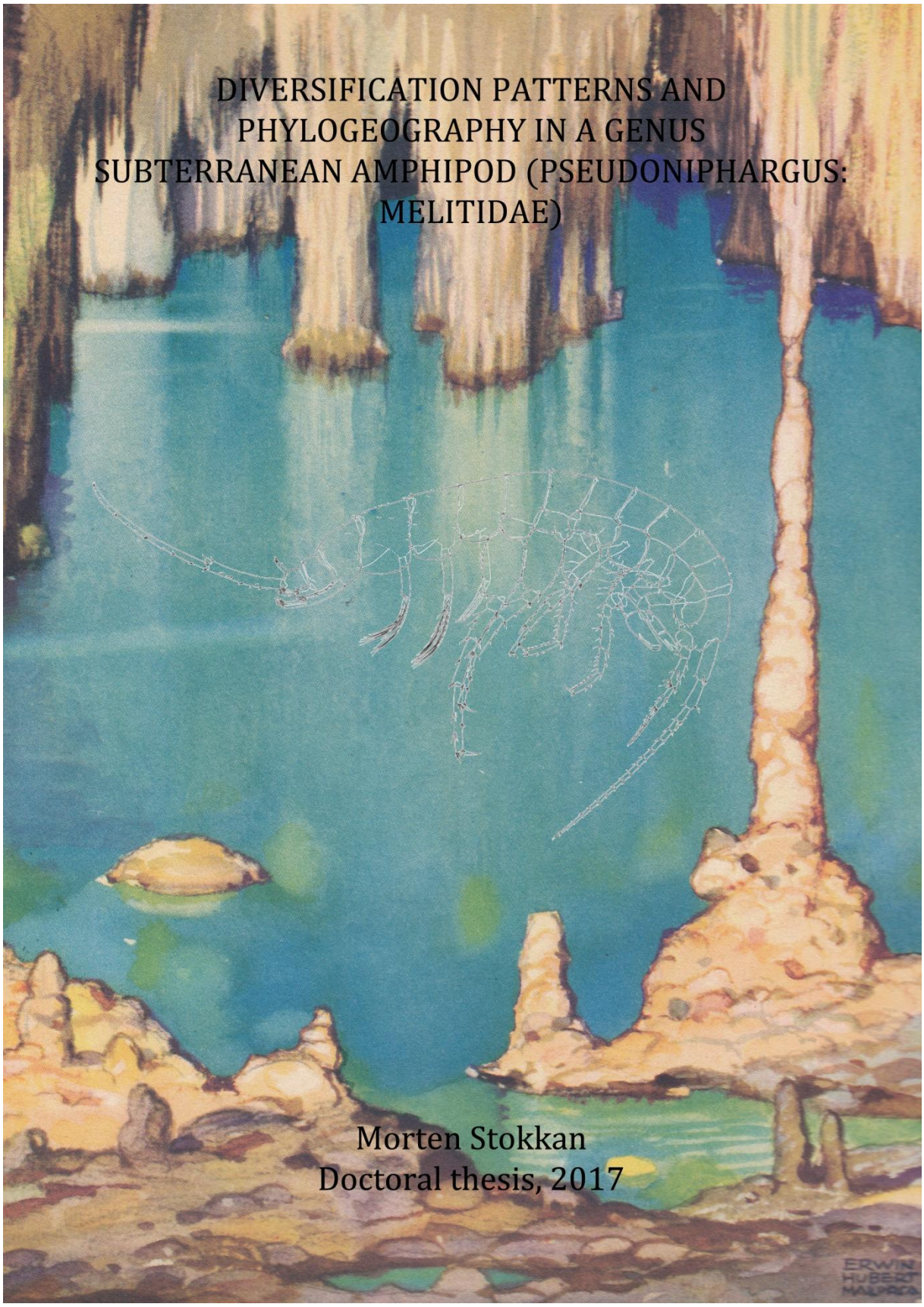

DIVERSIFICATION PATTERNS AND PHYLOGEOGRAPHY IN A GENUS SUBTERRANEAN AMPHIPOD (PSEUDONIPHARGUS: MELITIDAE)
Morten Stokkan
Doctoral thesis, 2017
ERWIN HUBERT MALPICA

Front cover By artist Erwin Hubert.

Universitat
de les Illes Balears

DOCTORAL THESIS
2017

DIVERSIFICATION PATTERNS AND PHYLOGEOGRAPHY IN A GENUS SUBTERRANEAN AMPHIPOD (PSEUDONIPHARGUS: MELITIDAE)

Morten Stokkan

Universitat
de les Illes Balears

DOCTORAL THESIS
2017

Doctoral Programme of *Marine Ecology*

DIVERSIFICATION PATTERNS AND PHYLOGEOGRAPHY IN A GENUS SUBTERRANEAN AMPHIPOD (PSEUDONIPHARGUS: MELITIDAE)

Morten Stokkan

Director/a: Damià Jaume Llabrés
Director/a: Joan Pons Pons
Tutor/a: Antoni Martínez Taberner

Doctor by the Universitat de les Illes Balears

Acknowledgements

There is a myriad of people that I would like to take this opportunity to thank for their inspiration and or influence in the process of this thesis. It has been a great part of my life and has been influenced in infinite different ways.

First and foremost I would like to take this opportunity to thank my supervisors Damià Jaume Llabrés and Joan Pons Pons for their endless support and guidance. You guys nurtured my interests and encouraged me to continue. And for listening to me and lending me your finite time to my work.

To José Antonio Jurado-Rivera for your friendship, showing me how to really work in a lab making me a better biologist and showing me how to be an honorary Spaniard.

Also to my loving wife and daughter for being there with me, holding out with me on these long days of work and for showing me the endless love that I feel for them. Thank you for teaching me so much and being a part of my life.

For my amazing parents for their unconditional support and care, as well as their help and kind words and the connection I feel with them.,and for my extended family both in Norway and Mexico for their love and support.

I would also like to take this opportunity to thank Dr. Lars Podsiadlowski for allowing me to visit the University of Bonn, being interested in my subjects and showing me “the unknowns” of the field and being a great inspiration.

For my friends and colleagues at the University of the Balearic Islands (UIB) and The Mediterranean Institute of Advanced Studies (IMEDEA) for the great and caring community we created, when most of us were far away from family we became our own family together.

For my friends and colleagues back in Norway that I have shared over a decade of friendship, learning, laughter and love with. I thank you for always cheering me on and continuing to broaden my knowledge with fantastic and interesting conversations.

Thanks to the Spanish National Research Council (CSIC) for realizing the FPI program and providing the funding, allowing me to embark on this wonderful, expanding and amazing journey.

And finally I would like to thank Spain and more specifically Mallorca and its people for their warmth and acceptance, for showing me the culture and traditions and expanding my mind both in science and as a person.

Thank you all

Published papers of this thesis:

STOKKAN, M., JURADO-RIVERA, J. A., JUAN, C., JAUME, D. & PONS, J. 2016. Mitochondrial genome rearrangements at low taxonomic levels: three distinct mitogenome gene orders in the genus Pseudoniphargus (Crustacea: Amphipoda). *Mitochondrial DNA***,** 1-11. (See **Appendix 2**)

Index

Summary 8
Chapter 1. Introduction 11
Chapter 2. Objectives 35
Chapter 3. Materials and Methods 36
Specimen sampling and morphological methods 36
Molecular methods 41
Chapter 4. Results and Discussion 51
Chapter 4.1 Diversity, species delimitation and crypticism in a genus of subterranean water amphipod crustaceans (*Pseudoniphargus*: Pseudoniphargidae) 51
Results 51
Discussion 57
Chapter 4.2 Taxonomical description of new species 62
Taxonomy 62
Chapter 4.3 Mitogenome rearrangement of 3 species of the genus *Pseudoniphargus* 80
Results and Discussion 80
Chapter 4.4 Phylogeny based on 32 mitogenomes 96
Results 96
Discussion 114
Chapter 5. Conclusions 121
Chapter 6. References 123
Appendix 142

Summary

This thesis deals with the phylogeny and phylogeography of the aquatic crustacean amphipod genus *Pseudoniphargus* Chevereux, 1901, an obligate inhabitant of continental subterranean waters. Special emphasis is placed on the origin of the genus, its extremely disjunct distribution on both sides of the Atlantic as well as in the Mediterranean Sea, and the processes underlying this distribution. Both morphological and molecular approaches are used to investigate the species diversity and detect possible cases of morphological convergence or of recently formed species hidden under cryptisism. Additionally this thesis attempts to place the evolution of the genus in a space-temporal framework by using both fossils and palaeogeographical events as calibration dating points. An ample sampling and collecting effort across the entire known distribution of the genus was undertaken and together with rigorous morphological analysis, molecular data was collected. Sanger sequencing and Next Generation Sequencing (NGS) was utilized to obtain both individual short marker and entire mitochondrial genomes covering all the major clades of the genus. By using four different methods of species delimitation based on mitochondrial *cox1*sequences, as well as morphology and geographical location we were able to assess the diversity attained by *Pseudoniphargus* across the vast majority of its current distribution. Several potential cryptic species were encountered, highlighting both the presence of strong morphological convergence and the necessity of using a broad multiple methodological approach to fully investigate species diversification. Two new species are formally described *P. morenoi* and *P. gevi*, both from gypsum caves in southern Spain. We used NGS on 31 species covering the major clades and distribution of the genus to recover their mitochondrial genome. The thesis presents a strongly supported, fully resolved phylogeny of the genus *Pseudoniphargus* identifying four main clades within the genus. By using both fossil data and palaeogeographical events we were able to place the genus in a temporal framework enabling to identify both vicariance and dispersal as contributing processes to explain the current distribution of the genus. Age and origin of the genus is estimated at 55 Ma, consistent with previous investigations and correlates well with the ancient (Eocene) Tethys Sea configuration and its subsequent regression. The mitogenomes also revealed the occurrence of an unprecedented level of gene rearrangements within the genome of various species, up to five, whereupon two of them involved protein coding genesthat is rare event in metazoans, and particularly within a genus. A detailed analysis of the nucleotide and aminoacid composition, and the secondary structure of both RNA and the control region shed light on the molecular patterns of evolution in mitochondrial genomes that in turn are crucial to implement accurate phylogenetic models including mitochondrial sequences. The new mitochondrial genomes presented herein offer an important possibility to establish further comparisons with other amphipods from identical habitats and showing similar distribution patterns to *Pseudoniphargus*.

Resumen

Esta tesis investiga el origen y la diversificación del género de crustáceos anfípodos *Pseudoniphargus,* Chevereux, 1901, exclusivo de aguas subterráneas continentales que presenta una distribución disjunta extrema con representantes a ambos lados del Atlántico Norte. Para ello se realizó un amplio muestreo que ha abarcado el área de distribución completa del género, que ha permitido abordar un riguroso análisis morfológico y genético de 410 especímenes. Mediante el uso de cuatro criterios diferentes de diferenciación de especies basados en un único locus de ADN, combinados con la morfología y la ubicación geográfica, se evaluó la diversidad de *Pseudoniphargus* a lo ancho de toda su área de distribución. Se demuestra que existe un elevado grado de concordancia entre especies morfológicas y especies definidas con criterios genéticos aunque esta última permitió descubrir varias especies potencialmente crípticas. Además, se describieron dos especies nuevas *P. morenoi* y *P. gevi*, ambas descubiertas en cuevas excavadas en yeso en Andalucía. La diversificación de los principales linajes de *Pseudoniphagus* se ha reconstruido a partir de una filogénia altamente soportada con 4 clados principales asociados a 4 zonas geográficas concretas construida con las secuencias de genomas mitocondriales de 31 especies obtenidas mediante *Next Generation Sequencing*. La calibración de la filogenia molecular obtenida, una vez detectada la presencia de tasas de substitución heterogéneas, ha permitido identificar tanto a la vicarianza como a la dispersión trans-oceánica como factores que han contribuido a la generación del patrón de distribución actual del género. Según la filogenia molecular elaborada en el presente estudio, el origen del género *Pseudoniphargus* se remonta a 55 millones de años, siendo esta fecha consistente con estudios previos basados únicamente en caracteres morfológicos de las especies, y que lo asociaban a la regresión de paleo-líneas de costa del Eoceno. El análisis de los mitogenomas obtenidos reveló un alto grado de reordenamiento de genes en varias especies, hasta 5, dos de ellos involucrando a genes codificantes para proteínas. Los genomas mitocondriales presentados brindan una gran oportunidad para establecer comparaciones con otras especies de anfípodos con hábitat y patrones de distribución similares a *Pseudoniphargus.*

Resum

Aquesta tesi investiga l'origen i la diversificació del gènere de crustacis amfípodes *Pseudoniphargus*, Chevereux, 1901, exclusiu d'aigües subterrànies continentals que presenta una distribució disjunta extrema amb representants a banda i banda de l'Atlàntic Nord. Per a això es va realitzar un ampli mostreig que ha abastat l'àrea de distribució completa del gènere, que ha permès abordar una rigorosa anàlisi morfològica i genètica de 410 espècimens. Mitjançant l'ús de quatre criteris diferents de diferenciació d'espècies basats en un únic locus d'ADN, combinats amb la morfologia i la ubicació geogràfica, es va avaluar la diversitat a *Pseudoniphargus* a l'ample de tota la seva àrea de distribució. Es demostra que hi ha un elevat grau de concordança entre espècies morfològiques i espècies definides amb criteris genètics encara que aquesta última va permetre descobrir diverses espècies potencialment críptiques. A més, es van descriure dues espècies noves *P. morenoi* i *P. gevi*, ambdues descobertes en coves excavades en guix a Andalusia. La diversificació dels principals llinatges de *Pseudoniphagus* s'ha reconstruït a partir d'una filogènia altament suportada amb 4 clados principals associats a 4 zones geogràfiques concretes que es va construir amb les seqüències de genomes mitocondrials de 31 espècies obtingudes mitjançant *Next Generation Sequencing*. El calibratge de la filogènia molecular obtinguda, un cop detectada la presència de taxes de substitució heterogènies, ha permès identificar tant a la vicariança com a la dispersió trans-oceànica com a factors que han contribuït a la generació del patró de distribució actual del gènere. Segons la filogènia molecular elaborada en el present estudi, l'origen del gènere *Pseudoniphargus* es remunta a 55 milions d'anys, sent aquesta data consistent amb estudis previs basats únicament en caràcters morfològics de les espècies, i que l'associaven a la regressió de paleo-línies de costa de l'Eocè. L'anàlisi dels mitogenomes obtinguts va revelar un alt grau de reordenament de gens en diverses espècies, fins a 5, dos d'ells involucrant a gens codificants per proteïnes. Els genomes mitocondrials presentats brinden una gran oportunitat per establir comparacions amb altres espècies d'amfípodes amb hàbitat i patrons de distribució similars a *Pseudoniphargus*.

Groundwater habitat

"Eventually, all things merge into one, and a river runs through it. The river was cut by the world's great flood and runs over rocks from the basement of time. On some of the rocks are timeless raindrops. Under the rocks are the words, and some of the words are theirs. I am haunted by waters."

— Norman Maclean, *A River Runs Through It and Other Stories*

Chapter 1. Introduction

Groundwater

The world is frequently called the blue planet based on the water it contains, although most of the water is found in the oceans. Fresh water is a more limited commodity, but is vital for much of life on earth. An estimated 30 % of the world freshwater is groundwater whereupon only 1.2 % represent lakes and rivers (U. S. Geological Survey, 2014). In some places the number has been listed as even higher where up to 97% out of total global freshwater is groundwater (Marmonier et al., 1993). Groundwater as a resource both now and in the future is a matter of great importance not only anthropologically, but also in questions of biodiversity as well as habitat and species conservation (Gibert et al., 2009; Griebler and Avramov, 2015; Griebler et al., 2014; Griebler et al., 2010; Tuinstra and van Wensem, 2014). In fact, the European Union relies on 75 % of the drinking water coming from groundwater (Tuinstra and van Wensem, 2014). However, the level of studies on groundwater compared to what has been done on lakes and rivers are quite disproportionate. Historically this has been due to its relative inaccessibility since the best gateways to these habitats are represented by cave systems, wells, bore holes and riverbeds, which are scarce, onerous and intricate. These factors make groundwater systems an important area for future studies in an array of different disciplines.

Groundwater refers to as all sub-surface waters and can be divided into several types based on intrinsic properties such as geology, depth, and water flow. However most influential for the subject of this thesis is groundwater contained in aquifers. The standard definition of an aquifer is "a rock unit capable of storing and transmitting water" (White, 2005). There are three main types of aquifers; fissured which is mainly comprised of granite or other types of insoluble rock; porous comprised of alluvium or loose soil and unconsolidated material; and finally karstic aquifers, made up of predominately limestone and dolomite, that can be developed also in evaporitic rock such as gypsum (Culver and Fong, 1994; Culver et al., 2009; White, 2005). These different aquifers vary in structure and hence the space to be taken up by water. The speed and water flow are also influenced by the geological structure, and as a direct consequence the level of nutrients and biological factors differ substantially, greatly affecting the biological composition and evolution of subterranean water dwellers (Marmonier et al., 1993).

History of groundwater research

Caves have likely been highly influential to human contact with groundwater as one of the more conspicuous access points to such an environment. Human history with caves goes back millennia, as these natural structures served humans, among others, as shelter, storage, rituals, burials and rock paintings and the very first cave explorers left only indirect evidences of their surveys. In more modern time written discussion and exploration of caves and cave fauna can perhaps be dated from the 17th century (Culver and White, 2005). Subterranean scientific studies of cave fauna may be said to begin in caves near Vrhinka in Slovenia where the cave salamander *Proteus anguinus* was first described by Laurenti (1768) and referred to as "a young dragon". Subsequent studies of caves followed also in Slovenia in the region Kras, which is located in the coastal region between Italy and Croatia, and hence giving origin to the name "Karst". Karst caves are distributed in different parts of the world, some with an extensively studied fauna (Culver and Fong, 1994; Culver et al., 2009; Trontelj et al., 2007; Zakšek et al., 2007).

The first record of scientific research focusing on groundwater ecology and evolutionary topics was by Racovitza in 1907 (Danielopol and Griebler, 2008), and for a long time these subterranean areas were considered from a biological perspective as fringe habitats, or anomalies from other types of habitats. Further interest and several studies over the subsequent 50 years, were mostly in a preliminary stage of identifying and cataloguing species from the hypogeal habitat or, below ground habitat, expanding knowledge of cave organisms not only in Europe but throughout the world (Culver and White, 2005; Danielopol and Griebler, 2008; Holsinger, 2005). New disciplines and discoveries have since vastly expanded the body of knowledge about groundwater as a whole. Many disciplines have been investigating the complex structure and nature of groundwater as a system, geology, geochemistry, physics, hydrology, hydrochemistry, microbiology and ecology to name a few. This has led to a plethora of terminology that is not always synchronized within the respective fields of research (Carapelli et al., 2007; Ginés and Ginés, 2007; Marmonier et al., 1993). The groundwater in karstic areas is complex and can often be subdivided into several sections of habitats. The epicarst, anchialine caves and marine caves are all different habitats with varying parameters, producing natural barriers for organisms, often within the same system of underground water flow (Bakalowicz, 2005; Brancelj and Culver, 2005; Danielopol et al., 2000; Iliffe, 2005; Schmidt and Hahn, 2012; White, 2005). The parameters in marine caves are highly influenced by the connecting sea (Suric et al., 2010), and anchialine caves despite being often situated further from the coastline are still connected to the adjacent sea. These caves may be influenced by the fluctuation of tides and precipitation, causing a high variation in salinity within the same cave (Iliffe, 1992; Iliffe and Kornicker, 2009; Sket, 2005). The epikarst is the porous zone of karst where climate and roots create permeable gateways for water flow coming from surface precipitation (Bakalowicz, 2005). Furthermore, the groundwater system has several types of ecotones adjacent to the main ground water system such as the hyporeic zone of rivers and streams (Arntzen et al., 2006), as well as the shift from the unsaturated to the saturated zone in lakes and rivers (Fraser et al., 1996; Griebler et al., 2014) This variety of types of groundwater habitat and ecotones provide different habitats for many organisms (Bakalowicz, 2005; Brancelj and Culver, 2005; Culver and White, 2005; Ginés and Ginés, 2007; Iliffe, 2005).

Groundwater conditions

Groundwater habitats are mostly oligothrophic i. e. with low organic content, meaning that most energy comes from an external source. This makes the system highly depended on

organic matter from the surface. The lack of primary producers and lack of energy is thought to favor omnivore habits instead of just predatory ones (Gibert and Deharveng, 2002). With limited resources, maximizing energy intake from various sources is an important ability. The influx of organic matter from the surface can be quite variable over time, season or region, depending mostly on precipitation. This in turn can highly influence species dynamics, particularly in shallow groundwater (Schmidt and Hahn, 2012). Although the air temperatures are quite stable over long periods of time especially in the deeper levels of caves, it can vary due to the latitude, altitude, cave size and ventilation in the cave system. This in turn, dictates the water temperature in the caves, that often follow that of the air, albeit with a difference of up to 1°C. Some groundwater animals like the isopod *Stenasellus virei* does not seem to be affected by large temperature changes. Even though original conditions in the groundwater habitat were temperatures of 5-7°C, laboratory experiments showed viable conditions for the species at up to 18°C (Magniez, 1975). Water flow can also vary greatly and is intrinsically linked to surface events such as precipitation, snow melting, etc. This factor can have an important impact on the groundwater habitats. For instance, in slow running or still water, sediment deposits can create anchor points for stable populations, as opposed to fast turbulent water.

Subterranean fauna

Another characteristic of the groundwater habitat is its frequent lower biodiversity compared to its epigean counterpart, meaning habitats above ground. There are three major terms for the categorization of the various levels of groundwater fauna (**Figure 1.1**). Stygobionts is the term for animals living permanently in groundwater. Stygophiles have lifecycles partially connected to, or frequent appearances in groundwater. Stygoxenic taxa are generally closely connected to the outside cave areas and are sometimes found inside (Barr and Holsinger, 1985; Gibert and Deharveng, 2002; Schmidt and Hahn, 2012). These terms refer to aquatic or marine animals from groundwater, and have their terrestrial hypogean counterparts with their specific nomenclature. Troglobionts are terrestrial taxa that are strictly found in caves and cave systems particularly Orthoptera, Coleoptera, Collembola and Araneae (Allegrucci et al., 2005; Barr and Holsinger, 1985). Troglophiles are land living animals that have significant but partial connection to cave life. Examples of this are bats, birds, snakes and some chilopods (Arita, 1996; Drda, 1968; Molinari et al., 2005; Speich et al., 1986). The term trogloxene and troglophile are difficult to separate and refer to which extent or gradient, the animal inhabits a cave. Due to the some times fragmented habitat and the often poor dispersal capabilities there is a high level of endemism in groundwater and caves (Christman et al., 2005; Culver and Sket, 2000; Iliffe and Kornicker, 2009; Lejeusne and Chevaldonne, 2006).

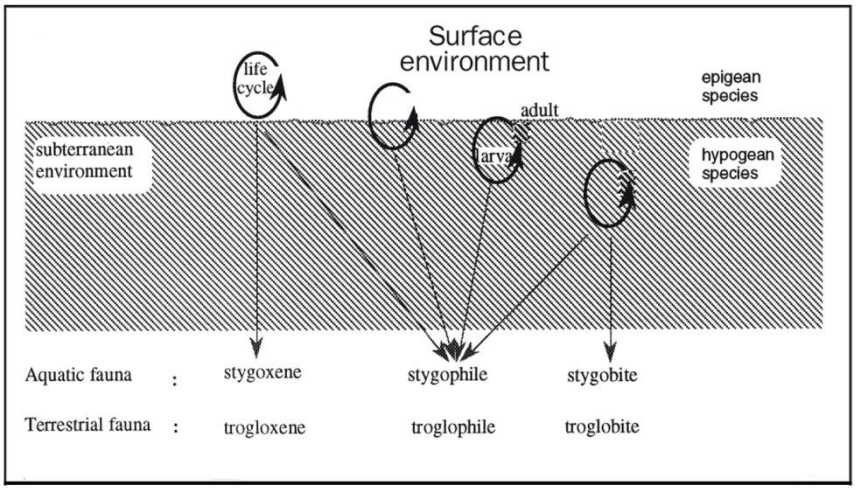


Figure 1.1 Graphic presentation from Gibert and Deharveng 2002, showing the life strategies and stages from stygoxene to stygobiont (aquatic) and trogloxene to troglobiont (terrestrial), and their various affiliations to the hypogean and epigean habitat.

Morphological characteristics of cave and groundwater fauna

Some attributes of groundwater and cave habitats such as perpetual darkness have often exerted a profound impact on the animal morphological structure and evolution. It is often mentioned the rudimentation or the regression of various morphological structures and sensory equipment (Botosaneanu, 2001; Juan et al., 2010). However equally important is the acquisition of novelties such as other sensory appendages and receptors (Botosaneanu, 2001). These regressive and progressive traits are often referred to as troglomorphy (Christiansen, 1962).

The most widely known and common regressive trait is the atrophy of the ocular sensory structures, the so-called anopthalmy. This phenomenon has been documented in virtually all cave fauna from fish to amphipods, as well as in terrestrial vertebrates and invertebrates. Other examples of reduction are the loss of body pigmentation and in some cases "smoothness" of cuticula (Botosaneanu, 2001). On the other hand, massive growth of some body structures is recurrent among most stygobiontic invertebrate fauna. Elongation of the body and appendages, and a higher developed chemo and mecano-receptors are the most commonly reported cases (Botosaneanu, 2001; Brancelj and Dumont, 2007; Holsinger, 1994; Poulson, 1963).

Groundwater biodiversity- a homage to Santa Rosalia

Hutchinson in 1959, wrote about his collecting trip to Monte Pellegrino (Palermo, Italy) to sample aquatic insects of the genus *Corixa* from the cave where the skeleton of Palermo´s patron Saint Rosalia was discovered. He contemplated there, why he found only two species of *Corixa* in the small pond and not 20 or 200. Based on this rather trivial observation he later developed several ideas and reason for numbers of species and the vast diversity of species around the world that ultimately inspired the question; "why are there so many kinds of animals?" This is a question that can be asked in a different manner "what are the limiting factors to biodiversity?", a question that perhaps still does not have a complete and satisfactory answer (Hutchinson, 1959; Stoch, 1995).

There are currently over 15,000 species described from subterranean habitats (Holsinger, 2005), but this number is and will probably increase in time to come as many species are yet to be described and discovered (Gibert et al., 2009; Holsinger, 2005). Many studies have

detected occurrence of high levels of cryptisism within stygobionts further elevating the species diversity (Camacho et al., 2011; Finston et al., 2007; Lefebure et al., 2007; Meleg et al., 2013; Zakšek et al., 2009). There is higher species richness in European groundwaters than in other parts of the world, though the reasons for this are still not fully understood. One part of the explanation may be sampling bias; but even in extensively studied groundwater systems outside of Europe, species richness levels are significantly lower than in European sites (Deharveng et al., 2009; Gibert et al., 2009). Groundwater was for decades considered to be a habitat of poor diversity and low species richness mostly due to lack of light and nutrients, although these views have slowly been changing, especially when Bacteria and Archea are introduced into the diversity concept (Danielopol et al., 2000; Gold, 1992; Griebler et al., 2014). The view has recently shifted from a fringe type habitat with poor diversity to higher levels than expected (Griebler et al., 2014). However there is still a distinct difference between epigean and hypogean fauna when it comes to species diversity numbers. Some groundwater habitats showing for example 20 species are considered to be species-rich for this type of habitat, albeit easily exceeded by even the poorest surface water diversity levels (Stoch and Galassi, 2010). On the European continent one major factor promoting species diversification is the Quaternary glaciations that have greatly affected the stygobiontic species richness. This pattern is also depicted in a North to South gradient of species richness. The limnic surface habitats are relatively recent, with most lakes and rivers in the area having a post-glacial origin (Foulquier, 2008; Stoch and Galassi, 2010).

There is also a greater representation of some taxonomic groups in stygobiontic faunas whereas others are almost absent. Crustaceans, and in particular amphipods, isopods and copepods are overrepresented in the groundwater fauna (Stoch, 1995). In fact, in Europe 1,111 of the 2,285 described freshwater crustacean species are stygobionts, meaning that half of the crustaceans in European freshwaters live in perpetual darkness. It is also important to take into account that the rate of newly described epigean species has already plateaued in the 1800, while no such indication is yet reached for the still increasing rate of new hypogean species described until now (Stoch and Galassi, 2010). On the other hand, there is a poor diversity of aquatic insects in hypogean habitats, as opposed to the epigean freshwater environment (Stoch, 1995). The processes behind this discrepancies are not fully understood, although pre-adaptation and exaptation enabled by the crustacean bauplan could be an important factor (Rouch and Danielopol, 1987; Stoch, 1995).
These are the main features heavily influencing subterranean biodiversity (Camacho 1992; Gibert and Deharveng, 2002):

-The absence of light, which creates an ecological filter that few lineages can surpass.
- The truncated food web, due to lack of primary producers, with food scarcity or its irregular availability shifting the normal pattern of primary producers, herbivores and predators, towards fewer predators and more omnivores.
- Severe habitat fragmentation resulting in a high level of endemism.
- Habitat persistence and stability over vast periods of time, enabling survival of a high number of relict taxa.

Marine relicts or freshwater refugees

A very natural question about the inhabitants of groundwater is “where did they come from”. Leading theories suggest that there are two main processes through which an organism becomes a stygobiont. The regression model proposes marine organisms, most commonly from the littoral zone, become eventually trapped within the interstitial medium associated to

the seashore or in costal cave systems as sea recedes (Notenboom, 1991)(**Figure 1.2**). Another hypothesis suggests that freshwater organisms have sought refuge or arbitrarily ended up in subsurface habitats following the infiltration from lakes and rivers, eventually being adapted and become a part of groundwater ecosystems (**Figure 1.3**). In the case of amphipods, most groups seem to have a marine origin, as they are distributed only in areas formerly occupied by the sea in past geological periods, or belong to families or genera typically marine (Botosaneanu and Holsinger, 1991; Holsinger, 1994; Notenboom, 1987b; Notenboom, 1991; Zakšek et al., 2007). This postulation especially holds true for *Pseudoniphargus* (Notenboom, 1991; Stock, 1980), where a marine origin is assumed since some species within the genus are still found in habitats still connected to or influenced by the sea such as anchihaline caves or the hyphoreic zone of freshwater springs connected to the littoral zone (Bréhier and Jaume, 2009; Sánchez, 1989; Sánchez, 1990; Sánchez, 1991; Stock, 1988; Stock and Abreu, 1992; Stock et al., 1986). Additionally, several species are found on oceanic islands that have not been a part of the continents (Notenboom 1991).

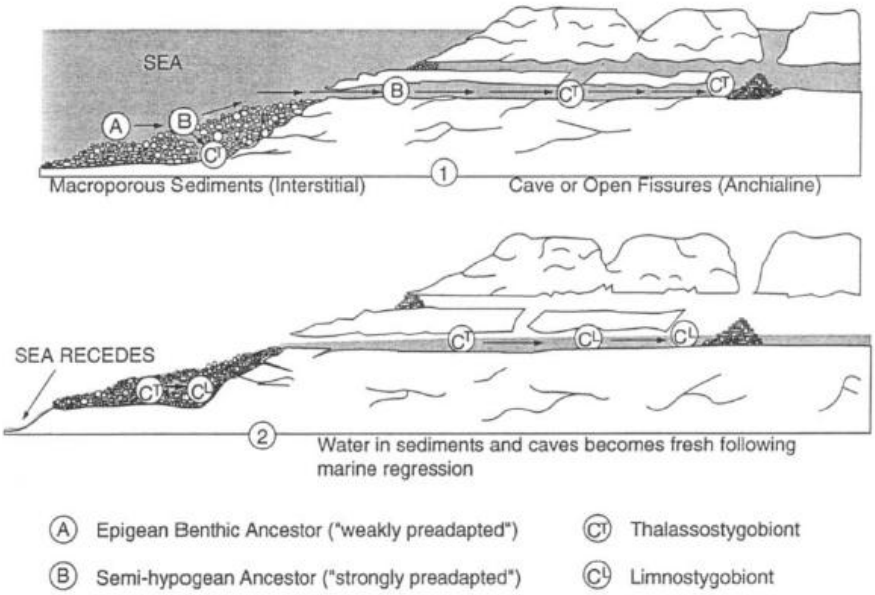


Figure 1.2 Illustration of the marine regression model. 1 Where thalassiosid organism from the littoral zone adapt and colonize underwater cave habitats through dispersal. 2) Where sea level declines over time entrapping and isolating littoral organisms. Image from Holsinger 1994.

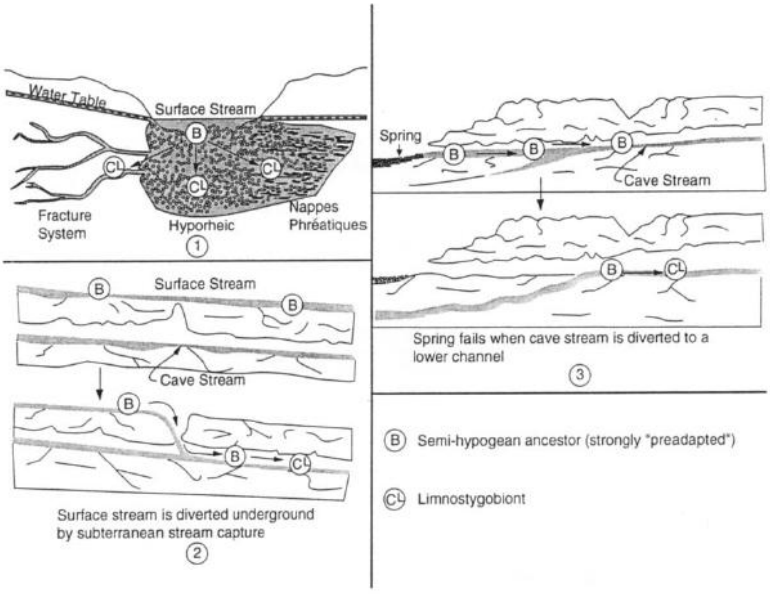


Figure 1.3 Illustration of the possible origin of limnic stygobionts through three different cenario. 1 Where hypogean waterways are used as refuges or in some way benefits an organism and colonization by dispersal. 2 Where a epigean stream converts to a hypogean stream by changing path or capture by a underlying cavity or stream. 3 Colonization by dispersal, with subsequently isolation caused by stream altering path. From Holsinger 1994.

Biogeography

Biogeography studies the geographical distribution of a given taxa and establishing a link between distributions over time. Large scale biogeography, explains distribution stretching over extremely long periods of time and explaining distribution mostly as caused by continental plate shifts or land emergence, while small scale biogeography often explains more local distribution caused by ecological effects within a smaller time-frame. (Holsinger, 2005). In general stygobiont organisms are exceptionally good model organisms to study biogeography. This is due to their ranges of distribution are relatively small, their levels of endemism are often high (Porter, 2007), and they frequently represent old lineages occupying subterranean refuges for long periods of time (Juan et al., 2010; Porter, 2007). Both adaptations to a subterranean habitat and stability of environmental conditions have led to morphological convergence and stasis, respectively, and hence precluding taxonomic distinction between closely related species. In the last two decades, phylogeographic studies of populations at the DNA level has allowed to analyze the molecular variance between and within populations to shed light on the evolutionary, phylogenetic, and geographical patterns and date their evolutionary history.

Vicariance or dispersal

One imperative question in biogeography of stygobiont species is determining whether vicariance or dispersal drove the present distribution of a particular lineage (i. e. diversification) although they are not mutually exclusive processes. Dispersal assumes a movement of an organism from one area to another across a preexisting geographical barrier, whereas vicariance envisages speciation as driven by the split of closely related taxa in disjunct areas due to the creation of a new geographical barrier (Holsinger, 2005). These natural boundaries can be a sea regression, terrain uplifts, plate tectonics, or rivers. Historically when speleology was in its infancy, the subterranean fauna was thought to be exclusively found in caves, and these were considered a sort of isolated islands since concepts of groundwater and connectivity between caves were still not developed. These facts implied

that all subterranean animals evolved isolated in a single cave with little or no dispersal (Culver et al., 2009). This point of view changed after the discovery of stygobiontic life in other habitats aside from caves, as in sediments of riverbeds and underground streams (Culver et al., 2009). As these habitats are more commonplace and not discontinuous, researchers realized that there was a greater potential for dispersal than previously presumed. This is pivotal for the stygobionts since groundwater is a continuous interconnected habitat occupying a greater area with pathways for dispersal especially in porous karstic grounds. In contrast, terrestrial troglobionts have in general a lower ability to disperse due to the discontinuous habitat (Holsinger, 2005). In later years, the advances in DNA sequencing and population biology reassembled vicariance as a possible and probable cause of speciation, based on the strong correlation found between current distributions and ancient geological events such as plate tectonics and the regression of the Tethys sea in the Mesozoic time era (Bauzà-Ribot et al., 2012; Culver et al., 2009; Holsinger, 1991; Holsinger, 2005; Juan et al., 2010; Kelly et al., 2006; Porter, 2007; Shih et al., 2011; Stutz et al., 2010). Rather than to argue for or against the motion of either of these two processes, it is more likely that the real picture is a result of both processes to a smaller or greater extent, and not an absolute dichotomy.

Amphipods have a documented ability to disperse in subterranean habitats (Lefebure et al., 2006; 2007). In addition to the connectivity of the aquatic subterranean realm, several studies show an ability for passive dispersal in different organisms such as planarians, copepods, isopods and amphipods, all smaller than 10 mm that can be transported from epikartic aquifers and drip via the cave ceiling into shallow subterranean pools (Holsinger, 2005). This circumstance renders it possible for different instars to disperse. Nevertheless, subterranean amphipods show a low level of larval dispersal due to eggs being retained and develop in a marsupium or brood pouch conformed by the female oostegites. The only exception could be the Caprellids and other amphipods closely related to algae that can disperse with algae rafts, and Hyperiids that live in pelagic symbiosis with Medusozoa (Auel and Werner, 2003; Cabezas et al., 2013a; Cabezas et al., 2013b; Gasca and Haddock, 2004; Laval, 1980). Disjunct distribution between congeneric species within stygobionts is not uncommon. Examples of this can be found in the remiped *Speleonectes* Yager, 1981, the anthurid isopod *Curassanthura* Kensley,1981 the shrimp *Typhlatya* Creaser, 1936 and the amphipods *Metacrangonyx* Chevreux 1901, *Spelaeonicippe* Stock and Vermeulen 1982, as well as *Pseudoniphargus,* all having a amphi-Atlantic distribution (Iliffe, 2005). This biogeographic paradigm has often been associated with plate tectonics via de aperature of the Atlantic. Molecular clock analyses with mitochondrial genomes sustained this hypothesis in *Metacrangonyx (*Bauzà-Ribot et al., 2012*),*but remains to be analyzed and corroborated for the focal taxon group of this thesis, the genus *Pseudoniphargus,* Karaman 1993.

Family Pseudoniphargidae Karaman 1993

Amphipod taxonomy, and specifically at the higher ranks, has a history of being everchanging and unsettled (Lowry and Myers, 2013; Bousfield, 1977). Amphipods are riddled with character reemergence, adaptive radiation and convergence or homoplasy. This is even more evident in stygobiontic fauna since they live in similar but fragmented habitats (Lefebure et al., 2006). They generally may cause great difficulties for traditional taxonomy, since distant relatives exhibit highly similar morphology. Thus molecular methods represent an independent source of evidence to substantiate classical taxonomy. Coexistence and mutual dependability of the two disciplines among others such as ecology and ethology have historically helped to identify, describe and classify species.

Pseudoniphargus was firstly placed within the family Gammaridae Leach 1814 (sens. lat.) as all stygobiontic amphipods at that time. This family was later revised by Bousfield in 1977 for the reason of being polyphyletic, and illogically large compared to other families within the superfamily Gammariodea, placing *Pseudoniphargus* within the family Niphargidae Karaman, 1962. This was contested by Barnard and Karaman (1980), arguing for its placement in Melitidae Bousfield, 1973 (Notenboom, 1988c). Later, Karaman in a paper in 1993 erected the family Pseudoniphargidae wherein *Pseudoniphargus* is one of two genera, the other being the monotypic genus *Parapseudoniphargus* Notenboom, 1988, represented by *Parapseudoniphargus baetis* endemic to the Guadalquivir river basin in southern Spain. Elevating Pseudoniphargids to the family level was done in a somewhat unorthodox manner where it was never explicitly mentioned as a family by Karaman, which does not concur with the ICZN (International Commission on Zoological Nomenclature) rules for establishing a family. Despite this, a type genus *Pseudoniphargus* was provided, and this is now accepted as a valid family (Lowry and Myers, 2013). However, we opted for the more conservative approach of retaining Pseudoniphargids in the Melitidae in the thesis title (see **appendix 1** for a list of all currently known members of Pseudoniphargidae).

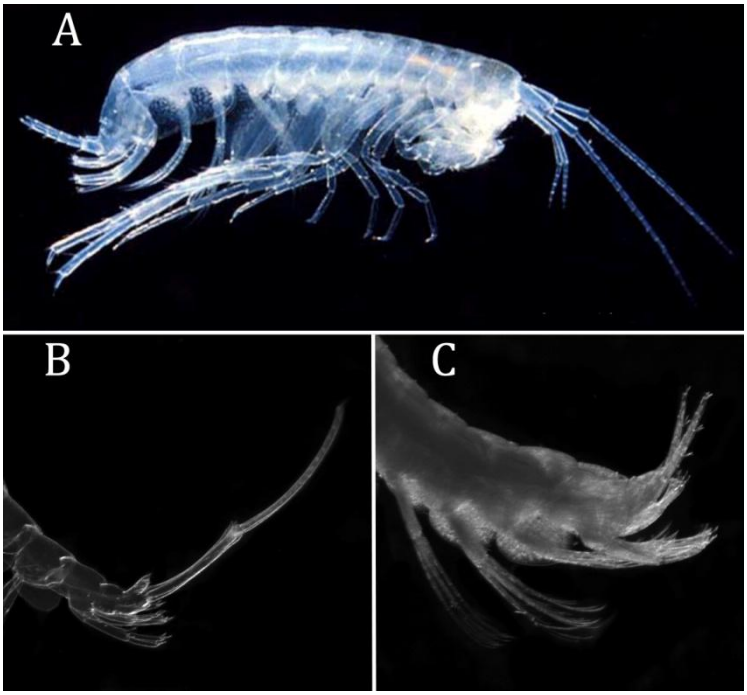


Figure 1.4 (A) Photo of habitus of *Pseudoniphargus* here represented by *P. grandimanus* from the island of Bermuda. (B&C) Detailed image of the 3rd uropod with presence and absence of an elongated uropod from two different species *P. mercadali* and P. sp (Pozo de Toni Martinez, Mallorca). One of the most conspicuous character differences between some of the *Pseudoniphargus* genera (Notenboom, 1987a; Notenboom, 1987b; Notenboom, 1988a; Notenboom, 1988c; Pretus, 1988). (Photo credits: A: Thomas Illife, B and C: Morten Stokkan)

Genus Pseudoniphargus Chevreux 1901

The genus *Pseudoniphargus* (**Figure 1.4**) was erected in 1901 by Edouard Chevreux with the type species *P. africanus* (Chevreux, 1901). It was firstly separated from *Niphargus* by the elongated articles on the outer ramus of uropods. *Niphargus* has a distoventral robust seta on urosomite 1 that is absent in *Pseudoniphargus*. Furthermore, the gnathopods 1 and 2 of *Niphargus* are in general more rounded and equal as opposed to *Pseudoniphargus* gnathopods that are strongly dissimilar in appearance, as well as gnathopod 1 being smaller than gnathopod 2. Another important feature to distinguish between both genera is that *Pseudoniphargus* often show sexual dimorphism (Lowry and Myers, 2013).
Parapseudoniphargus, the other genus recognized within the family differs from *Pseudoniphargus* in the lack of sexual dimorphism, display of a short exopodite on the 3rd uropod, the habitus of both head and body, the first coxal plates, and the lack of elongation of the posterior pereiopods (Notenboom, 1988b).
The taxonomy of stygobiontic amphipods is intricate and constantly altered by amendments. Subsequently, this issue also applies to the systematic position of the genus *Pseudoniphargus* that has undergone multiple changes since its discovery. It took 77 years after its original description until a second species was recognized when Karaman raised the then subspecies *adriaticus* to the species level (Karaman, 1978). Early publications stated that the species was distributed as widely as Portugal, Algeria, Tunisia, Yugoslavia, France, Corsica, Madeira, Azores and Spain, were all specimens were referred to as one species *Pseudoniphargus africanus.* Later Stock (1980) recognized the presence of several morphologically distinct species and elevated the number of species from 1 to 9 (Stock, 1980), due partially to its remarkable morphological plasticity and its wide distribution and occurrences from caves and wells at high altitudes above sea level as well as close to the sea shore. Since 1980, a cascade of new species were described until the current 69 species recognized today.

The 69 *Pseudoniphargus* species display an extremely disjunct distribution occurring as far north as Southern France, within the Mediterranean region, the Balearic Islands, and North Africa (Morocco and Algeria). The genus is also found adjacent to the Northern Atlantic in Northern Spain as well as in Portugal and the Azores archipelago, Madeira, the Canary Islands and as far west as Bermuda (Holsinger, 1994)(**Figure 1.5**).

Figure 1.5 Map showing the distribution of the genus *Pseudoniphargus*. Illustration from Holsinger 1994

Most of the 69 species known (v. Jaume et al. 2016) are very localized and often limited to occur on a single island or reduced portion of land (Stock 1980; 1988; Notenboom 1986; 1987a; 1987b; Boutin & Coineau 1988; Pretus 1988; 1990; Karaman & Ruffo 1989; Coineau & Boutin, 1996; Fakher el Abiari et al. 1999; Jaume 1991; Messouli et al. 2006; Bréhier & Jaume 2009). They mostly live in fresh inland waters ranging from the sea level to more than 1,000 m heigh (Notenboom 1987a). Aside a few exceptions (Stock et al. 1986; Notenboom 1987a; Stock 1988; Pretus 1990; Jaume 1991; Sánchez 1991) species behave as allopatric and show narrow and non-overlapping distributions, in many cases apparently reduced to a single cave or well.
Studies of this genus have been almost exclusively reduced to descriptive taxonomy and phylogenies using morphological characters. Only a single study has been performed on a population level (Mathieu et al., 1999), where the population structure of a *Pseudoniphargus sp*., was investigated and compared to *Niphargus rhenorhodanensis*.

Pseudoniphargus, Chevreux 1901 on the Iberian peninsula

In Iberia the genus is highly diversified, but far from being evenly distributed most of the species are concentrated in two nuclei that roughly correspond with the Cantabrian mountains on the north (13 species; Notenboom 1986), and the Betic ranges on the south and south-east (15 species; Notenboom 1987a). Aside of these territories, three species are known to occur on the western coast of the Peninsula (Notenboom 1987b), whereas there are two records (reported as *Pseudoniphargus africanus* Chevreux, 1901) from two caves at Tarragona, in NE Spain (Margalef 1970). As one might expect from a subterranean thalassoid lineage (*v.* Notenboom 1991), the genus is absent from the crystalline core of the Meseta (the central plateau of the Iberian Peninsula), permanently emerged since Palaeozoic times.

The *Pseudoniphargus* cluster from southern Spain was first studied by Notenboom (1987a), who reported the occurrence there of a minimum of 15 species. Notenboom (1988) related this high species diversity to the complex Cretaceous to Late Tertiary tectono-sedimentary history of the Betic ranges, and especially to the intricate and diachronous pattern of retreat of the sea from the numerous marginal and internal basins developed in the area during the Upper Miocene (Martín et al. 2014).
Here we describe two new species of *Pseudoniphargus* collected in two gypsum caves excavated in Triassic evaporites (Keuper) at Córdoba and Málaga provinces, respectively (southern Spain). These two localities do not show a physiographic connection and fall about 70 km apart, the cave at Córdoba being placed in a olistostrome detached from the Betic ranges and displaced into the Guadalquivir depression. That cave, "Cueva del Yeso", is 8 km north of the city of Baena and comprises 2,670 m of surveyed passages (Mora et al. 2011). It appeared quoted as "Cueva de las Palomas" in Margalef (1970), who reported the presumed occurrence there of *P. africanus* Chevreux, 1901, a species described from Algeria and with a distribution apparently limited to a narrow portion of land on the north of that country (*v.* Stock 1980). Since only specimens belonging to the new species appeared in our own collections from the cave, we guess the amphipods referred to by Margalef might correspond also to the new taxon. The second new species was collected in gours at "Complejo Romeral", a cave system comprising 600 m of surveyed passages located in the Gobantes karst (Antequera; Málaga; Disney et al. 2009).
The two new species described herein share the extreme elongation of the male third uropod, a striking troglomorphic feature that seems to have arisen independently in several of the lineages currently recognized within the genus (Notenboom 1988). These findings raise the number of species of *Pseudoniphargus* known in the area to 17.

Phylogenetic reconstruction

The main goal of phylogenetic analyses is the reconstruction of the evolutionary history of a set of organisms, generally species, to settle their ancestor-descendant relationships in a dichotomus tree. The basis of phylogenetic systematics is that only monophyletic groups are relevant for an adequate classification meaning those who include the ancestor and all of its descendants (Henning, 1999). In the early stages of phylogenetic systematic, this reconstruction was performed by comparing mostly morphological traits, found on extant species which are assumed to be homologous *a priori* in order to determine the tree with dichotomous connections of ancestor-descendant relationships among extant species and their putative ancestors. This approach is still the dominant method for taxonomists and systematics to recognize and establish relationships among organisms. Over the last decades, there is an ever-increasing base of research using DNA and protein sequences i. e. molecular markers, to attain the relationships among taxa. Molecular approaches have the benefits of larger numbers of independent characters available and their homology status *a priori* is generally more objective. Additionally, in many cases the morphology of male reproductive organs have been the only characters used for species identification and hence making significantly fewer samples viable for identification (Arnqvist, 1998), and obstacle bypassed by using molecular markers. For a long time, there were a high front between the two fields of using molecular data and classical morphology (DeSalle et al., 2005; Hillis, 1987). Ultimately what has proven to be a greatly beneficial procedure, is to work together with both types of data molecular and morphology as an integrated entity together with geographical distribution and ethology (Bateman et al., 2006; De Ley et al., 2005; Hughes, 2011; Moritz et al., 2000;

Nadler and De León, 2011). Either way, any character used in a phylogeny should be *a priori* homologous trait (i.e.originated from a common ancestor). Within the homologous characters, we distinguish between ancestral or pleisomorphic states, or later derived apomorphic states. Some derived characters are unique to a sample and are called autoapomorphic. On the other hand, derived character states shared between species are considered synapomorphies, which are the most important characters to set relationships between species under a cladistic criterion. On the opposite end, when a character trait has occurred independently more than once is called homoplasy, i. e. convergence (Page and Holmes, 1998).

As previously mentioned, phylogenetic relationships based on homologous traits can be represented graphically in a hierarchal tree format. This phylogenetic tree is the interpretation of this data in the context of evolution where the terminal nodes represent extant organisms, the internal ones their reconstructed putative ancestors and finally, branches the connecting steps of the underlying evolutionary process (Page and Holmes, 1998). Charles Darwin proposed a similar idea of lineages in a notebook in 1837 (**Figure 1.6**), 22 years before his iconic work "On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life" (Darwin, 1859).

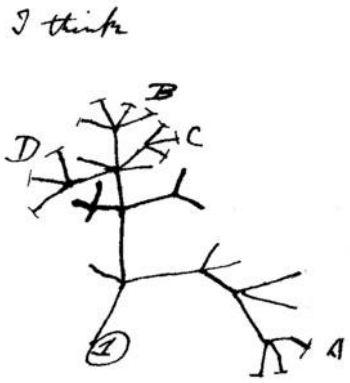


Figure 1.6 The first somewhat treelike structure illustrated by Darwin in one of his notebooks where he conceptualized the idea of common ancestry and the first ideas around the tree of life.

The tree of life

All organisms on this planet are related by common ancestry, and reconstructing these relationships or patterns is an important part of evolutionary biology. Generating evolutionary trees using mathematical algorithms are the most common approach to elucidate those evolutionary history and the relationships among organisms.
For molecular phylogenies except for parsimony, tree building is based on an underlying model of nucleotide or amino-acid substitution across the sequences analyzed. Those substitution models vary across both molecular markers, and across sites within single gene. For instance, the high A+T richness and the higher substitution rates found in third codon sites of mitochondrial genes, particularly in pancrustaceans, induce saturation of phylogenetic signals, i.e. loss of information due to multiple homoplastic changes in the same site (Arbogast et al., 2002) This fact together with difficulties of electing the correct evolutionary model, as well as differences in the substitution rate across sites, genes and species analyzed may altogether introduce possible errors in the phylogenetic construction and particularly regarding long branch attraction (Edwards, 2009). The first phylogenetic analyses were based on parsimony, which arguably is a derivative of the famous Occam´s razor, where the evolutionary tree or trees with the fewest possible evolutionary changes is selected as best. This approach holds two main advantages; their computation is fast and simple and has no

implicit evolutionary model; but cannot deal accurately with highly saturated data sets. Extremely fast algorithms estimating tree topology from genetic divergence called distance trees such as UPGMA (Unweighted Pair Group Method with Arithmetic mean) and NJ (Neighbor-Joining) were also abandoned since, despite taking into account homoplasy with complex evolutionary models, they build a single tree topology only without any further exploration of alternative topologies .

In recent times, statistical methods are widely preferred over parsimony and distance criteria to estimate accurately phylogenies. For instance, a likelihood function describes the joint probability of observing a set of given data under a specific model of evolution by an intense exploration of the tree topology space and an progressive optimization of the parameter values of the model (Holder and Lewis, 2003; Huelsenbeck and Crandall, 1997; Yang, 1994). This approach is called Maximum likelihood (ML) but any violation of that model implemented *a priori* may retrieve erroneous topology and branch lengths though ML is quite robust to model violation (Fukami-Kobayashi and Tateno, 1991). Besides, complex evolutionary models on large datasets have extremely long computational times tough in recent years several heuristic shortcuts and parallel algorithms have been developed (Guindon et al., 2010; Guindon et al., 2005; Tamura et al., 2011).

Bayesian theory is based on prior distribution for all parameters of the model, including tree topology which are then assessed and optimized with data formation to generate joint posterior probabilities, based on MC3 chains (Metropolis-coupled Markov chain Monte Carlo). After each cycle of generation, they are accepted or discarded based on Bayesian theorem (Alfaro et al., 2003; Holder and Lewis, 2003; Huelsenbeck and Ronquist, 2001). ML and Bayesian methods are by far the most utilized methods today in molecular phylogenetics and although demanding in computational power, today's rapid evolution in both algorithm and hardware advances have greatly improved speed.

Estimating divergence times

The concept of molecular clock can be dated back to the 1960 when Zuckerkandl and Pauling (1962) demonstrated that the accumulation of amino acid changes in the protein sequences of two sister species increases in a constant rate through time, the so called strict clock. These changes are ultimately the result of substitutions at the codon level (Page and Holmes, 1998). Discovering that the majority of substitutions are neutral and hence are not eliminated by natural selection together with the fact that the number of substitution across evolutionary lineages can be correlated with time, made it possible to construct ultrametric trees (Kimura, 1983; Page and Holmes, 1998).

The implementation of molecular clocks on phylogenetic studies had a deep impact on our understanding of evolutionary patterns since this incorporated a temporal vector in the evolutionary process. This allowed assessing how both geological and past climate events have influenced the diversification and evolution of species, as well as shed light on present phylogeographical and biogeographic patterns. However, the situation is rarely as simple as described since substitution rates are generally not constant over time. The rates of substitution over lineages and different genes and even within the same gene can be heterogeneous (Linder et al., 2005; Rutschmann, 2006; Rutschmann et al., 2007; Sanderson, 2003; Soubrier et al., 2012). Other factors thought to influence the accuracy of molecular clocks are saturation, which generally increases in the lineages including older splits, and presence of non-homogeneous nucleotide or amino acid frequencies across species. In addition, ancestral population processes such as changes in population size can affect the estimation of divergence both in ancient lineages and recent population events (Igawa et al., 2008; Linder et al., 2005; Rutschmann, 2006). This has led to a plethora of alternative

methods for molecular dating using relaxed clocks that have been developed on the Maximum likelihood and Bayesian framework, since the criterion for using strict clocks are violated in many if not most phylogenetic analyses (Drummond et al., 2006; Rutschmann, 2006; Yang, 2004). Clock relaxation has two main approaches; correlated relaxed clocks where the rate of the ancestor branch has an influence on descendants, while uncorrelated clocks ignore this fact (Drummond et al., 2006; Pybus, 2006). All phylogenetic analyses enforcing a molecular clock are using relative rates, i.e. the rate of a branch is *n* time faster than the root. To transform those relative rates to absolute geological ages, millions of years ago, at least one tree bifurcation (node) has to be calibrated with at least one dated fossil or a geological, biogeographic or paleoclimatic event. The fossil-record constitutes the main source of calibration points for dating the diverging lineages (Bibi, 2013; Forest, 2009). One of the main problems is that the fossil record is far from complete and researchers often are forced to use distant fossil species which increases the chance of a wrong assignation within the phylogeny. This can introduce a large factor of error in any given dataset. Geological data has also often been used as a calibration point as plate tectonics and the emergence of islands with endemic species (Forest, 2009; Renner, 2004) although they are also prone to large margins of error based on the interpretation of the data (Forest, 2009). A key factor to diminish the uncertainty and error associated to a single calibration point is using multiple separate and independent calibration points in the dataset to enhance the accuracy (Bibi, 2013; Renner, 2004). This approach is not without problems due to low sample size and a significant distance from the calibration points of the molecular clock can contribute to inaccuracies in the dating process (Linder et al., 2005).

"There are *n*+1 definitions of "species" in a room of *n* biologists."
John S. Wilkins

What is a species?

There are currently at least 27 different concepts of a "species" in the literature (Wilkins, 2011), making this is one of the most debated topics within biology (Morrison, 2011) both as a topic and a tool for frther investigation within several biological diciplines. (Barraclough and Herniou, 2003; Wilkins, 2011). Perhaps the most used definition of a species, is the so called biological species concept, i.e. a group of individuals that reproduce together and have fertile offspring. This however is somewhat inadequate for the vast majority of life, due to the fact that reproduction and life strategies have shown to be limitless within living organisms, and may be difficult and unworkable to authenticate in natural environment or in a controlled setting. The *concept* of a species is perhaps clearer than the way into which we define such a concept. Organisms are continuously evolving and although the species are quite well-defined it can be difficult to draw the line of what encompasses within a species and what is a subspecies or simply population plasticity, or a different specie altogether (Wiens, 2007). Most species concepts rely on certain criteria and to implement them requires following these definitions and placing species in encompassing groups and categories. In more recent time a different concept changing the idea from fixed concepts into where linear common ancestry is the focus and species are mere segments of groupings within this continuous evolutionary process, "General Lineage Concept" GLC (De Queiroz, 1999, 2007; Leavitt et al., 2015). Within this idea, the other species concepts may be used to divide this evolutive lineage and encounter species but defining the linear boundaries. Still, a working species definition encompassing all organisms and the plethora of reproduction strategies and codependent symbioses and regeneration has proven to be quite problematic. This holds especially for example asexual eukaryotes with no phenotypic identification possible. Birky et al. (2010) used DNA sequences and assigning them to species on simpler asexual organisms following some if not all the species concepts as criteria (Birky et al., 2010). Perhaps in particular lichens are a neat illustration of how complicated the species concepts can be. In the symbiotic relationship between fungi and cyanobacteria or chlorophyta that lichens constitute, the species are usually determined based on the morphology of fungi. Species have traditionally been determined by morphology, although in some instances the same fungi can use simultaneously a specific cyanobacteria or a specific green algae as their symbiont changes their morphology and hence potentially obscuring the true species boundaries (Leavitt et al., 2015; Nash III, 1996). In addition a recent study show the presence of a secondary fungus (both Ascomycetes and Basidiomycete) in the genus *Bryoria* Brodo and Hawksworth 1977 (Spribille et al., 2016), a discovery that illustrate only some of the difficulties in species delimitation across the living planet.

Subterranean fauna presents further taxonomic challenges. For instance,the high level of convergence driven by environmental factors (see above) on a global scale across diverse taxonomic groups, have caused misinterpretations and wrong classifications (Wiens et al., 2003). Furthermore, molecular data suggest that morphologically identical species were in fact extremely divergent, highlighting the difficulties of a accurate taxonomic diagnosis (Camacho et al., 2011; Finston et al., 2007; Fontaneto et al., 2011; King et al., 2012; Lefebure et al., 2006; Lefebure et al., 2007; Nygren, 2014; Proudlove and Wood, 2003; Trontelj et al., 2009; Trontelj and Fiser, 2009; ZakšEk et al., 2009). These species complexes are often referred to as cryptic speciation and although, it is somewhat a debated topic, has

been found as mentioned in many taxonomic groups (Bickford et al., 2007; Birky et al., 2010; De Queiroz, 2007; Hanage et al., 2005; Hey, 2001; Hey, 2006; Hey et al., 2003; Jörger et al., 2012; Jörger and Schrodl, 2013; Knowles and Maddison, 2002; Mallet, 1995; Mayden, 1997; Meleg et al., 2013; Reydon, 2005; Rousseau et al., 2001; Satler et al., 2013; Sato et al., 2005; Schultz and Wolf, 2009; Wilkins, 2011; Wu, 2001). The lines of what constitute a cryptic species may be unclear. Molecular data hinting cryptic species were corroborated by further morphological investigation (Saez and Lozano, 2005). On the other hand, in some cases formal description of species were made based on molecular data alone due to the absence of reliable diagnostic morphological characters (Jörger et al., 2012; Jörger and Schrodl, 2013). Cryptic speciation is important since it can cloud real levels of species diversity within a habitat (Nygren, 2014), which is a crucial issue in plans for conservation when of endangered and vulnerable species (Nygren, 2014).

Barcoding and species delimitation

Arnot et al. 1993 and Haliassos et al. 2001 perhaps coined the term barcoding meaning a specific universal DNA sequence used as an identifier for a diverse array of biological samples. Hebert et al. in 2003 launched the idea of a large library of DNA sequences, barcoding database, to rapidly identifying already described metazoan species. The elected gene was the 5' end of the Cytocrome Oxidase Subunit 1 or *cox1* based on two main features: 1) the ability to design "universal" primers to amplify this mitochondrial region in most eukaryotic species, 2) its moderate evolutionary rate enables both species identification even for close related species and the phylogenetic placement of distant species. It has also become a great tool to discover cryptic speciation, to assign small fragment of tissue to species in forensic studies, or as an additive to taxonomy where morphological characters are non-informative, as in nematodes, flatworms and nemerteans (Schander and Willassen, 2005). Despite a large success and an unprecedented amount of data that has come out of this concept caution was raised and perhaps rightly so for several reasons. Although *cox1* has been found to be a functional barcoding gene for a broad spectrum of metazoans (Frézal and Leblois, 2008; Waugh, 2007), it is not deemed useful for plant species identification (Rubinoff et al., 2006). Besides, this gene could not be amplified in some specific animal species or taxa, potentially producing gaps in datasets or variation in the alignment sequences due to using different primers. Perhaps there is no such thing as the perfect barcoding gene for universal species identification (Meyer and Paulay, 2005; Sundberg et al., 2010; Teletchea, 2010; Timmermans et al., 2010). Using the distance between sequences to determine a species is at best problematic due to the heterogeneous substitution rates both among and within genes (Teletchea, 2010).

"All models are wrong, some of them are useful."
Georg Box

Species delimitation based on DNA sequences

There is a plethora of different species delimitation methods created very early on before the blossom of DNA barcoding approaches (Cracraft, 1983; Davis and Nixon, 1992; Sites Jr and Marshall, 2003). However, these were unable to concretely delineate species and rather aggregate populations. Herbert´s DNA barcoding approach cannot be considered a delimitation method intrinsically since there is no *a priori* criterion other than morphology to delineate species. In fact, intra- and inter-specific genetic differences are estimated after morphological groups (species) are defined, and only the existence of a barcoding gap between intra- and inter-specific nucleotide divergence confirms or rejects the morphological hypothesis of species. For instance, the somewhat controversial "10x rule" of the DNA barcoding gap, requiring 10 times larger inter specific divergence than intra specific polymorphism to define a species (Hebert et al., 2004) has been proved to be unable to accurately delimit recently divergent species (Hickerson et al., 2006) In other cases, it is not certain that there is a gap between all inter-specific and intra-specific distances as overlaps have been detected (Meier et al., 2006; Puillandre et al., 2012). Ultimately every model using molecular markers as the input data is looking at nucleotide substitutions (i. e. genetic distances) between sequences in one way or another. However, the calculation and approaches to interpret the data can vary significantly. In the last decade, an array of methods based on DNA sequences has been developed in an attempt to accurately delimit distinct taxonomic operational units (OTUs) using single locus sequences (Choi 2016; Leavitt et al. 2015; Renner 2016). Here, we will focus on a few of the most recent and popular approaches that attempts to accurately delimit species based on DNA sequences and implementing algorithm using biological properties such as population dynamics and diversification patterns, and most importantly without the need for a prior definition of species groups.

The General Mixed Yule Coalescent model (GMYC) (Pons et al., 2006) is a maximum likelihood algorithm comparing waiting times of the branching pattern of an ultrametric tree built from a single loci sequence as input data to estimate the species boundaries. The underlying theory is that the branching rate is different between species and populations. For species, the bifurcation rate is determined by the diversification or speciation and extinctions (Nee et al., 1994), while for populations the branching rate is determined by coalescence processes (divergence between gene lineages over time)(Degnan and Salter, 2005; Hudson, 1991). These are two different models that can be distinguished and merged into the GMYC model where the transition between one of the processes to the other constitutes the species boundary, i.e. transition between population and species level (Pons et al., 2006). Thus, GMYC optimize the threshold between these merging concepts in an ultrametric tree by this maximum likelihood function.

$$b^* = \lambda_{k+1}\left(n_{i,k+1}\right)^{p_{k+1}} + \sum_{j=1,k} \left(\lambda_j \left(n_{i,j}\left(n_{i,j}-1\right)\right)^{pj}\right)$$

b^* is defined as the probability of an event of any type happening at the end of waiting interval. λ_{k+1} is the speciation rate, $k+1$ is the index for diversification process, $n_{i,j}$ is the number of lineages in waiting interval i belonging to process $j$and λ_j represent the branching rates for coalescent processes i. e population level, pj and p_{k+1}represent scaling parameters for optimization of the model fitting process for coalescent and diversification models, respectively. The algorithms test lower and upper values departing from constant population

size (pj=1) and constant speciation (p_{k+1}=1) to take into account for more complex models. Here the threshold between speciation and coalescence is arranged as a single parameter and hence assuming that the threshold is equal throughout the phylogeny under scrutiny. The algorithm compares the null hypothesis of a single species, i.e. all branches are compatible with a coalescent model with two parameters (λ1 and p1), versus a GMYC model with five parameters (λ1, p1, λ2, p2 plus the threshold time T) using a Likelihood Ratio Test (LRT). Moreover, it tests for the significance of LRT using three degrees of freedom that is the difference between the parameters to both models. In later papers it was suggested that this threshold should be changed to multiple thresholds in one dataset since these are not necessarily static throughout the dataset (i.e. through evolutionary time) (Monaghan et al., 2009; Powell, 2012). Later GMYC studies using simulated data show that single threshold yielded a more accurate rendition of the data than with multiple thresholds (Fujisawa and Barraclough, 2013; Fujita et al., 2012). In the study by Fujisawa and Barraclough in 2013, the parameter threshold time T was altered from being a parameter to be treated as a constraint of search space. In the same study the alteration of the algorithm permitted to estimate the support of each node to assess whether the node contains a population or a speciation event. Here a support value of 1 signifies that all coalescent models investigated indicate a species event. A support level lower than this or at zero, indicate that some or all the coalescent models do not support the threshold and thus it is less likely, or not likely at all, to have occurred a speciation event (Fujisawa and Barraclough, 2013). The main advantage of the GMYC algorithm is that *a priori* taxonomic identification of specimens under study is unnecessary. A larger number of specimens of a species increase the accuracy of the model while increasing the number of species and holding the number of samples constant might reduce the accuracy of the model (Fujisawa and Barraclough, 2013). An important limitation in the GMYC model is the assumption of monophyly. Species that has not yet had time to diverge sufficient to become monophyletic, and hence recent speciation or rapid adaptive radiation events might go undetected (Knowles and Carstens, 2007). Generally, in a group of organisms where the population sizes are low and the divergence rate is high then species delineation using GMYC algorithm is both accurate and conservative (Fujisawa and Barraclough, 2013). It is important to state that the GMYC model requires an ultrametric gene tree as the input data and any model misspecification during tree reconstruction may affect the species delimitation. In particular when it comes to the various methods of rate smoothing that can influence branchlengt and thus ultimately effect the GMYC estimates, making it vulnerable for erroneous estimations (Drummond and Suchard, 2010; Tang et al., 2014).

Poisson Tree Processes model (PTP) is a more recent ML approach inspired by the phylogenetic species concept that also employ rooted trees to delineate species from DNA sequences though trees are not ultrametric like in GMYC (Zhang et al., 2013). PTP, as well as GMYC, infers all parameter from the data, and hence not requiring an assumption of the taxonomic status of specimens or setting an *a priori* sequence similarity cutoff. This is achieved by estimating speciation rate directly from the number of substitutions, i.e. from branch lengths, between two branch events. This model also assumes that the mean number of substitutions within species is significantly higher than the mean number of mutations within a population (Zhang et al., 2013). Furthermore, it considers other assumptions: mutations are independent, and each mutation has a small probability to drive a speciation event. Thus, since the number of mutations in a population is generally large, the process follows a Poisson distribution. Thus, the model fits substitutions in two classes of Poisson processes. One of them designates speciation as the average number of mutations until next speciation event with an exponential distribution. The second one describes the coalescent branching events within species with another independent exponential distribution. Finally, PTP algorithms

conduct a search for transition points to find the optimal threshold that maximize the likelihood function where the branching pattern changes from inter-specific to intra-specific branching. By not using an ultrametric tree as the basis eliminates one possible factor of error introduced by and inaccurate tree-estimate or selecting suboptimal population models as well as greatly reduces computational time needed. Since computational powers needed are relatively low it is well suited to be implemented on large datasets. Studies have shown that PTP is possibly vulnerable to a low sampling numbers or inequality of sampling numbers, incomplete lineage sorting, depending on the organism in focus (Tang et al., 2014; Zhang et al., 2013).

Automatic Barcode Gap Discovery (ABGD) (Puillandre et al., 2012) identifies the so called barcode gap when the intra specific divergence is smaller than inter specific divergence. The algorithm finds the first occurrence of a gap in the dataset to partition the data into subgroups (i.e. MOTUS or putative species) based on a genetic distance threshold (i.e. gap). This can be problematic since it has been shown that the divergence between intra specific and inter specific samples may overlap and hence a clear visual "gap" might even not exist (Collins and Cruickshank, 2013; Meyer and Paulay, 2005; Zhang et al., 2013), even though the ABGD software should accommodate for this (Puillandre et al., 2012). In an initial step, the ABGD clusters together sequences into species from minimum and maximum starting intra-specific values which are used to maximize the barcoding gap. These clusters are then further attempted to be split up by optimizing barcode gap and this process continues recursively until no further splitting can form based in these criteria. The method compares the results of the initial partition and the recursive partition. The benefit this method is that virtually no prior information about the data is needed and the threshold for the barcode gap is inferred by on the basis of the current dataset although both the width of this threshold and the limit to intraspecific diversity is set arbitrarily *a priori* and might not be representative for the given dataset at hand (Puillandre et al., 2012). This is a computationally efficient method, perhaps best implemented in preliminary analysis. The authors of this method state that it is meant as a supplementary line of investigation into species delimitation. They also stressed that sampling distribution and a potentially incomplete dataset may affect the accuracy of the method. For instance, few samples per location might not detect real divergence in a population and hence skew the "barcode gap" detection (Meier et al., 2006).

TSC is a network building program based on statistical parsimony to represent genealogical relationships at the population level but also to study phylogenies *a posteriori (Clement et al., 2000)* This method is based on that, at the population level, the time scale is generally much smaller than at the species level (Posada and Crandall, 2001). It assumes that an ancestral haplotype is not deemed extinct in the population and two different haplotypes might be descendants from an ancestral but still existing haplotype. On the contrary, the traditional methods such as Neighbor Joining and ML assume that the common ancestor of a linage is extinct, contrasting the very real possibility that this is not the case (Posada and Crandall, 2001). The development of network methods like TCS accommodates for these population events and are not only bifurcating haplotype trees (Clement et al., 2000). These network methods might therefore be more accurate on the population level than traditional phylogenetic methods. In particular, TCS is a network model based on genetic distances estimated under statistical parsimony criterion (i.e. the hypothesis with the least amount of steps in the cladistic tree is the most plausible one) (Templeton et al., 1992). It compares the pair-wise distances and calculates the maximum number of differences between the halotypes on the basis of the substitutions encountered in the dataset with a 95 % confidence level, or the parsimony limit (Posada and Crandall, 2001). Then all haplotypes are connected together

in a single network based on the substituting differences among them, until and slightly further than the 95% parsimony limit set as default (Clement et al., 2000). This can be used to at least tentatively to delineate independent networks (i.e. gene genealogies within species) without any other prior information. Computationally, this is a fast method of evaluating data but mostly used to infer intra-specific difference at the population level, and hence might be prone to errors as sampling per population is low and the scope of the study is broader.

The delimitation models presented above are all in principal single loci approaches. The first two methods, GMYC and PTP, are tree based approaches while ABGD is a distance based method. TCS is based on the networks formed by the genotypic differences on a population level.

Gene evolution is not necessarily equivalent to species evolution in given taxon. Some studies have intended to remedy this by concatenating dataset of various genetic markers and delimiting species based on mutilocus data (John et al., 2013). Despite this the accuracy of this method is reported to outperform most of the single loci methods when compared on simulated and real datasets, we will not since is out of the scope of the present study (Camargo et al., 2012; Toussaint et al., 2015; Yang and Rannala, 2010).

Here we have summarized a few of the most common methods and approaches to species delimitation each with their strengths and advantages but also with their own limitations and drawbacks. Most of the reasons for this are mentioned above however, in general, all models require a simplification to real life data or implement assumptions that do not always reflect real life (Carstens et al., 2013). One way to overcome this is to utilize a broad range of these methods to the same dataset and only conclude about the congruent taxon groups across the various models and approaches as described by Carstens et al. 2013, (Carstens et al., 2013; Fontaneto et al., 2015; Fujita et al., 2012; Leavitt et al., 2015; Sukumaran and Knowles, 2017) in order to accurately delimit species boundaries.

In this thesis we take advantage of a broad faunistic sampling of *Pseudoniphargus* performed across its entire distributional range to assess the biodiversity and pattern of species diversification of the genus, emphasizing on the occurrence of cryptic species. We will compare our results with the condition found in other stygobiontic crustaceans displaying similar life habits and general distribution to *Pseudoniphargus*, e.g. *Metacrangonyx* Chevreux, 1909 (Bauzà-Ribot et al. 2011; 2012), *Niphargus* Schiödte, 1847 (Foulquier et al. 2008; Lefébure et al. 2006; 2007), *Typhlatya* Creaser 1936, (Hunter et al. 2008), or *Tethysbaena* (Cánovas et al. 2016), to eventually derive generalizations regarding criteria of species delimitation, occurrence of cryptic speciation, and size of geographic range in stygobiont crustaceans.

Marker selection and mitogenomics

The markers and method implemented in this thesis were selected for several and equally important reasons. Firstly, Cytocrome Oxcidase subunit1 (*cox1*) the most widely used marker in Metazoa and also in amphipods, and it is the selected marker to be used as DNA barcode sequence (Hebert et al., 2004). This lends a basis for having compatible comparisons within the amphipods and the assumed closer relatives to the study genus *Pseudoniphargus*. Several other markers both of nuclear and mitochondrial origin were also amplifiedbut they were discarded as amplification was very limited compared to *cox1*. In fact, we only obtained positive amplification in several *Pseudoniphargus* populations with a newly designed primer

pair based on *cox1* sequences obtained from other sites. In other cases more conservative markers such as Histone 3, *H3* showed no discernible phylogenetic signal. Moreover, we opted for having a broad and large dataset rather than eliminating taxa for not having a suitable second marker since one of the main goals of this thesis is the assessment of the biodiversity of *Pseudoniphargus* across its geographical range. The models or approaches for species delimitation in this thesis were also carefully selected based on cost/benefit assessment of the various model presented above and their advantages and possible problems as well as recognizing the need for a broad approach using multiple models and implementing as much as possible information to increase the accuracy of the species delimitation.

The possibility to produce large dataset with entire mitochondrial genomes have increased dramatically the last decade due to recent advances in molecular sequencing and next generation sequencing (NGS) (Duchêne et al., 2011; Shen et al., 2015). The latter has rapidly improved the share amount of recoverable data for challenging samples (Duchêne et al., 2011) as the NGS prices decline and increase the amount of reads. Altogether, it makes mitogenome sequences a useful molecular marker to study any metazoan groups and for most researchers. It has been widely used both within vertebrates and invertebrates to investigate phylogeny and phylogeography (Botero-Castro et al., 2013; Li et al., 2017; Shen et al., 2013; Shen et al., 2015). Comparing to traditional mitochondrial markers, the longer sequences has contributed to higher topological resolution, and stronger node support as well as an increased accuracy in branch lengths and hence determining divergence times (Wielstra and Arntzen, 2011). Mitogenomes also enables a comprehensive estimation of evolution rates both as a whole but also individually across multiple markers in the mitochondrial genome, as well as giving insight into complex structures like gene arrangements and secondary structures of ribosomal RNA genes and transfer RNA genes (Bauza-Ribot et al., 2009; Botello and Alvarez, 2010; Fahrein et al., 2007; Kilpert and Podsiadlowski, 2010a; Podsiadlowski, 2010b). This has proven to be a possible alternative to implementing nuclear genes in a study since they can be highly difficult to amplify in unexplored non-model organisms. Using mitochondrial genomes is not without its problems as compared to nuclear markers the levels of saturations are elevated and may cause problems in a deeper phylogeny (Hassanin, 2006). However for the scope of this thesis mitochondrial genomes is well within the estimated range for this type of phylogenetic marker, and mitogenomes have been assessed to work well up until intra-orden levels (Talavera and Vila, 2011). A typical metazoan mitochondrial genome (mitogenome) contains 13 protein-coding genes (PCGs), two ribosomal RNA genes (rRNA), 22 transfer RNA genes (tRNA), and a non-coding control region (CR) also called the D-loop that includes the origin of replication (Boore 1999). However, gene order differs remarkably across higher taxonomic ranks (family and above) due to the occurrence of gene rearrangements such as transpositions and reversals (Arndt and Smith 1998; Dowton and Austin 1999; Bensch and Harlid 2000; Hickerson and Cunningham 2000; Bauza-Ribot et al. 2009; Irisarri et al. 2014). Rearrangements of this kind, especially those affecting tRNA genes, are frequently detected when comparing animal mitogenomes belonging to different taxonomic families or even between genera within the same family (Arndt and Smith 1998; Kurabayashi et al. 2008; Dowton et al. 2009; Kurabayashi et al. 2010). But rearrangements within the same genus, especially those involving PCGs, are rather rare (Rawlings et al. 2001; Matsumoto et al. 2009). Mitochondrial gene rearrangements can occur as a product of one of four major events: reversals, transpositions, reverse transpositions, or tandem duplications with subsequent random loss (TDRL) (Moritz and Brown 1987; Chaudhuri et al. 2006; San Mauro et al. 2006). Reversals consist of one or several genes that switch from one DNA strand to the other, while transpositions involve a shift to another location without changing the sense of the strand. Both processes are deduced to have operated in reverse transpositions.

TDRLs are more complex processes where a segment including more than one gene undergoes duplication and subsequently, suffers the deletion of particular gene copies at random, producing novel gene arrangements that can differ considerably from the ancestral gene order. The latter is an interesting mechanism from a phylogenetic point of view, since it is the only shift that can be considered with confidence to be irreversible (Chaudhuri et al. 2006). The rearrangement events that took place in a particular mitogenome with respect to an ancestral gene order can be heuristically determined using strong interval trees (Berard et al. 2007; Bernt et al. 2007; Perseke et al. 2008). The ultimate causes of rearrangement events in the mitochondrial genome remains unclear, although nuclear substitution rate has been shown to correlate with rearrangement frequency (Shao et al. 2003). Furthermore, the origin of replication –that is located in the control region– also has been postulated to be a hot-spot for gene rearrangements (Macey et al. 1997; Kraytsberg et al. 2004; Arunkumar and Nagaraju 2006; San Mauro et al. 2006). Besides, some life history and ecology traits such as founder effect and parasitism have been postulated to play a role in the gene rearrangement rate within animal lineages (Tsaousis et al. 2005). Despite gene content of mitogenomes is mostly conserved across metazoans, nucleotide and aminoacid composition, as well as codon usage, also vary greatly across major taxonomic groups. Thus, insects and other arthropods display very AT-rich mitogenomes with PCGs showing heavily biased codon usages (Sheffield et al. 2008; Yang 2008), while vertebrates show lower AT-richness (Sano et al. 2005).
The amphipod crustacean family Pseudoniphargidae Karaman 1993 contains only two genera, The mitogenome of *P. daviui* (Jaume 1991), of around 15 kb, was sequenced and used as outgroup in a broad scope biogeographic study (Bauzà-Ribot et al. 2012) although its gene order, composition and secondaries structures were not addressed. Mitochondrial genome sequences have been broadly used as a phylogenetic marker since: (1) Duplication and recombination events are extremely rare or easily detected in the maternally inherited mitochondria, and hence establishing gene orthology is here clear-cut as opposed to in most nuclear genes; (2) There are many conserved regions on which to design "universal" primers to amplify mitochondrial genes, even across distant taxa; and (3) Advances in molecular methods such as long-PCR amplification and Next Generation Sequencing allow for the fast acquisition of complete mitogenomes at low cost. Altogether these features make mitochondrial sequences first-class molecular markers to assess phylogenetic relationships and to establish a time framework for diversification, the so-called phylo-mito-genomics (Jurado et al. in press; Bauzà-Ribot et al. 2012; Shen et al. 2013; Talavera and Vila 2011; Pons et al. 2010).

However, reconstructing tree topologies and estimating branch lengths and node ages are compromised by many factors. Namely: Fitness of nucleotide substitution model on data (Posada and Krandall 1998), type of data partitioning implemented across both genes and sites (Brown and Lemon 2007), level of saturation of third codon sites (Hassanin 2006), occurrence of compositional bias (Lartillot and Philippe 2004), constancy of rates across species (Phillips 2009), and the particular method implemented for molecular clock relaxation (Baele et al. 2012), just to mention the more relevant.

Another crucial point is related to the transformation of rates (i.e. number of substitutions) to absolute (geological) ages in million years by constraining a particular node to the age of a dated fossil or geologic/palaeogeographic event. Problems arise here not only because the fossil record is usually incomplete (Donoghue and Yang 2016), but also because palaeogeographic events driving the separation of populations generally do not happen suddenly but progressively over long periods of time. Moreover, fossils are often difficult to associate to a specific lineage (node) of a particular phylogeny since they frequently do not display the diagnostic features of extant lineages (Paul 1992). To overcome all these drawbacks and date nodes with confidence, several studies have proposed to combine several

fossils or geologic/palaeogeographic events in the analyses and assess their congruence (Forest 2009; Near et al. 2005). This approach reduces at least the errors associated to rely on single calibration points only.

Finally, there are multiples example in which mitogenome sequences have resolved phylogenetic questions that were enigmatic using other markers such as resolving the decapod crustacean phylogeny (Shen et al., 2013), resolving the evolutionary history of leafy-nosed bats (Botero-Castro et al., 2013) as well as in crested newts (Wielstra and Arntzen, 2011) and phylogenetic studies of the Delphinidae family (Vilstrup et al., 2011).

Chapter 2. Objectives

This thesis aims at shedding light on the diversification of the stygobiont genus *Pseudoniphargus* and its potential drivers. Specifically:

1) To assess the biodiversity of the genus by extensive sampling performed across its entire distribution area, and assign the material collected to anyone of the known taxonomic species or eventually describe the new ones.

2) To analyze the genetic diversity of populations and species to recover their phylogeographic pattern and investigate the probable causes of their current distribution.

3) To delimit *Pseudoniphargus* species using advanced computing methods performed on DNA sequences, and assess their congruence with the species recognized by classical taxonomy. Eventually it will enable to detect the occurrence of hidden biodiversity (i.e. cryptic species).

Select representatives of the most divergent lineages and major geographic areas where the genus is present to obtain their complete mitochondrial genome using next generation sequencing to:

4) Describe the evolution of several features of the mitogenome such as gene order, nucleotide and amino-acid composition, and the secondary structure of both RNA genes and the control region.

5) Build a strongly supported and resolved phylogenetic tree of *Pseudoniphargus* based on 13 mitochondrial protein-coding genes to reconstruct the systematic relationships of species.

6) Set a temporal framework for the phylogeny based on strong fossil and/or biogeographic evidences to estimate the age of the most common ancestor of *Pseudoniphargus* and to test whether the amphi-Atlantic distribution of the genus is compatible with the estimated age for the opening of this oceanic basin.

Chapter 3. Materials and Methods

Specimen sampling and morphological methods

Specimen sampling

We sampled *Pseudoniphargus* at 62 different sites (wells, caves and river interstitial medium), that appear summarized in **Table 3.1**. and **Figure 3.1**. Samples were collected and preserved in 95% ethanol just after collection and were kept as cold as possible until reaching the laboratory.

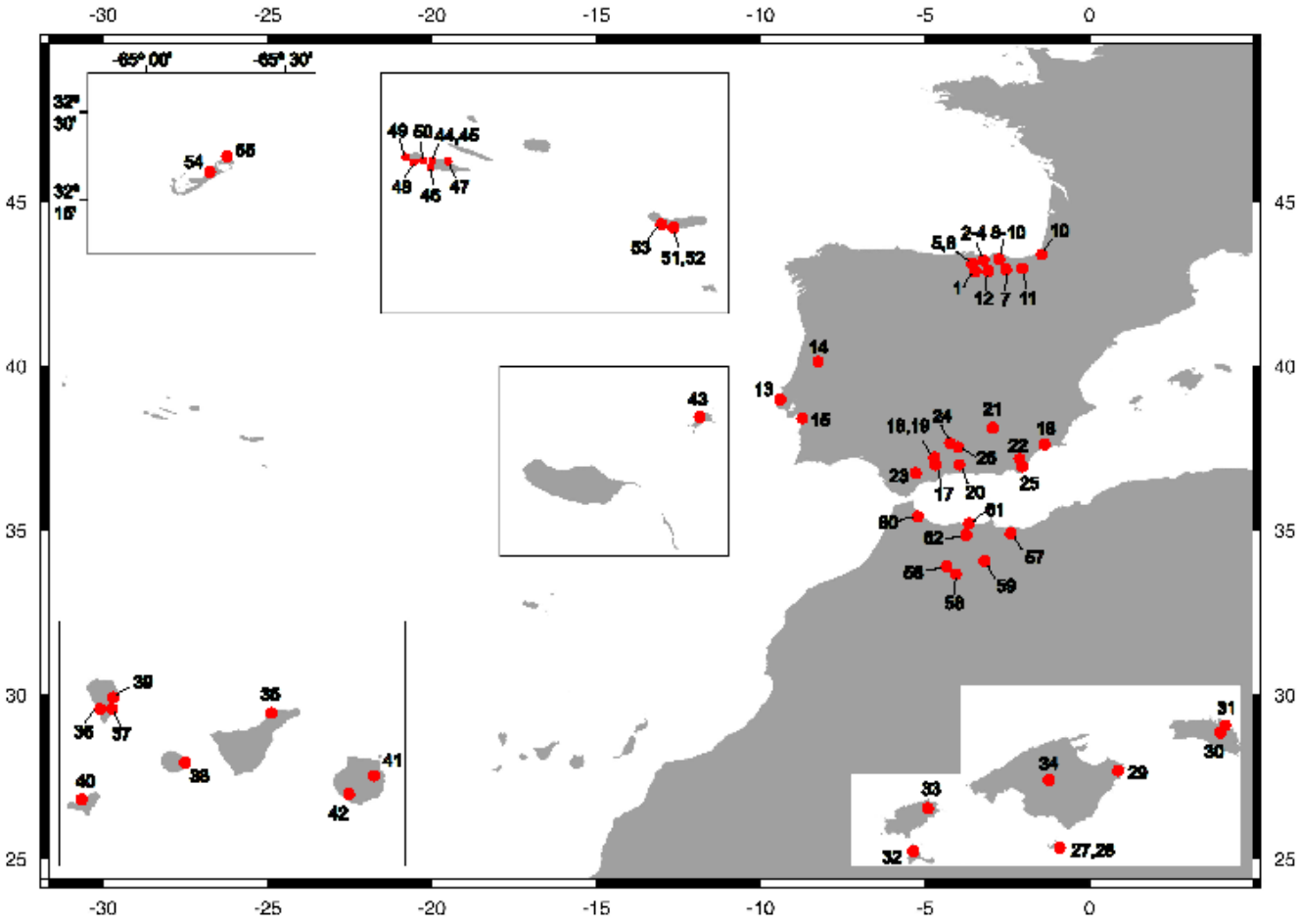


Figure 3.1. Map showing approximate geographic location of the 62 different sampling sites of *Pseudoniphargus*. See **Table 3.1 for** precise localaties for each taxon.

Table 3.1 Showing all species collected and their subsequent localities as illustrated in **Figure 3.1**., together with species codes used in the analysis. Number of individuals per site, number of individuals per species as well as number of haplotypes per species is also shown. G- represent the group or clade.

	Species	Locality	Latitude	Longitude	Code	# in Fig1.1	# ind_site	# ind species	# hapl species
Northern Spain									
G1	*P. burgensis* Notenboom, 1986	Burgos: La Torcona, Ojo Guareña (Hornillayuso; Merindad de Sotoscueva) [TYPE LOCALITY]	30T 450555	4762369	BUBUOG01-02	1	2	2	2
G2	*P. elongatus* Stock, 1980	Santander: Cueva del Valle (Rasines)	30T 465984	4794036	ELSAVR01-02	2	2	10	6
		Santander: Cueva Fresca (Asón)	30T 451401	4785631	ELSACF01	3	1		
		Santander: Cueva del Gándara (Asón)	(ED50) 30T 452465	4782348	ELSACG01	4	1		
		Burgos: Cueva de Imunía (Portillo de la Sía)	(ED50) 30T 451580	4777420	ELBUCI01-05	5	5		
		Burgos: Torca de Lunada (Puerto de Lunada)	(ED50) 30T 447282	4780101	SEBUTL01	6	1		
G1	*P. gorbeanus* Notenboom, 1986	Alava: Artzegi'ko Koba (Cigoitia)	N 43º01.121'	W 2º45.242'	GBAAKC01-21, GOALKC01	7	22	22	5
G1	*P. guernicae* Notenboom, 1986	Vizcaya: Cueva de San Pedro (Axpe; Busturia)	N 43º22.925'	W 2º42.038'	GUVISP01; 03-08; 10; 12-13	8	10	14	8
		Vizcaya: Goiko Etxe (Busturia)	N 43º21'46.25"	W 2º42'08"	GUVIGE01-04	9	4		
G1	*P.* sp1-Basque	Vizcaya: Cueva de San Pedro (Axpe; Busturia)	N 43º22.925'	W 2º42.038'	GUVISP02; 09; 11	8	3	4	2
		Vizcaya: Goiko Etxe (Busturia)	N 43º21'46.25"	W 2º42'08"	GUVIGE05	9	1		
G1	*P. incantatus* Notenboom, 1986	Navarra: Cueva de las Brujas (Zugarramurdi) [TYPE LOCALITY]	N 43º16.141'	W 1º32.831'	INNAZU01-02	10	2	2	1
G1	*P. unisexualis* Stock, 1980	Guipúzcoa: Túnel de San Adrián (Zegama) [TYPE LOCALITY]	N 42º56.134'	W 2º18.944'	UNGUZE01-05	11	5	5	2
G1	*P. jereanus* Notenboom, 1986	Alava: Cueva SI-44 (río Kabata; Sierra Salvada)	(ED50) 30T 493144	4762031	PSALKS01-02, 05-06, 08-09	12	6	6	3
Portugal									
G4	*P. mateusorum* Stock, 1980	Setúbal: "Lapa dos Morcegos" (Costa da Arrábida) [TYPE LOCALITY]	N/A	N/A	MATEPO01	13	1	1	1
G3	*P.* sp1-Portugal	Gruta de Legaçao (Sicó)	N/A	N/A	SPPOGL01	14	1	1	1
G4	*P.* sp2-Portugal	Lisboa: Gruta de Assafora (Sintra)	N 38º54'24.80"	W 9º25'17.31"	SPPOGA01-04	15	4	4	1
Southern Spain									
G4	*P.* sp1-Murcia	Murcia: Cueva del Agua (Isla Plana; Mazarrón)	30S 657222	4160295	SPMUPC01-03	16	3	3	3
G3	*P. "gevi"*	Málaga: Complejo El Romeral (Antequera)	(ED50) 30S 366968	4100145	PSMARO01-07	17	7	7	3

G3	*P.* sp1-Andalusia	Málaga: Cueva del Yeso III (Antequera)	(ED50) 30S 345817	4095033	PSMAYE01-07	18	7	11	7
		Málaga: Complejo La Zarza	(ED50) 30S 349090	4094700	SPMACZ01; 03-05	19	4		
G3	*P. grandis* Notenboom, 1987	Málaga: Fuente de la Fájara (Canillas de Aceituno; Sierra de Tejeda y Almijara)	30S 402267	4082380	GRMAFF01-04	20	4	4	2
G3	*P. latipes* Notenboom, 1987	Jaén: Interstitial of river down Tranco de Beas reservoir [TYPE LOCALITY]	N 38º11.055	W 2º53.897	LAJATB01-08	21	8	8	3
G3	*P. sorbasiensis* Notenboom, 1987	Almería: Cueva del Agua (Sorbas)	(ED50) 30S 585118	4107318	SOALAS02-12	22	11	11	2
G3	*P. stocki* Notenboom, 1987	Cádiz: "El Pozo Blanco" (Villaluenga del Rosario; Sierra de Grazalema)	30S 286334	4063509	STCAPB01-03	23	3	3	1
G3	*P.* "*morenoi*"	Córdoba: Cueva del Yeso (Baena)	(ED50) 30S 380474	4170957	SPCOCY01-06	24	6	6	3
G4	*P.* sp2-Andalusia	Almería: Noria at Pozo de los Frailes (Níjar; Cabo de Gata)	30S 579451	4071809	PSALCG01-05	25	5	5	2
G3	*P.* sp3-Andalusia	Córdoba: well at property of family Corpas (paraje La Almozara; Priego)	30S 393069	4141670	JACOPR01-05	26	5	5	5
Balearic Islands									
G3	*P. daviui* Jaume, 1991	Cabrera: Font d'Enciola	31S 493662	4331321	DACAFE01-12	27	12	19	1
		Cabrera: Font de s'hort de can Feliu [TYPE LOCALITY]	31S 495001	4332427	DACAFH01-07	28	7		
G4	*P. mercadali* Pretus, 1988	Mallorca: Cova de na Barxa (Capdepera)	N 39º41.033'	E 3º27.445'	MEMACA01-09; 11-13	29	12	18	8
		Menorca: Well at Binicreixent (Es Mercadal)	N 40º01.121'	E 4º 05.929'	MEMNBI01-04; 06	30	5		
		Menorca: Well in dry stream down Es Frare (sa Albufereta Peninsula, Fornells)	N 40º03.481'	E 4º09.848'	MEMNAF01	31	1		
G3	*P. pedrerae* Pretus, 1990	Formentera: Coves de sa Pedrera (St. Ferrán) [TYPE LOCALITY]	(ED50)31S 364917	4285874	PDFOSF01-12	32	12	12	5
G3	*P. pityusensis* Pretus, 1990	Formentera: Coves de sa Pedrera (St. Ferrán) [TYPE LOCALITY]	(ED50) 31S 364917	4285874	PIFOSF01-02	32	2	5	2
		Ibiza: Well at Sant Joan de Llabritja	(ED50) 31S 370263	4326249	PIEVSL01-03	33	3		
G3	*P. triasi* Jaume, 1991	Cabrera: Font de s'hort de Can Feliu [TYPE LOCALITY]	31S 495001	4332427	TRCACF01, triasiCA (mitogenome)	28	2	2	1
G3	*P.* sp1-Balearics	Mallorca: well at Mr. Toni Martínez house, between Cas Canar and Ruberts	31S 493792	4387159	PSMAPT01-08	34	8	8	2
Canaries									
G4	*P. associatus* Sánchez, 1991	Tenerife: Manantiales de Jóver	28R 365663	3158455	FOTEMJ01-07	35	7	7	3
G4	*P. cupicola* Stock, 1988	La Palma: Charco Verde (S of Puerto Naos)	28R 216486	3164089	CULPCV01-06; 08;10-14;16-18; 20-22	36	18	26	9

		La Palma: Charco de Los Chochos (El Porís) [TYPE LOCALITY]	28R 226631	3159004	CULPCC01-08	37	8		
G4	*P. gomerae* Stock, 1988	La Gomera: well at Playa del Avalo (N of San Sebastián), almost on shoreline, behind cliff [TYPE LOCALITY]	28R 293129	3111276	GMLGPA01-25	38	25	25	5
G4	*P. multidens* Stock, 1988	La Palma: well within military zone beside UNELCO (electricity plant), just S of Sta. Cruz	28R 229408	3174090	MULPSC01-22	39	22	22	7
G4	*P. salinus* Stock, 1988	El Hierro: Pozo de Las Calcosas	28R 20976	308294	SAEHPC01-07	40	7	7	2
G4	*P.* sp1-Canaries	Gran Canaria: Mina de los Llanetes (Valsequillo)	28R 45272	309590	PEGCML01-04	41	4	4	2
G3	*P.* sp2-Canaries	Gran Canaria: Mina Los Roques (Barranco de Arguineguín; El Sao)	N/A	N/A	PDGCRO01-02	42	2	2	1
Madeira									
G4	*P. portosancti* Stock & Abreu, 1993	Porto Santo Island: Fonte do Tanque [TYPE LOCALITY]			POMAPO01-11	43	11	11	7
Azores									
G4	*P. brevipeduculatus* Stock, 1980	Pico Island: Areia Larga	26S 366033	4264583	BRPIAL01-02	44	2	60	42
		Pico Island: Rua Joao Lima, Madalena, Areia Larga	N/A	N/A	BRPILM01; 03-06	45	5		
		Pico Island: Calhau	26S 3658	42609	BRPICA01-03; 05-07; 10-14	46	11		
		Pico Island: house of Mr. Melo; Furna de Sao Antonio, Sao Roque	N/A	N/A	BRPISR01-17	47	17		
		Faial Island: Feteira	26S 35399	426526	BRFAFE01-04	48	4		
		Faial Island: Varadouro	26S 3457	42702 ??	BRFAVA01-02	49	2		
		Faial Island: Horta-Conceiçao	26S 35832	426730	BRFAHC01-09; 12-13; 15-16;18-23	50	19		
G4	*P.* sp1-Azores	Sao Miguel Island: Poço E (Lagoa)	N/A	N/A	SPSMPE01-04	51	4	7	3
		Sao Miguel Island: Poço F (Lagoa)	N/A	N/A	SPSMPF01	52	1		
		Sao Miguel Island: Praia do Populo, poço B (S. Roque)	N/A	N/A	SPMPPB02-03	53	2		
Bermuda									

G4	*P. carpalis* Stock, Holsinger, Sket & Iliffe, 1986	Bermuda: Red Bay cave	20S 336598	3578502	CARPBE01; 02; 04; 06	54	4	4	1
G4	*P. grandimanus* Stock, Holsinger, Sket & Iliffe, 1986	Bermuda: Red Bay cave	20S 336598	3578502	CARPBE03; 05	54	2	3	3
		Bermuda: well at Fort St. Catherine	20S 342492	3584965	GRBESC01	55	1		
Morocco									
G3	*P. longipes* Coineau & Boutin, 1996 (= *Ps.* Sidi Abdellah-1 (large, male U3 elongate, U1 with basofacial spine)	Taza: Sidi Abdellah, well beside road to Fes, in front of mosque	30S 377115	3781486	SPMASF01;06; SPMASA02-03; 06-08; 10-12 PSTASA01-04	56	14	14	3
G3	*P. ruffoi* Coineau & Boutin, 1996	Berkane: Source below Grotte de Chameau, at Gorges du Zeghzel	30S 558800	3855270	PSMEBE01, SPBEZE01, BERKAN01	57	3	5	2
		Taza: well at Friouato polje, 1361 m a.s.l.	30S 400242	3772805	PSTAFU03	58	1		
		Guerzif: Code: P2 Guerzif	N/A	N/A	SPGUGU01	59	1		
G3	*P.* sp1-Morocco	Oued Laou: Ounsed Tamda	N/A	N/A	PSMAOL01-03	60	3	3	2
G3	*P.* sp2-Morocco	Al Hoceima: Bni Boufrah	N/A	N/A	PSMAHO02-04	61	3	3	2
G3	*P.* sp3-Morocco	Al Hoceima: Bni Boufrah	N/A	N/A	PSMAHO01	61	1	1	1
G3	*P.* sp4-Morocco	Taza: well at Friouato polje, 1361 m a.s.l.	30S 400242	3772805	PSTAFU01	58	1	1	1
G3	*P.* sp5-Morocco	Aknoul: P.S. Ali Boubarg	N/A	N/A	SPMAAK01	62	1	1	1
G3	*P.* sp6-Morocco	Taza: Sidi Abdellah, well beside road to Fes, in front mosque	30S 377115	3781486	SPMASF02-05; SPMASA04; 09	56	6	6	3

Morphology

Specimens that were fixed in the field in 95% ethanol and later deemed to be used for morphological diagnosis were treated in the laboratory with lactic acid to remove internal tissues to facilitate observation. Drawings were prepared using a camera lucida on Olympus BH2 and Leica DM 2500 microscopes equipped with Nomarski differential interference contrast. Material preserved on slides was mounted in lactophenol and the coverslips sealed with nail varnish. Body measurements were derived from the sum of the maximum dorsal dimensions of head, pereionites, pleonites and urosomites including telescoped portions, and exclude telson length. Material is deposited in the Crustacea collection of The Natural History Museum, London (BMNH). Antennule and antenna appear abbreviated elsewhere in descriptions as A1 and A2, gnathopods I-II as G1-G2, pereiopods III-VII as P3-P7, pleopods I-III as PL1-PL3, and uropods I-III as U1-U3, respectively.

Molecular methods

Extraction

DNA of whole specimens was individually extracted using the Quiagen blood & tissue kit (Quiagen, Hilden, Germany) following factory's protocol. Due to low yields of genetic material caused by the inherent small size of the specimens (1-14 mm), for most of the material used for mitochondrial genome amplification were pooled together up to 7 individuals always assuring that the samples were from the same localities and shared the same morphotype.

Barcoding amplification, sequencing and data asembly

Sequence amplification of cytochrome oxidase subunit 1 (*cox1*) was performed using the primer-pair HCO/LCO (Folmer et al. 1994), or a degenerated primer set especially designed for this study to amplify recalcitrant *Pseudoniphargus* individuals (PseuF 5'-GCTCATGCTTTTGTTATGATTTTYTTYATRGT-3' and PseuR 5'-CAAAACAGATGTTGATAAAGAATTGGRTCNCCNCC-3'). PCR amplification, purification, and Sanger sequencing followed standard barcoding protocols (Bauzà-Ribot et al., 2011). Amplification was done using TaKaRa Taq (Takara Bio. Inc. Tokyo) in a 25µl volume per sample. PCR protocol used were as follows; [94°C for 2 min; (94°C for 30 s, 46-55°C for 30 s, 72°C for 1 min)x 35-40 cycles, 72°C for 10 min, 4°C hold]. Results were identified on a 1% agarose gel Purification was done using spin columns from Invitec (Invitek GmbH, Berlin, Germany). Amplicons were sequenced in both directions using the ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction kit v. 2.0. An ABI 3100 automated sequencer (Applied Biosystems, Foster City, CA, USA) was used for the electrophoresis and detection of the sequences. Alignment of *cox1* sequences in MAFFT (Katoh 2013) was trivial due to the absence of insertions or deletions.

Species delimitation criteria using barcodes

We removed identical sequences from the dataset with the uniqeseq perl script by Takabayashi to retain only unique haplotypes since the GYMC method cannot handle identical sequences. We also used this reduced dataset for PTP, TCS and ABGD species delimitations to be consistent despite that the latter can handle identical sequences. The dataset was partitioned into two sets, one partition including 1st and 2nd positions, and 3rd positions in a second partition since in *Pseudoniphargus* their composition and rates differ greatly (Stokkan et al. 2016). Models were evaluated by MrAIC.pl v1.4.6 (Nylander 2004) yielding TRN+G and HKY+G as best-fitting models. An Ultrametric tree was estimated in BEAST v1.8.2 (Drummond and Rambaut 2007) under the models and partitions described above. Analysis also implemented a coalescent constant population size as tree model, and a single uncorrelated relaxed clock with a log-normal distribution. The rate was arbitrarily set with a mean of 0.0115 nucleotide substitutions per site and per lineage and Ma (log mean -4.465, stdev 0.1, 95% confidence interval 0.0095-0.014) since GMYC only needs relative rates/ages to optimize the threshold. This mean rate (0.115) matches the standard pair-wise distance of 2.3% widely implemented in *cox1* sequences from invertebrate species (Brower 1994). The run included 50 million generations sampled every 10,000th. Convergence of the run was assessed in Tracer v1.6 (Rambaut et al. 2014) with a 10% burnin, which ensured a large statistical confidence for the estimated parameters (ESS > 200). Species delimitation with the GMYC algorithm was performed with the R library splits v 1.0-19 (Ezard et al. 2009) using a single threshold and the required R packages ape, paran, and MASS. We also estimated the support value of each branching event (node) to hold a transition from species branching to population branching with statistical probability p=0.95. The GMYC support values are defined as the sum of Akaike weights of candidate delimitation models where the node is included (Fujisawa & Barraclough 2013). Support values vary from full support (1.0), that indicates that all coalescent models tested validate a threshold for that node, to zero where no models support such boundary.

Trees for the PTP species delimitation were estimated using a ML criterion in RaxML v.8.2.4 (Stamatakis 2006) implementing two partition (1st + 2nd vs 3rd codon sites) and two independent GTR+CAT models. We also estimated one thousand tree topologies with branch lengths using the fast bootstrap algorithm as implemented in RaxML. They were used to delineating species boundaries and their support under Poisson Tree Process and a single threshold under a ML criterion as implemented in the python version of PTP v2.2 (Zhang et al., 2013). The dataset was run through TCS (Clement et al. 2000) based on the cladistic method described by Templeton (Templeton et al. 1992). Statistical parsimony analysis partitions data into independent networks of haplotypes connected by changes that are non-homoplastic with a 95% probability. Finally, the species were delimited based on genetic distances in ABGD (Puillandre et al. 2012). This program finds recursively the slope above a cut-off value (X) that splits the data in intra- and interspecific distances starting from an arbitrary minimum (p) and maximum (P) value for intraspecific distances. We implemented a p value of 0.0001 enabling the existence of nearly identical samples, and a P value of 0.1 that is the maximum intraspecific distance found in *Metacrangonyx longipes*, a stygobiont amphipod species that was densely sampled in the Balearic islands (Bauzà-Ribot et al. 2012). The nucleotide substitution model implemented to calculate distances was the Tamura-Nei model, with a lower cut-off value (X=1.0) than the default one (X=1.5), since the latter produced a single species, i.e. the default slope threshold was not

significant. Those values were used to estimate the inflection point between two distributions of ranked frequencies of pairwise distances using 10 recursive steps, i.e. the minimum optimal gap between intra- and interspecific distances.

Genetic distances

The Molecular Operational Taxonomic Units hereby referred to as MOTUs defined by each of the delimitation approaches were set as groups and mean p-distances were calculated in MEGA v7 (Tamura et al. 2013) at both intra- and inter-group levels. Boxplots were generated by R statistics (R Development Core Team 2008).

Mitogenomes from outgroup species

The sequences of the following outgroups were downloaded from GenBank and metAMiGA (Feijão et al. 2006) databases: Mysidacea (*Neomysis orientalis* - KC995119); Isopoda (*Asellus aquaticus* -GU130252; *Eophreatoicus* sp. - FJ790313; *Ligia oceanica* - DQ442914; *Armadillidium vulgare* - GU130251; *Sphaeroma serratum* - GU130256; *Eurydice pulchra* - GU130253; *Glyptonotus antarcticus* - GU130254; *Idotea balthica* - DQ442915); and Amphipoda (*Bahadzia jaraguensis* - NC_019661; *Caprella mutica* - GU130250; *Gondogeneia antarctica* - JN827386; *Onisimus nanseni* - FJ555185; *Parhyale hawaiensis* - AY639937; *Gammarus duebeni* - JN704067; *Eulimnogammarus verrucosus* - KF690638; *Metacrangonyx boveei* - HE860498; and *M. remyi* - HE860512). The sequence of the outgroup species *Metacrangonyx dhofarensis* was also obtained in the frame of this study.

Mitogenome amplification, sequencing and annotation

Three of the mitogenomes used in this thesis are already published by us (Stokkan et al. 2016): *Pseudoniphargus sorbasiensis* (LN871175), *P. gorbeanus* (LN871176), and *P. daviui* (FR872383). Mitogenome sequences were obtained following three different approaches. The mitogenomes of 10 *Pseudonipargus* species were obtained with a long ranged PCR amplification method. Two large amplicons usually arround 9 and 6 kb constituting the entire mitochondrial genome were amplified using Herculase™ II Fusion DNA polymerase (Agilent Technologies, Santa Clara, CA), and following the recommended protocol, pluss adding 10µl of 50M betain per sample in several of the samples to enhance amplification. For this approach, species-specific primers were designed in cytochrome oxidase subunit 1 (*cox1*) and cytochrome oxidase b (*cob*) (Stokkan et al. 2016; Bauzà-Ribot et al. 2012; **Table 3.2** below and **Table 4.4.1** chapter 4.4 page 97). Purification was done by using Invitek MSB® Spin PCRapace purification kit (Invitek GmbH, Berlin, Germany), and quantification was estimated using (Qubit Thermo Fisher Scientific Inc., Waltham, MA). Each batch were pooled in equimolar concentration of 100 ng/ml. Finally pooled fragments were sheared by nebulization, adapters ligated, fragments amplified by clonal emPCR, and finally sequenced unidirectionally in a Roche GS Junior using Titanium chemistry. Reads were assembled in contigs in CodonCode Aligner v.5.1.5 (CodonCode Corp., Denham, MA, USA).

Table 3.2 Primers used in this study, both specific primers designed for long range PCR, as well as primers used for specific gene regions of mitochondrial genes.

Gene	Name	Forward/ Reverse	5' to 3' sequence	Reference
cox1	Sorbasiensis_cox1_F	Forward	TGATAGAAAGAGGGGTGGGAACTGG	Current study
cob	Sorbasiensis_cob_R	Reverse	CGGTTTGGTGGAGGAGGGTAATA	Current study
cox1	Sorbasiensis_cox1_F2	Forward	GGTATGATAGAAAGAGGGGTGGGAAC	Current study
cob	Sorbasiensis_cob_R2	Reverse	CAGCAATAGTGGGTAGAAGAGGCAAA	Current study
cob	Sorbasiensis_cob_F	Forward	ACTCCAGCAATACAAGCAGTTACCCCTCATC	Current study
cox1	Sorbasiensis_cox1_R	Reverse	TAAACAGTTCAACCAGTTCCCACCCCTCT	Current study
cox1	Gorbeanus_cox1_F	Forward	GGTATAATAGAAAGAGGAGTTGGTACAGGCT	Current study
cob	Gorbeanus_cob_R	Reverse	CTAAGGATAATGGTATGTGTAGGCAAAG	Current study
cob	Gorbeanus_cob_F	Forward	TTATTATCTTTGCCTACACATACCATTATCC	Current study
cox1	Gorbeanus_cox1_R	Reverse	CATACTATCATAACCTATTCCCGCTGTCC	Current study
cox1	Daviui_cox1_F2	Forward	CAGACATAGCCTTCCCTCGCATAAATAAC	Bauzà-Ribot et al. (2012)
rrnL	Daviui_16s_R1	Reverse	AAATACTATAATAAAATAAAGTGACGATAAGACCCTATAAGC	Bauzà-Ribot et al. (2012)
rrnL	Daviui_16s_F1	Forward	ATCTTATGCTCATCCATCAATGTAAAACTCAC	Bauzà-Ribot et al. (2012)
cox1	Daviui_cox1_R1	Reverse	GTTATTTATGCGAGGGAAGGCTATGTCTG	Bauzà-Ribot et al. (2012)
cox1	PSNI_cox1_F_A39	Forward	GCTCATGCTTTTGTTATGATTTTYTTYATRGT	Current study
cox1	PSNI_cox1_R_A71	Reverse	CAAAACAGATGTTGATAAAGAATTGGRTCNCCNCC	Current study
cox1	F_LCO	Forward	GGTCAACAAATCATAAAGATATTGG	(Folmer et al., 1994)
cox1	R_HCODaviui	Reverse	TAAACTTCAGGGTGACCAAAAAATCA	(Folmer et al., 1994)
rrnL	M14_meta	Forward	RGTATTTGACCGTGCTAAGG	Xiong and Kocher 1991
rrnL	M74_meta	Reverse	TGTAAAAAATTAAARGTTGAACAAAC	Xiong and Kocher 1991
cob	CB3	Forward	AAAAGAAARTATCATTCAGGTTGAAT	Barraclough et al. 1999
cob	CB4	Reverse	GAGGAGCAACTGTAATTACTAA	Barraclough et al. 1999

Another 18 mitogenomes were obtained by the sequencing of full genomes at low coverage in Illumina platforms. Full mitogenomes could be posteriorly retrieved bioinformatically at reasonable coverage since eukaryotic genomes harbor multiple mitochondrial copies (Burger et al. 2003). For most species, we pooled DNA of several specimens from the same population since yields from single individuals were extremely low due its tiny size (often less than 8 mm). Individual shotgun genomic libraries were generated by fragmentation and end repair of DNA and prepared with the Hyper Library construction kit from Kapa Biosystems (Wilmington, Massachusetts, USA) using single indexed Illumina TruSeq adapters. Adapter of read1 included a specific index sequence of 6 bp for each library that was unable to discriminate reads by species. The DNA fragment size of shotgun libraries was around 300-450 bp (ranging from 250 to 900 bp) depending on fragmentation behavior. Individually indexed libraries were quantified by qPCR, pooled in equimolar concentration, and sequenced from each end of the fragments, paired-ends, in a Miseq v2 (2 x 150) or one HiSeq2500 lane with TruSeq Rapid SBS sequencing kit (2 x 150). For 5 individually indexed libraries we obtained longer reads on a HiSeq2500, 2 x 160, using the HiSeq SBS kit v4 (Illumina). Fastq files were generated and demultiplexed with bcl2fastq software (Illumina). Adapters and low quality bases were removed from reads using AdapterRemoval v.2.0.0 (trim Ns, trim minimum quality 4, minimum read length 100 bp, maximum mismatch in adapters 0.1), or Trimmomatic v.0.33 (Bolger et al. 2014) (minimum overall quality 30, minimum read length 100, minimum quality base 33). Overlapping paired-end reads were collapsed in a consensus sequence. Contigs were

assembled *de novo* with Trinity v.2014-04-13p1 (Grabherr et al. 2011), and Ray v.2.3.1 (Boisvert et al. 2010), by using collapsed and truncated paired-end reads only.

For 4 *Pseudoniphargus*, we increased the amount of genomic DNA by multiple displacement amplification (MDA; Blanco et al. 1989) since we obtained an extremely low yield. This procedure exploits the highly processive DNA polymerase phi29 to add complementary base pairs to the initial strand while displacing the other one. Afterwards, secondary priming is initiated on the displaced DNA strand, finally resulting in multiple branches that produce a large amount of long DNA fragments (> 10 kb). Amplification reaction was performed following the protocols of the REPLI-g Mitochondrial DNA Kit (Qiagen, Hilden, Germany). It uses a set of primers that specifically target the human mitochondrial genome instead of the set of random hexamer primers that amplify the whole genome (Alsmadi et al. 2009). We took advantage of this procedure but replacing human mitogenome primers by the 7 primers specifically designed to amplify the mitochondrial genome of crustaceans (**Table 3.3**).

Table 3.3 List of primer set used to perform multiple displacement amplification (MDA).

name	gene	direction	sequence	target
FOR12S	*rrnS* (12S)	forward	AGGGTATCTAATC	All species
REV12S	*rrnS* (12S)	reverse	CTGGAAGGTG	Typhlatya
FOR16S	*rrnL* (16S)	forward	GAACTCAAATCAT	All species
REV16S	*rrnL* (16S)	reverse	AAGGTAGCATAAT	All species
FORtrnR	*trnR*	forward	CTGAAAATAGGC	Pseudoniphargus
REVCB4	*cob* (cytB)	reverse	TCATTCTGGTTGG	from Barraclough
FORLCO	*cox1* (COI)	forward	CATAAAGATATTGG	from Folmer

Primers used along with DNA polymerase phi29 are generally short (10-14 bp) and synthesized with phosphothioate bonds on the second and third base from the 3' end, which protect them from the 3' exonuclease activity of phi29. Primers for this study were designed on conserved regions from the sequences available for the amphipod species of the genera *Pseudoniphagus* and *Metacrangonyx*, and the shrimp genus *Typhlatya*. These sequences were obtained with "universal" primers for rrnL and rrnS genes (Bauzà-Ribot et al. 2012), and from mitogenome sequences obtained using previous approaches for trnR gene. We also included a short version of the "universal" primers LCO for cox1 (Folmer et al. 1994) and CB4 for cob (Barraclough et al. 1999). MDA protocol targeting mitochondrial genomes of *Pseudoniphagus* was as follows: First, we added 10 ng of genomic DNA to 29 µl of amplification mix that included our 7 primers at a final concentration of 0.4 µM each, and the mix was denatured at 75ºC for 5 minutes. After cooling down, we added 1 µl of polymerase to the mix, and it was incubated at 33ºC for 12 hours to finally inactivate polymerase by incubating mix for 3 minutes at 65ºC. DNAs amplified were purified using Invitek MSB Spin PCRapace purification kit (Invitek GmbH, Berlin, Germany), size and DNA integrity checked in 1% agarose gel, and quantified in Qubit Fluorometer (Thermo Fisher Scientific Inc.,

Waltham, MA). Genomic DNA amplified by MDA from several species were pooled in equimolar concentrations, libraries constructed by shotgun fragmentation, and the latter sequenced with MiSeq v2 to obtain 2 x 150 bp paired-ends. The contigs within each pool were associated to species by using species-specific bait sequences such as cox1, rrnL and cob, all obtained with "universal" primers (Timmermans et al. 2010; Pons et al. 2014). Reads were assembled *de novo* into contigs using Trinity, idba_ud v.1.1.1 (Peng et al. 2012) and Newbler 2.7 (Roche, Basel, Switzerland.

Short contigs and control regions were extended or completed by mapping reads at the 5' and 3' ends of the consensus sequence of contig using Bowtie v.2.1.0 (Langmead et al. 2009) in CodonCode Aligner.

Gene annotation and sequence analysis

Gene annotation was performed with DOGMA web server (Wyman et al. 2004) and the 5' and 3' gene ends manually refined by comparing with mitogenome annotations of other crustaceans. Secondary structures of tRNAs were corroborated using Arwen (Laslett and Canbäck 2008) and tRNAscan-SE (Schattner et al. 2005) search servers in parallel. Finally, annotations were checked with MITOS web server (Bernt et al. 2013), particularly to estimate the secondary structures of rRNA sequences. These were manually refined with Mfold web server (Zuker 2003) and graphically visualized using VARNA v3.9 (Darty et al. 2009). Palindromes in non-coding spacer regions and the control region were identified using the Mfold. Nucleotide and aminoacid composition, as well as relative synonymous codon usage (RSCU), were calculated using MEGA v5.05 (Tamura et al. 2011). The effective number of codons (ENC) (Wright 1990) was estimated with INCA V2.1 (Supek and Vlahoviček 2004). AT and GT skews were estimated as described in Pons et al.(2014), and ggskew options within the EMBOSS v6.6.0 package (Rice et al. 2000).

Mitochondrial genome rearrangements

Gene rearrangements with respect to the putative ancestral Pancrustacean and other known amphipod gene orders were investigated using CREx (Bernt et al. 2007). This software calculates and creates "strong interval trees", to heuristically disentangle the possible processes involved (Berard et al. 2007; Bernt et al. 2007; Perseke et al. 2008). A "common interval" is a subset of genes that appears successively in two or more input orders. A "strong interval" is defined by Berard et al. (2007) as:"A common interval I of a permutation P is a strong interval of P if it commutes with every common interval of P." This means both original and derived mitogenomes share a particular subset of consecutive genes or any rearrangement of the original one (strong common intervals), though they must not overlap, and be either disjoint or completely contained within the new arrangement. Their analysis produces a strong interval tree whose nodes can be defined as linear increasing if they appear as in the original gene order, or linear decreasing if they appear in reverse position to the original gene order (Berard et al. 2007). Otherwise, they are defined as prime (shown with rounded shapes, **see Figures 4.3.3-4.3.5** chapter 4.3 pages 83-85**)**. These nodes can again be interpreted as produced by different rearrangement events such as those caused by transpositions, reverse transpositions, reversals or by more complex TDRL events.

Mitogenomic phylogenetic analyses

Nucleotide sequences of individual PCGs excluding terminal stop codons were aligned in Seaview v.4.4.2. First, they were translated into aminoacids, aligned with Muscle v.3.8.31 with default parameters, and finally back translated to nucleotides conserving the integrity of codon triplets. Poorly aligned regions, mainly at 5' and 3' ends due to uneven sequence lengths, were removed with Gbloks v.0.91b under relaxed conditions and working at the codon level, i.e. allowing smaller final blocks, gap positions within final blocks, and less strict flanking positions. Nucleotide composition and bias, as well as proportion of gaps and ambiguous positions were estimated in IQTREE multicore v.1.3.12 (Nguyen et al. 2015).

The best partition scheme and substitution model for nucleotide sequences were estimated in PartitionFinder v.1.1.1 (Lanfear et al. 2012) using Bayesian Information Criterion (BIC), implementing the greedy algorithm and unlinking branch length across

partitions. We only included those models implemented in Beast (number of substitution parameters 1, 3 and 6) with a gamma parameter represented by 4 categories to take into account among-site rate variation. Models were implemented without invariants parameter as previous studies have raised concern about simultaneous use of invariants and gamma parameters in evolutionary models of nucleotide sequences (Bidegaray and Arnedo 2011; Yang 2006). This fact is particularly relevant at third codon sites where frequency of invariant sites is negligible. Besides, invariant positions constitute already a fraction of the gamma distribution defined by the alpha parameter. Previous studies showed that positions within codons differ in nucleotide composition, substitution model and rate (e.g. Pons et al. 2010; 2014; Jurado-Rivera et al. in press). Hence, input partitions for PartitionFinder were defined using a mechanistic criterion by splitting each of the 13 mitochondrial PGGs by the 3 codon positions, i.e. 39 partitions in total. The best partition scheme was also estimated at the aminoacid level using the 13 PCGs as initial partitions although models also included proportion of invariant sites since the value of this parameter is generally large at the aminoacid level. Alternative aminoacid models not implemented in PartitionFinder such as mtZOA were calculated in IQ-TREE software v.1.5.1 (Nguyen et al. 2015) with TEST command. Alternative partition schemes were compared using BIC values estimated in IQ-TREE.

Phylogenetic trees were built using Maximum Likelihood (ML) criterion at both nucleotide and aminoacid levels in IQ-TREE, with the best partition scheme and models estimated in PartitionFinder and branch lengths calculated independently for each partition (i.e. unlinked branch lengths). Tree topologies were also assessed under parsimony criterion in PAUP v.4.0 beta10 with 1,000 random addition replicates, saving 50 trees per replicate, and heuristic searches with tree-bisection-reconnection branch-swapping (Swofford 2002). Node support was assessed by 1,000 bootstrap replicates in ML and parsimony. Tree searches were also performed under the Bayesian criterion in Phylobayes mpi v.1.5a implementing the GTR+CAT model, that is less prone to phylogenetic errors caused by compositional bias or saturation (Lartillot and Philippe 2004). Two independent runs were sampled every 1,000 generations until convergence was reached, i.e. maximum difference < 0.1, and ESS values of all parameters were > 50 (mostly > 200).

Nucleotide substitutions of mitochondrial PCGs are usually saturated as genetic distance increases, particularly at the deepest nodes and on third codon sites due to genetic code degeneracy. Hence, the Xia saturation test was implemented on the whole dataset and for first, second and third codon sites independently, to assess the saturation of the phylogenetic signal as implemented in DAMBE v.6.0.0 (Xia and Lamey 2009; Xia et al. 2003). Saturation was also visualized by 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 et al. 2006).

Calibration

Tree topology, model parameter values and node ages were co-estimated and optimized in Beast v.1.8.2 (Drummond et al. 2012) implementing the best partition scheme and models described above. Bayesian analyses were run for 150 million generations on Tesla C2050 and Tesla C2075 graphic cards using Beagle library v.2.1 (Ayres et al. 2012) to speed up analyses. Convergence of the run was assessed in Tracer v.1.6 (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 (Drummond et al. 2012). We implemented the default priors for all parameters except for clock rates and Yule rate priors since the default ones did not allow for convergence during path sampling. The ucld.mean prior for first codon sites was set as a log normal distribution in real space with a mean (M) 0.005 and standard deviation (S) 1.5 (95% confidence interval 8.58x10e-05 - 3.07x10e-02); M 0.1 and S 1.5 (1.72x10e-04 - 6.14x10e-02) for second codon positions; and M 0.5 and S 1.0 (4.27x10e-3 - 0.21) for third 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 S 3.0 (95% confidence interval 1.55x10e-05 - 1.98). We estimated node ages under different clock models: Strict clock, two types of uncorrelated clocks where rates on descendant branches are independent of the rate at the parent branch, relaxed log-normal clock and local random clocks. The first relaxed model estimates a log-normal distribution of rates for the whole tree, then it is split in n bins as the number of branches in the tree, to finally assign one bin (rate) to each branch. The latter allows different and independent rates across the tree whose number is estimated from a Poison distribution. Runs implementing random local clock 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 the stepping stone model as implemented in Beast. We performed 100 steps of 5 million generations using a path scheme with a betaQuantile 0.33 (Baele et al. 2012), discarding 25% of the run as burnin.

Since the fossil record of the Amphipoda is extremely limited and does not extend beyond the Eocene (45-50 Ma; Jażdżewski et al., 2014; Starr et al., 2016), we had to rely on relevant fossils of other peracarids to estimate phylogenetic divergence dates. We used two isopods –a group with a rich fossil record that extends back to the Palaeozoic– to assign a minimum age constraint on the divergence event at the base of their respective clades. Namely, the oldest fossil isopod known, the Carboniferous (Middle Pennsylvanian) phreatoicidean *Hesslerella shermani* Schram, 1970 (see Wilson & Edgecombe, 2003), and the oldest undisputed record of the Sphaeromatidea, *Elioserolis alpina* Basso & Tintori, 1994, from the late Norian (Triassic) of the southern Alps (Basso & Tintori, 1994; Etter, 2014). Since there is no information about precisely when these divergence events may have happened within the geological subperiod/age assigned originally to the fossils, we defined a log-normal distribution as hard constraints for these two nodes with a mean (M) 5.71 and Standard deviation 0.01 (CI: 296.0-307.8) for *Hesslerella* and M 5.35 and S 0.01 (CI: 206.5-214.8) for *Elioserolis*. Despite the membership of these two fossils to any modern family cannot be assured, their subordinal ascription seems undeniable (Etter 2014) and enable their node assignment as suggested.

In lieu of ingroup fossils, we combined the foregoing calibration bounds with 3 others derived from geological evidence to constrain 3 shallow nodes in the *Pseudoniphargus* tree. In doing so, we assumed that phylogenetic divergence between sister species occurring only on the same island or narrow portion of land took place after the emergence of such territories from the sea. Thus, and contrary to the age of the fossils commented above (minimum age constraints), these ages should be considered as maximum age constraints.

We have relied only on geological age estimates of oceanic islands harboring sister species (see below). Islands inhabited by a single species of *Pseudoniphargus* or

accounting with representatives of different lineages are inadequate for calibration since the age of a lineage may bear no relation to the age of an island. This is because colonization may have occurred well after the formation of the island, or indeed beforehand –when the direct ancestor still inhabited the marine environment–, as there are older, submerged seamounts in many archipelagos, and therefore simple island ages cannot properly constrain a node (Page et al. 2016).

The three geological events selected are as follows:

A) Emersion of the Basque Country (western edge of the Pyrenees; northern Spain) at the Early Oligocene (33.7 Ma; Rögl 1998). This territory harbors the sister species triplet *P. gorbeanus*/*P. unisexualis*/*P*. sp1-Basque. We constrained this node with a strong prior by implementing a log-normal distribution with a mean of 32.73 Ma and stdev 0.1 (95% confidence interval 26.77-39.62).

B) Establishment of the so-called Southern Rifian 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. 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 hard constraint for this node with a mean of 7.05 Ma and stdev 0.1 (5.77-8.53).

C) Emersion of La Palma Island (Canaries), which harbors the sister species pair *P. cupicola*/*P. multidens*. The age of the split between these two taxa cannot be older than the first subaerial exposure of the island (2.0 Ma; Carracedo et al. 2001). We defined a log-normally distributed function as a hard constraint for this node with a mean of 2.0 Ma and stdev 0.1 (1.64-2.42).

Chapter 4. Results and Discussion

Chapter 4.1 Diversity, species delimitation and crypticism in a genus of subterranean water amphipod crustaceans (*Pseudoniphargus*: Pseudoniphargidae)

Results

Taxonomic analyses based on morphological characters indicated that the specimens analyzed corresponded to 30 formally described species plus other 18 recognized herein as new morphotypes/species (see **Figure 3.1 and Table 3.1.** Chapter 3 page 36 and 37 respectively). Most species were collected on a single site and just a few ones came from different but adjacent localities (viz. *P. brevipedunculatus* (10), *P. elongatus* (5), *P. mercadali* (3) and *P. ruffoi* (3)). As expected for a stygobiont crustacean genus, most caves and wells harbored a single species but in a few sites two species coexisted; namely *P. triasi* and *P. daviui* on Cabrera Island (Balearics), *P. pedrerae* and *P. pityusensis* on Formentera Island (Balearics), *P. guernicae* and *P.* sp1-Basque both cohabiting in Goiko Etxe caves (Vizcaya; North Spain), *P. carpalis* and *P. grandimanus* in Red Bay cave in Bermuda, *P. longipes* and *P.* sp6-Morocco in a well at Sidi Abdellah (Taza; Morocco), *P.* sp2-Morocco and *P.* sp3-Morocco in a well at Al Hoceima (Morocco), and finally *P. ruffoi* and *P.* sp4-Morocco in a well at the Friouato Polje (Taza; Morocco).

After a preliminary morphological assessment, we generally sequenced five (3-7) individuals per species/morphotype and site. In some cases, we sequenced more specimens since sampling was extremely successful so that we could get a better representation of intraspecific genetic variation and increase the chance of picking up cryptic species or rare haplotypes; e.g. *P. gorbeanus* from Cigoitia (22), *P. brevipedunculatus* from Horta-Conceiçao (19) and Sao Roque (17), *P. cupicola* from Charco Verde (18), *P. longipes* from Sidi Abdellah (14), and *P. daviui* from Font d'Enciola (12). In total, we gathered the mitochondrial *cox1* sequences of 410 *Pseudoniphargus* individuals out of the ca. 450 individuals screened from the 62 collecting sites. The final alignment contained 657 nucleotide positions with no indels, stops codons or rare non-synonymous substitutions that could suggest the presence of numts (i.e. mitochondrial pseudo-copies inserted in the nuclear genome). The average sequenced region of *cox1* was 462 bp with a standard deviation of 121 bp and a mode of

294 bp (62 sequences), corresponding to the fragment amplified by the primer pair PseuF-PseuR specifically designed for this study. Other differences in length were due to the trimming of bad quality bases at the 5' and 3'. The percentage of constant sites was 53.272% and the fraction of parsimony-informative positions 43.683%. No shared haplotypes were found between localities except for two cases where the same, single haplotype was foundin sites placed a few kilometers apart: *P. daviui* in the two known localities of the species on Cabrera Island (Balearics); and *P. brevipedunculatus* in two wells, Areia Larga and Calhau, on Pico Island (Azores). After removal identical sequences, the remaining 182 sequences were used to perform further analyses.

The widely used single threshold of the GMYC algorithm estimated that the two-parameter model including both coalescent and Yule branching (ML 295.907) fit better our *cox1* dataset than the null model with only a single coalescent parameter (ML 245.244): likelihood ratio test 101.326, with high statistically significance ($p < 1e^{-15}$). The single threshold date that optimized the switch from diversification to coalescence was estimated at 2.740 Ma (Million years ago), and retrieved 63 MOTUs (entities in GMYC) with a confidence interval ranging from 59 to 67. The absolute value of the threshold was older than expected suggesting that the arbitrary rate implemented is probably too slow for *Pseudoniphargus*. There were 21 MOTUs represented by a single specimen, whereas the remaining 42 clusters were composed of multiple individuals (**confidence interval 39-44, Figure 4.1.2. Table 4.1.1**). Most taxonomic species studied, 30 out of 47, corresponded to a single MOTU defined by GMYC, with most of the latter strongly supported by GMYC support values ranging between 0.9-1.0 (i.e. most of the coalescent models tested supported them). Nonetheless, a few others were weakly supported, such as most of the MOTUs estimated for *P. brevipedunculatus*, *P. elongatus, P. gorbeanus, P. longipes* and *P.* sp1 and *P.* sp3-Andalusia.

The other tree-based method PTP estimated less putative species than GMYC, between 38 and 78 MOTUs, with a mean of 57.066. The most supported partition scheme by simple heuristic search included 55 MOTUs that were mostly supported with high values (> 0.90) except for *P.* sp1-Azores, *P. cupicola*, *P. grandimanus, P. carpalis*, *P. mercadali, P. elongatus*, *P.* sp1 and *P.* sp3-Andalusia, *P. longipes*, and *P. gorbeanus*. Forty-four MOTUs were identical between PTP and GYMC, with 36 of them corresponding to morphological species (**Table 4.1.1**). The most striking differences found corresponded to the four GMYC MOTUs of *P. brevipedunculatus*, that PTP merged in a single one. In fact, PTP showed a better correspondence between morphological species and MOTUs compared to GMYC, e.g. in *P. cupicola*, and *P.* sp1 and *P.* sp3-Andalusia aside from the above mentioned *P. brevipedunculatus*. On the other hand, PTP over-split *P. gorbeanus* even more, but merged the sequences from two morphological species, *P. guernicae* and *P.* sp1 Basque, in a single MOTU.

The algorithms based on the number of parsimony changes below homoplasy (TCS) and on genetic distances (ABGD) produced less MOTUs than the tree-based methods, 56 and 49 respectively (**Table 4.1.1**). In TCS, 42 MOTUs corresponded to a single species and only a few were split into many MOTUs such as *P. elongatus* (5), *P. mercadali* (3), and *P.* sp6-Morocco, *P.* sp1-Azores, and *P. gorbeanus* with two MOTUs each (**Figure 4.1.2; Table 4.1.1**). Finally, ABGD with the recursive analysis of the sequences estimated the maximal intraspecific distance (P) as 4.64%, which produced

49 MOTUs. Forty out of the 47 morphological species corresponded to a single MOTU except *P. grandimanus* and *P. carpalis*, that were merged in a single one as well as *P. guernicae* and *P.* sp1-Basque, whereas *P. mercadali* was split in three MOTUs, and *P.* sp6-Morocco and *P. elongatus* were split into two each.

Genetic diversity

Mean p-distances within MOTUs were quite low as expected, with a median of 0.1% and half of values falling below 0.4-0.7% except for ABGD, whose median was 0.3% and many values were above 1.15% (**Figure 4.1.1**). PTP, TCS, and ABGD analyses showed few outliers, in the range of 1-3%, but two values were higher in PTP (6.5%) and TCS (4.1%), respectively. Median distant values among MOTUs, corrected by Tamura-Nei model to take into account saturation, were around 24.5% in all four methods, with few upper outliers (**Figure 4.1.1**). The main differences among the four methods were on the lowest values: GMYC (2.1-6.5%), PTP (1.5-6.5%), TCS (3.1-6.5%) and ABDG (>7%).

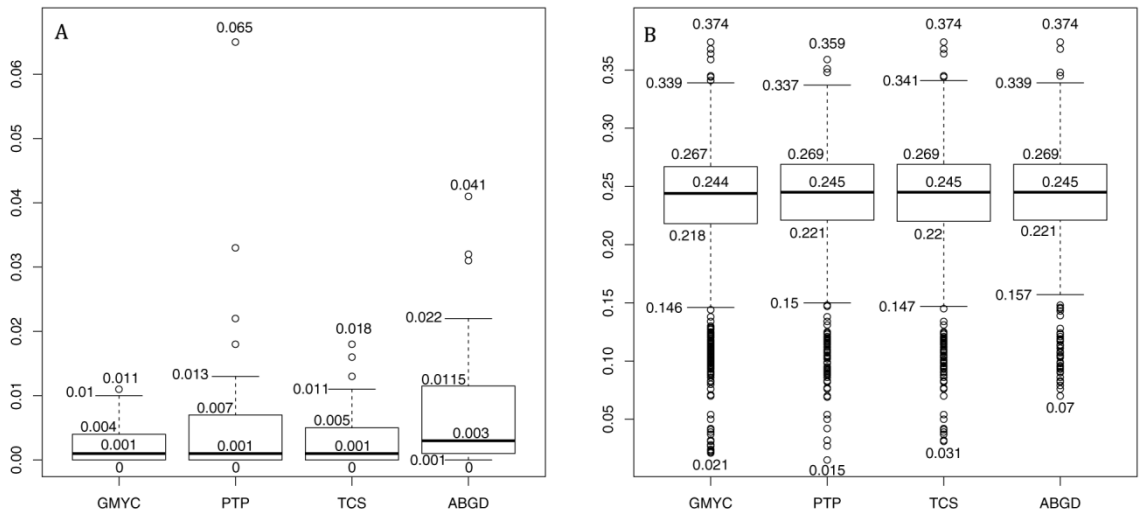


Figure 4.1.1 Boxplot showing A) intraspecific and B) interspecific genetic p-distance for the four molecular methods of species delimitation implemented

Species groups

We divided the samples into four groups (Group 1 to Group 4) for operational descriptive purposes based on Bayesian tree topology (**see Figure 4.1.2**), and considered four broad geographic areas as described below and in chapter 3 (**Table 3.1** chapter 3 page 37). Overall the four methods yielded consistent results in almost half of the MOTUs on the dataset where all the methods concurred, and even were more congruent with morphology if we take into account GMYC and PTP support values. There were however several occasions where one or more of the four methods did not render coincident results with the rest. *P. burgensis*, *P. gevi*, *P. gorbeanus*, *P. guernicae*, *P. incantatus*, *P. unisexualis*, *P. jereanus*, *P. mateusorum*, *P. grandis*, *P.*

latipes, *P. sorbasiensis*, *P. stocki*, *P. daviui*, *P. pedrerae*, *P. triasi*, *P. associatus*, *P. gomerae*, *P. multidens*, *P. salinus*, *P. portosancti*, *P. brevipedunculatus*, *P. grandimanus*, *P. carpalis*, *P. longipes* and *P. ruffoi* are all formally described species that throughout all methods resulted in one MOTU per species, whereas *P.* sp1-Andalusia and *P.* sp1-Azores showed to represent complexes of cryptic species since each embraced two non-sister MOTUs. Other formally described species, such as *P. elongatus*, *P. cupicola*, *P. carpalis*, the populations of *P. mercadali* from Menorca and *P. gorbeanus* embrace two or more MOTUs each, although showing sister kinship. All these cases, plus other 17 corresponding to not yet formally described species discovered in the course of our samplings, and also included in the analyses –some of which embracing several MOTUs–, are dealt with in detail in the forthcoming paragraphs.

Table 4.1.1 Number of MOTUs delimitated by the various methods presented here.. Values 0.5 means that the sister species *P. guernicae* and *P.* sp1-Basque, and *P. carpalis* and *P. grandimanus* were merged in a single MOTU.

Clade / Species	GMYC	PTP	TCS	ABGD
P. mateusorum (s)	1	1	1	1
P. mercadali	4	4	3	3
P. sp1-Murcia	1	1	1	1
P. sp2-Andalusia	1	1	1	1
P. multidens	1	1	1	1
P. cupicola	2	1	1	1
P. brevipedunculatus	4	1	1	1
P. carpalis	1	1	1	0.5
P. grandimanus	1	1	1	0.5
P. associatus	1	1	1	1
P. gomerae	1	1	1	1
P. salinus	1	1	1	1
P. azores	2	2	2	1
P. sp1-Canaries	1	1	1	1
P. sp2-Portugal	1	1	1	1
P. portosancti	1	1	1	1
P. daviui	1	1	1	1
P. grandis	1	1	1	1
P. sp4-Morocco (s)	1	1	1	1
P. sp3-Morocco (s)	1	1	1	1
P. sp2-Morocco	1	1	1	1
P. sp1-Balearics	1	1	1	1
P. sp1-Andalusia	2	1	1	1
P. "gevi"	1	1	1	1
P. stocki	1	1	1	1
P. sp3-Andalusia	2	1	1	1
P. "morenoi"	1	1	1	1
P. sp5-Morocco (s)	1	1	1	1
P. sp1-Portugal	1	1	1	1
P. longipes	1	1	1	1
P. sp1-Morocco	1	1	1	1
P. ruffoi	1	1	1	1
P. sorbasiensis	1	1	1	1
P. sp6-Morocco	2	2	2	2
P. pityusensis	1	1	1	1

P. triasi	1	1	1	1
P. pedrerae	1	1	1	1
P. sp2-Canaries (s)	1	1	1	1
P. latipes	1	1	1	1
P. elongatus	5	3	5	2
P. jereanus	1	1	1	1
P. gorbeanus	2	3	2	1
P. unisexualis	1	1	1	1
P. burgensis	1	1	1	1
P. guernicae	1	0.5	1	0.5
P. sp1-Basque	1	0.5	1	0.5
P. incantatus	1	1	1	1

Group 1

The species from the Basque Country in northern Spain conform a monophylum (G1 in **Figure 4.1.2** and **Table 3.1** chapter 3 page 36 and 37) representing the most basal cluster differentiated within the genus. Its distribution embraces the upper reaches of the Ebro river basin in the portion of the Spanish province of Burgos adjacent to the Basque Country, plus valleys and massifs of the three Basque provinces (Alava, Guipúzcoa and Vizcaya) (see **Figure 3.1** chapter 3 page 36). Out of the nine species formally recognized in the area (Notenboom, 1986), we have been able to recover molecular data from six of them (see **Table 3.1** chapter 3 page 37). In addition, we have discovered a new species at Ría de Guernika on the Biscay (Vizcaya) coast (*P.* sp1-Basque; **Figure 4.1.2, Table 3.1** chapter 3 page 37 and **Table 4.1.1**), where *P. guernicae* Notenboom, 1986, was supposed to be the only representative of the genus. The new species appears in our analyses as sister to *P. guernicae*, but can be readily differentiated based on the proportions of the G2 (Gnathopode2) in both sexes. However, based on the molecular data only two of the four methods were able to detect this as a separate species, namely GMYC and TCS. This is the only case of sympatry that has been recorded within this northern group of species, involving the two formerly commented taxa from Ría de Guernika, at Cueva de San Pedro (Axpe; Busturia). *P. gorbeanus* is another example of a possible species complex where GMYC and TCS shows 2 MOTUs and PTP indicate 3 MOTUs, and only ABGD concurs with the morphologically defined species as 1 MOTU. Although there are some genetic differences between the groups, it still forms a monophyletic entity and the support for the splitting is not particularly high. This might be indicative of the existence of some level of variety within an old lineage.

Group 2

The species from the Cantabrian mountains of northern Spain, also form a well-defined clade in our analyses (**Figure 4.1.2**). Only one out of the four formally recognized species present in the area was included in the study, viz. *P. elongatus* Stock, 1980 (see Notenboom, 1986). It is endemic to the Asón river valley and surrounding areas of Santander and Burgos. Molecular species delimitation criteria suggests the presence of three MOTUs, with quite high support, within this otherwise morphologically well characterised species (it is the only *Pseudoniphargus* in northern Spain exhibiting an elongate male third uropod exopod), although the paucity of material from two of them (from Cueva Fresca and Cueva del Gándara, respectively, placed in two adjacent valleys at the headwaters of the Asón River; see **Figure 3.1** and **Table 3.1** chapter 3 page 36 and 37) does not enable going further on the study of their morphological

differentiation. This could be either a current speciation event or sign of strong convergence; however perhaps the former appear to be more accurate considering the close distance between localities. This again is an old lineage and might show elevated levels of variety compared to younger lineages for that very reason. This clade G2 appears as sister to the clade G4, described below.

Group 3

This is mainly a Mediterranean clade embracing species from S and SE Spain, the Balearic Islands as well as northern Morocco, but includes also two species from extra-limital areas, *viz*. one from Gran Canaria (Canary Islands; *P*. sp2-Canaries; **Table 3.1** chapter 3 page 37), and another one from a cave at Sicó (mainland Portugal; *P*. sp.1-Portugal; **Table 3.1** chapter 3 page 37). The various methods of species delimitation show here high levels of concordance with the species diagnosed based on morphology, save in two cases where presence of cryptic species cannot be ruled out. Namely, a population of a not yet described species *P*. sp1-Andalucia, from 2 localities (Cueva del Yeso and Complejo Zarza) where GMYC indicates that the two localities possibly harbour two separate MOTUs, however the support is not particularly strong (0.65) and all other methods indicate that this is only one species. Another possible cryptic species complex might be found in the third species from Andalusia, the not yet described *P*. sp3-Andalusia. Here a single specimen is detected as a separate MOTU by the GMYC method with some support (0.77), but other methods do not recognize this MOTU.

There is also a different case for a not yet described species, that appears in sympatry with *P. longipes* in a single well at Sidi Abdellah (Morocco; *P*. sp6-Morocco; Table 1). Contrary to the latter species, *P*. sp6-Morocco displays a short male third uropod and the basofacial spine on the protopod of the first uropod is absent, thus enabling a straightforward distinction between both taxa and hence is not a case of cryptic speciation. In addition, our analyses show that the morphology of *P*. sp6-Morocco covers two distinct, sister and sympatric MOTUs (see **Figure 4.1.2**. and **Table 4.1.1**.).

Cases of sympatry among members of this clade readily distinguishable morphologically are frequent and include the species-pairs *P. triasi*/ *P. daviui* in their type locality on Cabrera Island (Balearics); *P. pityusensis*/ *P. pedrerae* in their type locality on Formentera Island (Balearics); *P*. sp2-Morocco/ *P*. sp3-Morocco, from the same well at Al Hoceima (Morocco); *P. ruffoi*/ *P*. sp4-Morocco recorded in the same well at Friouatto Polje (Taza; Morocco); and *P. longipes*/ *P*. sp6-Morocco in the well at Sidi Abdellah commented above.

Group 4

A fourth clade includes most of the species from Atlantic Islands but also some from Portugal, the Balearics and SE Spain. Molecular species delimitation criteria suggest the occurrence of up to 4 MOTUs within *P. mercadali* from three localities from Menorca and Mallorca islands (Balearics). Both GMYC and PTP indicate a four MOTUs complex while TCS and ABGD suggest three MOTUs although this complex shows quite low support in the deeper nodes. This is suggesting a possible cryptic species complex since no morphological differences were found. This may be caused by relative short speciation time. As indicated above, this species may have also reached its current location by dispersal, and might have suffered a rapid differentiation caused by the dispersal event (i.e. founder effect or bottleneck genetic drift).

Furthermore, another possible cryptic species complex is found within *P. cupicola* from La Palma Island (Canaries), which embraces two sister MOTUs according to GMYC, corresponding to the two localaties Charco Verde and Charco de Los Chochos with high support (0,96). However, these two populations are recovered as a single species by the three other molecular methods. This also might be an early sign of an ongoing speciation event.

Other, more prominent cases of presumed cryptic species than the former ones since the MOTUs involved do not show sister kinship, include *P.* sp1-Azores discovered on the Island of Sao Miguel in the Azores with respect to the populations of *P. brevipedunculatus* from the islands of Faial and Pico in the same archipelago. With regard to the difference between the two island-populations from Pico and Faial, the GMYC method indicates 4 species but only 3 of them with high support (0.86-0.93) (**Figure 4.1.2**). The other methods indicate a single species. It is worth mentioning here that the populations from the latter two islands share some haplotypes, which is not surprising since both islands formed a single composite-island during the last glaciations. The Canary Island species would form a monophyletic group except for the taxa from Sao Miguel (Azores), which appears embedded within the group, and the placement of *P.* sp2-Canaries from Gran Canaria, which appears nested within the above mentioned Mediterranean clade.

Discussion

The analysis of 410 specimens of *Pseudoniphagus* revealed the existence of a larger biodiversity in this lineage of subterranean amphipods than previously known, since we detected 17 new species corroborated by both morphology and genetic diversity of *cox1* sequences. Interestingly, the species *P.* sp1-Basque and *P.* sp1-Balearics were only recognized as morphological species because previously molecular analyses have recognized them as highly supported MOTUs. A closer look to the specimens allowed the discovery of diagnostic traits that supported their taxonomic status based on morphology. We also found a few additional putative species (MOTUs) at single collecting sites but just supported by DNA data alone (e.g. Yeso III, La Zarza, Na Barxa, Binicreixent, Charco Verde, Charco de los Chocos, etc). Most sites show a set of unique haplotypes that even differ from those found in localities placed nearby, suggesting that groundwaters are evolutionary traps enabling only limited dispersal in *Pseudoniphargus*. In a detailed inspection we found from close geographic sites with low genetic distances to close localities with high divergences suggesting that this pattern has arisen multiple times and at different ages during the diversification of the genus, assuming that substitution rate in *cox1* is constant. The geographic pattern discovered in *Pseudoniphargus* corroborates the narrow and disjunct distribution of species/morphotypes found in most stigobiont species studied thus far, implying extremely limited dispersal abilities (Holsinger 1994; Notenboom 1988, 1991; Bauzà-Ribot et al. 2011; 2012; Foulquier et al. 2008; Lefébure et al. 2006; 2007; Hunter et al. 2008, Cánovas et al. 2016) An opposite geographic pattern is observed in Macaronesia where localities on islands separated by hundreds of kilometers harbor closely related species with low genetic distances between them (up to 12% only). *P.* sp2-Canaries revealed another intriguing pattern since despite being located on Gran Canaria Island,

it is closely related to species from Morocco, Andalusia and the Balearic Islands instead of to *P*. sp1-Canaries another species from the same island.
This allopatric pattern also hints that, far for being evolutionary dead-ends (Prendini et al. 2010), underground habitats such as caves and wells acted as reservoirs of relict lineages for millions of years in *Pseudoniphargus*. Once adapted to the subterranean environment, they survived with limited morphological change because environmental conditions there are extremely stable. We detected the existence of at least four possible cryptic species complexes: those of *P. gorbeanus* and *P. elongatus* on the northern clade; *P. brevipedunculatus* through the islands of Sao Miguel, Faial and Pico in the Azores; and *P. mercadali* through Menorca and Mallorca (Balearics). The high level of differentiation found within both *P. gorbeanus* and *P. brevipedunculatus* might be partially due to sampling bias related to the comparatively high number of samples from these two taxa (Fujisawaand and Barraclogh, 2013; Zhang et al., 2013). However, other species densely sampled such as *P. multidens* and *P. portosanti* did not show such over-splitting. An alternative hypothesis is that they are true species complexes with members only diagnosable at the genetic level due to strong and constant environmental constraints do not fasten selective pressure on the already strongly modified and adapted morphologies. An opposite pattern to crypticim was discovered in two different monophyletic groups *P. guernicae* / P. sp1-Basque and *P. carpalis* / *P. grandimanus* since they attained diagnostic morphological characters despite showing low genetic divergence, 3.8% and 5.3%, respectively. There are other species showing at least some variance among molecular data, such as *P*. sp1-Andalusia (Yeso III and La Zarza), and *P. cupicola* (Charco Verde and Charco de los Chochos), but in these cases the support as putative species is lower than in the preceding cases. The low divergence found between those localities is more congruent with the occurrence of highly structured populations without gene flow between them.

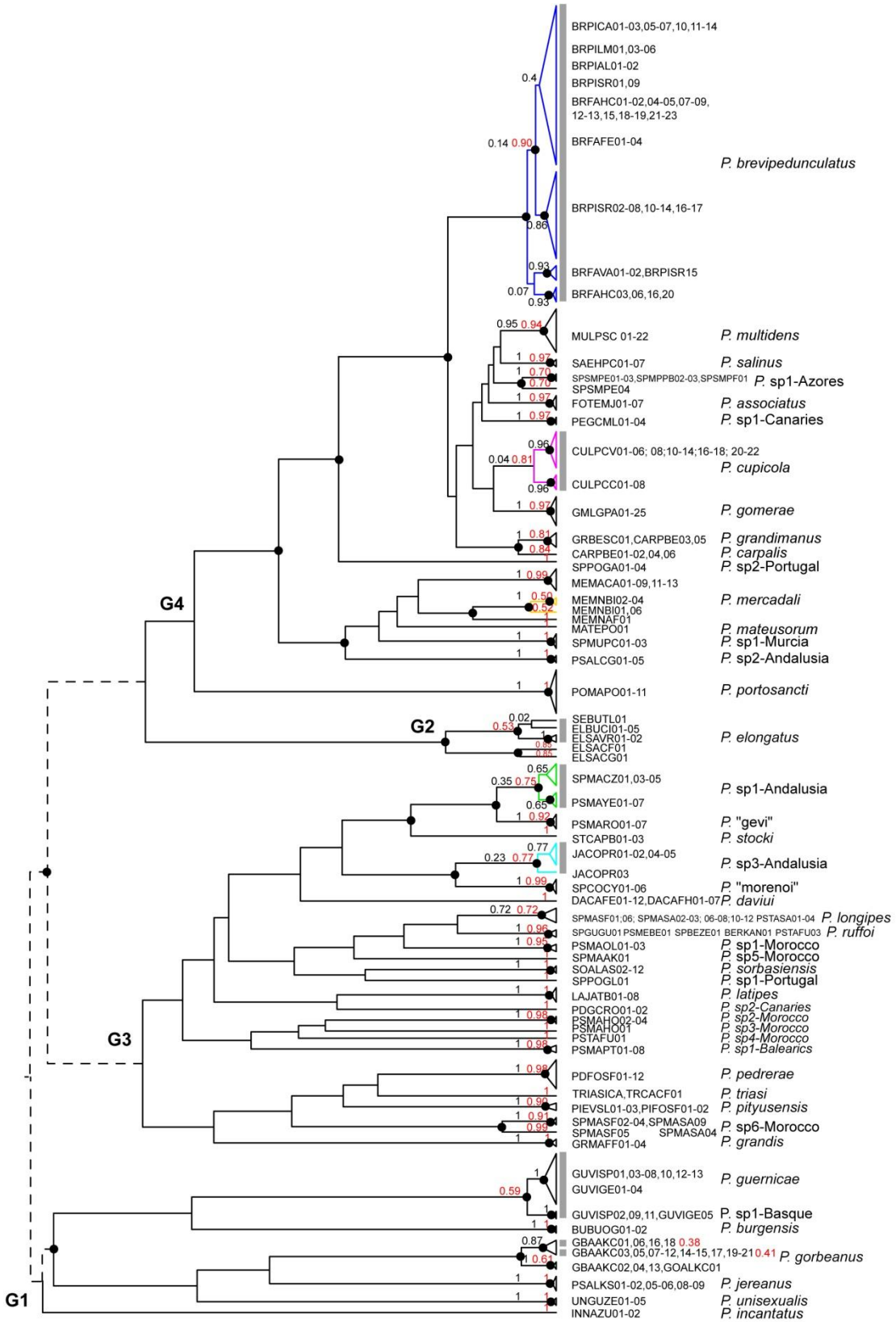


Figure 4.1.2 Summary of 3 species delimitation methods (GMYC, PTP and TCS) plotted on the bayesian phylogenetic tree used to estimate GMYC MOTUs. Each tree terminal (branch or triangle) summarizes a GMYC MOTU with triangles merging the different haplotypes corresponding to the same MOTU. Vertical grey bars indicates PTP MOTUs that differ from GMYC results. Branches with different colours indicates GMYC MOTUs recognized as single MOTUs for TCS algorithm. Black numbers on nodes

indicated support values for GMYC, and red ones support for PTP delimitation. Black dots on nodes indicated a Bayeasian credibility support value above 0.95. Basal branches with dotted lines were trimmed in length to fit the figure in a single page.

There is a large agreement among morphological species and the MOTUs delimited by ABDG and TCS, but not so much with those delimited by GMYC and PTP. However, the correspondence between morphology and DNA increases for GMYC, and is nearly identical for PTP if we take into account the values of GMYC and PTP support. For instance, if those monophyletic sister MOTUs with a support below 0.90 are merged until their added GMYC support is equal or above 0.90, then the number of highly supported MOTUs decreased from 63 to 58. Doing so, we merged two MOTUs in *P. brevipedunculatus* reducing from four to three MOTUs, *P. elongatus* from five to three, and lumped *P.* sp1-Andalusia and *P.* sp3-Andalusia in a single MOTU. This is an arbitrary but stringent cut-off that avoids over-splitting, and hence is a conservative approach. We applied the same procedure for PTP, and the number of MOTUs dropped from 55 to 48 reducing the two MOTUs found in *P.* sp1-Azores to a single one, *P. mercadali* from four to three, and *P. gorbeanus* and *P. elongatus* both from three to one. Surprisingly, this approach also merged the morphological species *P. grandimanus* and *P. carpalis* composed of two independent MOTUs in a single one. Comparing the two tree-based methods GMYC and PTP, the latter is the more conservative, which is congruent with previous studies that indicate GMYC tends toover-split taxa (Zhang et al 2013; Kekkonen et al., 2015; Esselstyn et al., 2012; Talavera et al., 2013).

Whereas PTP allows for higher intra-specific divergences, GMYC might be better in detecting early speciation events (i.e. lower intra-specific divergences). This fact is reflected in the boxplot comparing the intra-specific divergences between models **(Figure 4.1.1**). For instance, *P.guernicae* is detected as two MOTUs by GMYC with a low intraspecific p-distance of 0.8% and 0.1%, and also a low interspecific distance between these two MOTUs of 3.5 %. On the other hand, the PTP model retrieved *P. guernicae* as a single MOTU with a larger intra-specific p distance of 1.8%. Interestingly, if we considered the genetic distance between the monophyletic sister-groups in the Bayesian tree (**Figure 4.1.2**), then PTP showed a broader range of p-distance values despite delimiting more MOTUs than the GMYC and TCS approaches, suggesting that PTP is sensible to larger instraspecific variation reducing the number of speciation events. Notwithstanding that, this implies that recent speciation events will not be noticed using the PTP algorithm.
The disagreement found between the different DNA methods seems to be related to the number of mutations that are required to occur after species diversification to detect such a speciation event. Generally, GMYC detects very early speciation events such as in *P.* sp1-Andalusia, MOTUs Yeso III and La Zarza, that are not detected by other methods since other delimitation models consider/detect speciation events after more mutations are fixed in the derived lineages. The other DNA-based methods treat them only as geographic structure between populations, and only are recognized as true species when morphological diagnostic characters are fixed, which is an extremely slow process in stygobiont species since there are no environmental changes enforcing selective pressure towards morphological change.

The comparison of methods of species delimitation presented above illustrates on the potential sources of conflict that could arise, and the importance of conducting several and not relying on only a single one. Some additional potentially problematic factors to be taken into account when applying any of these methods are the time scale and the evolutionary rate, meaning that more recent clades might not be sufficiently divergent as to accurately delimit species boundaries. This is supported by the fact that the different delimitation criteria rendered more consistent results when applied on taxa from the older lineages, which already account with a long evolutionary history behind. By conducting a broad-scale species delimitation process using both morphology and DNA we discovered that, based on the results from the molecular data, reevaluation of the morphological characters helped in detecting morphological differences and thus enabling species description.

Chapter 4.2 Taxonomical description of new species

Taxonomy

Order AMPHIPODA Latreille, 1816
Family Pseudoniphargidae Karaman, 1993
Genus *Pseudoniphargus* Chevreux, 1901
***Pseudoniphargus morenoi* sp. nov.**
Figs 4.2.1–4.2.6

Material examined. "Cueva del Yeso" (Baena; Córdoba; Spain). UTM coordinates (Datum ED50): 30S 380474, 4170957; 288 m a.s.l. HOLOTYPE: male 7.7 mm preserved in 70% ethanol vial [**BMNH-XXXX**]; collected by Agustín Castro, 5 August 2001. PARATYPES: 16 males, all in 70% ethanol vial, same data as holotype [**BMNH-XXXX**]. 20 females, all in 70% ethanol vial, same data as holotype [**BMNH-XXXX**]. 11 specimens, both sexes, in single 70% ethanol vial, preserved at IMEDEA; collected by Manuel Baena, 11 August 2001.

Table 4.2.1. Species of *Pseudoniphargus* with strongly elongated male U3 exopod (>10 times longer than broad) and protopod (4 or more times longer than broad).

Species	**Distribution**	**Male U3** [length: width] ratio		**U1** basofacial robust seta	**Male pleosomite II** with dentate dorsal spur	**Male P5-P7** posterodistal lobe	**Male G1** carpus to propodus length	**Epimeral plates** armature formula	**Male G2 propodus** [posterior margin: palm margin] length ratio
		expopod	**protopod**						
P. morenoi sp. nov.	S Spain	19.5	7	+	-	not developed	C>P	0-(1 or 2)-(1 or 2)	>1
P. adriaticus S. Karaman, 1955	Mediterranean	19.7-23.2	4-6.2	+/-: variable among populations	+	narrow and strongly overhanging	C=P	1-1-1	>1
P. affinis Notenboom, 1987	S Spain	24	7.5-9.5	+	-	Not developed	C=P	0-2-2	<1
P. africanus Chevreux, 1901	N Algeria	27.5	9.1-10.3	+	-	broad and slightly overhanging	C<P	0-1-1	>1
P. branchiatus Stock, 1980	S Spain	22	3-4	+	-	not developed	?	0-(1 to 3)-(1 to 3)	<1
P. elongatus Stock, 1980	N Spain	19.8	7.1	+	-	broad and slightly overhanging	C=P	0-1-1	>1
P. granadensis Notenboom, 1987	S Spain	18	4-5	+	-	not developed	C>P	0-(2 or 3)-(2 or 3)	=1
P. grandis Notenboom, 1987	S Spain	23	6	+	-	not developed	C=P	0-(2 or 3)-(2 or 3)	<1
P. illustris Notenboom, 1987	S Spain	22.5	4	+	-	not developed	C>>P	0-1-1	>1
P. inconditus Karaman & Ruffo, 1989	Sicily	>15	4	+	+	narrow and strongly overhanging	C>P	2-4-5	>1
P. macrotelsonis Stock, 1980	N Algeria	>15	6>	-	-	broad and slightly overhanging	C=P	0-(1 or 2)-(1 or 2)	>1
P. mercadali Pretus, 1988	Mallorca; Menorca (Balearics)	20	3.2-4.6	+	+	narrow and strongly overhanging	C=P	(0 or 1)-(1 or 2)-(2 or 4)	>1
P. obritus Messouli et al., 2006	Corsica	21.8	8.2	- (male)/ + (female)	-	narrow and strongly overhanging	C<P	1-2-1	>1
P. pedunculatus Sánchez, 1989	Gran Canaria (Canaries)	15.7	7.4	-	-	not developed	C=P	(0 or 1)-(1 or 2)-(1 or 2)	=1
P. sodalis Karaman & Ruffo, 1989	Sicily	20.9	7.1	-	-	not developed	C<P	0-1-2	=1
P. stocki Notenboom, 1987	S Spain	20	6.5	+	-	not developed	C=P	0-(1 or 2)-(1 or 2)	<1
P. vomeratus Notenboom, 1987	S Spain	29	5-5.5	+	-	not developed	C=P	1-1-2	<1

Table 4.2.2. Species of *Pseudoniphargus* with male U3 exopod strongly elongated (more than 10 times as long as broad), but with protopod only moderately elongated (2.1-3.5 times as long as broad).

Species	**Distribution**	**Male U3** [length: width] ratio		**Female U3** [length: width] ratio		**Male pleosomite II** with dentate dorsal spur	**Male P5-P7** posterodistal lobe on basis	**Epimeral plates** armature formula	**Male G1** [carpus to propodus] length ratio	**G1** Palm angle armature
		expopod	**protopod**	**exopod**	**protopod**					
P. gevi sp. nov.	S Spain	22.4	3.3	7.6	1.8	-	not developed	0-2-2	C>P	7
P. associatus Sánchez, 1991	S Spain	16.2	3.1	6.3	1.6	-	not developed	(0 or 1)- (1 or 3)- (1 or 2)	C<P	4
P. callaicus Notenboom, 1987	NW Spain	11-12	2.5-3	11-12	2.5-3	-	Broad and overhanging in both sexes	0-1-2	C<<P	5
P. fragilis Notenboom, 1987	S Spain	16.5	2-2.5	9.6	2.1	-	narrow to broad, but non-overhanging	0-1-2	C<<P	?
P. gibraltaricus Notenboom, 1987	S Spain	14	2.3-3	14	2.3-3	-	broad and slightly overhanging	0- (0 or 1)- (1 or 2)	C>P	?
P. leucatensis Bréhier & Jaume, 2009	S France	11.4	3.5	8.8	1.9	-	narrow and strongly overhanging	0-1-1	C<P	6
P. longicauda Stock, 1988	Tenerife (Canaries)	18-20	2.2	10.3	1.8	-	broad and overhanging in both sexes	1- (3 or 4)- 3	C<P	6
P. longipes Coineau & Boutin, 1996	N Morocco	16	2.2	15.2	2	-	lobe developed and overhanging only on P5	1-4-5	?	?
P. macrurus Stock & Abreu, 1992	Madeira	20	2.2	14.5	2.1	-	narrow and strongly overhanging	(0 to 2)- (2 or 3)- (1 to 4)	C<P	4
P. nevadensis Notenboom, 1987	S Spain	24	3	14	2	-	not developed	0- (1 or 2)- 2	C<P	4
P. planasiae Messouli et al., 2006	Pianosa Is. (Tuscan Archipelago)	13.5	2.6	10	2	+	broad and strongly overhanging	0-4-(3 or 4)	C<P	6

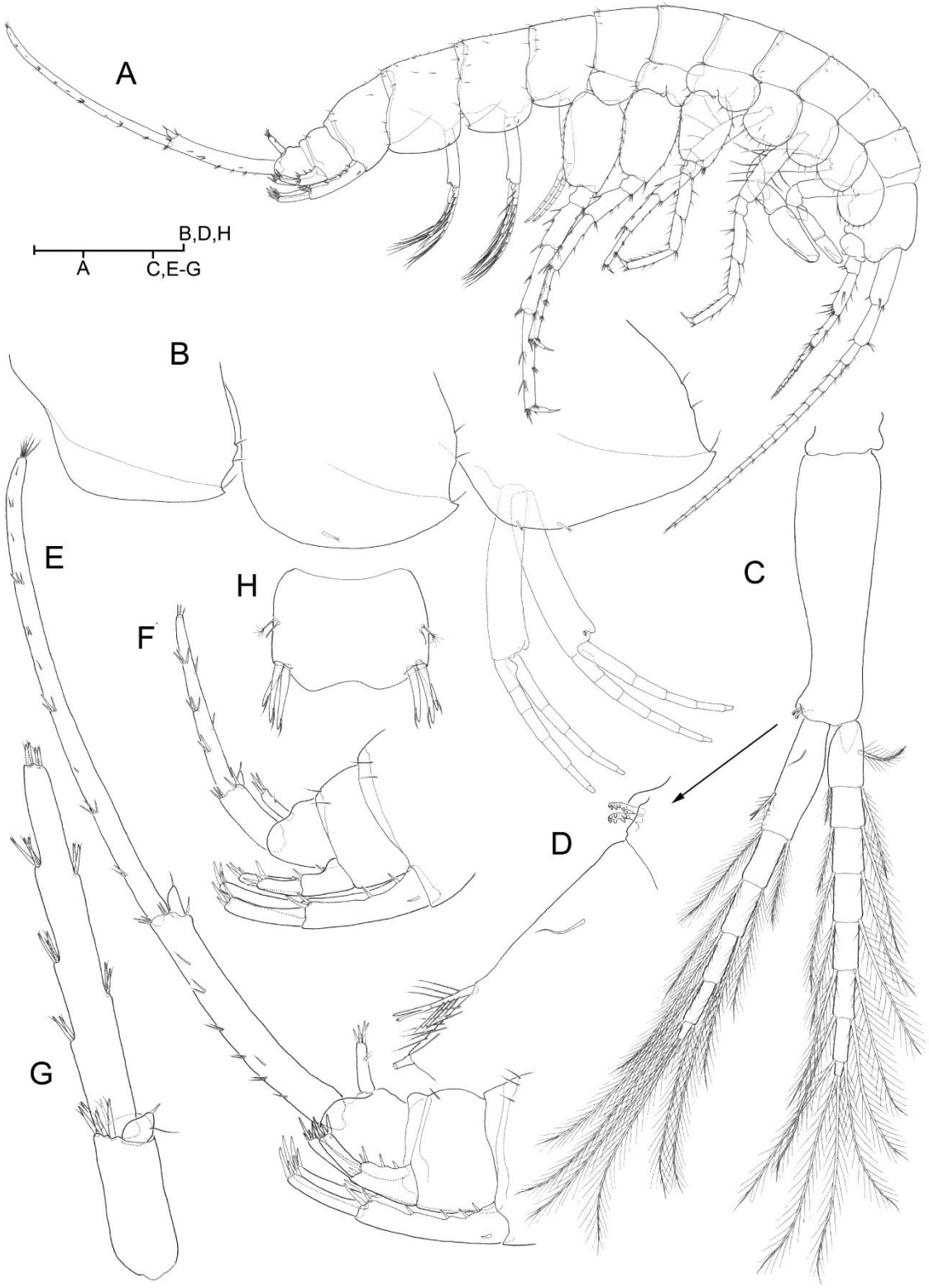


Figure 4.2.1. *Pseudoniphargus morenoi* sp. nov., male paratype (A-D); female paratype (E-G) **A** habitus **B** epimeral plates, lateral view **C** left PL1, anterior **D** detail of distal angle of protopod of PL1 and of armature on medial margin of proximal article of endopod, anterior **E** male urosome, lateral **F** female urosome, lateral **G** female U3, posterior (= dorsal) **H** female telson. [Scale bar: 0.5 mm (A-B, E-F); 0.25 mm (C, G-H); 0.125 mm (D)].

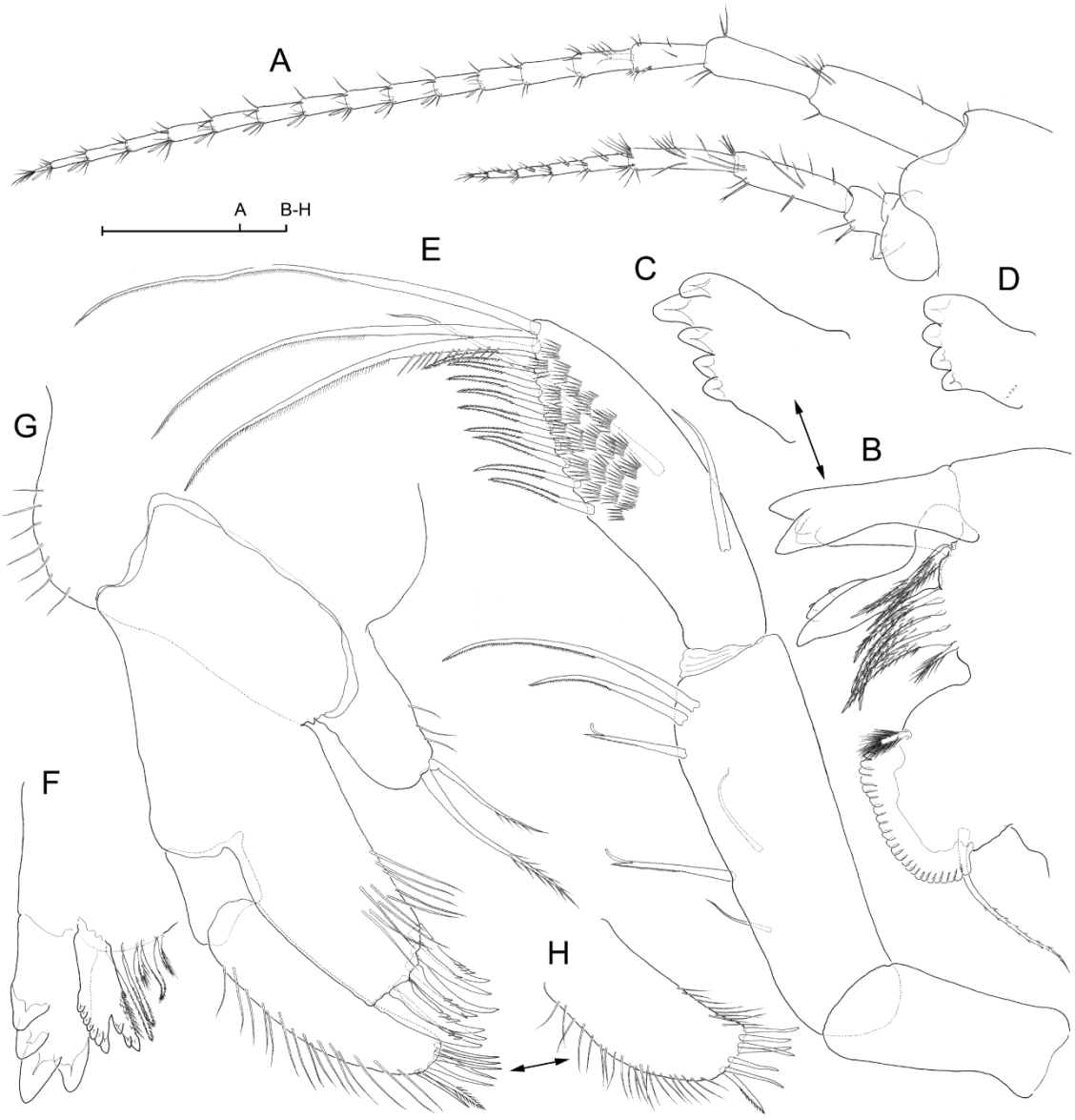


Figure 4.2.2. *Pseudoniphargus morenoi* sp. nov., male paratype **A** head, left A1 and left A2, lateral **B**, left mandible **C** inset of incisor of latter **D** inset of lacinia **E** left mandibular palp **F** right mandible **G** maxillule **H** detail of distal segment of endopod (= palp) of latter. [Scale bar: 0.05 mm (A); 0.1 mm (B-H)].

Figure 4.2.3 *Pseudoniphargus morenoi* sp. nov., male paratype **A** maxilla **B** left maxilliped, anterior (= dorsal) **C** detail of armature of basal endite (= inner plate) of latter **D** inset of distal part of ischial endite (= outer plate) **E** inset of distal segments of maxillipedal palp (carpus-dactylus). [Scale bar: 0.05 mm (A); 0.1 mm (B-E)].

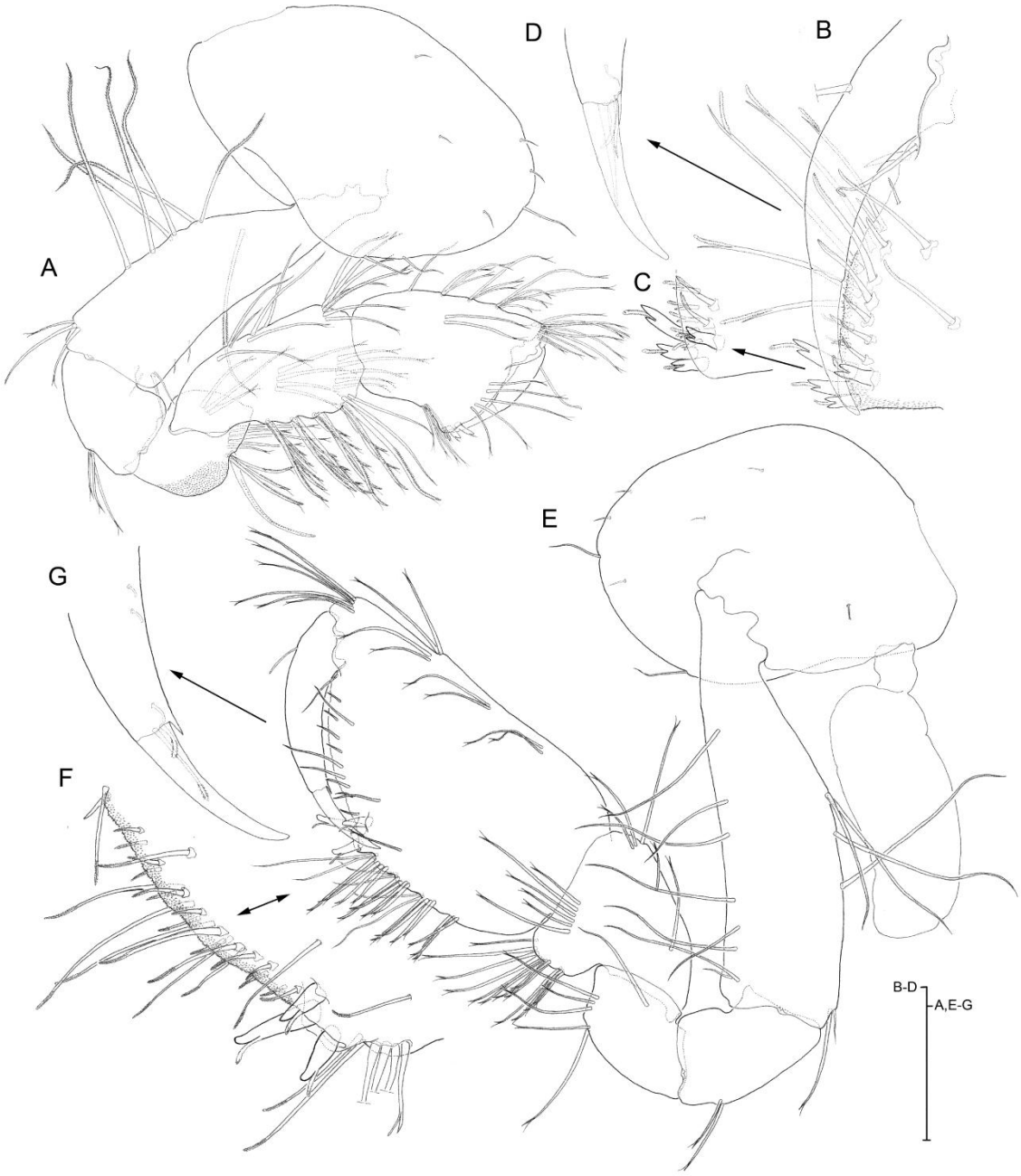


Figure 4.2.4 *Pseudoniphargus morenoi* sp. nov., male paratype **A** right G1 with palm margin armature partially omitted, lateral **B** detail of palm margin armature of latter, medial **C** detail of armature on palm angle, medial **D** detail of armature of nail, medial **E** right G2 with armature of palm margin partially omitted, medial **F** detail of armature of palm margin of latter, medial **G** detail of armature of nail, medial. [Scale bar: 0.25 mm (A, E); 0.1 mm (B, D); 0.125 mm (F, G)].

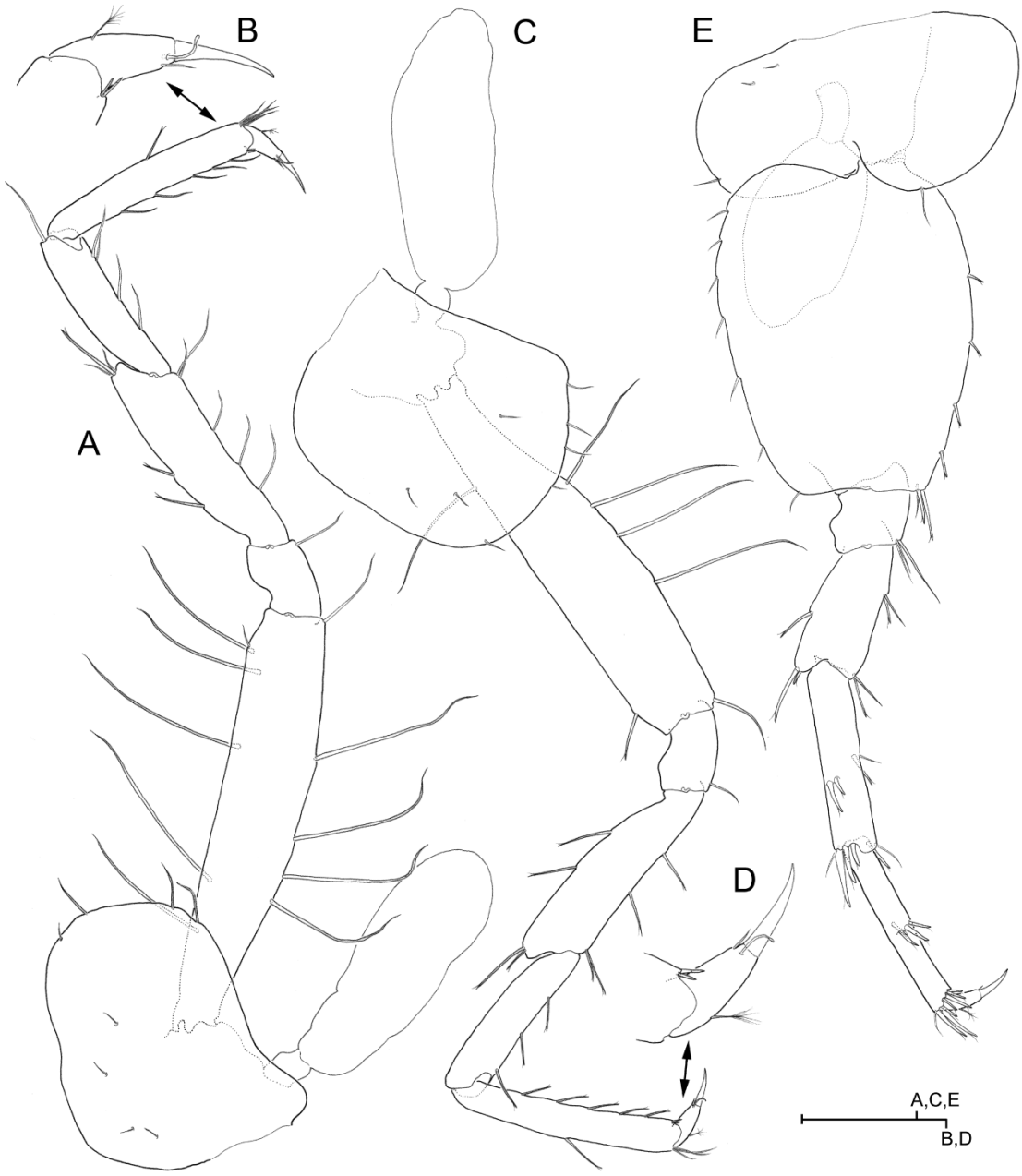


Figure 4.2.5. *Pseudoniphargus morenoi* sp. nov., male paratype **A** right P3, lateral **B** detail of nail of latter **C** left P4, lateral **D** detail of nail of latter **E** right P5, lateral. [Scale bar: 0.25 mm (A, C, E); 0.125 mm (B, D)].

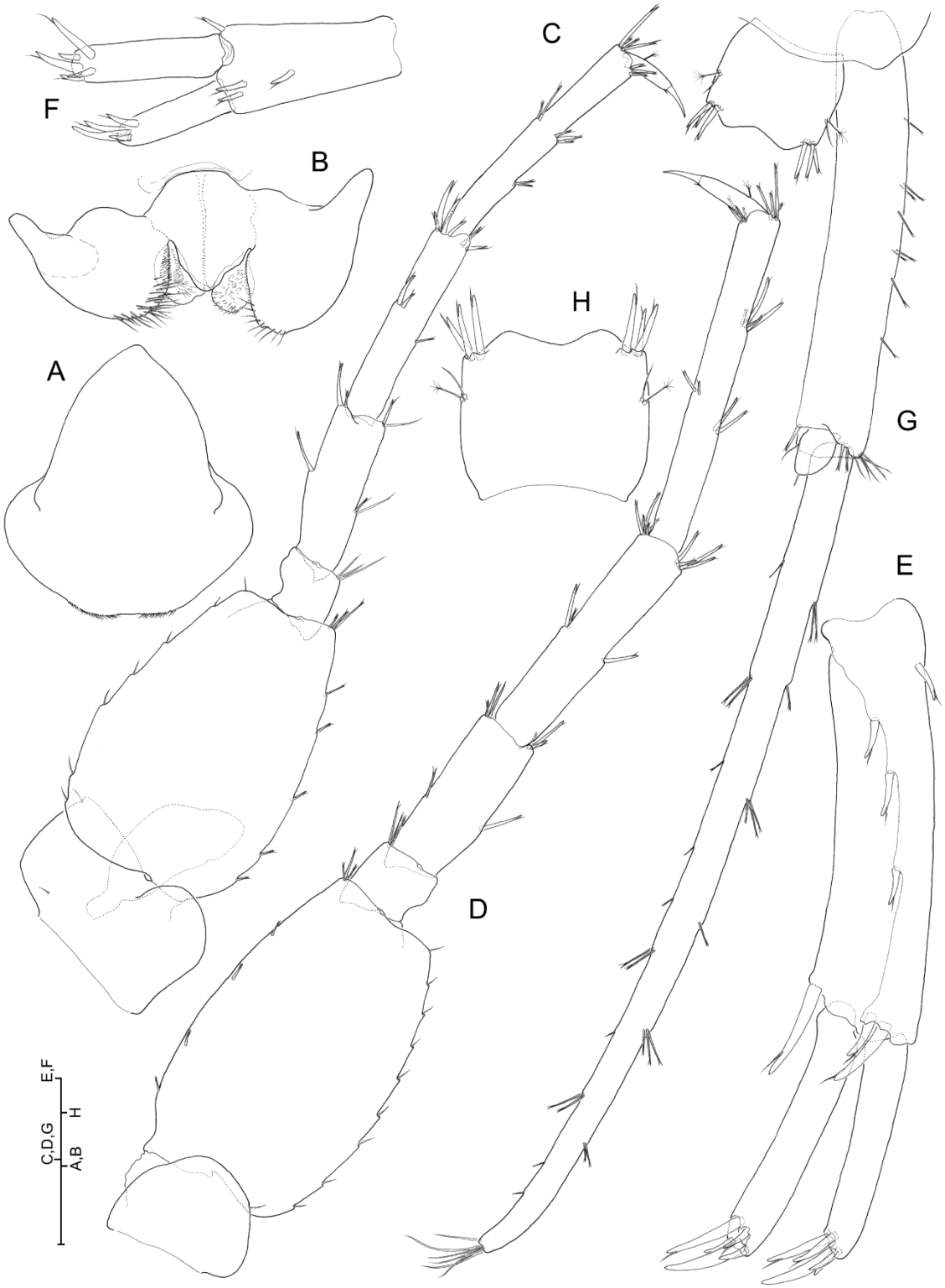


Figure 4.2.6. *Pseudoniphargus morenoi* sp. nov., male paratype **A** labrum, anterior view **B** paragnaths, anterior (= dorsal) **C** left P6, lateral **D** right P7, lateral **E** right U1, posterolateral **F** right U2, posterior **G** right U3 and telson, posterior (= dorsal) **H**, detail of telson. [Scale bar: 0.1 mm (A, B); 0.25 mm (C-H)].

Diagnosis. Male U3 protopod very elongate (up to 7 times longer than broad) and exopod extremely elongate (up to 19.5 times longer than broad) and upcurved. Male pleosome smooth, lacking dorsal spur on pleosomite II. Male G1 carpus slightly longer than propodus. Male G2 propodus posterior margin longer than corresponding palm margin. Posterodistal lobe on basis of P5-P7 not developed. Male U1 with basofacial robust seta present on protopod. Posterodistal angle of epimeral plates not strongly produced into sharp, pointed process. Robust setae on basal endite (= outer lobe) of maxillule coarsely denticulate. A1 not extremely elongated, shorter than body length. P5 shorter than P3-P4. Female telson 1.2 times broader than long, with distal robust setae shorter than telson itself. U1 and U2 devoid of lateral armature on rami.

Etymology. Species name after the Cordovan speleologist Antonio Moreno de la Rosa, in recognition of his contribution to the knowledge of the subterranean environment of the province.

Male. *Body* unpigmented, eyeless, up to 8.7 mm long (**Figure 4.2.1A**). *Head* lacking rostrum, with broadly rounded lateral lobe (**Figure 4.2.2A**). *Antennule* and *antenna* (**Figure 4.2.2A**) ordinary. Aesthetascs on articles of main flagellum of antennule each shorter than corresponding article. Relative length of segments 4 and 5 of peduncle of antenna as 1: 0.87; setae on these segments longer than width of corresponding segment. *Labrum* and *paragnaths* ordinary (**Figure 4.2.6A, B**).

Left mandible (**Figure 4.2.2B**) with 5-dentate incisor (**Figure 4.2.2C**); lacinia 4-dentate with row of tiny denticles proximally on anterior surface (**Figure 4.2.2D**); spine row not continuous, comprising 2+5 pappose elements separated by diastem, most proximal reduced, distalmost 2 placed beside each other and not at same plane as rest of members of spine row. Molar process columnar, triturative, provided with ordinary molar seta and with additional short frayed seta placed adjacent to grinding surface on distal margin. Palp (**Figure 4.2.2E**) relative length of segments as 0.48: 0.94: 1; second segment provided with up to 6 unequal setae on medial surface; distal segment with distomedial patch of spinules and with 3 long E-setae, up to 9 D-setae, 1-2 A-setae and 1-2 B-setae (*sensu* Stock 1974); one of E-setae with distinct proximal row of long setules on one side; ornamentation of rest of setae as figured.

Right mandible (**Figure 4.2.2F**) with bifid lacinia 7+3-dentate; one of margins covered with short spinules; spine row comprising 5 unequal elements as figured; rest of mandible as left counterpart.

Maxillule (**Figure 4.2.2G**) coxal endite (= inner lobe) with two unequal pinnate setae on tip. Basal endite (= outer lobe) with 7 robust setae, 5 of them provided with stout denticles, other two smooth. Endopod (= palp) with 6 smooth and 1 pinnate setae on distal segment (**Figure 4.2.2H**).

Maxilla (**Figure 4.2.3A**) normal, with coxal endite (= inner lobe) lacking both medial and oblique facial row of setae, and with basal endite (= outer lobe) displaying two separate groups of setae on distal margin; ornamentation of setae as figured.

Maxilliped (**Figure 4.2.3B**) basal endite (= inner plate) with 3 short, robust simple setae, 3 pinnate and 1 simple seta distally (**Figure 4.2.3C**). Ischial endite (= outer plate) with up to 17 blade-like, robust pectinate setae along distal and distomedial margin; setae progressively longer and more slender towards distal (**Figure 4.2.3D**). Carpus slender, about 3.2 times as long as broad (**Figure 4.2.3E**). Distolateral surface of propodus and dactulus covered with short denticles. Unguis about as long as dactylus.

Gnathopod I (**Figure 4.2.4A**) carpus slightly longer than propodus. Propodus 1.4 times as long as broad, with maximum width attained at protruding palm angle, placed at 46% of maximum (=anterior margin) length of segment. Palm angle armature comprising 1+3 unequal bifid flagellate robust setae (**Figure 4.2.4B, C**). Palm margin

convex, with armature comprising series of long flagellate stiff setae plus short flagellate robust setae distributed along margin as in **Figure 4.2.4B, C**. Dactylus: unguis length ratio 1.6; dactylus provided with three distal and one tiny subdistal seta as in **Figure 4.2.4D**.

Gnathopod II (**Figure 4E**) carpus short, about 55 % length of propodus. Propodus 1.8 times as long as broad, with sub-parallel anterior and posterior margins; palm angle placed at 56 % of maximum (= anterior margin) length of segment, marked with 3 unequal flagellate robust setae. Palm margin oblique, convex, with microtuberculate integument, armed as in **Figure 4.2.4F**. Dactylus: unguis length ratio 2.0. Dactylus with pointed process on postero-distal angle, and with 5 short unequal setae placed as in **Figure 4.2. 4G**.

Pereiopod III-IV (**Figure 4.2.5A, C**) subsimilar except for outline of coxal plates, that of P4 with shallowly excavated posterior margin. Unguis shorter than dactylus in both limbs (both with dactylus: unguis length ratio as 1.2; *v*. **Figure 4.2.5B, D**). Armature on segments of both limbs as figured.

Pereiopod V (**Figure 4.2.5E**) clearly shorter than P4. Basis broad, about 1.3 times as long as broad, with convex anterior and posterior margins; postero-distal angle not produced into overhanging lobe. Dactylus: unguis length ratio 1.1.

Pereiopod VI (**Figure 4.2.6C**) basis broad with convex anterior and posterior margins, although more slender than basis of preceding limb (1.5 times as long as broad *vs*. 1.3 times in P5); postero-distal angle not produced into overhanging lobe. Dactylus: unguis length ratio 1.8.

Pereiopod VII (**Figure 4.2.6D**) basis more slender than those of P5 and P6, 1.7 times as long as broad, with postero-distal angle not produced into overhanging lobe. Dactylus: unguis length ratio 2.0.

Epimeral plates (**Figure 4.2.1B**) with 0-1(2)-2(1) submarginal small robust setae on distal (= ventral) margin; postero-ventral angle of each plate not strongly produced; posterior margin of plates broadly convex.

Pleopods with protopod conspicuously constricted subdistally (**Figure 4.2.1C**), each provided with two retinacles partially covered by rounded postermedial outgrowth of segment (**Figure 4.2.1D**). Rami multi-articulated, proximal article of endopod longer than rest and with proximo-medial seta transformed into robust flagellate element **(Figure 4.2.1D**).

Uropod I (**Figure 4.2.6E**) protopod with basofacial robust seta and row of 3-4 robust setae along posterolateral ridge; medial margin unarmed; 2 unequal robust setae on distolateral corner, and longer robust seta on distomedial corner of segment not reaching midway of endopod. Rami devoid of marginal armature except in two specimens, which displayed single robust seta on exopodi; each ramus with 5 unequal robust setae on tip.

Uropod II (**Figure 4.2.6F**) protopod with 1-2 robust setae on posterolateral ridge, 2 robust setae (exceptionally 3 in one specimen) on distolateral corner and more slender robust seta on distomedial corner. Rami devoid of marginal armature; 4 and 5 robust setae on tip of exopod and endopod, respectively.

Uropod III (**Figure 4.2.1E**) with elongated protopod, up to 6.4 times as long as broad in larger specimens; lateral margin provided with row of slender robust setae; medial margin unarmed. Exopod extremely elongated and upcurved, up to 20.8 times as long as broad in larger males, marginal armature of clusters of slender robust setae and cluster of short simple setae on tip. Endopod with reduced simple seta on tip.

Telson (**Figure 4.2.6H**) slightly (1.1 times) broader than long, with shallow to markedly excavated (in larger specimens) distal margin; all specimens with 3+3 robust

setae implanted subdistally except three specimens with 4+5, 3+2 and 2+2 robust setae, respectively; tips of lobes unarmed.

Brooding female. Body up to 7.9 mm long. As male except for U3, which is considerably shorter (**Figure 4.2.1F, G**): Protopod slightly elongated, 2.5 to 2.9 times as long as broad, and exopod 7.8 to 11.2 times as long as broad. Telson comparatively broader (about 1.2 times broader than long; **Figure 4.2.1H**) than in male; distal excavation varying as in male; all specimens with 3+3 robust setae except 2 specimens displaying 2+2, and 1 with 4+4 arrangement.

Remarks. Species of *Pseudoniphargus* are described mainly based on a set of very simple morphological features that appear in a species-specific diagnostic combination of states. Irrespective of their taxonomic value, character states shared among species might be non-homologous but in turn be the result of parallel evolution or less probable state reversal (Notenboom 1988). Out of the 69 species of *Pseudoniphargus* formally described thus far (Jaume et al. 2016), three are know only from the female. Out of the remaining 66, only 16 share with the new species from Baena the display of a extremely sexually-dimorphic U3 where the exopod appears strongly elongated in the male (>10 times longer than its maximum width), and where elongation affects also strongly the protopod, which is 4 or more times as long as broad (see Table 4.2.1). Members of this cluster of species can be readily told apart from each other based on the differential expression of the following features: (1) presence of basofacial robust seta on protopod of U1; (2) presence of a dentate dorsal spur on pleosomite II in the male; (3) presence and degree of development of a posterodistal lobe on basis of male P5-P7; (4) relative length of carpus-to-propodus in male G1; (5) outline of male G2 propodus (determined by the relative length of posterior margin to palm margin); (5) by the armature formula of the distal margin of epimeral plates; and (6) by the much shorter male U3 protopod (up to 4 times as long as broad, *vs*. 7x in the new species), among other features (*v*. **Table 4.2.1**). Based on these features, *Pseudoniphargus morenoi* sp. nov. is phenetically closest to *P. illustris* Notenboom, 1987, a species known only from the hyporheic habitat at river Guadalbullón –an affluent of the Guadalquivir–, in Jaén (Notenboom 1987a). Nevertheless, the latter species differs in several striking features such as the sharply pointed, produced posterodistal angle of epimeral plates, the more strongly elongate carpus of male G1, and the finely denticulate condition of robust setae on basal endite (= outer lobe) of maxillule.

As previously stated, three species of *Pseudoniphargus* are known only from the female. They are readily differentiated from the female of *P. morenoi* sp. nov. as follows: *P. duplus* Messouli, Messana and Yacoubi-Khebiza, 2006, from Sicily, differs in the display of an extremely elongated A1, as long as body length (cf. Messouli et al.: Figure 1 and our **Figure 4.2.1**); P5 longer than P3-P4 (*vs*. P5 shorter than preceding limbs in *P. morenoi* sp. nov.; basis of P5-P7 with distinct, slightly overhanging postero-distal lobe; protopod of U3 non-elongated (1.7x; *vs*. 2.7 to 2.9x in the new species); and telson distinctly broader than long (1.9x; *vs*. 1.2x in the new species), among other features (*v*. Messouli et al. 2006).

Pseudoniphargus unispinosus Stock, 1988 from Tenerife (Canary Islands) differs, among other features, in the non-elongated protopod of U3 and the distinctly broader-than-long telson (1.9x), which displays a single terminal robust seta on each lobe (*v*. Stock 1988).

Finally, *Pseudoniphargus italicus* Karaman and Ruffo, 1989, known also only from the female, differs in the display of lateral armature on rami of U1 and U2 (*vs*. rami devoid of lateral armature in the new species) and in the comparatively longer

distal robust setae on telson (as long as telson itself; *vs*. distinctly shorter in the new species; *v*. Karaman and Ruffo 1989).

***Pseudoniphargus gevi* sp. nov.**

Figures 4.2.7–4.2.10

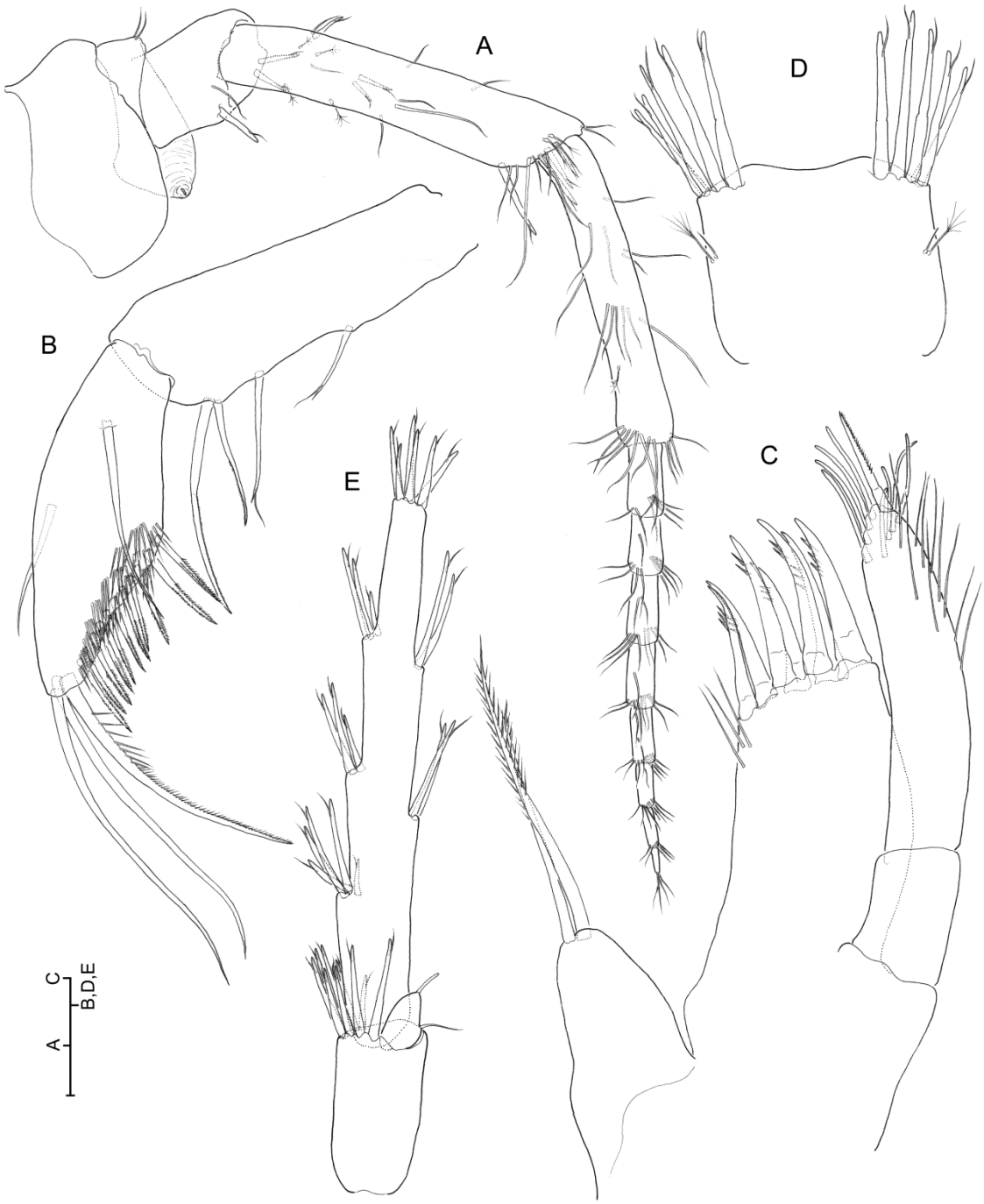


Figure 4.2.7. *Pseudoniphargus gevi* sp. nov., male paratype 7.9 mm (A-D); female paratype 5.7 mm (E) **A** right A2, lateral **B** distal segments of mandibular palp **C** maxillule **D** telson, dorsal **E** right U3, dorsal. [Scale bar: 0.25 mm (A, D, E); 0.125 mm (B); 0.1 mm (C)].

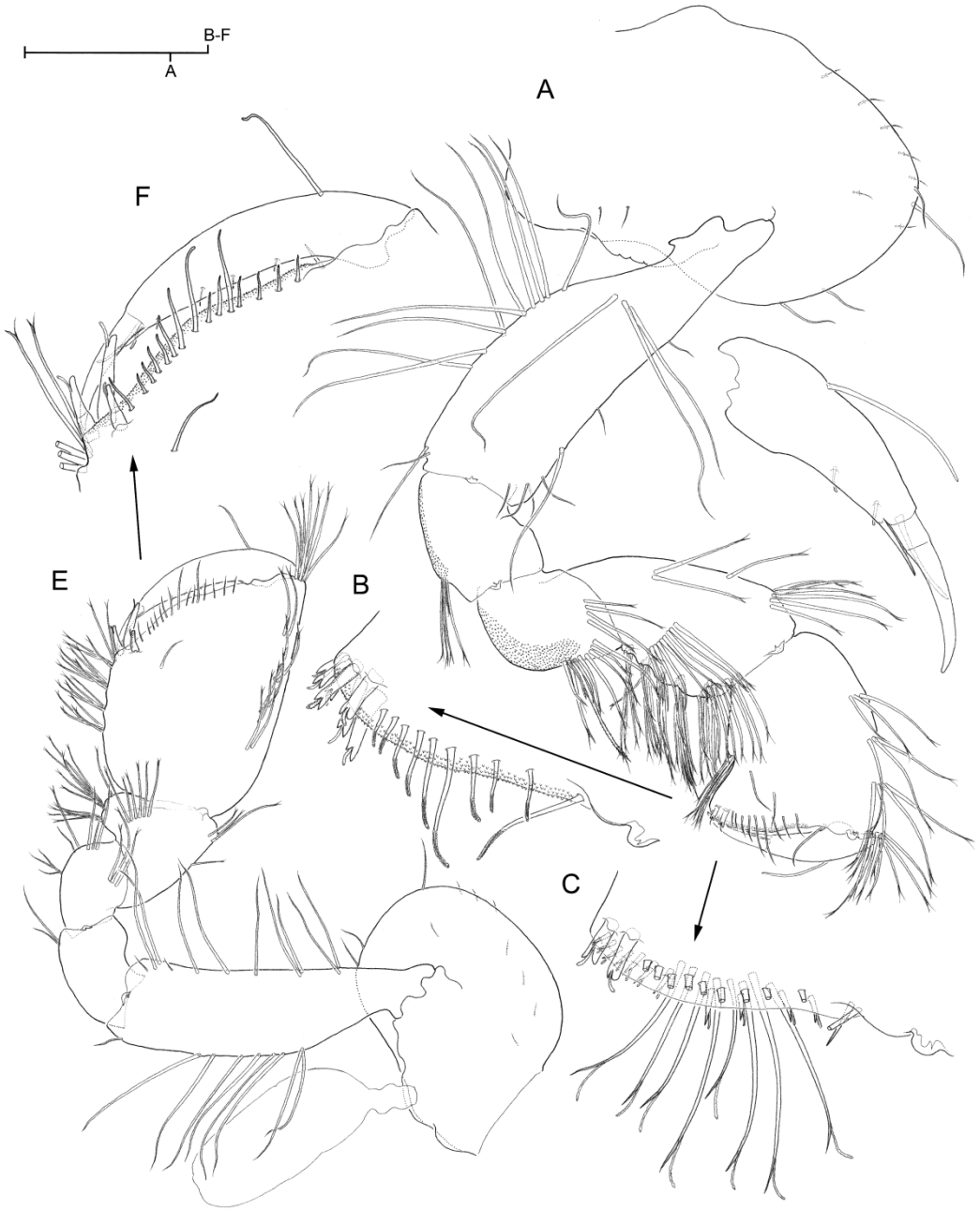


Figure 4.2.8. *Pseudoniphargus gevi* sp. nov., male paratype 7.9 mm **A** left G1, medial **B** detail of armature along palm margin and palm angle, medial (armature on lateral margin omitted) **C** same, with medial armature omitted and showing lateral armature **D** nail, medial **E** right G2, medial **F** detail of palm margin, palm angle and nail of latter, medial. [Scale bar: 0.25 mm (A); 0.125 mm (B-D, F); 0.5 mm (E)].

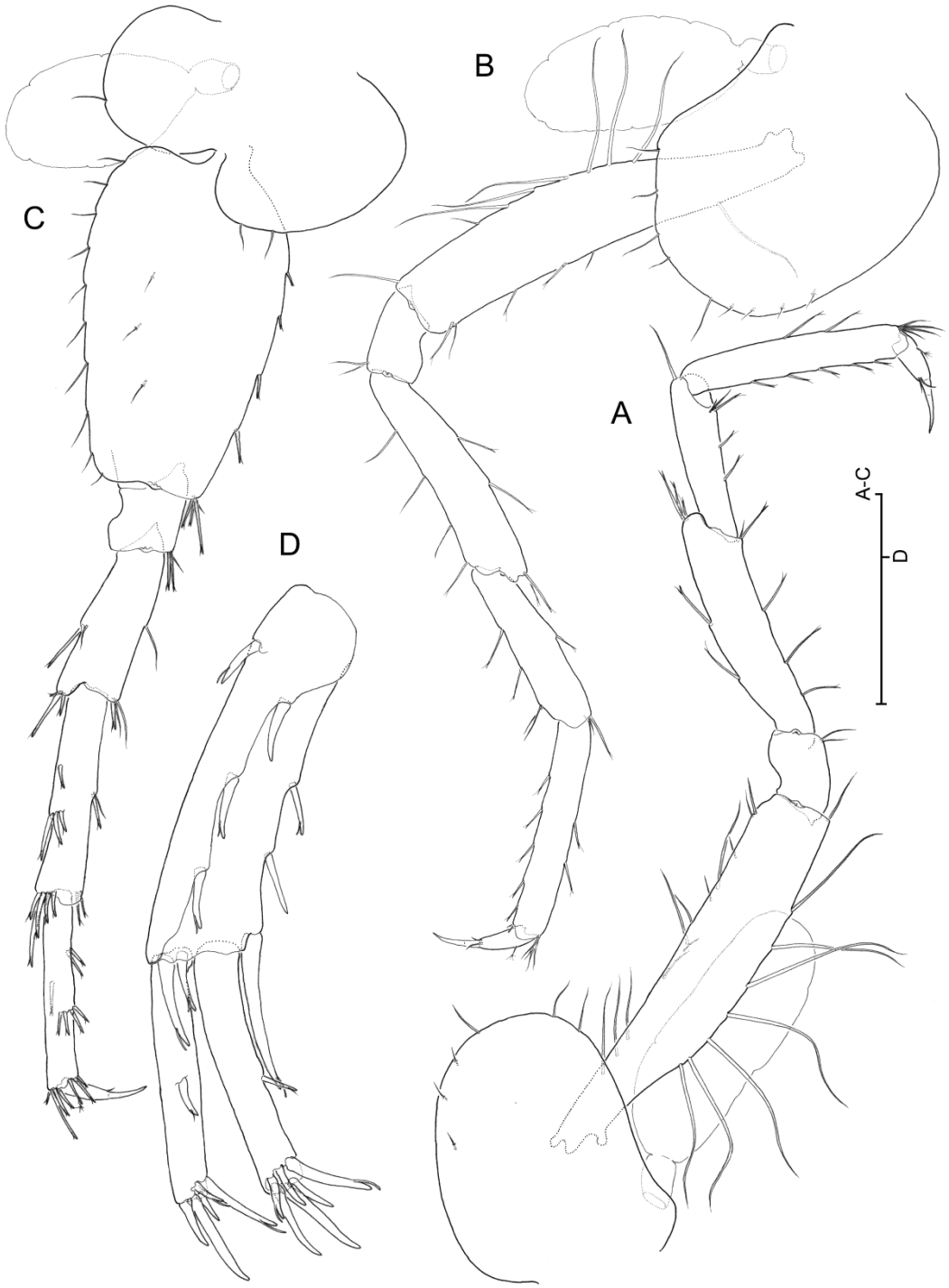


Figure 4.2.9. *Pseudoniphargus gevi* sp. nov., male paratype 7.9 mm **A** right P3, lateral **B** right P4, lateral **C** right P5, lateral **D** left U1, posterolateral. [Scale bar: 0.5 mm (A-C); 0.25 mm (D)].

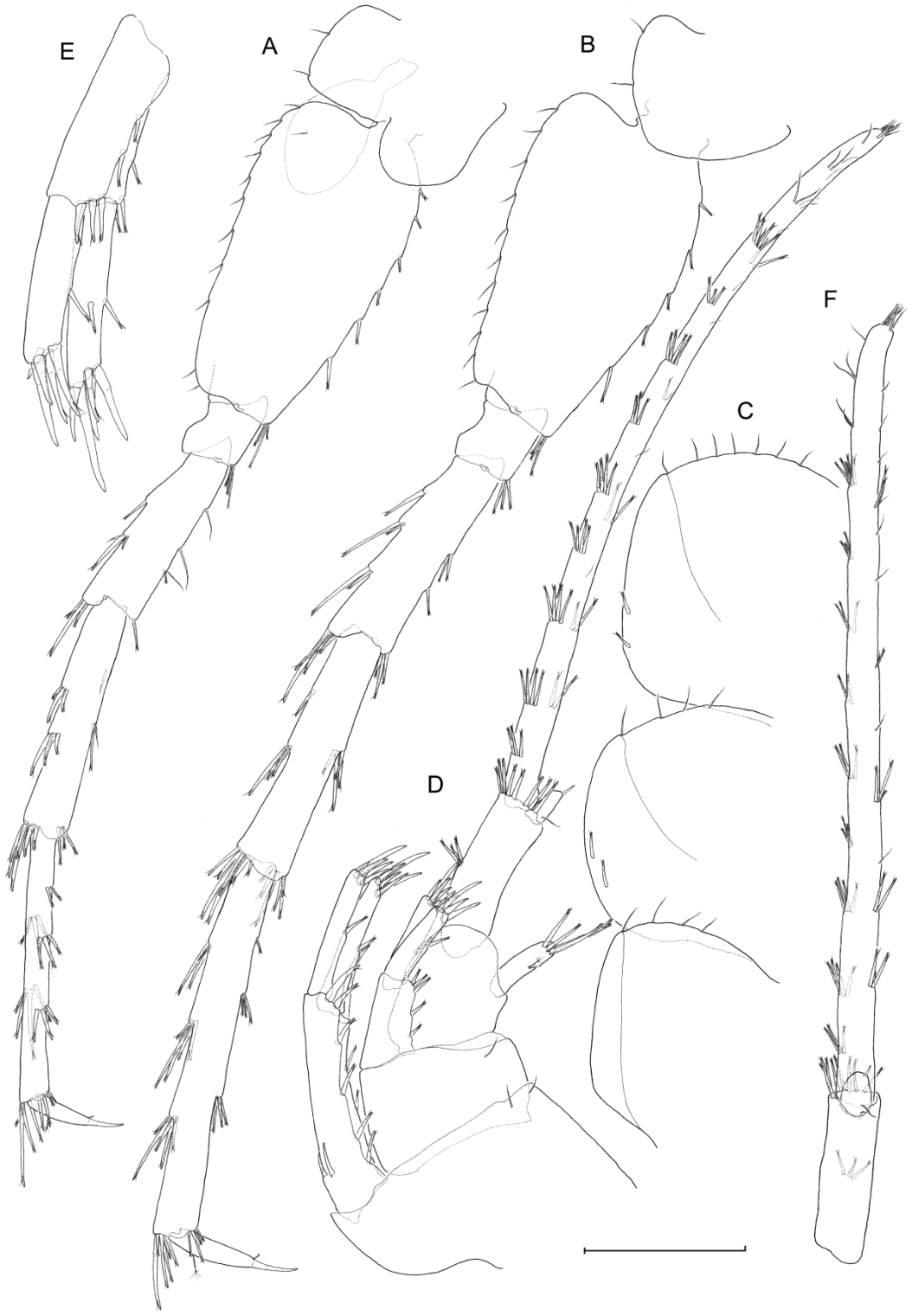


Figure 4.2.10. *Pseudoniphargus gevi* sp. nov., male paratype 7.9 mm. **A** right P6, lateral **B** right P7, lateral **C** detail of epimeral plates **D** detail of urosome, lateral **E** left U2, lateral **F** left U3, dorsal. [Scale bar: 0.5 mm (A-D, F); 0.25 mm (B)].

Material examined. "Complejo Romeral" (Antequera; Málaga; Spain). UTM coordinates (Datum ED50): 30S 366968, 4100145. HOLOTYPE: male 8.9 mm preserved in single 70% ethanol vial [**BMNH-XXXX**]. PARATYPES: five males between 8.0 and 6.3 mm long, and five females between 6.0 and 5.7 mm, preserved in single 70% ethanol vial [**BMNH-XXXX**]. Collected by Antonio Pérez Fernández, 23 February 2008.

Diagnosis. Male U3 exopod extremely elongated (up to 22.4x) and upcurved, but with protopod only moderately elongated (3.3x). Female U3 protopod non-elongated (1.8x). Male pleosome smooth, lacking dorsal spur on pleosomite II. Postero-distal lobe not developed on basis of male P5-P7. Male G1 carpus longer than corresponding propodus; palm angle with 7 bifid robust setae. Epimeral plates each with posterodistal angle evenly rounded, not strongly produced into pointed process. U1 and U2 rami poorly armoured, with only 1-2 marginal robust setae per ramus. Telson slightly broader than long, subquadrate, distal margin with 5+5 robust setae.

Etymology. Species name honouring Grupo de Espeleología de Villacarrillo (G.E.V.), whose members collected the specimens on which the present description is based on.

Remarks. This new species, discovered in a cave close to Antequera (Málaga), conforms with other 10 species scattered across the entire geographic range of the genus (*v*. **Table 4.2.2**) a cluster characterised by the display of a strongly sexually dimorphic U3 where the male exopod is strongly elongated (more than 10 times as long as broad), but the protopod is only moderately elongated (2.1 to 3.5 times as long as broad). Species within this group can be differentiated based on presence/absence of: (1) posterodistal lobe on basis of male P5-P7; (2) dentate dorsal spur on male pleosomite II; (3) relative elongation of carpus of male G1; (4) number of bifid robust setae on G1 palm angle; and (5) armature of epimeral plates (*v*. **Table 4.2.2**). Within this group, the new species is phenetically closest to *P. fragilis* Notenboom, 1987, and *P. nevadensis* Notenboom, 1987, two species of southern Spain that nevertheless differ from the new species in the display of a male G1 where the carpus shorter than the propodus (*vs*. carpus longer than propodus in the new species). *Pseudoniphargus nevadensis* is a species found only in hyporheic habitats of the southern slopes of the Sierra Nevada up to 1,420 m a.s.l. (*v*. Notenboom 1987a); it is easily differentiated from the new species based on additional features such as its epimeral plates with strongly produced and pointed posterodistal angles and its male telson longer than broad. *Pseudoniphargus fragilis* is known only from wells and hyporheic habitats at Tolox (Granada), and differs from the new species in its much more strongly armoured (robust setae) rami of U1 and U2 (*v*. Notenboom 1987a: Figure 34f, g), among other features.

As regard the three species of *Pseudoniphargus* known only from the female (*v*. above), *P. duplus* and *P. unispinosus* differ from the new species in the display of a telson much broader than long and much less armoured. Moreover, *P. duplus* displays a fewer number of bifid robust setae on the palm angle of propodus of G1 (4-5; *vs*. 7 in the new species). *Pseudoniphargus italicus*, the third species known only from the female, displays a slightly elongated U3 protopod (2.8x; *vs*. only 1.8x in the female of the new species), aside of displaying also a fewer number of bifid robust setae (4 or 5) on the palm angle of G1.

Chapter 4.3 Mitogenome rearrangement of 3 species of the genus *Pseudoniphargus*

Results and Discussion

Genome organization, gene order and rearrangements

The mitochondrial genomes of *P. daviui* (15,155 bp), *P. gorbeanus* (14,190 bp), and *P. sorbasiensis* (15,462 bp) are circular as found in most metazoans and include the canonical 13 PCGs, 22 tRNAs and two rRNAs (Figure. 4.3.1). Mitogenome sequences of *P. sorbasiensis* and *P. gorbeanus* are deposited in the EMBL database under accession number LN871175 and LN871176, respectively. The sequence coverage was as follows: 73x in *P. daviuvi*, 138x in *P. gorbeanus*, and 270x in *P. sorbasiensis*. The comparison of the mitogenomes of three *Pseudoniphargus* species at both nucleotide and aminoacid levels and taking into account the secondary structure of RNAs allowed a more accurate annotation of gene boundaries and to detect some errors in the published sequence of *P. daviui* (Bauzà-Ribot et al 2009), particularly on the 3' end of *atp8*, *nad4*, *trnQ* and *rrnS* genes. These corrections were reported to European Nucleotide Archive (ENA), and they now are included in the FR872383 entry.

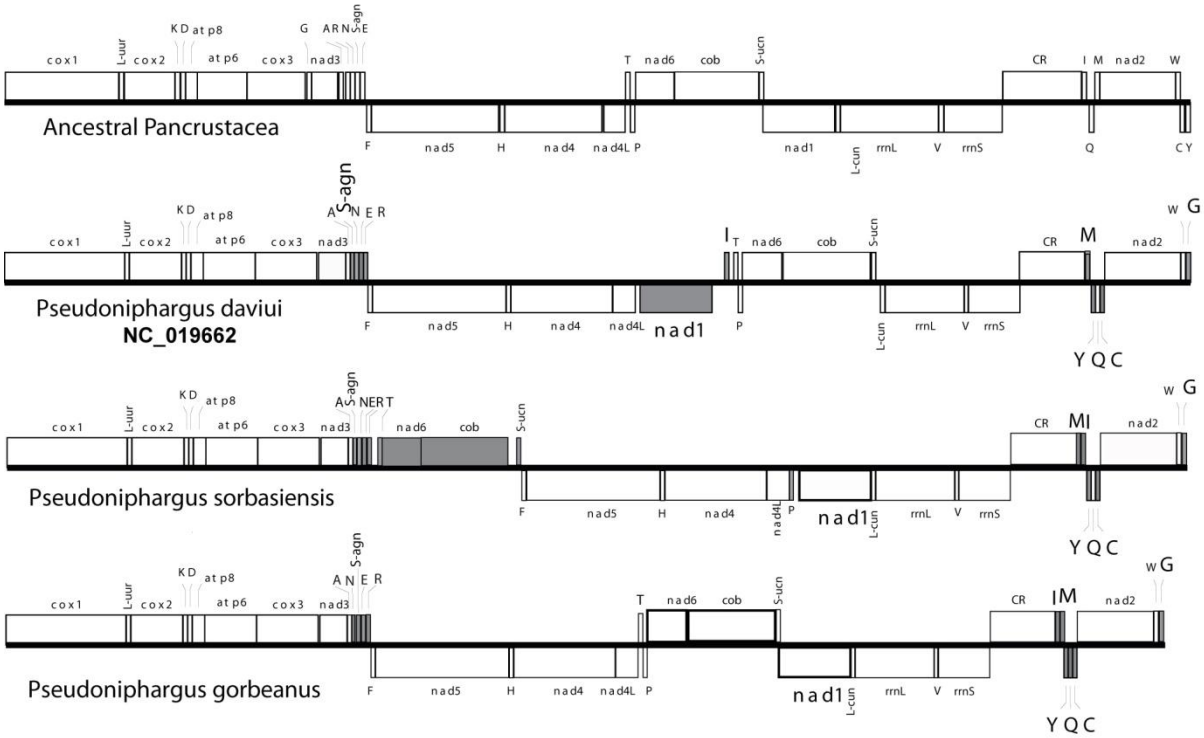


Figure 4.3.1. Comparison of mitochondrial gene order in the three *Pseudoniphargus* species studied herein with respect to the presumed Pancrustacean gene order. Genes appearing in a different arrangement than the Pancrustacean order are remarked in grey. Genes placed under the line are coded in the negative strand while genes above are coded in the positive strand.

The discrepancy in size among the three mitogenomes is largely due to the fact that we were unable to fully recover the control region as well as a small part of *rrnS* (approximately 40 bp) of *P. gorbeanus*, and to the presence of intergenic non-coding regions differing slightly in length. The three mitogenomes show several short non-coding spacers: *P. gorbeanus* (2-5 bp), *P. sorbasiensis* (2-16 bp), and *P. daviui* (2-18 bp). Besides, the last two species display three relatively large non-coding regions of 112 bp, 184 bp and 450 bp (*P. sorbasiensis*), and 54 bp, 56 bp and 156 bp (*P. daviui*). These non-coding regions have been compared all along their respective mitogenomes, finding no sequence similarity to other regions.

The three *Pseudoniphargus* mitogenomes show different gene arrangements that in turn differ from both the putative Pancrustacean pattern and the gene orders of amphipods published elsewhere (Bauza-Ribot et al. 2009; Kilpert and Podsiadlowski 2010; Pons et al. 2014) (**Figure 4.3.1**). It is noteworthy to remark that *Pseudoniphargus* gene orders differ in the position of one or several PCGs while rearrangements involving PCGs are extremely rare at the genus level in other metazoans (Rawlings et al. 2001; Matsumoto et al. 2009). The unexpected length of two PCR amplicons in *P. sorbasiensis* compared to *P. gorbeanus* and *P. daviui* already suggested the occurrence of a putative gene rearrangement (**see Figure 4.3.2**).

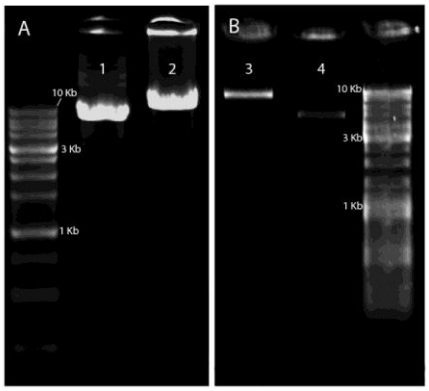


Figure 4.3.2 Agarose gel electrophoresis of long PCR amplicons of *P. sorbasiensis* (A) and *P. gorbeanus* (B). Lanes 1 and 3 show amplicons including the mitochondrial DNA fragment from 5' end of *cox1* to 3' end of *cob* genes, and lanes 2 and 4 the opposite fragment from 5' end of *cob* to 3' end of *cox1* genes. The difference in length between the two species is due to the rearrangement of *cob* and *nad6* genes in *P. sorbasiensis* (see Figure 1 and Results and Discussion section).

Sequence analyses show that all three genomes have undergone several gene shifts with respect to the putative ancestral Pancrustacean gene order (Boore et al. 1995, 1998; see **Figure 4.3.1** and **Figure 4.3.3-4.3.5**). Some of these rearrangements found in *Pseudoniphagus* are shared throughout all currently known amphipod mitogenomes (Pons et al. 2014). For instance, *trnG* gene has undergone a transposition from its ancestral position between *cox3* and *nad3* to occupy an adjacent position to *nad2* and *trnW* genes. The *trnR* gene embedded in the tRNA complex between *nad3* and *nad5* has been transposed to the left of *trnE* in the same complex. Finally, *trnC* gene has relocated from *trnW* to occupy an adjacent position to *nad2* in all amphipods except *Caprella mutica* and *C. scaura* (Ito et al. 2010; Kilpert and Podsiadlowski 2010). Other rearrangement events appear to be unique to *Pseudoniphargus* and even species-specific. In *P. daviui*, the gene coding for *nad1* has relocated downstream of *nad6* and

cob, possibly through a tandem duplication subsequently affected by a TDRL (**Figure 4.3.4**). In addition, in *P. sorbasiensis* the position of both *nad6* and *cob* have changed through another TDRL event which placed them before *nad5* (**Figure 4.3.4**). This rearrangement explains the above mentioned opposite sizes observed in the agarose gel after electrophoresis of the PCR fragments in *P. sorbasiensis*. *P. gorbeanus* has apparently undergone one TDRL event shared with the two other *Pseudoniphargus* and other amphipods such as *Bahadzia jaraguensis* (Bauzà-Ribot et al. 2012). *P. gorbeanus* displays three transpositions and one TDRL event that differ from the presumed ancestral Pancrustaean arrangement (**Figure 4.3.5**), although none of them are unique. Moreover, the mitogenome of this species shows no trace of the transposition events involving *trnN* evident in the two other *Pseudoniphargus* species as well as in the rest of amphipods except *Parhyale hawaiiensis*. Finally, *P. daviui* shows four transpositions and two TDRL events differing from the ancestral genome, whereas *P. sorbasiensis* displays five transpositions and two TDRL events.

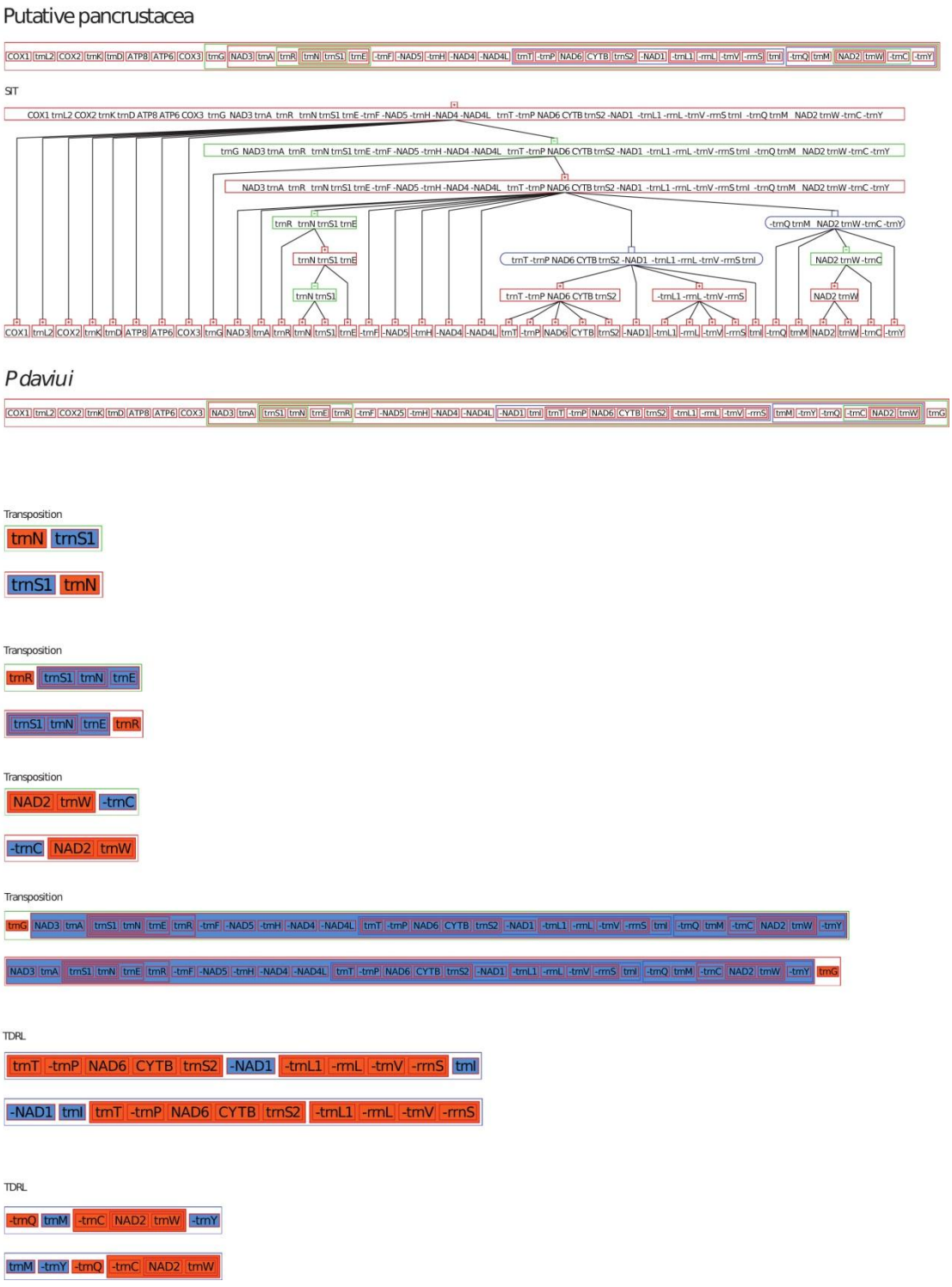


Figure 4.3.3 Rearrangement events in *P. daviui* mitogenome deduced by strong interval tree (SIT) analyses with respect to the presumed Pancrustacean gene order. Strong intervals are shown in rectangles and trivial strong intervals in elliptic boxes. If the product of a node appears in the same order in both gene orders (called linear increasing), it is indicated with a (+), while if the product order is opposite to the input gene order (called linear decreasing), it is indicated with a (-); otherwise, the term prime is used.

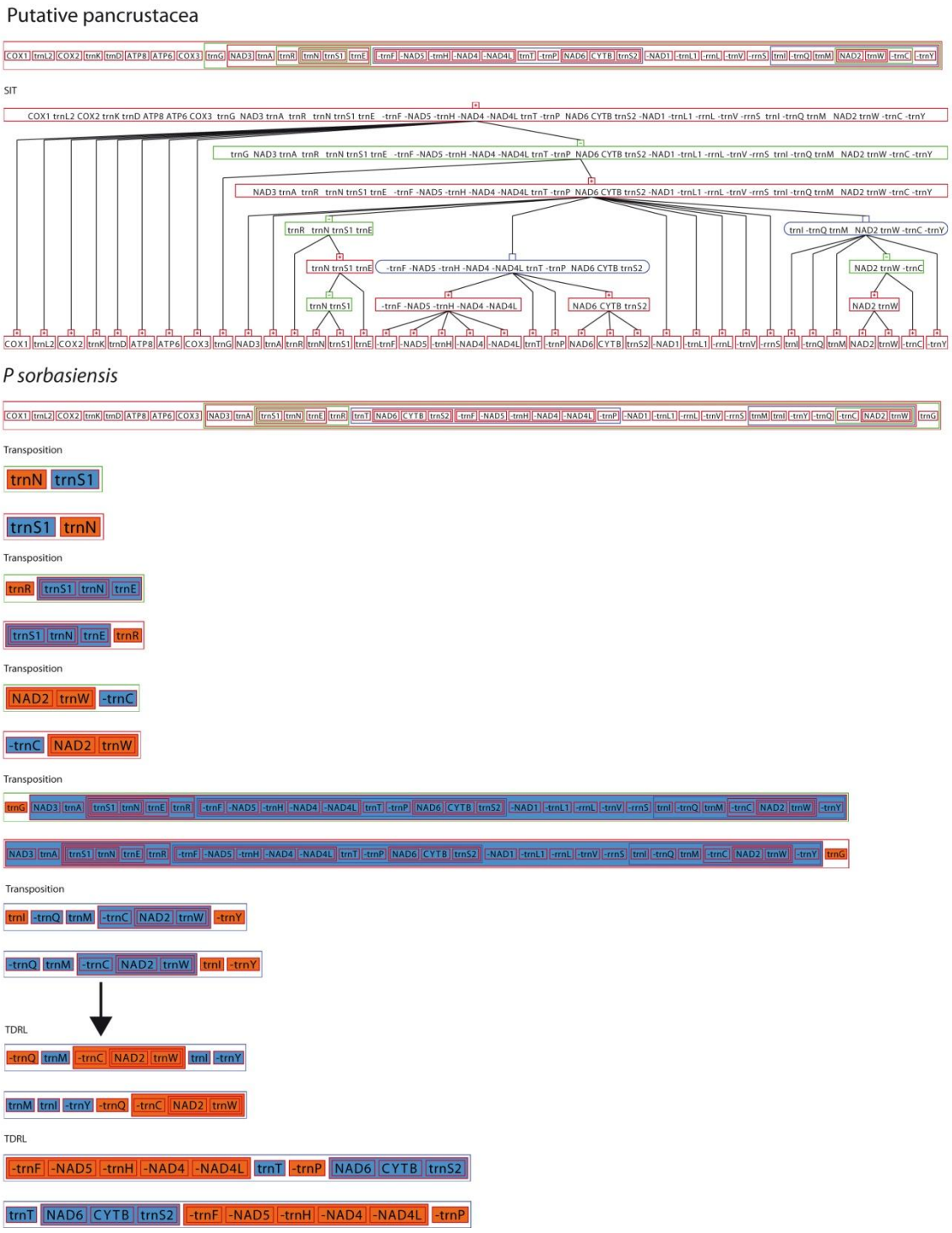


Figure 4.3.4 Rearrangement events in *P. sorbasiensis* deduced by strong interval tree (SIT) analyses with respect to the presumed Pancrustacean gene order. Details as in **Figure 4.3.3**.

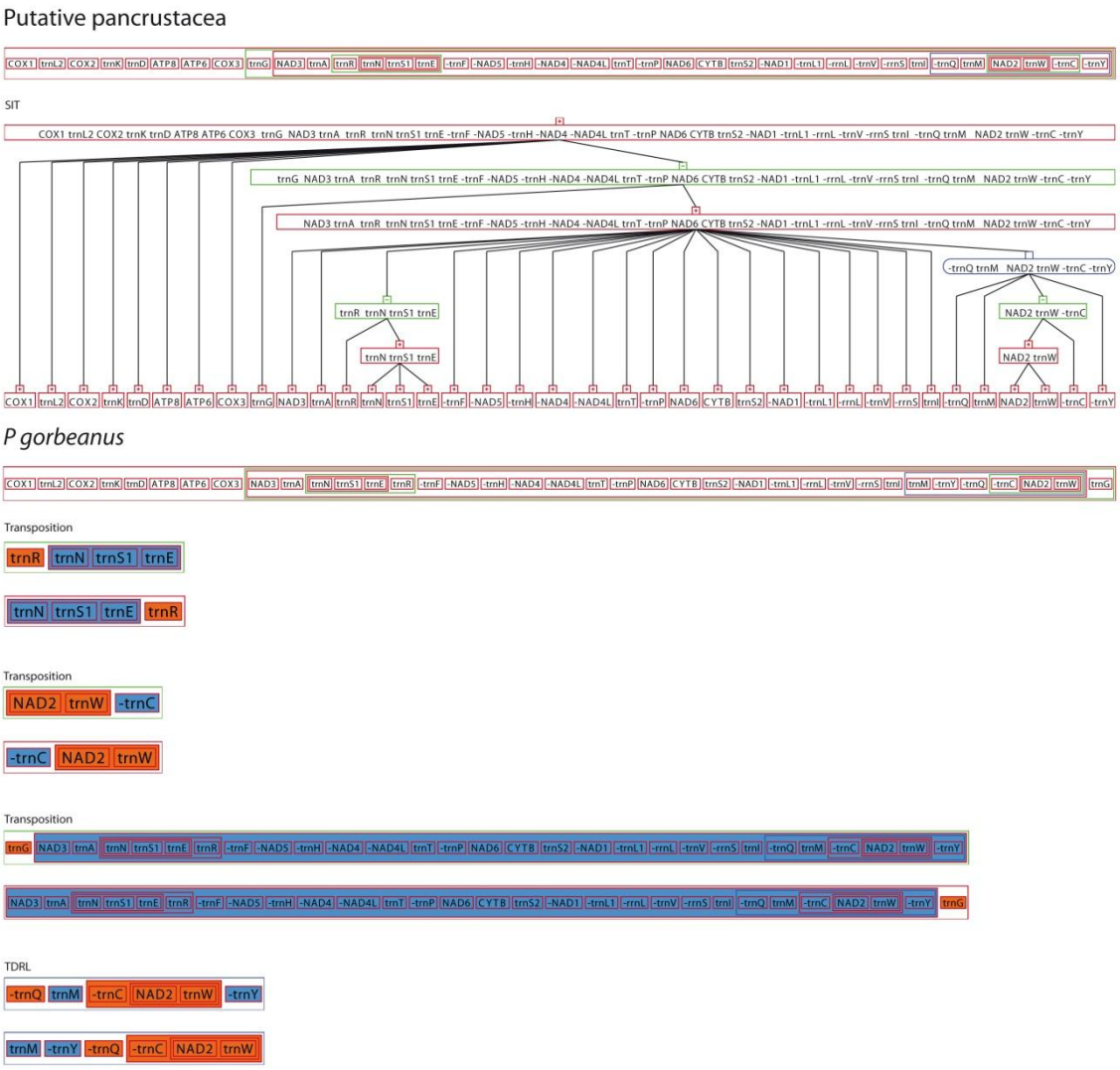


Figure 4.3.5 Rearrangement events in *P. gorbeanus* deduced by strong interval tree (SIT) analyses with respect to the presumed Pancrustacean gene order. Details as in **Figure 4.3.3**.

An additional observation is that the PCGs rearranged by TDRL display large non-coding regions flanking both the novel and prior placements in both *P. daviui and P. sorbasiensis* (see **Figure 4.3.1**). Similar large non-coding areas adjacent to new and old gene placements have been shown to occur in *Culicoides* (Diptera, Ceratopogonidae; Matsumoto et al. 2009), whereas smaller non-coding spacers (from 12 bp to 60 bp) has been reported in tRNA rearrangements of provannid and vermetid marine gastropods (Hidaka et al. 2013; Rawlings et al. 2001). It is likely that these non-coding spacers represent residual sequence artifacts of prior rearrangement events (Boore et al. 1998).

Table 4.3.1 Gene length, AT content, and GC and AT skews for the species *P. daviui*, *P. gorbeanus*, and *P. sorbasiensis*. Asterisk identifies the rRNA of *P. gorbeanus*, where rrnS is partial, lacking about 40 bp.

		P. daviui				*P. gorbeanus*				*P. sorbasensis*			
Gene	Strand	Size bp	AT %	AT-skew	GC-skew	Size bp	AT %	AT-skew	GC-skew	Size bp	AT%	AT-skew	GC-skew
Complete	+	15 155	71.2	0.065	−0.292	14 190	68.7	−0.001	−0.314	15 460	69.7	0.016	−0.338
rDNA	–	1717	73.90	−0.077	0.435	1676*	76.9	0.108	−0.357	1705	75.9	−0.065	0.394
tRNA	n.a	1311	70.3	0.036	0.077	1311	71.4	0.022	0.091	1326	70.5	0.041	0.098
PCG	n.a	11 007	66.4	−0.175	0.006	11 001	70.0	−0.149	0.030	11 001	67.5	−0.167	0.012
cob	+	1137	64.8	−0.086	−0.335	1131	69.6	−0.011	−0.283	1137	65.7	−0.102	−0.265
nad2	+	975	72.3	−0.203	−0.307	975	74.0	−0.192	−0.305	975	71.5	−0.180	−0.333
cox1	+	1536	62.4	−0.176	−0.114	1536	64.7	−0.097	−0.020	1536	63.7	−0.149	−0.129
cox2	+	675	62.3	−0.114	−0.284	675	65.9	−0.026	−0.226	675	62.9	−0.107	−0.229
atp8	+	156	66.0	−0.088	−0.322	156	72.4	−0.044	−0.348	156	70.6	−0.091	−0.349
atp6	+	669	67.3	−0.204	−0.323	669	69.6	−0.086	−0.399	669	65.6	−0.143	−0.442
cox3	+	786	63.4	−0.177	−0.153	786	65.5	−0.056	−0.122	786	64.6	−0.130	−0.209
nad3	+	348	67.2	−0.205	−0.104	348	65.4	−0.151	−0.148	348	67.3	−0.180	−0.122
nad5	–	1707	66.7	−0.148	0.441	1707	71.1	−0.221	0.384	1707	68.6	−0.152	0.457
nad4	–	1299	68.1	−0.181	0.420	1299	71.7	−0.250	0.493	1299	70.8	−0.232	0.555
nad4L	–	288	72.9	−0.218	0.535	288	78.4	−0.301	0.546	288	73.9	−0.240	0.571
nad6	+	507	69.8	−0.181	−0.503	507	75.6	−0.130	−0.492	501	70.1	−0.005	−0.453
nad1	–	924	65.9	−0.278	0.431	924	70.2	−0.316	0.470	924	67.6	−0.290	0.463
Control region	+	790	85.5	0.074	0	n.a	n.a	n.a	n.a	572	83.4	0.031175	−0.0964

Sequence composition and codon usage

The AT-content in these mitogenomes is quite high as expected for arthropods: 71.2 %, 68.7 % and 69.7% in *P. daviui*, *P. gorbeanus*, and *P. sorbasiensis*, respectively (**Table 4.3.1**). Using metAMiGA (Feijão et al. 2006), we establish that these AT-content values are well within the normal range of crustaceans and in the middle range of peracarids (60.8% - 76.9%). The highest AT content occurs in the control region (85.5% - 83.4 %) as is common in arthropods (Yang and Yang 2008; Ki et al. 2010; Wang et al. 2013; Liu et al. 2014), and in a lesser extent in ribosomal and tRNA genes (**Table 4.3.1 and Table 4.3.2**). PCGs showing the highest AT content were *nad4L* (72.9% -78.4%) and *nad2* (71.5%-74.0%) in all three species, with minor differences in regard to their coding strand as reported in Pons et al. (2014). Nonetheless, if codon position is taken into account, third positions are considerably AT richer (70.9-77.6) than first and second sites (63.2-66.4; **Table 4.3.2**). This composition bias stresses the importance of data partitioning in phylogenetic studies using mitochondrial PCGs (Hassanin 2006; Pons et al. 2010). There is a significant C over G content, i.e. negative GC-skew, in the PCGs on the positive strand from -0.019 up to -0.503, and an opposite trend on the negative strand (0.384-0.555 in *nad1, nad5, nad4 and nad4L*) (**Table 4.3.2**). These results are similar to the values reported for most malacostracan mitogenomes (Krebes and Bastrop 2012; Pons et al. 2014). In addition, GC-skew varies slightly depending on the codon position with first sites being positive (from 0.190 to 0.208), second ones being negative (from -0.113 to -0.095) and third positions being closer to zero (from -0.087 to -0.027) **(see Table 4.3.2**). AT-skew is negative for all PCGs and positive in tRNAs, rRNAs and the control region. AT-skew shows similar trends although it appears to be less prominent **(Table 4.3.2**).

Table 4.3.2 Presenting the AT frequency as well as both AT and GT-skew in 1st, 2nd and 3rd position as well as inpositive and negative strand of the total genes.

Species	A+T 1st	A+T 2nd	A+T 3rd	A+T Positive 1st	A+T Negative 1st	A+T Positive 2nd	A+T Negative 2nd	A+T Positive 3rd	A+T Negative 3rd
P. daviui	63.2	64.9	70.9	61.7	65.6	64.4	66.0	71.0	70.9
P. gorbeanus	65.9	66.4	77.6	64.7	67.9	65.1	68.4	77.0	78.4
P. sorbasiensis	64.6	65.3	72.4	62.5	67.9	64.5	66.7	71.5	73.8

Species	GC skew 1st	GC skew 2nd	GC skew 3rd	GC skew Positive 1st	GC skew Negative 1st	GC skew Positive 2nd	GC skew Negative 2nd	GC skew Positive 3rd	GC skew Positive 3rd
P. daviui	0.190	-0.105	-0.093	0.034	0.471	-0.348	0.302	-0.497	0.560
P. gorbeanus	0.208	-0.113	-0.027	0.068	0.452	-0.335	0.285	-0.432	0.668
P. sorbasiensis	0.198	-0.095	-0.087	0.013	0.539	-0.341	0.327	-0.509	0.650

Species	AT skew 1st	AT skew 2nd	AT skew 3rd	AT skew Positive 1st	AT skew Negative 1st	AT skew Positive 2nd	AT skew Negative 2nd	AT skew Positive 3rd	AT skew Negative 3rd
P. daviui	-0.009	-0.436	-0.080	-0.037	0.027	-0.416	-0.464	-0.042	-0.142
P. gorbeanus	-0.002	-0.431	-0.031	0.029	-0.052	-0.401	-0.477	0.101	-0.240
P. sorbasiensis	-0.012	-0.436	-0.064	-0.011	-0.016	-0.417	-0.466	0.001	-0.165

Accumulative AT-skew plotted for *P. sorbasiensis* indicates that a maximum positive value is reached at around 8,000 bp and lowest at around 12,000 bp, clearly indicating the shifts from positive to negative strands in the sequence, where the decline represents the genes on the negative strand **(see Figure 4.3.6**). A similar trend is found in *P. daviui*, the maximum positive value starting here around 5,500 bp, whereas a maximum negative value is reached around 10,000 bp. In *P. gorbeanus* the maximum value being reached at around 5,200 bp, and experiences a shift in rate from 8,800 to 10,800, indicating the falling trend of the curve on the negative-stranded genes. Accumulative GC-skew has been used previously to discover the origin of replication in bacterial circular chromosomes. In our case, it can be identified at the end of the plateau around 14,000 bp in both *P. daviui* and *P. sorbasiensis* **(Figure 4.3.6).** Finally, the effective number of codons (ENC) estimated for the mitogenomes of the three species were 48.9, 45.7 and 48.4 for *P. daviui*, *P. gorbeanus and P. sorbasiensis*, respectively. No clear correlation is found between GC content in third positions and the ENC values as reported in other studies (Bauza-Ribot et al. 2009; Kilpert and Podsiadlowski 2006), perhaps due to the lower A+T bias occurring in *Pseudoniphargus*.

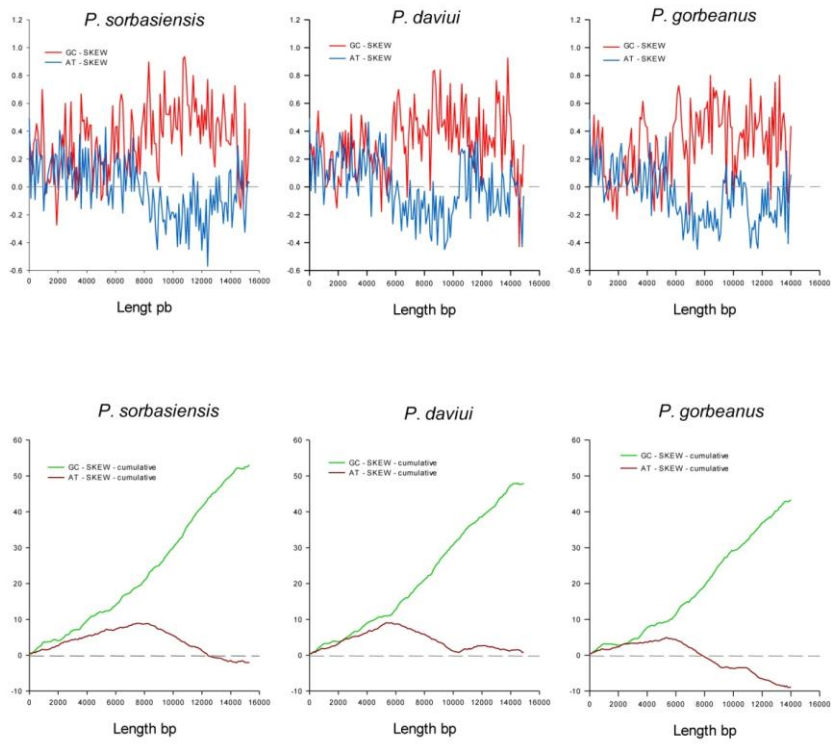


Figure 4.3.6 plus AT–skews and cumulative GC and AT skews for the three *Pseudoniphargus* species

Start and stop codons

Eight out of the 13 protein coding genes (*nad2*, *cox1*, *cox2*, *nad3*, *atp8*, *nad5*, *nad4* and *nad1*) show unusual start codons (ATY, ATT, TTG, TTG, ATT, TTG, TTT, and GTR, respectively) and in four of them the stop codons are truncated to either T or TA (*nad2*, *cox1*, *nad5* and *nad4*). Most of these non-canonical start codons had been already reported in previous studies, including amphipod mitogenomes (Bauza-Ribot et al. 2009; Ito et al. 2010; Ki et al. 2010; Kilpert and Podsiadlowski 2010). In addition, *P. daviui* and *P. sorbasiensis* display a truncated stop codon in *cox2*. These truncated stop codons are likely to translate to UAA terminal codons that will be completed in later stages by post-transcriptional polyadenylation, as described elsewhere (Ojala et al. 1980; 1981). The large amount of mitogenomes published in recent years corroborates the initial hypothesis that non-canonical start and stop codons are common features in mitochondrial genomes (Boore at al. 2005) which are generally detected due to large overlapping sequence with upstream or downstream genes. Most of these translation exceptions are already codified in the translation tables available at GenBank, ENA, and DNA Data Bank of Japan (DDBJ) or can be manually added as translation exception notes.

Transfer RNA genes

We identified all 22 expected tRNA genes in the three mitogenomes, with fourteen genes coded on the positive and eight on the negative strand. This strand pattern is the same as in the putative Pancrustacean pattern, although tRNA gene order varies across *Pseudoniphargus* as shown above (**Figure 4.3.1**). Secondary structure is well conserved within all three species, with *P. sorbasiensis* showing the most divergent primary

sequences (**Figure 4.3.7**). The tRNAs sequence length varies slightly across the three species as follows: 52-66 bp (*P. sorbasiensis*), 50-64 pb (*P. daviui*), and 51-61 bp (*P. gorbeanus*). The tRNA Threonine lacks the TΨC arm as already reported to be the case in *Metacrangonyx* (Pons et al. 2014), whereas the tRNA genes for Valine and Serine 1 (codon UCN) lack the DUH arm as described for all metazoans (Ki et al. 2010; Kilpert and Podsiadlowski 2006). The tRNA Glutamine lacks the TΨC arm, as reported also in *Caprella mutica* (Kilpert and Podsiadlowski 2010). The overall structure of the remaining 19 tRNAS is as follows: 7 pairs in the aminoacid acceptor (AA) arm, two bases joining AA and *dihydrouridine (DUH) arms*, 2-4 pairs in the DUH stem, 1-7 nucleotides in the DUH loop, 5 pairs in the anticodon stem (AC), 7 bases in the anticodon loop including the three anticodon nucleotides, 4-5 bases in the variable loop, 2-4 pairs in the pseudo (TΨC) stem, and finally 4-6 nucleotides in the TΨC loop (Figure 2). The presence of two nucleotides between AA and DUH arms, a single nucleotide joining DUH and AC arms, and the absence of nucleotide link between TΨC and AA arms are constant throughout all the tRNA genes. Most of the nucleotide substitutions found in the tRNAs of the three *Pseudoniphargus* mitogenomes are compensatory mutations located on stem regions (AU/GC 35, GU/GC 13, GU/AU 11, and UA/AU three times) although there are also 18 mismatches. These results are congruent with the pattern found in a previous analysis of 21 *Metacrangonyx* species (Pons et al. 2014), suggesting that compensatory mutations play a key role in the evolution of tRNA sequences.

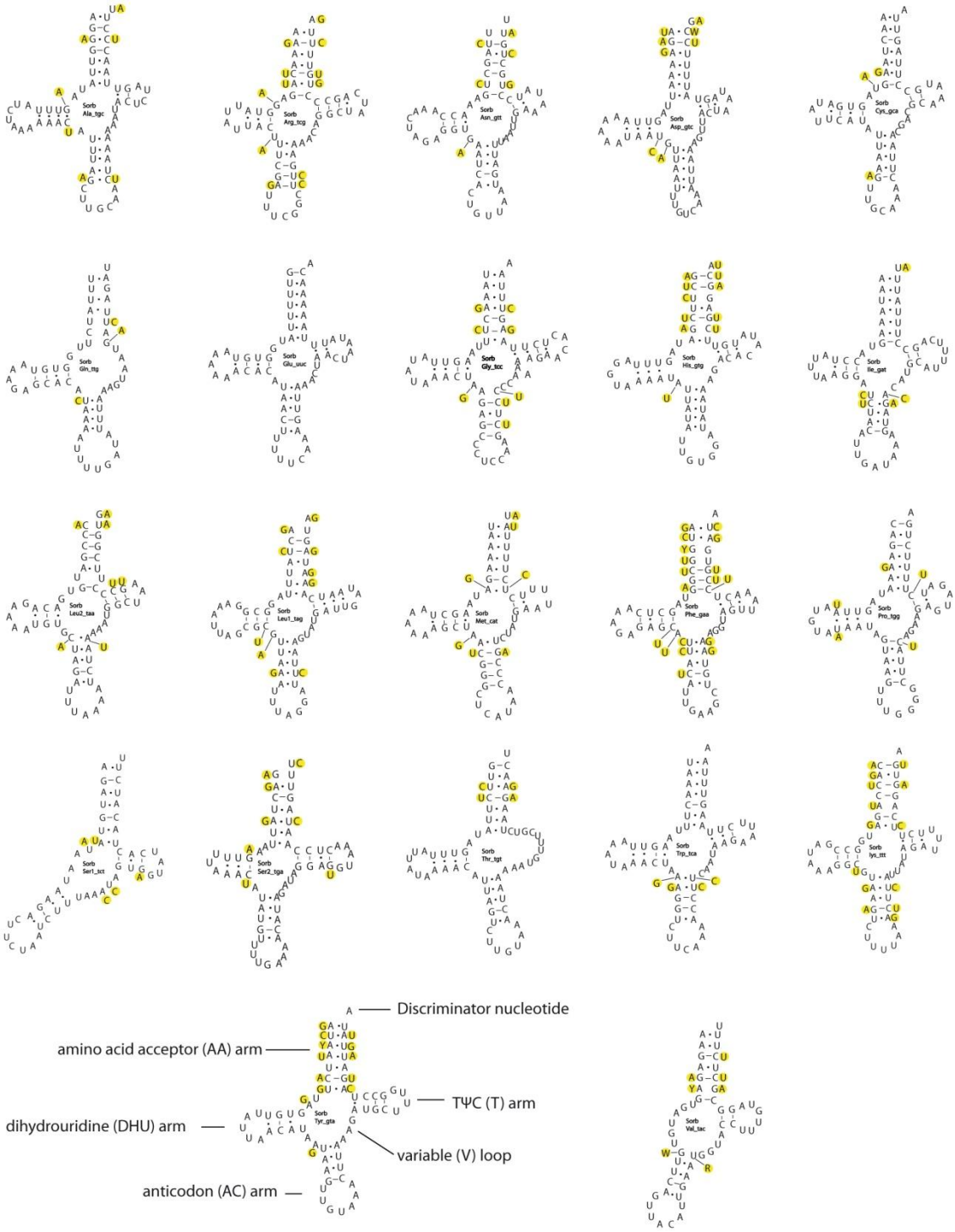


Figure 4.3.7 Mitochondrial tRNAs secondary structures in *P. sorbasiensis*. Variable positions are highlighted in yellow.

Ribosomal genes

The *rrnL* and *rrnS* genes show a similar length across the three species, *P. daviui* (*rrnL* 1012 bp; *rrnS* 700 bp), *P. gorbeanus* (*rrnL* 1001 bp; partial *rrnS* 675 bp), and *P. sorbasiensis* (*rrnL* 1008 bp; *rrnS* 698 bp). These are the shortest *rrnL* sequences reported for the Crustacea, although *Metacrangonyx* has a shorter *rrnS* (Pons et al. 2014). The predicted secondary structures are shown in **Figures 4.3.8 and 4.3.9**. They were tentatively constructed using MITOS and compared to the already known secondary structures of amphipods and other arthropods (Carapelli et al. 2004; Negrisolo et al. 2011; Wang et al. 2013; Pons et al. 2014). The secondary structure of the small ribosomal unit of *P. sorbasiensis* (**Figure 4.3.8)** was very similar to that described for the amphipod *Metacrangonyx boveii* (Pons et al. 2014), the branchiopod crustacean *Artemia franciscana* (http://www.rna.ccbb.utexas.edu/RNA/Structures/b.16.m.A.franciscana.bpseq) and the neuropterid insect *Libelloides macaronius* (Negrisolo et al. 2011). However, the primary sequence is only conserved in some stretches of domain II and in most of the domain III. In fact, secondary structure of domains I and II had to be reconstructed by finding small anchor regions (i.e. conserved regions on the aligned sequences), and then folding primary sequences in small sections of 30-100 bp using Mfold. Furthermore, the secondary structures shown here, despite being significantly suboptimal in Mfold, were selected because they closely resemble those previously published elsewhere (Carapelli et al. 2004; Negrisolo et al. 2011; Wang et al. 2013; Pons et al. 2014). The first domain of the small subunit of *P. sorbasiensis* (209 bp) is larger than those previously calculated for the crustaceans *M. boveii* (186bp) and *A. franciscana* (181 bp), but slightly shorter than that of the insect *L. macaronious* (217 bp). The large ribosomal unit shows a similar pattern, with the most divergent primary sequences also found in domain I (**Figure 4.3.9).** As in any other arthropods, *P. sorbasiensis* also lacks the domain III, and the last domain (V) is the most conserved (Carapelli et al. 2004; Negrisolo et al. 2011; Wang et al. 2013; Pons et al. 2014). Both secondary structures seem to be folded correctly since all bindings related to tertiary folding are conserved (yellow lines in **Figures 4.3.8 and 4.3.9**). Although MITOS is a very useful tool for PCG annotation and building tRNA secondary structures, it appears to be less effective for determining *rrnS* and *rrnL* secondary structures, particularly in the most variable domains. This is probably because it focuses on reconstructing global secondary structures of the ribosomal molecules instead of local motifs that can be better defined throughout conserved stretches from the aligned sequences.

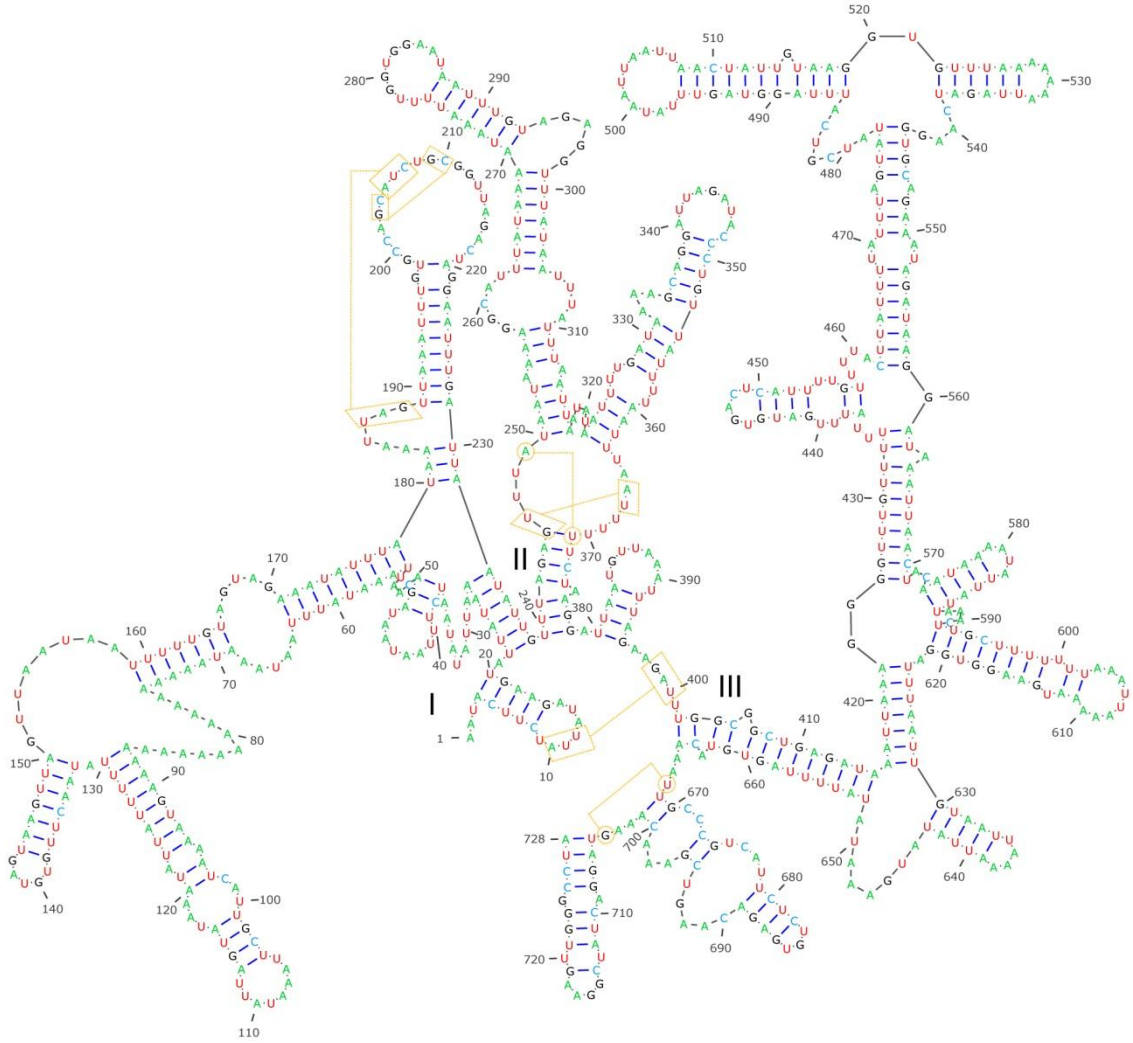


Figure 4.3.8 Predicted secondary structure for the small ribosomal mitochondrial RNA (12S) of *P. sorbasiensis*. Different domains are labeled with roman numerals, and sites involved in tertiary folding are remarked with yellow dotted lines.

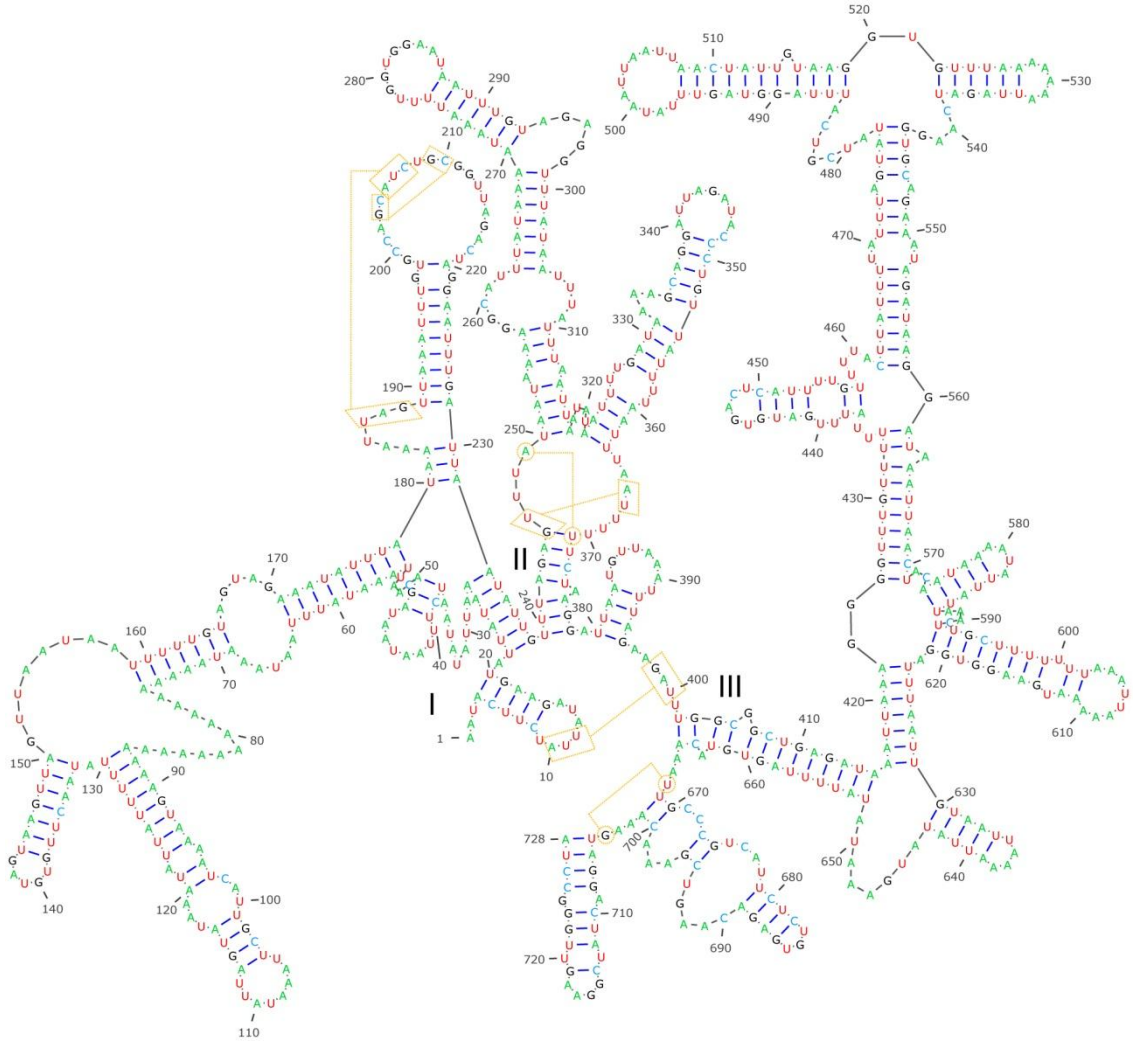


Figure 4.3.9 Predicted secondary structure of the large ribosomal mitochondrial RNA (16 S) of *P. sorbasiensis*. Details as in **figure 4.3.8.**

Control region

Both *P. daviui* and *P. sorbasiensis* mitogenomes have a large non-coding region of 790 bp and 572 bp, respectively, situated between *rrnS* and *trnM* genes. These sequences display characteristic features of the control region (CR), such as its high AT-content (85.5 % in *P. daviui*; 83.4 % in *P. sorbasiensis*) and repetitive motifs (Arunkumar and Nagaraju 2006). Both control regions were compared using blastn, revealing a 75% sequence identity with 5% gapped positions, suggesting a relative conservation of this fast evolving sequence. In the amphipod genus *Metacrangonyx*, the comparison of ten control regions found a remarkable conservation only between two very closely related species (Pons et al. 2014). We were unable to detect AT-rich palindromes associated with TATA and GA(N)T motifs in other non-coding areas except at the CR, thus discarding that *Pseudoniphargus* had more than one CR as described in some amphipods (Ito et al. 2010; Kilpert and Podsiadlowski 2010). The mitogenomes of *P. daviui* and *P. sorbasiensis* display candidate sequence regions for the origin of replication within the CR (**Figure 4.3.10.**). In *P. daviuvi*, we identified the motifs

related to the origin of replication (5´end ATAT and 3´ end GANT motifs) plus a large palindrome within positions 14,501-14,703. We also identified these motifs in *P. sorbasiensis* occurring within positions 8,106-8,219, but in this species the 3' end motif differs (GAAA or GTA). Several studies have shown that these motifs appear sometimes modified or truncated (Black and Roehrdanz 1998; Fahrein et al. 2007). Interestingly, the sequences of both palindromes are relatively conserved (59,23 % identity; **Figure 4.3.10B.**), suggesting the operation of selective evolutionary constrains in these regions.

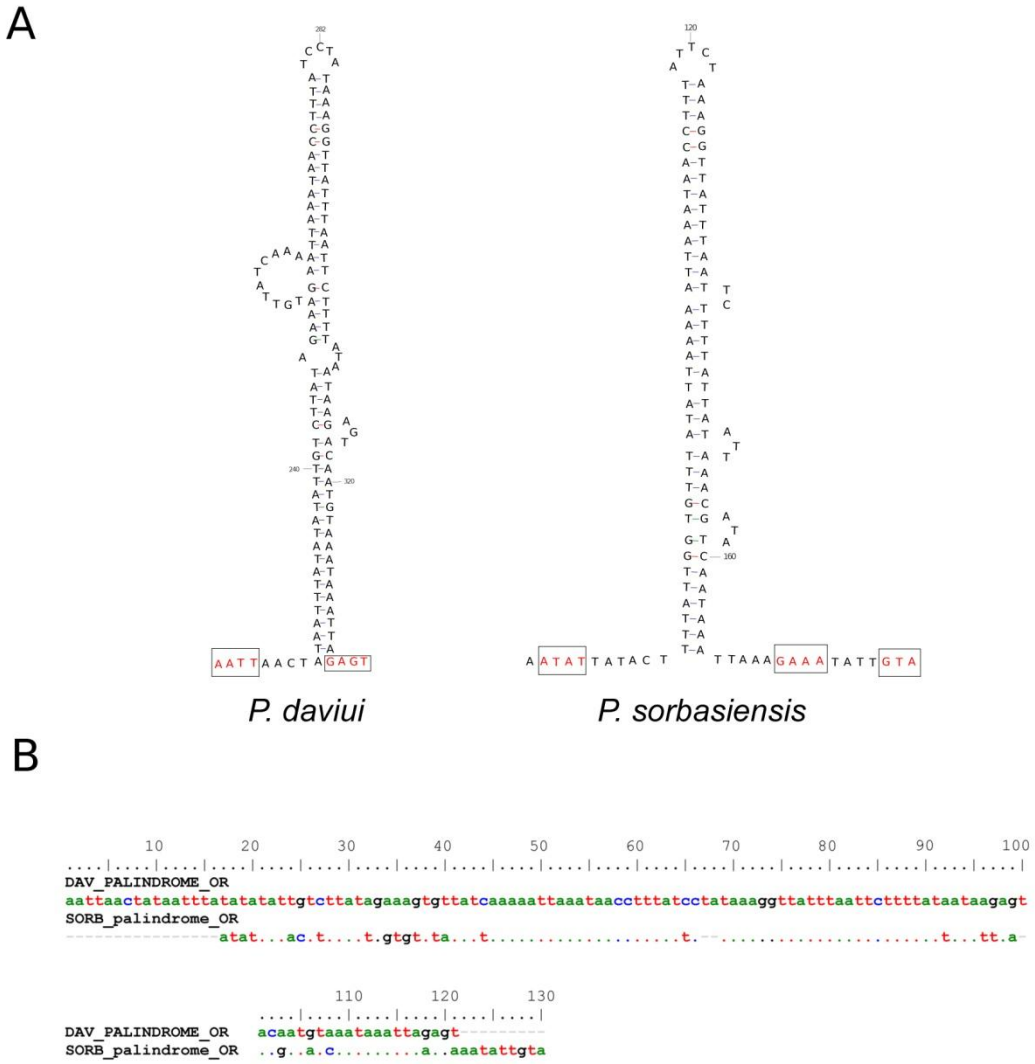


Figure 4.3.10 Putative origin of replication for *P. daviui* and *P. sorbasiensis*, with the start and stop motifs indicated in red (A). DNA alignment of *P. daviui* and *P. sorbasiensis* palindromes (B).

Chapter 4.4 Phylogeny based on 32 mitogenomes

Results

Mitochondrial genome assembly

We obtained the complete or nearly complete sequence of the mitochondrial genome of 30 *Pseudoniphargus* species representing the major lineages recognized within *Pseudoniphargus* (14-15 kb; **Table 4.4.1** and **Figure 4.4.1**). The main lineages were identified in a tree previously built by us based on cox1 sequences of 410 individuals collected throughout the entire known geographic distribution of the genus. The species *P.* sp2-Portugal (12.7 kb) and *P. salinus* (10.5 + 4 kb) with two non-overlapping fragments were the exception due to the presence of long polyA/polyT runs. Reads including these poly-runs are generally discarded as low quality sequences and cause low or no coverage in such particular regions. We did not attempt to circularize mitochondrial genomes on the control region because sequencing duplications and poly AT runs present therein requires of a huge effort. Besides, control regions are not included in phylogenetic analyses due to their fast evolutionary rate, which precludes their accurate alignment. The coverage obtained with the different amplification and sequencing methods varied broadly both within and between approaches, but was generally high, with 22 out of 30 mitogenomes reaching above 100x coverage and just 4 species falling below 50x (**Table 4.4.1**). The assembly of the genomic DNA library of individuals from Sidi Abdellah (Morocco), apparently corresponding to a single species based on morphology, produced 2 large contigs with more than 10% of genetic distance, both matching the mitogenomes of *Pseudoniphargus*. They did not show stop codons or any other typical features of nuclear mitochondrial copies, i.e. numts. Since the presence of several species on the same well is not rare in *Pseudoniphargus*, we included both in further analyses. It is quite common to find a few polymorphic sites when a library is constructed from genomic DNA of several individuals of the same population but, surprisingly, this is the first time that we retrieved two complete and differentiated mitogenomes.

Table 4.4.1 List of the 31 *Pseudonipahrgus* species plus the outgroup *Metacrangonyx dopharensis* included in this study with information about collecting locality on map Figure 4.4.1, GenBank accession number, sequencing method (SM), mitogenome length, number of reads obtained ($x10^3$), depth coverage and average length of mapped reads. Sequencing approaches: 1) Long PCR Fragments and FLX/454 with tagging, 2) Long-PCR amplification and GS JUNIOR without tagging, 3) MDA and Miseq 2x150 bp without tagging, 4) WGS low coverage Miseq 2x150 bp with tagging, 5) WGS low coverage Hiseq2500 2x150 bp with tagging, and 6) WGS low coverage Hiseq2500 2x160 bp with tagging.

Species name	Locality	map	Genbank	SM	length	n reads	Coverage	Av. read pb
P. carpalis	Bermuda Red Bay Cave	1		2	14861	15	258	522
P. ruffoi	Morocco Berkane	2		5	14055	2380	8	134
P. longipes	Morocco Taza Sidi Abdellah	3		4	14658	2480	23	152
P. sp6-Morocco A	Morocco Taza Sidi Abdellah	3		6	14180	2240	275	138
P. sp6-Morocco B	Morocco Taza Sidi Abdellah	3		6	12746	2240	253	128
P. brevipedunculatus	Portugal Azores Pico Calhau	4		5	14424	2710	110	147
P. sp1-Azores	Portugal Azores Sao Miguel S. Roque	5		5	14981	2480	185	144
P. mateosorum	Portugal Lapa dos Morcegos	6		3	14248	1	114	152
P. portosancti	Portugal Madeira Fonte do Tanque	7		2	14093	20	99	526
P. sp1-Portugal	Portugal Sicó Gruta de Legaçao	8		6	15728	2850	890	148
P. sp2-Portugal	Portugal Sintra Gruta de Assafora	9		6	14150	2360	97	141
P. gorbeanus	Spain Alava Cigoitia Artzegi'ko Koba	10	LN871176	2	14188	29	148	481
P. sorbasiensis	Spain Almeria Sorbas Cueva del Agua	11	LN871175	2	15460	22	271	401
P. sp2-Andalusia	Spain Almeria Níjar Pozo de los Frailes	12		6	14191	1930	74	147
P. daviui	Spain Baleares Cabrera Font de s'hort de Ca'n Feliu	13	FR872383	1	15155	N/A	73	410
P. triasi	Spain Baleares Cabrera Font de s'hort de Ca'n Feliu	13		5	15109	2940	419	148
P. pedreae	Spain Baleares Formentera St. Ferrán Coves de sa Pedrera	14		3	14740	1.1	32	152
P. pityusiensis	Spain Baleares Eivissa Sant Joan de Llabritja	15		5	14748	2690	495	148
P. mercadali	Spain Baleares Mallorca Capdepera Cova de na Barxa	16		5	14402	4510	53	138
P. sp1-Balearics	Spain Baleares Mallorca Sencelles Ruberts	17		5	15256	2020	83	152
P. stocki	Spain Cádiz Villaluenga del Rosario El Pozo Blanco	18		2	14694	12	364	381
P. salinus	Spain Canarias El Hierro Pozo de Las Calcosas	19		5	9748+4056	2610	17	150
P. sp2-Canaries	Spain Canarias Gran Canaria El Sao Barranco de Arguineguín Mina Los Roques	20		3	15321	1	476	152
P. gomerae	Spain Canarias La Gomera San Sebastián Playa del Avalo	21		2	14150	21	136	246
P. cupicola	Spain Canarias La Palma Charco Verde Puerto Naos	22		2	15351	18	196	268
P. multidens	Spain Canarias La Palma Santa Cruz	23		2	14504	27	206	354
P. "morenoi"	Spain Córdoba Baena Cueva del Yeso	24		5	14810	2710	112	145
P. unisexualis	Spain Guipúzcoa Zegama Túnel de San Adrián	25		2	14969	16	181	412
P. grandis	Spain Málaga Canillas de Aceituno Fuente de la Fájara	26		2	14800	23	247	397
P. sp1-Murcia	Spain Murcia Isla Plana Cueva del Agua	27		6	14170	230	156	149
P. elongatus	Spain Burgos Portillo de la Sía Cueva de Imunía	28		5	15356	3330	212	146
P. sp1-Basque	Spain Vizcaya Axpe Busturia Cueva de San Pedro	29		3	14187	0.98	1005	152
M. dhofarensis	Oman Salalah			4	14399	3490	35	152

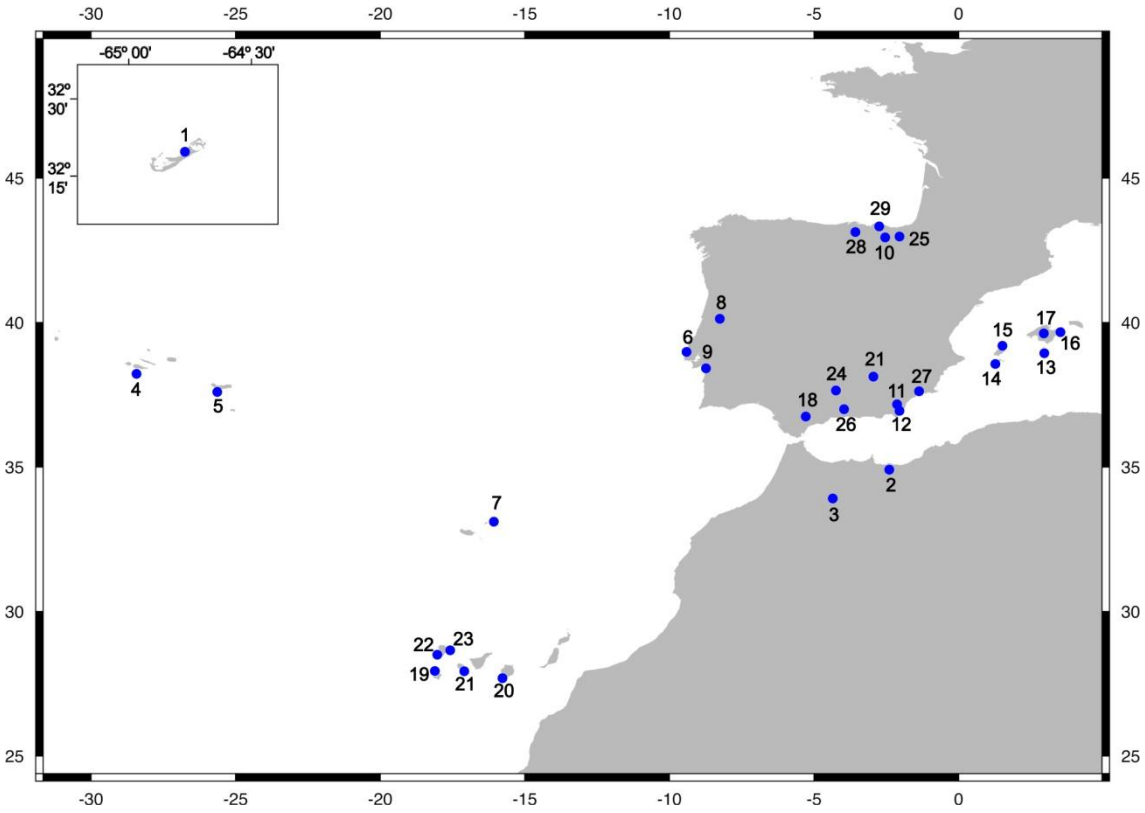


Figure 4.4.1 Map showing the localities where *Pseudoniphargus* were collected for mitochondrial genome amplification and analysis. See **Table 4.4.1** for list of species and subsequent localities.

Multiple displacement amplification (MDA)

The amplification by MDA using the 7 primers designed for this study increased the DNA yield in 19 species from 10 ng to 2661 ± 1056 ng, i.e. 106-472 fold. The agarose gel electrophoresis revealed that MDA amplification produced large DNA fragments above 10 Kb as expected (**Figure 4.4.2**). Human DNA amplified by specific primers for human mitochondrial genome included in the kit was used as a positive control, yielding a DNA amount falling within the lower end of the range (170 fold, **Figure 4.4.2**). We also included negative controls (i.e. reactions without primers) that produced no amplification (**Figure 4.4.2**).

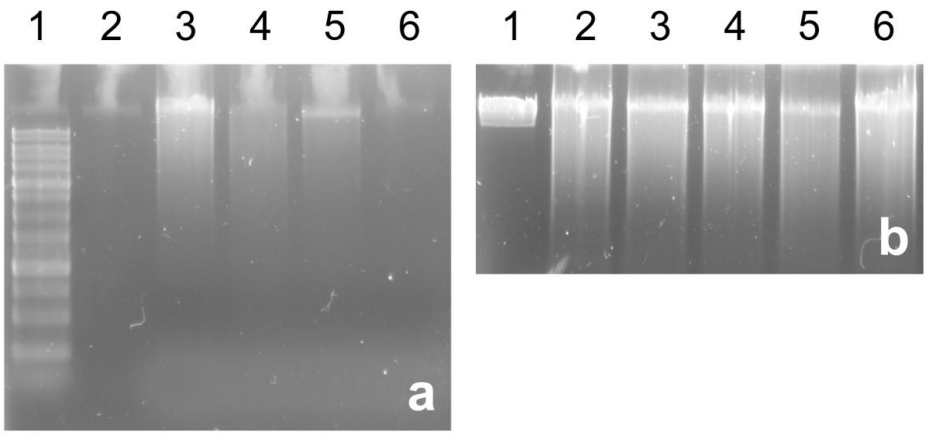


Figure 4.4.2 Agarose gel electrophoresis of genomic DNAs of several species amplified by multiple displacement amplification (MDA). a) Ladder composed of 100 pb multimers (1), negative control with human DNA but without primers (2), *P. sp1*-Basque (3), *P. pityusensis* (4), *P. pedrerae* (5), and *P. sp2*-Canaries (6). b) Positive control with human DNA plus primer for human mitogenome (1), *P. mateusorum* (2), *P. mercadali* (3), *P. associatus* (4), *P. brevipedunculatus* (5), and human DNA plus primer designed for crustacean mitogenomes (6).

We had positive amplification in 23 out of the 26 crustacean species targeted including 9 *Pseudoniphargus* species: *P. salinus*, *P. pityusensis*, *P. mateusorum*, *P. gorbeanus*, *P. brevipedunculatus*, *P.* sp2-Canaries, *P.* sp1-Basque, *P. mercadali*, and *P. associatus*. Genomic DNAs of 8-9 crustacean species amplified by MDA were pooled in equimolar concentration to construct a TruSeq library, and then 3 libraries were pooled and analyzed in a single Illumina Miseq run. Using bioinformatic tools, we could only retrieve the complete mitogenome of 4 *Pseudoniphargus* species included in the Miseq run: *P. salinus* in the first library, *P. pedrerae* and *P.* sp2-Canaries in the second one, and *P.* sp1-Basque in the last one. We also retrieved small contigs of around 2-3 Kb each, though only 1 fragment could be assigned to a particular *Pseudoniphargus* species (*P. associatus*). To test the efficiency of our method, we included in the Miseq run both human DNA amplified with the specific primers for human mitochondrial genome provided in the REPLI-g kit and human DNA amplified by our primers. The number of reads obtained from DNA amplified with human primers was higher than that produced with crustacean primers, 397,950 vs. 74,969, respectively. Despite the primers used, both human mitogenome sequences were identical (16,570 bp) and showed only 4 transitions (A/G position 203, T/C 8959, G/A 13396, and T/C 16190) and 1 deletion (position 310) relative to the sequence KF451815 from GenBank.

The amount of reads that formed part of mitogenome contigs in the GS Junior sequencing of long PCR products was around 85-90%, while it was about 0.1-1% and 0.003-0.333%, for MDA amplification in Miseq and direct WGS in Hiseq2500 and Miseq, respectively. Most of the variation found among libraries assembled using the same method might be due to quantification and pipetting errors produced at every pooling.

All 29 mitogenomes obtained in this study showed the 37 genes characteristic of Metazoans: 13 PCGs, 22 tRNA genes, plus 2 rRNA genes. Gene arrangement varied across species as described previously in three *Pseudoniphargus* species (Stokkan et al. 2016). In addition, we found 2 new rearrangements: one common in most species and another unique for *P.* sp1-Portugal (Gruta de Legaçao; Sicó). Altogether, they add up to

five distinct mitochondrial gene orders in *Pseudoniphargus*, upon which only two involve the rearrangement of protein-coding genes. We do neither discuss this topic, nor primary and secondary structures of RNA coding genes because it is beyond the goal of the present study, and most findings were already described in a previous publication (Stokkan et al. 2016).

Phylogenetic results

Each one of the 13 mitochondrial PCG sequences was aligned separately. Sequences of a few outgroup species were trimmed at the 5' and 3' ends because they differed greatly from the rest at the protein level due to the presence of unnoticed truncated stop codons. After Gblocks trimming, most of the 13 mitochondrial PCGs retained more than 94% of the original positions of the alignment except nad4L (91.26%), nad2 (90.57%), nad6 (85.41%), and the highly variable atp8 (74.55%). The individual alignments were concatenated in a single dataset comprising 10,833 bp (i.e. 3611 aminoacids or codon triplets). Gaps and ambiguities found corresponded to less than 0.1% of the total sites, mostly belonging to outgroup species. Sequences showed compositional bias towards A+T nucleotides, especially at third codon sites, as expected for mitochondrial genes. The average nucleotide composition was similar in ingroup and outgroup sequences (62.62-60.64%, 64.95%-63.44%, 70.56-72.10% for first, second, and third codon positions, respectively). Nucleotide composition was non-homogeneous across species. Twenty out of the 51 species included failed in the Chi square homogeneity test in first codon sites, 5 taxa in second positions, and 25 sequences in third codon bases. We assessed occurrence of saturation on the third codon sites since many studies suggest that this issue can compromise an accurate estimation of tree topology but particularly of branch lengths and node ages (Arbogast et al. 2002). The Xia test estimates that Iss values for a symmetrical tree topology are statistically lower than Issc ones ($p<0.001$), and showed there is little saturation on first and second codon sites (0.3 and 0.4 vs 0.8, respectively). A similar pattern was found on third codon sites though with higher Iss values (0.79 vs. 0.83). This analysis was performed including outgroup species and hence saturation level within the ingroup should be lower. The best partition scheme starting from 39 partitions (13 PCGs per three codon positions) was by grouping them in 4 sets: (1) all first codon sites plus second and third codon positions of atp8 gene; (2) remaining second codon sites; (3) third codon sites of genes coded on the positive strand except atp8; and (4) third codon sites of genes coded on the negative strand (nad1, nad4, nad4L, nad5). PartitionFinder also identified GTR+G as the best fitting model for each of the 4 partitions. BIC scores improved further when second codon positions of atp8 gene were included together with the other second codon sites (partition 2), and third codon sites of atp8 with the third codon sites on positive strand (partition 3). This alternative scheme was selected as the preferred one for further analyses (**Table 4.4.2, Figure 4.4.3a)**. We also assessed the impact of different partition schemes on the phylogenetic signal using BIC values (**Table 4.4.2**). The overall trend suggested a major improvement of BIC scores while genes were split by codon positions rather than by genes. The split of the third codon sites by strand reduced the BIC scores but at a slower pace. The tree topologies estimated with lower fitting partition schemes were very similar to that retrieved with the best BIC score (**Figure 4.4.3a**). Support values were high for most nodes irrespective of the partition scheme implemented in the ML analyses. Similar results were also found when tree searches were run under parsimony criterion, treating data as 62 codon categories (GY98), and

using complex Bayesian models allowing each nucleotide position to fit in one of the multiple categories with different composition and rate (CAT model; **Figure 4.4.3a**). At the aminoacid level, the best partition scheme estimated by PartitionFinder was grouping the 13 PCGs in 2 sets (**Table 4.4.2**), though slightly better BIC scores were obtained when they were analyzed as a single partition with mtArt or mtZOA models (**Table 4.4.2**). The tree topologies and node support values retrieved at the aminoacid level were congruent to those achieved with nucleotide sequences (**Figure 4.4.3a**). Bayesian analyses at the aminoacid level with the CAT model were extremely congruent as well (**Figure 4.4.3b**).

We also assessed the occurrence of saturation by plotting branch lengths estimated under the simple substitution model F81+G, which does not resolve multiple substitutions per site, versus the complex GTR+G one that is generally capable to retrieve multiple hits. We used the tree topology retrieved with the preferred partition scheme as a fixed constrain. The branch lengths estimated on first codon sites with a simple F81+G were used as a non-saturated model since they had stronger phylogenetic signal than second codon positions. Branch length estimated for first codon sites under the GTR+G model showed a slight increment in the number of substitutions, but nearly perfect correlation ($R2 > 0.99$) and no branch length saturation (**Figure 4.4.4**). Second codon sites showed lower phylogenetic signal than first codon positions not showing any signal of saturation ($R2 > 0.99$) . Third codon sites showed a different pattern in where the complex GTR+G model retrieved a larger number of substitutions than the simpler F81+G. Moreover, the simpler model showed high saturation on the longest branches that decreased by using GTR+G. Correlation of branch lengths estimated for third codon sites (GTR+G) versus first codon sites (F81+G) was also high ($R2 = 0.95$) with little saturation (i.e. strong phylogenetic signal). In fact, the use of a complex model had little impact on first and second codon sites but high on third ones since correlation between first codon sites (F81+G) vs. third codon sites (F81+G) was lower ($R2=0.85$).

Table 4.4.2 Comparison of different partition schemes based on Bayesian information Criterion (BIC, BIC = -2Ln + Kln(n), where Ln is the tree Likelihood, K the number of free parameters of the models including the number of branch lengths 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 abreviatures GY indicate the Goldman and Young codon model. * 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 *atp8*, *coob*, *nad2*, *nad3*, *nad4*, *nad5*, *nad6*, *nad4l* genes and the model JTT+F+I+G. Plus and minus refers on which strand are genes coded.

partition	Ln	TL	K	n partitions	BIC	% Δ BIC
Nucleotide Models						
Partition Finder Best (see results)	-326434.277	36.724	432	4	656881.987	0.000
codon1,codon2,codon3plus,codon3minus	-326362.065	36.936	432	4	656737.561	-0.022
codon1,codon2,codon3	-327496.221	31.639	324	3	658002.517	0.171
codon1&codon2,codon3	-329269.870	31.654	216	2	660546.456	0.558
bygene & bycodon	-322049.672	42.583	4212	39	683230.307	4.011
bygene & bycodon (HKY+G)	-323077.401	42.482	4056	39	683836.470	4.103
bygene	-331861.427	32.601	1404	13	676766.508	3.027
single	-335344.159	25.314	108	1	671691.677	2.255
Codon Model						
codon GY(HKY)+F3X4+G	-324804.524	244.292	111	1	650518.330	NA
Aminoacid Models						
Partition Finder Best *	-141189.818	20.615	238	2	284329.270	0.000
single mtART+F+G	-142123.033	27.136	119	1	285220.882	0.314
single mtZOA+F+I+G	-141325.081	23.948	120	1	283633.171	-0.245
Genes plus, Genes minus (mtZOA+F+I+G)	-140867.106	23.931	240	2	283700.230	-0.221

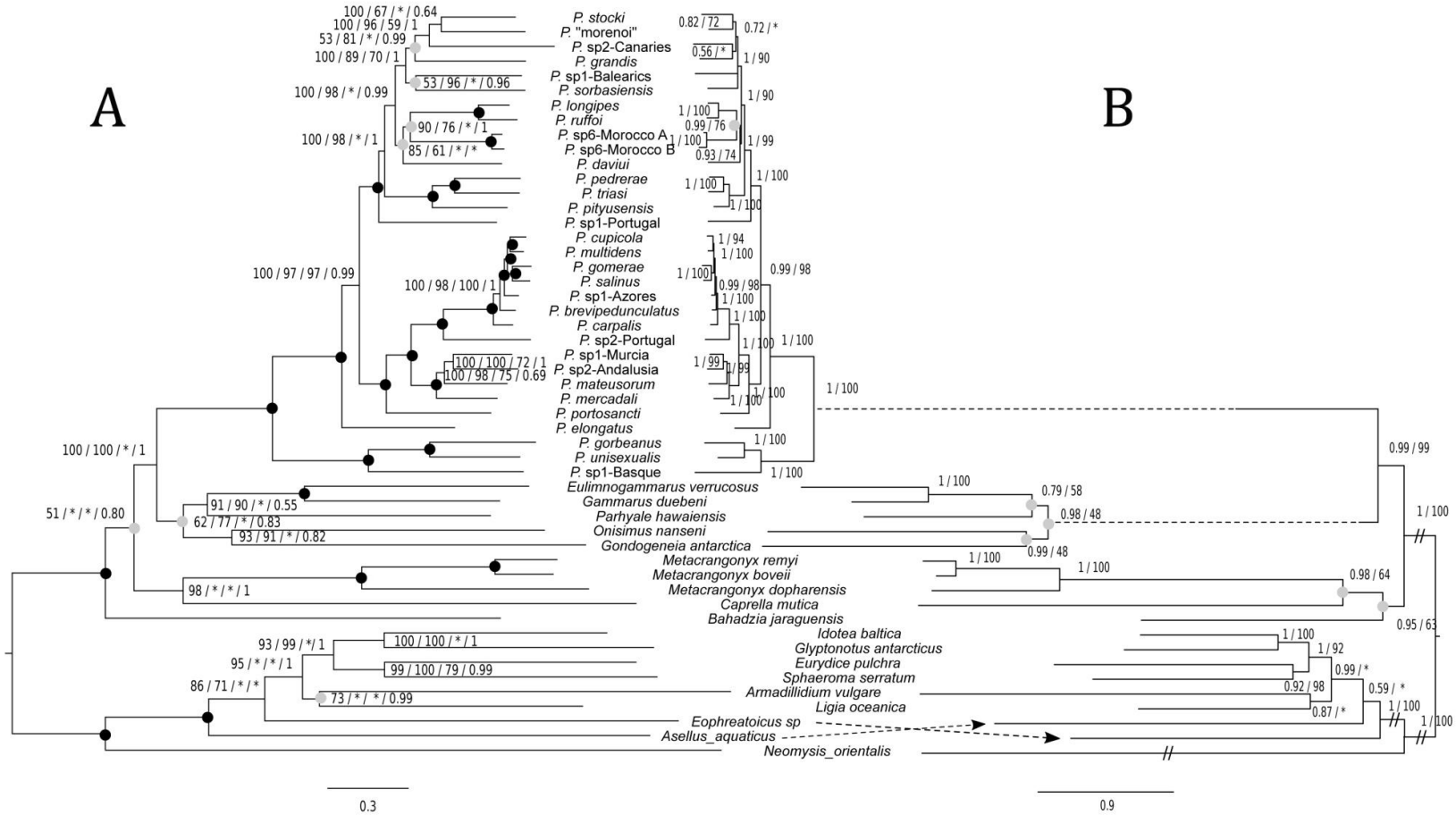


Figure 4.4.3 A) ML topology obtained from mitochondrial protein coding genes at the nucleotiode (nt) level using the best partitioning scheme (1st codon, 2 codon, 3rd codon of genes coded on plus strans , 3rd codon sites of genes coded on the minus strands). Numbers on nodes indicate support values: for best nt /ML codon GY model / parsimony / Bayesian with GTR+CAT model. Nodes with full support for all methods are marked with a black dot and grey one indicates that those nodes wererecovered by 60-85% of the ML analyses with 8 different partition schemes (**Table 4.4.2**). **B)** Bayesian tree topology estimated at the protein level in PhyloBayes GTR+CAT. Support values as follow: Bayesian GTR+CAT / ML with a single mtZOA model. Grey dots highlights the few nodes not supported in all 4 ML analyses made at the protein level (See **Table 4.4.2**)

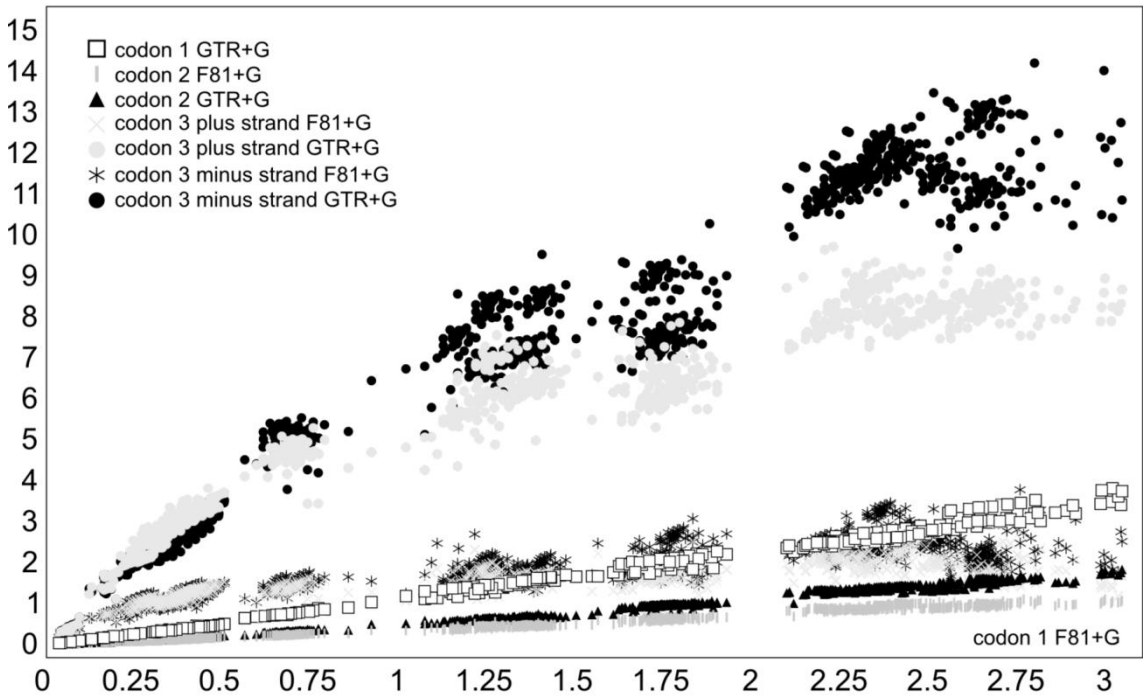


Figure 4.4.4 Assesment of saturation levels in the different codon sites as well as with different substitution models.

Age estimation

We estimated node ages in a Bayesian framework using 2 fossil constrains under different partitions, clock and tree (i.e. diversification) models that were statistically compared using Bayes Factors (**Table 4.4.3**). Since the use of third codon sites has been heavily criticized due to saturation, we also estimated node ages using only first and second codon positions. However, results cannot be compared with those obtained with the full dataset since the number of positions differs. As shown above, splitting data by codon site, and by strand on the third codon positions increased the fitness of the model, and this trend was also found when implementing an independent clock for each partition. The best clock model fitting the data was a relaxed clock with an uncorrelated log-normal distribution, followed by allowing a more relaxed framework with local random clocks, and finally under a strict clock. The best scheme by Bayes Factors was dividing nucleotide sequences in 4 sets with and independent uncorrelated log-normal clock for each one and a Yule diversification model (1st, 2nd, 3rd positive and 3rd negative). The most common recent ancestor of the genus *Pseudoniphargus* was estimated to live 165.73 Ma (146.43-185.15), and the age of the lineage to be 252.98 (228.35-279.1; **Figure 4.4.5**). The use of the sphaeromatidean isopod fossil as single node constrain rendered nearly identical results than using the phreatoicidid isopod fossil as unique constrain. The Birth-Death model was rejected over the Yule model though estimated age differences between them were low (**Table 4.4.3**). The analyses including the first and second codon sites showed the same trend than above but retrieved slightly older ages **(Table 4.4.3**; **Figure 4.4.5**)

Table 4.4.3. Comparison of 20 bayesian analysis under different partition schemes, clock relaxations and models of diversification based on Bayes factors. Marginal likelihood were estimated by pathsampling in BEASt v1.8.2. Specific local clocks were implemented isopods, *Pseudoniphargus* and remaining amphipods (*9 clocks since we differentiated them by codon sites) thouhgh the later grouping was split as *Metacrangonyx* vs remaining amphipods (**12 local clocks). Models without third codon sites or recoded as RY can not be compared with full data set analyses since they do not have the same number of nucleotide positions. The best substitution model for each individual partition was a GTR+G except for third codon position recoded as RY that was F81+G. All analyes are based on two fossil ages defined as hard constraints with a log-normal distribution with a mean (M) 5.71 and Standard deviation 0.01 (CI: 296.0-307.8) for *Hesslerella* and M 5.35 and S 0.01 (CI: 206.5-214.8) for *Elioserolis*. Abreviatures: independent (ind), confidence interval (CI).

partition scheme	N partitions	clock model	Diversification model	marginal Ln	Bayes Factors
no partitioning	1	single relaxed log-normal clock	Yule	-335444.44	16571.83
codon1+codon2,codon3	2	single relaxed log-normal clock	Yule	-333495.87	12674.70
codon1+codon2,codon3	2	2 ind relaxed log normal clocks	Yule	-329809.05	5301.05
codon1,codon2,codon3	3	single relaxed log-normal clock	Yule	-331975.01	9632.98
codon1,codon2,codon3	3	3 ind relaxed log-normal clocks	Yule	-327274.55	232.06
codon1,codon2,codon3	3	3 ind relaxed log-normal clocks	Birth-Death	-328174.88	2032.71
codon1,codon2,codon3	3	3 ind strict clocks	Yule	-328916.40	3515.75
codon1,codon2,codon3	3	9 ind fixed local clocks *	Yule	-328582.46	2847.88
codon1,codon2,codon3	3	3 ind local random clocks	Yule	-328321.98	2326.91
codon1,codon2,codon3+codon3-	4	4 ind relaxed log-normal clocks	Yule	-327158.52	0.00
codon1,codon2,codon3+codon3-	4	4 ind relaxed log-normal clocks	Birth-Death	-327164.86	12.66
codon1,codon2,codon3+,codon3-	4	4 ind strict clocks	Yule	-327908.61	1500.18
codon1,codon2,codon3+codon3-	4	12 ind fixed local clocks **	Yule	-327552.65	788.26
codon1,codon2,codon3+codon3-	4	4 ind local random clocks	Yule	-327324.93	332.81
codon1,codon2,codon3+codon3-	4	4 ind relaxed log-normal clocks	Yule Cirolanidae fossil only	-327169.83	22.61
codon1,codon2,codon3+codon3-	4	4 ind relaxed log-normal clocks	Yule Isopoda fossil only	-327168.33	19.62
codon1,codon2,(codon3 sites removed)	2	2 ind relaxed log normal clocks	Yule	-151432.12	0.00
codon1,codon2,(codon3 sites removed)	2	2 ind local random clocks	Yule	-151455.41	46.58
codon1,codon2,(codon3 sites removed)	2	single local random clock	Yule	-151796.58	728.92
codon1,codon2,codon3 recoded as RY	3	single relaxed log-nornal clock	Yule	-151734.92	N/A

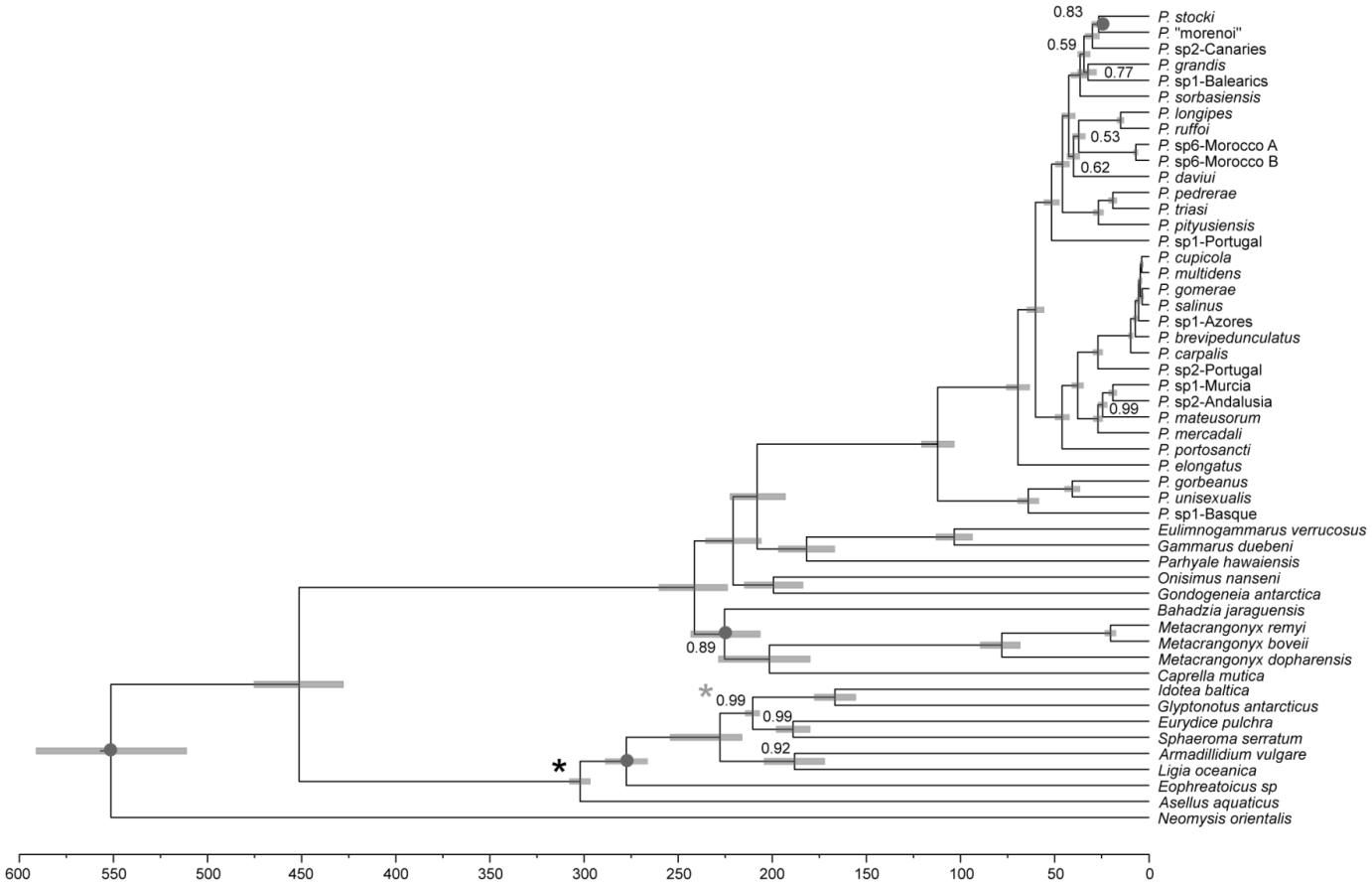


Figure 4.4.5 Divergence time estimates using two fossil calibration constrains (*) based on mitochondrial protein-coding genes of 31 *Pseudoniphargus* species plus 19 outgroups. This analysis implemented a Random Local Clock (RLC) instead of a relaxed log-normal clock (rc) although partition schemes and substitution models were identical to the best fitting model. Two fossil ages were defined as hard constraints with a log-normal distribution with a mean (M) 5.71 and Standard deviation 0.01 (CI: 296.0-307.8) for *Hesslerella* and M 5.35 and S 0.01 (CI: 206.5-214.8) for *Elioserolis*. Nodes with grey pointdots are nodes that were not recovered in all 16 analyses including all positions. Bars across nodes represent the 95% credibility interval.

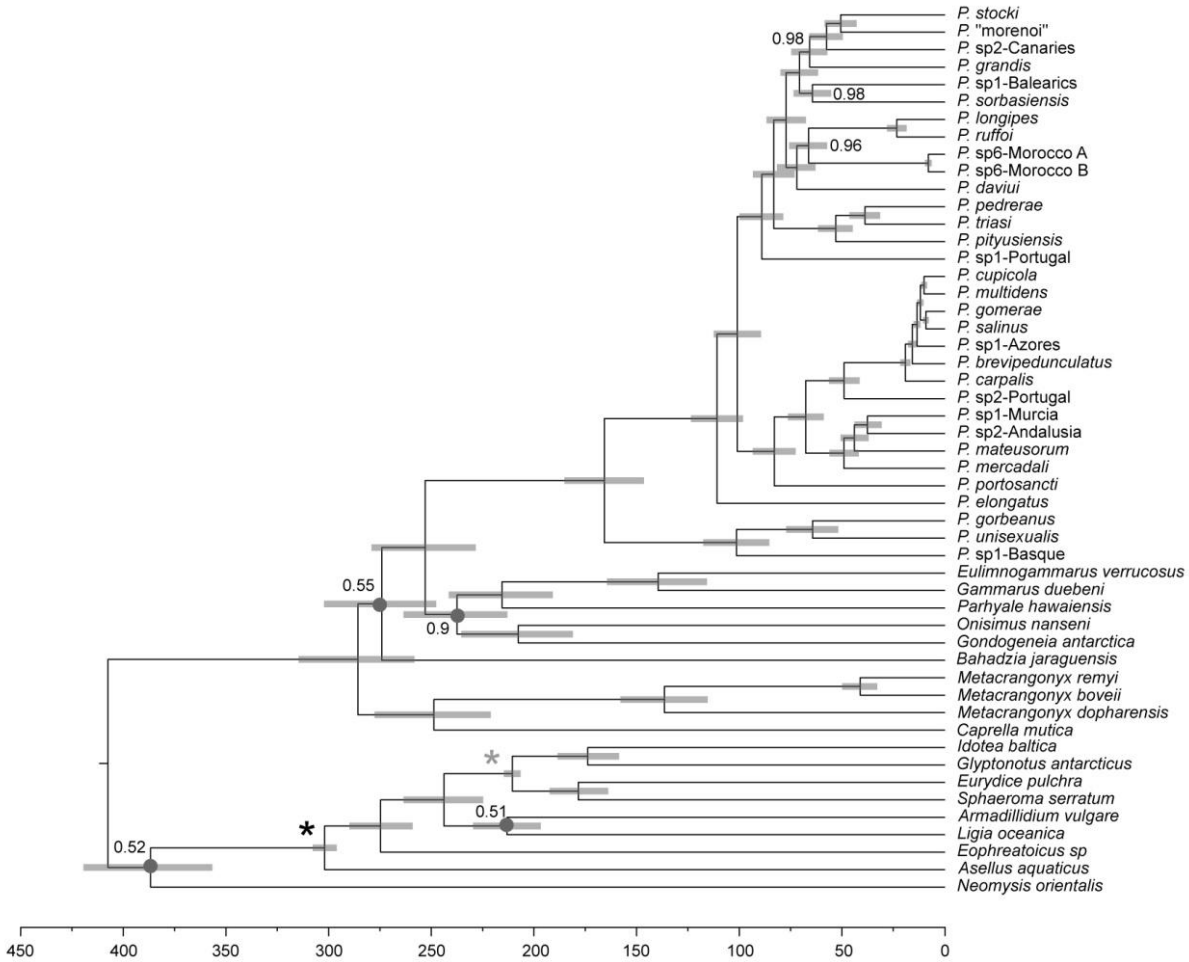


Figure 4.4.6 Divergence time estimates using two fossil calibration constrains (*) based on mitochondrial protein-coding genes of 31 *Pseudoniphargus* species plus 19 outgroups. This analysis implemented the best partitions, and substitutions and clocks models fitting the data (see Suppl. Table S4.2.2) . Two fossil ages were defined as hard constraints with a log-normal distribution with a mean (M) 5.71 and Standard deviation 0.01 (CI: 296.0-307.8) for *Hesslerella* and M 5.35 and S 0.01 (CI: 206.5-214.8) for *Elioserolis*. Nodes with grey dots were not recovered in all 16 analyses including all positions. Bars across nodes represent the 95% credibility intervals.

The global rates estimated for ingroup and outgroup were lower than expected for a mitochondrial genome (0.00575 vs. 0.0115 nucleotide substitutions per site, per Ma and per lineage, NSSML, for standard rate) with even lower rates for the ingroup species. Moreover, several splits within the genus *Pseudoniphargus* were fully incompatible with the geological events that seem to have driven the diversification of several sister species. On the other hand, both fossil constraints resulted in an age of about 40 Ma (32.7-49.9) for the mrca of *Metacrangonyx remyii* and *M. boveii* which is congruent with the uplift of the Marrakech High Atlas in Morocco (37.2–25.0 Ma) that presumably caused the split of these species located in valleys in the opposite slopes on this mountain range (Bauzà-Ribot et al. 2012). In fact, the rate estimated for the 21mitochondrial sequences of *Metacrangonyx* was 5.45 NSSML, which is about 5 times higher than the "standard" rate for mitochondrial invertebrate species (0.0115 NSSML; Brower, 1994). These results suggested the existence of branches or lineages with higher divergent rates, hence, we estimated tip-to-root lengths for all species on a non-ultrametric ML tree based on 3 partitions (by codon) and independent GTR+G models including the 21 *Metacrangonyx* species considered in a previous study (Bauzà-Ribot et al. 2012). We did not differentiate between third codon positions from genes coded on positive and negative strands since the cytochrome b gene is coded on the negative strand in *Metacrangonyx* but in the opposite strand in the remaining taxa. The box plot of branch lengths showed that number of nucleotide substitutions (i.e. rates) in outgroup isopod species (including fossil constraints) and the species of the amphipod genus *Metancragonyx* are 40% longer that those from *Pseudoniphargus* and other amphipods (**Figure 4.4.7**).

This bimodal or multimodal distribution of the rates cannot be handled appropriately neither by relaxed log-normal clocks, random local clocks nor even by local clocks in Beast. Hence, node ages cannot be estimated with confidence when all age fossil constraints are located within a clade whose rates are extremely different from those exhibited by our focal ingroup clade.

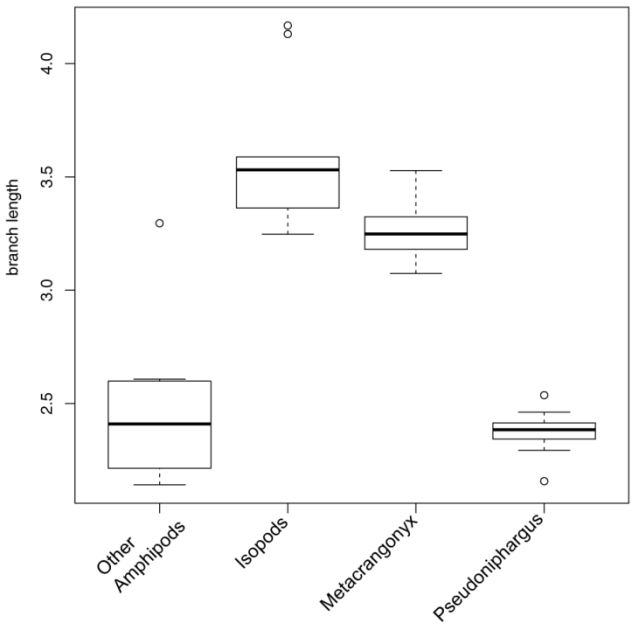


Figure 4.4.7 Boxplot illustrating the different substitution rates between *Pseudoniphargus*, *Metacrangonyx*, Isopods (calibration point) and other amphipods.

In order to overcome this Gordian knot, we selected 3 geological events (see Material and Methods) as node constraints to estimate the remaining node ages due to the lack of relevant amphipod fossils. We also performed a cross-validation of the results by implementing individual constraints. We did not include outgroup species in these analyses since isopods and *Pseudoniphargus* display a very different rate and the remaining amphipods considered are not the focus of the present study. We also analyzed how different partition schemes and clock models affected age estimates and rates by ranking them using Bayes Factors. The best partition scheme was by partitioning data by codon site plus differentiating third codon sites from genes coded on the positive and negative strands. The merging of partitions decreased the fitness of the models (**Table 4.4.3**). On the other hand, the inclusion of independent clocks for each partition improved the fitness except for the differentiation of third codon sites by coding strand. The enforcing of random local clocks or Birt-Death models instead of log-normal clocks and Yule model did not improve the fitness of the model based on Bayes Factors. The topologies and ages derived were very similar despite the model implemented though those with lower Bayes Factors retrieved the most divergent topologies and ages (**Figure 4.4.8**). The 3 palaeogeographic events implemented as constraints estimated very similar ages except La Palma constraint, that retrieved younger ages. The best model with 4 independent partitions (first, second, third positive, third negative), 3 independent clocks (first, second, third) and a Yule diversification parameter estimated an age of 52.8 Ma (CI 46.5-59.5) for the mrca of *Pseudoniphargus* species, and a diversification age for the mrca of the North Spain clade and its sister clade at ca. 32 Ma (**Figure 4.4.9**).

Table 4.4.3 List of Bayesian analyses constrained under different molecular clocks, partition schemes, and diversification models ranked on Bayes Factors. Marginal likelihood were estimated using a pathsampling method which values were used to calculate Bayes Factors. Data was partitioned as four partitions (1st codon, 2nd codon, 3rd codon plus strand, 3rd codon minus strand), three partitions (1st, 2nd ,3rd), and two partitions (1st+2nd, 3rd). Abbreviations: rc (relaxed clock with a log-normal distribution), RLC (random local clocks), and strict (strict clock).

model	Marginal Ln	Bayes Factors
4part 3rc Yule	-162220.017	0.000
4part 3rc Bith-Death	-162235.267	-30.499
4part 4rc Birth-Death	-162237.696	-35.357
4part 4rc Yule	-162243.669	-47.304
4part 3RLC Yule	-162302.005	-163.976
4part 4RLC Yule	-162331.782	-223.529
4part 2rc Yule	-162416.241	-392.448
4part 4strict Yule	-162441.814	-443.593
4part 3strict Yule	-162450.010	-459.986
3part 3rc Yule	-163814.156	-3188.278
4part 1rc Yule	-164341.140	-4242.246
2part 2rc Yule	-164447.132	-4454.230
2part 1rc Yule no3rd	-63545.708	0.000
2part 2rc Yule no3rd	-63605.540	-119.663
2part 2rRLC Yule no3rd	-63637.950	-184.484
2part 1RLC Yule no3rd	-63833.142	-574.868

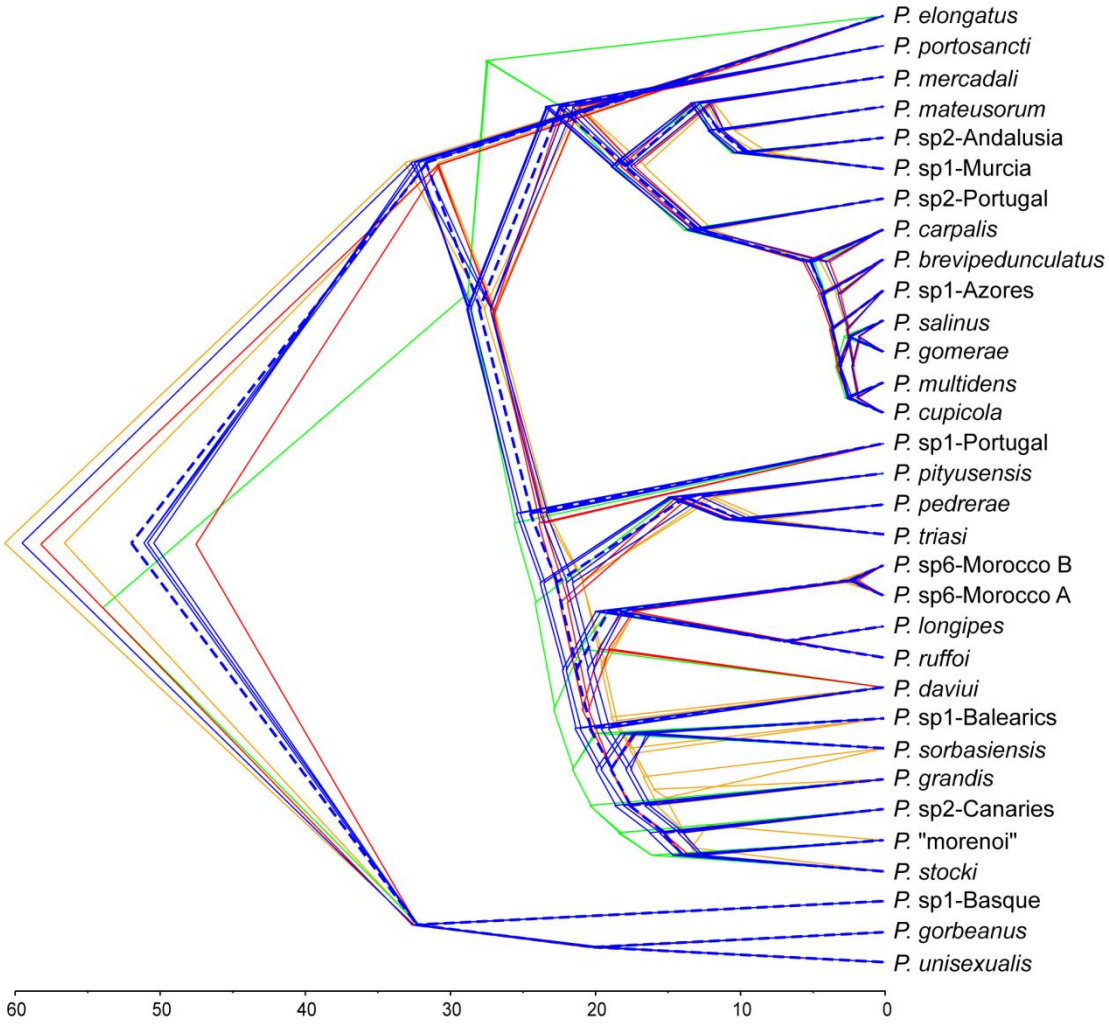


Figure 4.4.8 Comparison of tree topologies (in different colors)) and node ages for the different analyses in table 4.4.3. Blue dashed line best tree with 4 codon partitions 3 rc, blue whole line: 4 partitions 3 RCL, first orange (left): 2 partitions 1 RLC but without 3rd codon sites, second orange (right): 2 partitions 2 rc without 3rd codon sites, first red (left): 4 partitions 4 RLC, and second red (right): 2 partitions 2 rc. All model implemented a Yule diversification model, Abbreviations: rc (relaxed clock with a log-normal distribution), and RLC (random local clocks)

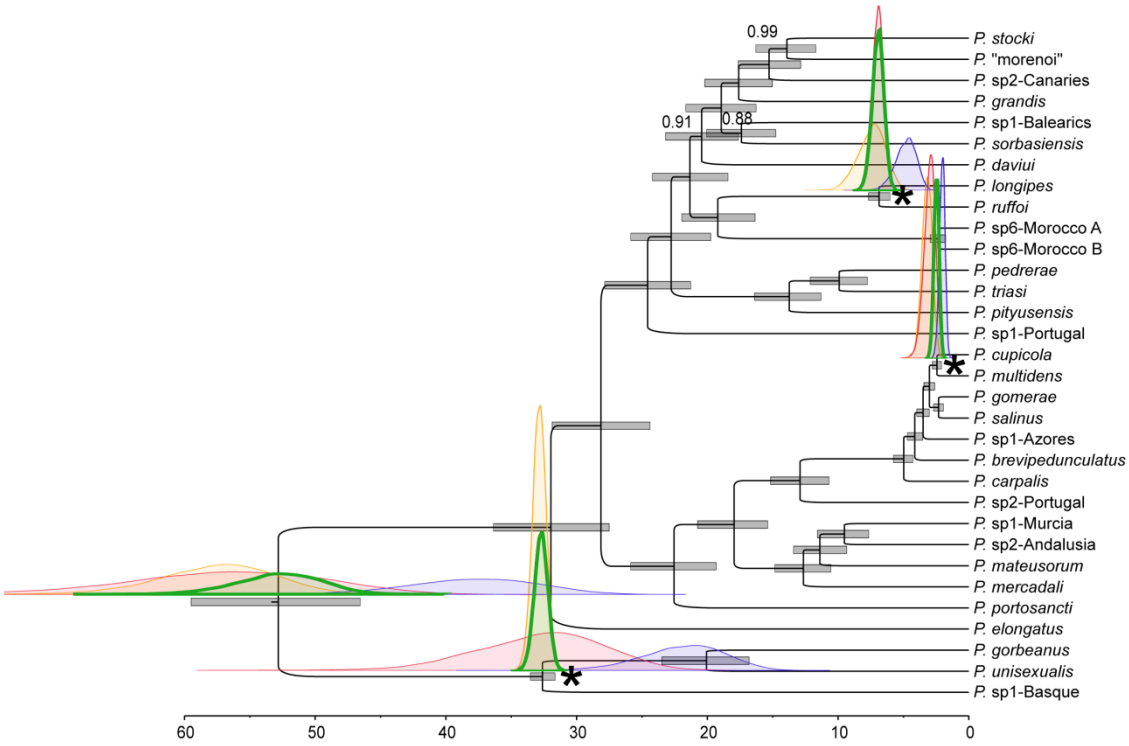


Figure 4.4.9 Divergence times for *Pseudoniphargus* species estimated from Bayesian analysis of all mitochondrian protein-coding genes using three paleogeographic events as constrains (*) and the best partitions, and substitutions and clocks models. Bars across nodes represet the 95% highest probability density intervals. Orange only the Emersion of the Basque Country constrain was only implemented, red the Southern Rifian Corridor only and blue only the emersion of La Palma and green included all three constrains.

The rates estimated based on these 3 palaeogeographic events for the best scheme was of 0.0223 NSSML (CI 0.0197-0.0251 95%), or a 4.46% pairwise distance. Results did not vary significantly using random local clock, strict clock or a Birth-Death model (0.0224, 0.0212, and 0.0223 NSSML, respectively). Adding or merging clocks and/or partitions had a slight impact on rates (0.0182-0.0223) except when using 4 partitions under single log-normal clock (0.0377). The analysis of single constraints under the best model retrieved higher rates only for La Palma (0.0308) but not for the Basque Country (0.0201) and Morocco (0.0206). The analysis of the most commonly used phylogenetic marker, the mitochondrial cox1, as a single partition with the GTR+I+G model and a single log-normal clock –as most phylogenetic analysis do– with closely related species retrieved a rate (0.0156 NSSML or 3.12% pairwise divergence) that is closer to the classical Browser's clock (0.0115 or 2.3%).

Most of the phylogenetic signal resided in the third codon sites as expected, followed by first and second codon positions, which are about 50% and 25% of 3rd codon positions, respectively (**Figure 4.4.7**). The signal of third codon sites peaked very fast at about 6 Ma, and then slowed down very fast, being about 50% of the peaking signal at about 24 Ma to descend to 25% at about 50 Ma. Saturation is extremely slow at first codon sites, reducing their signal by half after 50 Ma, and being insignificant at second codon positions.

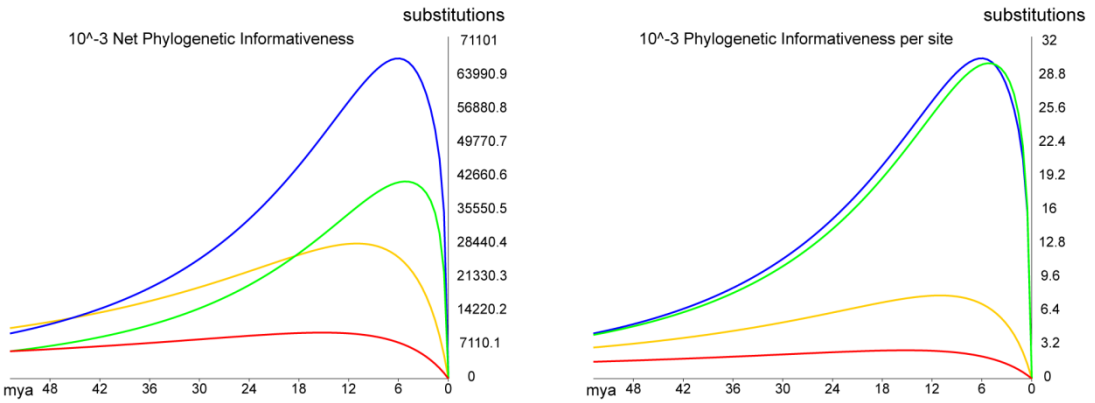


Figure 4.4.10 Net (left) and per site (right) phylogenetic signal (i.e. number of substitutions) over time. Blue 3rd codon sites coded on plus strand, Green: 3rd codon sites coded on minus strand, Yellow: 1st codon sites , Red 2nd codon sites .

Discussion

Comparing sequencing methods

As expected, we obtained the best results with WGS at low coverage with Illumina sequencers, although tagging libraries with species-specific indexed adapters allowed a faster assembly with less effort, particularly in those regions with low coverage (Timmermans et al. 2010). Control regions were more difficult to complete due to the presence of long poly-A stretches but there was a huge improvement compared with 454 sequencing. The accuracy of Illumina reads also improved quality and speed of the annotation, resulting in less manual curation. Despite MDA was only used once, we got better performance with high quality and fresh DNA rather than with old samples. This could be explained because fresh DNA has longer fragments, and hence more multi-priming sites that increase the exponential amplification, i.e. more DNA yield (Blanco et al. 1989). Our MDA results are incipient but promising, though amplification, rather than specific for mitochondrial sequences only, was rather unspecific instead. In fact, in order to get an exponential and specific amplification of mitochondrial genome by MDA such as that from human DNA from kit, the design of many short species-specific primers evenly distributed throughout mitogenome are needed.

Fossil calibration and rate heterochaty

Our results pointed out that in presence of strong heterogeneous substitution rates between outgroup (isopods) and ingroup species (*Pseudoniphargus*), and with fossil constraints present in outgroups only, the estimated rates for the ingroup clade are extremely slow (0.57%) compared with the standard mitochondrial rate (1.15%), and in turn ages are older than expected. However, ages estimated in *Metacrangonyx* species were similar to those predicted in a previous study based on 2 palaeogeographic events (Bauzà-Ribot et al. 2012). The extreme difference found in rates between isopods and *Pseudoniphargus* but close similarity, despite being statistically different, between isopods and *Metacrangonyx* could explain the observed pattern. Both underestimation

of rates and overestimation of ages in *Pseudoniphargus* could not be overcome neither by implementing UCLN nor even using RLC or several local clocks as implemented in Beast. The occurrence of Heterotachy, i.e. of heterogeneous (clade-specific) rates of nucleotide substitution in different lineages of a particular phylogeny has been reported to occur in groups such as viruses (HIV-1; Wertheim et al. 2012), angiosperms (Beaulieu et al. 2015) or dolphins (Dornburg et al. 2012) and can blur the accurate estimation of lineage ages within a particular taxonomic group.
Similar drawbacks could not be solved in cetaceans using similar clock relaxations since rates significantly change between dolphins and whales, and even several times across whales (Donburg et al. 2012). These authors also corroborated using simulations that an abrupt change in rate in a particular or a few lineages can render misleading age and rate estimates despite the focal node was calibrated. This study also suggested that the effect of overestimation could be alleviated if several calibration points spanning both clades were included (Donburg et al. 2012). These results suggest that datasets including heterochateous lineages should implement relaxation of the molecular clock with multimodal or joint distributions rather than unimodal distributions such as exponential, log-normal, or Poisson distributions. The analysis of the crown age of the angiosperms (Beaulieu et al. 2015) also revealed that the presence of heterogeneous (clade-specific) rates of nucleotide substitution in different lineages of flowering plants could explain differences found between ages assumed from the fossil record and those estimated from phylogenetic analysis enforcing relaxed molecular clocks. In fact, many studies showed that differences between herbaceous and woody clades such as in growth habit are correlated with variability in rates of nucleotide substitution (Beaulieu et al. 2015, and references therein).

In viruses, the analysis of different subtypes of the virus HIV-1 showed that the estimation of the age of each subtype was different if they were analyzed in a single tree and single relaxed clock or dealing with each subtype alone in a separate analysis (Wertheim et al. 2012). These authors assessed several approaches and simulations to overcome an overestimation of the root in the combined analyses of heterotacheous lineages since current dating methods are unable to deal with heterotachy. However, none of these approaches succeeded in overcoming this drawback, and confirming that current models overestimate the age of the root.

Our results corroborate that an optimal partitioning of the mitochondrial genes by codon site, and the implementation of an independent relaxed clock per partition, generally improved the fitness of the model based on Bayes Factors estimated from marginal likelihood and path-sampling, as previously shown elsewhere (Pons et al. 2010; Jurado-Ribera et al. in press). The partitioning by coding strand also improved fitness further, but at lower pace. Besides, the implementation of a relaxed clock with UCLN distribution instead of a strict clock improved also the fitness of the model. Surprisingly, RLC and tree models using Birth-Death diversification instead of pure birth (Yule) diminished fitness rather than improving it. The models including RLC generally estimated older ages on the root but not on the other nodes. On the other hand, the removal of third codon sites, the most saturated positions, did not affect node ages, corroborating the low level of saturation in *Pseudoniphargus* as suggested by Xia test.

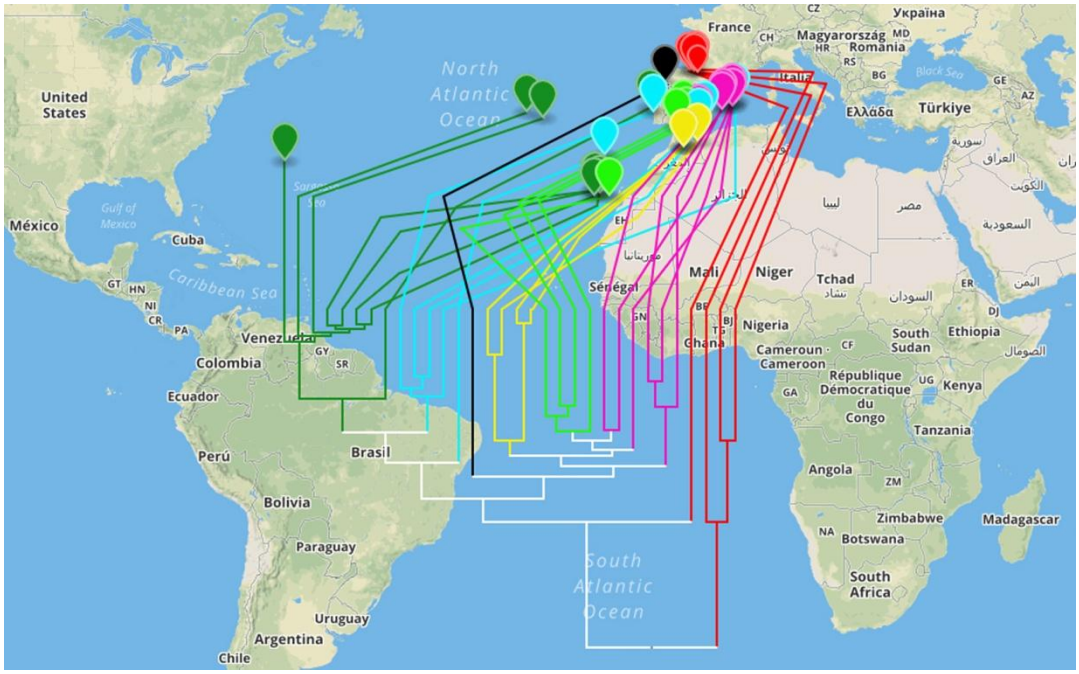


Figure 4.4.11 Map showing main clades and distribution of the 32 mitogenomes of Pseudoniphargus. This is best read as a functional interactive map at the following web address: http://bl.ocks.org/anonymous/raw/1939d97ec1fcaa96b499b8cc9daf1beb/

Biogeography of Pseudoniphargus

The tree recovered in our study strongly supports the existence of 3 main clades within the genus *Pseudoniphargus*, recognizing the one including the species distributed around the Gulf of Biscay in northern Spain as the oldest lineage (See **figure 4.4.11**). This is in agreement with former studies on the phylogenetic relationships of species based on morphological traits (Notenboom 1986; 1988b).

Gulf of Biscay

The area of the Iberian Peninsula adjacent to the Gulf of Biscay has been colonized at least twice. The first colonization episode was carried out by the ancestor of the cluster of species from the Basque Country (western edge of the Pyrenees), which represent the oldest offshoot in the *Pseudoniphargus* tree. The three species of this group whose mitogenomes were sequenced (viz. *P.* sp1-Basque, *P. unisexualis* and *P. gorbeanus*) conform a solidly supported monophylum that presumably diversified within the area currently corresponding to the Basque Country once it emerged from the sea at the Early Oligocene (ca. 33.7 Ma; Rögl 1998). Consequently, species diversification could not be older than that age.

A second colonization episode of the area involved the ancestor of *P. elongatus*, which is distributed throughout Santander and the north of Burgos provinces in northern Spain.

Mainland Portugal

The *Pseudoniphargus* species recorded in mainland Portugal derive from at least 2 independent colonization episodes. These involved the respective ancestors of *P. mateusorum* (Setúbal) and *P*. sp2-Portugal (Gruta de Assafora; Sintra) on the one side, and of *P*. sp1-Portugal (Gruta de Legaçao; Coímbra) on the other.

Macaronesia + Bermuda

Our phylogeny suggests the past occurrence of at least 3 independent colonization episodes of Bermuda and the Macaronesian archipelagoes (Canaries, Madeira and Azores). One led to the occupation of Porto Santo Island in Madeira by the ancestor of *P. portosancti*. This island is under subaerial exposure since 22 Ma and represents, together with the island of Fuerteventura in the Canaries (21 Ma) –which is apparently devoid of *Pseudoniphargus*–, the oldest continuously emerged land in this cluster of North Atlantic archipelagoes (Fernández-Palacios et al. 2011). This correlates well with our dating for the origin of the species *P. portosancti* at around 23 Ma.

The second episode may have led to the colonization and diversification of the genus on Bermuda (*P. carpalis*) and the islands of La Gomera (*P. gomerae*), La Palma (*P. cupicola* and *P. multidens*) and el Hierro (*P. salinus*) in the Canaries; and Faial (*P. brevipedunculatus*) and Sao Miguel (*P*. sp1-Azores) in the Azores. The ages of emergence of these territories vary broadly between the 12 Ma of La Gomera to only 1.1 Ma for El Hierro in the Canaries, whereas those of the Azorean islands is much more recent: 0.73 Ma for Faial, 0.25 Ma for Pico, and 4.01 Ma for Sao Miguel (Azevedo & Portugal Ferreira 2006). Since our phylogeny points out that the species present on Sao Miguel Island is derived from *P. brevipedunculatus*, an inhabitant of the younger Islands of Faial and Pico, the occurrence of a transoceanic dispersal event between these islands cannot be ruled out. The age of the Sao Miguel species correlates well with the emergence of the island. The age of Bermuda is significantly older (47-34 Ma; Vogt and Young 2007) than the date assigned in our analysis to the speciation there of *P. carpalis*.

The presence of two sister taxa on La Palma Island (Canaries) enables to calibrate this node as no older than the age of emergence of this island, i.e. 3.5 Ma (Carracedo et al. 2001). It is worth mentioning here that our analyses based on cox1 sequences including additional taxa from all these Atlantic archipelagoes and whose mitogenomes were not sequenced show that they should be included in this colonization episode. Namely: *P. grandimanus* (Bermuda); *P. fontinalis* (Tenerife); the populations of *P. brevipedunculatus* from Pico in the Azores; *P*. cf. *cupicola* (sensu Stock 1988) from Charco de los Chochos at La Palma; and *P.* sp1-Canaries from Mina de los Llanetes at Gran Canaria.

The island of Gran Canaria harbors representatives of a third colonization wave that involved the ancestor of *P*. sp2-Canaries from Mina Los Roques. The island is exposed subaerially at least since 15 Ma. As previously stated, it harbors also a representative of the second colonization wave (*P*. sp1-Canaries from Mina Los Llanetes).

Southern Spain

The area corresponding to southern Spain was presumably colonized independently at least twice. These 2 colonization waves were carried out by the common ancestor of the species from Isla Plana (Murcia; *P*. sp1-Murcia) and Pozo de los Frailes (Almería; *P*. sp2-Andalusia) on the one side; and the ancestor of *P. sorbasiensis* (Almería), the common ancestor of *P. stocki* (Cádiz) and *P. "morenoi"*(Baena; Córdoba), and the ancestor of *P. grandis* (Málaga) on the other.

North Africa (Morocco)

Our phylogeny infers the past occurrence of a single colonization episode to explain the derivation of the 4 taxa from northern Morocco whose mitogenomes where analyzed. Namely: *P. ruffoi* (Berkane), *P. longipes* (Sidi Abdellah) plus 2 undescribed, sympatric species found in the latter locality (*P*. sp6-Morocco A; Psp6-Morocco B).

The area occupied by *Pseudoniphargus* in Morocco remained emerged since the Cretaceous until the Upper Tertiary, when the so-called Rifian Corridors were established across the Rif domain, representing the last connection between the Mediterranean and the Atlantic before the onset of the Messinian Salinity Crisis. The placement, extent and age of these corridors are fairly well established (Achalhi et al. 2016) and enable to date the diversification of this species assemblage at between 8 and 6.1 Ma.

Balearic Islands

The *Pseudoniphargus* assemblage of the Balearic Islands is polyphyletic and seems to be the outcome of at least two different colonization events. These involved the ancestor of *P. mercadali* from Mallorca on one side, and the common ancestor of the rest of species on the other: *P. triasi* (Cabrera), *P. pityusensis* (Ibiza), *P. pedrerae* (Formentera), *P. daviui* (Cabrera) and *P*. sp1-Balearics (Mallorca). The topology of the tree suggests that at least one of the species from Cabrera Island (*P. triasi*) derives directly from an ancestor coming from the Pityusic Islands (Ibiza + Formentera). In the Balearic Promontory, the landmass corresponding to Mallorca+Menorca+Cabrera is separated by a broad strip of deep water from Ibiza+Formentera at least since the beginning of the Pliocene, when the Mediterranean refilled after the Salinity Crisis. Since the speciation event producing *P. triasi* is apparently older than Pliocene in age, it is more plausible the colonization of Cabrera took place by a non-marine ancestor already established on land when all the Balearics conformed a single landmass during the Messinian, and not to be the result of direct transmarine dispersal from the Pityusics to Cabrera.

Several species dealt with in this study (viz. *P. mercadali*, *P. portosancti*, *P. carpalis*, *P.* sp2-Canaries) occupy territories that seem to have been colonized after episodes of long-range, trans-oceanic dispersal. Ecologically, little is known about the life-habits of

Pseudoniphargus, and no satisfactory explanation for this level of dispersal has been reached. Other species indicate quite clearly they are the result of vicariant events.

In general, we see evidence for a much higher dispersal rate than previously assumed for this genus. This is surprising due to the lack of a free swimming larval stage often expected or even required for this level of dispersal. It is particularly remarkable the case of *P. mercadali* from Mallorca (Balearic Islands), which derives from an Atlantic ancestor. Or of *P.* sp2-Canaries, found on the island of Gran Canaria but belonging to a clade mostly Mediterranean.

This study reveals that the lineages of *Pseudoniphargus* are younger than some other amphipods displaying a similar disjunct distribution embracing both sides of the Atlantic Ocean (Bauzà-Ribot et al. 2012). The presence of *Pseudoniphargus* on Bermuda (W side of the Atlantic) renders mere vicariance by plate tectonics improbable to explain its presence on that island since our analysis renders an age of only 10 to 15 Ma for this species. Additionally many of the speciation events found in *Pseudoniphargus* are younger than the age of the Tethys Sea (which lasted until 20 Ma), contrary to other taxa studied showing a similar distribution. Also, and as a general trend, the clearest evidence for dispersal is found in localities placed on islands, particularly in the Atlantic, while on mainland vicariance seems to have been the mechanism more probable to explain the diversification and distribution of current species.

Chapter 5. Conclusions

1) The broad sampling performed across most of the known distribution of the genus *Pseudoniphargus* together with the subsequent genetic analysis of the DNA sequence of the mitochondrial gene cytochrome oxidase 1 of the collected specimens showed the occurrence of hidden biodiversity in the genus, with at least 18 new species discovered, of which two are formally described herein.
2) Most species found are endemic to a single well or cave and show species-specific diagnostic mutations and low levels of intraspecific divergence, although in a few cases two or three divergent lineages/species were found to coexist in a single site.
3) There is a nearly complete agreement between species boundaries delimited based on morphological characters and those based on DNA sequences, though several cases of crypticism were found.
4) The four molecular methods of species delimitation tested in this study detected the occurrence of different levels of genetic divergence during the speciation process, with GMYC detecting early stages of speciation based on a few diagnostic mutations whereas ABGD considered as belonging to a single species those populations separated by a higher number of fixed mutations. In GMYC and PTP methods, more important than delimiting putative species is to consider the statistical support of those molecular entities.
5) Five different gene arrangements have been detected in the mitochondrial genome of *Pseudoniphargus*, with two rearrangements involving protein-coding genes. These are rare events at the intra-generic level in metazoans, and the mechanisms behind remain unknown. Anyway, our results contradict the main theories around the subject that suggest that high levels of gene rearrangement in a lineage are associated to higher rates of nucleotide substitution.
6) The detailed study of *Pseudoniphargus* mitogenomes corroborates the condition found in most metazoans, particularly the occurrence of a high AT bias and high substitution rates at third codon sites, with differences found depending on coding strand. Moreover, the last domains of both the large and small ribosomal subunits are the most conserved at both primary and secondary structures.
7) We present a fully resolved phylogeny of the main *Pseudoniphargus* lineages based on 32 newly sequenced mitogenomes, with high statistical support for most nodes. The phylogenetic signal was strong and did not vary analyzing sequences at the nucleotide or the amino-acid level, or by implementing different partition schemes and substitution models.
8) The presence of a high level of heterotachy between isopods and the amphipods *Pseudoniphargus* precluded the use of isopod fossils to calibrate our phylogeny. This rare event demonstrates that it is crucial to detect the existence of heterotachy in a phylogeny since it induces older age estimations than expected.
9) The three palaeogeographic events used to calibrate the phylogenetic tree dated the origin of the genus *Pseudoniphargus* about 55 Ma, which is congruent with the regression of the ancient Tethys Sea in the continental area currently occupied by the oldest representatives of the genus (Basque Country).
10) The age of several diversification events were found to be congruent with paleogeographical vicariant events, but other were extremely young indicating that dispersal also played an important role to establish the current distribution of *Pseudoniphargus*, such as the presence of representatives of the genus in several geologically young islands of Macaronesia.

Future work will include the comparison of the large mitochondrial genome dataset from other amphipods and crustacean found in similar habitat and with the same distribution pattern as *Pseudoniphargus* to construct a general biogeographic model that explains their present extreme disjunct distribution.

Chapter 6. References

ACHALHI, M., MÜNCH, P., CORNÉE, J-J., AZDIMOUSA, A., MELINTE-DOBRINESCU, M., QUILLÉVÉRÉ, F., DRINIA, H., FAUQUETTE, S., JIMÉNEZ-MORENO, G., MERZERAUD, G., BEN MOUSSA, A., EL KHARIM, Y., FEDDI, N. 2016. The late Miocene Mediterranean-Atlantic connections through the North Rifian Corridor: New insights from the Boudinar and Arbaa Taourirt basins (northeastern Rif, Morocco). *Palaeogeography, Palaeoclimatology, Palaeoecology* 459: 131–152.

ALFARO, M. E., ZOLLER, S. & LUTZONI, F. 2003. Bayes or Bootstrap? A Simulation Study Comparing the Performance of Bayesian Markov Chain Monte Carlo Sampling and Bootstrapping in Assessing Phylogenetic Confidence. *Molecular Biology and Evolution,* 20**,** 255-266.

ALLEGRUCCI, G., TODISCO, V., SBORDONI, V. 2005. Molecular phylogeography of Dolichopoda cave crickets (Orthoptera, Rhaphidophoridae): A scenario suggested by mitochondrial DNA. *Molecular Phylogenetics and Evolution,* 37.

ALSMADI, O., ALKAYAL, F., MONIES, D., MEYER, B. F. 2009. Specific and complete human genome amplification with improved yield achieved by phi29 DNA polymerase and a novel primer at elevated temperature. *BMC Research Notes,* 2**,** 48.

ARBOGAST, B. S., EDWARDS, S. V., WAKELEY, J., BEERLI, P. & SLOWINSKI, J. B. 2002. Estimating divergence times from molecular data on phylogenetic and population genetic timescales. *Annual Review Ecology Evolution Systematics,* 33**,** 707-740.

ARITA, H. T. 1996. The conservation of cave-roosting bats in Yucatan, Mexico. *Biological Conservation,* 76**,** 177-185.

ARNDT, A., SMITH, M. J. 1998. Mitochondrial gene rearrangement in the sea cucumber genus Cucumaria. *Molecular Biology and Evolution* 15:1009–1016.

ARNOT, D. E., ROPER, C., BAYOUMI, R. A. L. 1993. Digital codes from hypervariable tandemly repeated DNA sequences in the Plasmodium falciparum circumsporozoite gene can genetically barcode isolates. *Molecular and Biochemical Parasitology,* 61**,** 15-24.

ARNQVIST, G. 1998. Comparative evidence for the evolution of genitalia by sexual selection. *Nature,* 393**,** 784- 786.

ARNTZEN, E. V., GEIST, D. R., DRESEL, P. E. 2006. Effects of fluctuating river flow on groundwater/surface water mixing in the hyporheic zone of a regulated, large cobble bed river. *River Research and Applications,* 22**,** 937-946.

ARUNKUMAR, K. P., NAGARAJU, J. 2006. Unusually long palindromes are abundant in mitochondrial control regions of insects and nematodes. *PLoS One* 1:e110.

AUEL, H., WERNER, I. 2003. Feeding, respiration and life history of the hyperiid amphipod Themisto libellula in the Arctic marginal ice zone of the Greenland Sea. *Journal of Experimental Marine Biology and Ecology,* 296**,** 183-197.

AUNINS, A. W., NELMS, D. L., HOBSON, C. S., KING, T. L. 2016. Comparative mitogenomic analyses of three North American stygobiont amphipods of the genus Stygobromus (Crustacea: Amphipoda). *Mitochondrial DNA Part B,* 1**,** 560-563.

AYRES, D. L., DARLING, A., ZWICKL, D. J., BEERLI, P., HOLDER, M. T., LEWIS, P. O., HUELSENBECK, J. P., RONQUIST, F., SWOFFORD, D. L., CUMMINGS, M. P., RAMBAUT, A. & SUCHARD, M. A. 2012. BEAGLE: An Application Programming Interface and High-Performance Computing Library for Statistical Phylogenetics. *Systematic Biology,* 61**,** 170-173.

AZEVEDO, J. M. M., PORTUGAL FERREIRA, M. R. 2006. The volcanotectonic evolution of Flores Island, Azores (Portugal). *Journal of Volcanology and Geothermal Research* 156: 90–102.

BAELE, G., LEMEY, P., BEDFORD, T., RAMBAUT, A., SUCHARD, M. A., ALEKSEYENKO, A. V. 2012. Improving the Accuracy of Demographic and Molecular Clock Model Comparison While Accommodating Phylogenetic Uncertainty. *Molecular Biology and Evolution,* 29**,** 2157-2167.

BAKALOWICZ, M. 2005. *Epikarst*, Elsevier Academic Press, 220-223.

BALAZUC, J., ANGELIER, E. 1951. Sur la capture a Banyuls-sur-mer (Pyrenees-orientales) de Pseudoniphargus africanus Chevreux 1901 (Amphipodes Gammaridae). *Bulletin de la Société Zoologique de France.,* 76**,** 309-312.

BARRACLOUGH, T. G., HOGAN, J. E. & VOGLER, A. P. 1999. Testing whether ecological factors promote cladogenesis in a group of tiger beetles (Coleoptera: Cicindelidae). *Proceedings of the Royal Society of London. Series B: Biological Sciences,* 266**,** 1061-1067.

BARRACLOUGH, T. G. & HERNIOU, E. 2003. Why do species exist? Insights from sexuals and asexuals1. *Zoology,* 106**,** 275-282.

BARNARD, J. L., KARAMAN, G. S. 1980. Classification of Gammarid Amphipoda. *Crustaceana. Supplement***,** 5-16.

BARR, T. C. J., HOLSINGER, J. R. 1985. Speciation in Cave Faunas. *Annual Review of Ecology and Systematics,* 16**,** 313-337.

BASSO, D., TINTORI, A. 1994. New Triassic isopod crustaceans from northern Italy. *Palaeontology*, 37: 801–810.

BATEMAN, R. M., HILTON, J., RUDALL, P. J. 2006. Morphological and molecular phylogenetic context of the angiosperms: contrasting the 'top-down' and 'bottom-up' approaches used to infer the likely characteristics of the first flowers. *Journal of Experimental Botany,* 57**,** 3471-3503.

BAUZÀ-RIBOT, M. M., JAUME, D., JUAN, C., PONS, J. 2009. The complete mitochondrial genome of the subterranean crustacean Metacrangonyx longipes (Amphipoda): a unique gene order and extremely short control region. *Mitochondrial DNA,* 20**,** 88-99.

BAUZÀ-RIBOT, M. M., JAUME, D., FORNÓS, J. J., JUAN, C., PONS, J. 2011. Islands beneath islands: phylogeography of a groundwater amphipod crustacean in the Balearic archipelago. *BMC Evolutionary Biology,* 11**,** 221.

BAUZÀ-RIBOT, M. M., JUAN, C., NARDI, F., OROMÍ, P., PONS, J. & JAUME, D. 2012. Mitogenomic Phylogenetic Analysis Supports Continental-Scale Vicariance in Subterranean Thalassoid Crustaceans. *Current Biology,* 22**,** 2069-2074.

BEAULIEU, J. M., O'MEARA, B. C., CRANE, P., DONOGHUE, M. J. 2015. Heterogeneous Rates of Molecular Evolution and Diversification Could Explain the Triassic Age Estimate for Angiosperms. *Systematic Biology,* 64**,** 869-878.

BENSCH, S., HARLID, A. 2000. Mitochondrial genomic rearrangements in songbirds. *Molecular Biology and Evolution* 17:107–113.

BERARD, S., BERGERON, A., CHAUVE, C., PAUL, C. 2007. Perfect sorting by reversals is not always difficult. IEEE/ACM Trans *Journal of Bioinformatics and Computational Biology* 4:4–16.

BERNT, M., DONATH, A., JÜHLING, F., EXTERNBRINK, F., FLORENTZ, C., FRITZSCH, G., PÜTZ, J., MIDDENDORF, M., STADLER, P. F. 2013. MITOS: Improved de novo metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution* 69:313–319.

BERNT, M., MERKLE, D., RAMSCH, K., FRITZSCH, G., PERSEKE, M., BERNHARD, D., SCHLEGEL, M., STADLER, P., MIDDENDORF, M. 2007. CREx: inferring genomic rearrangements based on common intervals. *Bioinformatics* 23:2957–2958.

BIBI, F. 2013. A multi-calibrated mitochondrial phylogeny of extant Bovidae (Artiodactyla, Ruminantia) and the importance of the fossil record to systematics. *BMC Evolutionary Biology,* 13**,** 1-15.

BICKFORD, D., LOHMAN, D. J., SODHI, N. S., NG, P. K. L., MEIER, R., WINKER, K., INGRAM, K. K., DAS, I. 2007. Cryptic species as a window on diversity and conservation. *Trends in Ecology & Evolution,* 22**,** 148-155.

BIDEGARAY-BATISTA, L. & ARNEDO, M. A. 2011. Gone with the plate: the opening of the Western Mediterranean basin drove the diversification of ground-dweller spiders. *BMC Evolutionary Biology,* 11**,** 1-15.

BIRKY, C. W., JR., ADAMS, J., GEMMEL, M., PERRY, J. 2010. Using Population Genetic Theory and DNA Sequences for Species Detection and Identification in Asexual Organisms. *PLoS One,* 5**,** e10609.

BLACK, W. C., ROEHRDANZ, R. L. 1998. Mitochondrial gene order is not conserved in arthropods: prostriate and metastriate tick mitochondrial genomes. *Molecular Biology and Evolution* 15:1772–1785.

BLANCO, L., BERNAD, A., LÁZARO, J. M., MARTÍN, G., GARMENDIA, C., SALAS, M. 1989. Highly efficient DNA synthesis by the phage phi 29 DNA polymerase. Symmetrical mode of DNA replication". *The Journal of Biological Chemistry* 264 (15): 8935–40

BOISVERT, S., LAVIOLETTE, F., CORBEIL, J. 2010. Ray: Simultaneous Assembly of Reads from a Mix of High-Throughput Sequencing Technologies *Journal of Computational Biology,* 17**,** 1519-1533.

BOLGER, A. M., LOHSE, M., USADEL, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics,* 30**,** 2114-2120.

BOUSFIELD, E. (1977). A New Look at the Systematics of Gammaroidean Amphipods of the World. *Crustaceana. Supplement,* (4), 282-316

BOUTIN, C., COINEAU, N. 1988. Pseudoniphargus maroccanus n. sp. (subterranean amphipod), the first representative of the genus in Morocco. Phylogenetic relationships and paleobiogeography. *Crustaceana Supplement (Leiden)***,** 1-19.

BOORE, J. L. 1999. Animal mitochondrial genomes. *Nucleic Acids Research* 27:1767–1780.

BOORE, J. L., BROWN, W. M. 1994. Complete DNA sequence of the mitochondrial genome of the black chiton, Katharina tunicata. *Genetics,* 138**,** 423-443.

BOORE, J. L., COLLINS, T. M., STANTON, D., DAEHLER, L. L., BROWN, W. M. 1995. Deducing the pattern of arthropod phylogeny from mitochondrial DNA rearrangements. *Nature* 376:163–165.

BOORE, J. L., LAVROV, D., BROWN, W. 1998. Gene translocation links insects and crustaceans. *Nature* 392:667–668.

BOORE, J. L., MACEY, J. R., MEDINA, M. 2005. Sequencing and comparing whole mitochondrial genomes of animals. *Methods Enzymology* 395:311–348.

BOTELLO, A., ALVAREZ, F. 2010. Genetic variation in the stygobitic shrimp Creaseria morleyi (Decapoda: Palaemonidae), evidence of bottlenecks and re-invasions in the Yucatan Peninsula. *Biological Journal of the Linnean Society,* 99**,** 315-325.

BOTERO-CASTRO, F., TILAK, M.-K., JUSTY, F., CATZEFLIS, F., DELSUC, F., DOUZERY, E. J. P. 2013. Next-generation sequencing and phylogenetic signal of complete mitochondrial genomes for resolving the evolutionary history of leaf-nosed bats (Phyllostomidae). *Molecular Phylogenetics and Evolution,* 69**,** 728-739.

BOTOSANEANU, L., HOLSINGER, J. R. 1991. Some aspects concerning colonization of the subterranean realm - Especially of subterranean waters - A response. *Stygologia,* 6**,** 11-39.

BOTOSANEANU, L. 2001. Morphological rudimentation and novelties in stygobitic Cirolanidae (Isopoda, Cymothoidea). *Vie Et Milieu-Life and Environment,* 51**,** 37-54.

BRANCELJ, A., CULVER, D. C. 2005. *Epikarstic communities*, Elsevier Academic Press, 223-229.

BRANCELJ, A., DUMONT, H. J. 2007. A review of the diversity, adaptations and groundwater colonization pathways in Cladocera and Calanoida (Crustacea), two rare and contrasting groups of stygobionts. *Fundamental and Applied Limnology / Archiv für Hydrobiologie,* 168**,** 3-17.

BREHIER, F., JAUME, D. 2009. A new species of Pseudoniphargus (Crustacea, Amphipoda, Melitidae) from an anchialine cave on the French Mediterranean coast. *Zoosystema,* 31**,** 17-32.

BUHAY, J. E., MONI, G., MANN, N. & CRANDALL, K. A. 2007. Molecular taxonomy in the dark: Evolutionary history, phylogeography, and diversity of cave crayfish in the subgenus Aviticambarus, genus Cambarus. *Molecular Phylogenetics and Evolution,* 42**,** 435-448.

BURGER, G., GRAY, M. W., FRANZ LANG, B. 2003. Mitochondrial genomes: anything goes. *Trends in Genetics,* 19**,** 709-716.

CABEZAS, M. P., NAVARRO-BARRANCO, C., ROS, M., GUERRA-GARCÍA, J. M. 2013. Long-distance dispersal, low connectivity and molecular evidence of a new cryptic species in the obligate rafter Caprella andreae Mayer, 1890 (Crustacea: Amphipoda: Caprellidae). *Helgoland Marine Research,* 67**,** 483-497.

CABEZAS, M.P., CABEZAS, P., MACHORDOM, A., GUERRA-GARCÍA, J. M. 2013. Hidden diversity and cryptic speciation refute cosmopolitan distribution in Caprella penantis (Crustacea: Amphipoda: Caprellidae). *Journal of Zoological Systematics and Evolutionary Research,* 51**,** 85-99.

CAMACHO , A. (ed.) 1992. *The Natural History of Biospeology,* Madrid: National Museum of Natural Sciences.

CAMACHO, A. I., DORDA, B. A., REY, I. 2011. Identifying cryptic speciation across groundwater populations: First CO1 sequences of Bathynellidae (Crustacea, Syncarida). *Graellsia,* 67**,** 7-12.

CAMARGO, A., MORANDO, M., AVILA, L. J., SITES, J. W. 2012. Species delimitation with ABC and other coalescent-based methods: A test of accuracy with simulations and an empirical example with lizards of the *Liolaemus darwinii* complex (Squamata: Liolaemidae). *Evolution,* 66**,** 2834-2849.

CÁNOVAS, F., JURADO-RIVERA, J. A., CERRO-GÁLVEZ, E., JUAN, C., JAUME, D., PONS, J. 2016. DNA barcodes, cryptic diversity and phylogeography of a W Mediterranean assemblage of thermosbaenacean crustaceans. *Zoologica Scripta,* 45**,** 659-670.

CARAPELLI, A., LIO, P., NARDI, F., VAN DER WATH, E., FRATI, F. 2007. Phylogenetic analysis of mitochondrial protein coding genes confirms the reciprocal paraphyly of Hexapoda and Crustacea. *BMC Evolutionary Biology,* 7**,** S8.

CARAPELLI, A., SOTO–ADAMES, F., SIMON, C., FRATI, F., NARDI, F., DALLAI, R. 2004. Secondary structure, high variability and conserved motifs for domain III of 12S rRNA in the arthropleona (Hexapoda; Collembola). *Insect Molecular Biology* 13:659–670.

CARRACEDO, J. C., BADIOLA, E. R., GUILLOU, H. DE LA NUEZ, J., PÉREZ TORRADO F. J. 2001. Geology and volcanology of La Palma and El Hierro, western Canaries. *Estudios Geológicos,* 57: 175-273.

CARRACEDO, J. C., BADIOLA, E. R., GUILLOU, H., DE LA NLARTILLOT, N., PHILIPPE, H. 2004. A Bayesian Mixture Model for Across-Site Heterogeneities in the Amino-Acid Replacement Process. *Molecular Biology and Evolution*. 2004 21(6):1095-1109.

CARSTENS, B. C., PELLETIER, T. A., REID, N. M. & SATLER, J. D. 2013. How to fail at species delimitation. *Molecular Ecology,* 22**,** 4369-4383.

CASTRO, A., DELGADO, J. A. 2001. Iberoporus cermenius, a New Genus and Species of Subterranean Water Beetle (Coleoptera: Dytiscidae) from Spain. *Aquatic Insects,* 23**,** 33-43.

CHAUDHURI, K., CHEN, K., MIHAESCU, R., RAO, S. 2006. On the tandem duplication-random loss model of genome rearrangement. Proceedings of the seventeenth annual ACM-SIAM symposium on Discrete algorithm, 2006 Miami, Florida. Philadelphia PA, USA: Society for Industrial and Applied Mathematics, 564-570.

CHEN, H., STRAND, M., NORENBURG, J. L., SUN, S., KAJIHARA, H., CHERNYSHEV, A. V., MASLAKOVA, S. A., SUNDBERG, P. 2010. Statistical Parsimony Networks and Species Assemblages in Cephalotrichid Nemerteans (Nemertea). *PLoS One,* 5**,** e12885.

CHEVREUX, E. 1901. Amphipodes des eaux soutteraines de France et d'Algerie. Bulletin de la Société Zoologique de France, 26**,** 211-216.

CHOI, S. C. 2016. Methods for delimiting species via population genetics and phylogenetics using genotype data. *Genes & Genomics,* 38**,** 905-915.

CHRISTIANSEN, K. A. 1962. Proposition pour la classification des animaux cavernicoles. *Spelunca* 2**,** 76–78.

CHRISTMAN, M. C., CULVER, D. C., MADDEN, M. K., WHITE, D. 2005. Patterns of endemism of the eastern North American cave fauna. *Journal of Biogeography,* 32**,** 1441-1452.

CLEMENT, M., POSADA, D. & CRANDALL, K. A. 2000. TCS: a computer program to estimate gene genealogies. *Molecular Ecology,* 9**,** 1657-1659.

COINEAU, N. BOUTIN, C. (1996). Age and origin of the stygobiontic Amphipod Pseudoniphargus (Crustacea) in Morocco, with the description of three new species. *Bolletino del Museo Civico di Storia Naturale di Verona,* 20, 2, 503-520

COLLINS, R. A. & CRUICKSHANK, R. H. 2013. The seven deadly sins of DNA barcoding. *Molecular Ecology Resources,* 13**,** 969-975.

CRACRAFT, J. 1983. Species Concepts and Speciation Analysis. *In:* JOHNSTON, R. (ed.) *Current Ornithology.* Springer US.

CULVER, D. C., FONG, D. W. 1994. Small-scale and large -scale biogeography of subterranean crustacean faunas of the Virginias. *Hydrobiologia,* 287**,** 3-9.

CULVER, D. C., SKET, B. 2000. Hotspots of subterranean biodiversity in caves and wells. *Journal of Caves and Karst Studies,* 62**,** 11-17.

CULVER, D. C., WHITE, W. B. (eds.) 2005. *Cave, Definition of*: Elsevier Academic Press.

CULVER, D. C., PIPAN, T., SCHNEIDER, K. 2009. Vicariance, dispersal and scale in the aquatic subterranean fauna of karst regions. *Freshwater Biology,* 54**,** 918-929.

DANIELOPOL, D. L., GRIEBLER, C. 2008. Changing Paradigms in Groundwater Ecology – from the 'Living Fossils' Tradition to the 'New Groundwater Ecology'. *International Review of Hydrobiology,* 93**,** 565-577.

DANIELOPOL, D. L., POSPISIL, P., ROUCH, R. 2000. Biodiversity in groundwater: a large-scale view. *Trends in Ecology & Evolution,* 15**,** 223-224.

DARTY, K., DENISE, A., PONTY, Y. 2009. VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics* 25:1974–1975.

DARWIN, C. 1859. *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life,* London.

DAVIS, J. I., NIXON, K. C. 1992. Populations, Genetic Variation, and the Delimitation of Phylogenetic Species. *Systematic Biology,* 41**,** 421-435.

DE LEY, P., DE LEY, I. T., MORRIS, K., ABEBE, E., MUNDO-OCAMPO, M., YODER, M., HERAS, J., WAUMANN, D., ROCHA-OLIVARES, A., JAY BURR, A. H., BALDWIN, J. G., THOMAS, W. K. 2005. An integrated approach to fast and informative morphological vouchering of nematodes for applications in molecular barcoding. *Philosophical Transactions of the Royal Society of London B: Biological Sciences,* 360**,** 1945-1958.

DE QUEIROZ, K. 2007. Species Concepts and Species Delimitation. *Systematic Biology,* 56**,** 879-886.

DE QUEIROZ, K. (ed.) 1999. *The General Lineage Concept of Species and the Defining Properties of the Species Category,* Cambridge: MIT Press.

DEGNAN, J. H., SALTER, L. A. 2005. Gene tree distributions under the coalescent process. *Evolution,* 59**,** 24-37.

DEHARVENG, L., STOCH, F., GIBERT, J., BEDOS, A., GALASSI, D., ZAGMAJSTER, M., BRANCELJ, A., CAMACHO, A., FIERS, F., MARTIN, P., GIANI, N., MAGNIEZ, G., MARMONIER, P. 2009. Groundwater biodiversity in Europe. *Freshwater Biology,* 54**,** 709-726.

DELLICOUR, S., MARDULYN, P. 2014. spads 1.0: a toolbox to perform spatial analyses on DNA sequence data sets. *Molecular Ecology Resources,* 14**,** 647-651.

DESALLE, R., EGAN, M. G., SIDDALL, M. 2005. The unholy trinity: taxonomy, species delimitation and DNA barcoding. *Philosophical Transactions of the Royal Society of London B: Biological Sciences,* 360**,** 1905-1916.

DISNEY, H., PÉREZ, A. M., PÉREZ, T., TORRES, D. 2009. Nueva especie de díptero (Fam. Phoridae) descubierta en el Complejo del Romeral (Antequera, Málaga). *Espeleo (Bio-Espeleo)* [Online], 21. Available: https://drive.google.com/file/d/0B4JcDrAba_RPdGE1UzZxWlBkVTA/edit.

DONOGHUE, P. C. J., YANG, Z. 2016. The evolution of methods for establishing evolutionary timescales. *Philosophical Transactions of the Royal Society B: Biological Sciences,* 371.

DORNBURG, A., BRANDLEY, M. C., MCGOWEN, M. R., NEAR, T. J. 2012. Relaxed Clocks and Inferences of Heterogeneous Patterns of Nucleotide Substitution and Divergence Time Estimates across Whales and Dolphins (Mammalia: Cetacea). *Molecular Biology and Evolution,* 29**,** 721-736.

DOWLING, T. E., MORITZ C., PALMER, J. D., RIESBERG, L. H. (eds.) 1996. *Nucleic Acids III: Analysis of Fragments and Restriction Sites,* Sunderland: Sinauer Associates Inc.

DOWTON, M, AUSTIN, A. D. 1999. Evolutionary dynamics of a mitochondrial rearrangement "hot spot" in the Hymenoptera. *Molecular Biology and Evolution* 16:298–309.

DOWTON, M., CAMERON, S., DOWAVIC, J., AUSTIN, A., WHITING, M.2009. Characterization of 67 mitochondrial tRNA gene rearrangements in the Hymenoptera suggests that mitochondrial tRNA gene position is selectively neutral. *Molecular Biology and Evolution* 26:1607–1617.

DRDA, W. J. 1968. A Study of Snakes Wintering in a Small Cave. *Journal of Herpetology,* 1**,** 64-70.

DRUMMOND, A. J., RAMBAUT, A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology,* 7.

DRUMMOND, A. J., HO, S. Y. W., PHILLIPS, M. J., RAMBAUT, A. 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biology,* 4.

DRUMMOND, A. J., SUCHARD, M. A., XIE, D., RAMBAUT, A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution,* 29.

DUCHÊNE, S., ARCHER, F. I., VILSTRUP, J., CABALLERO, S., MORIN, P. A. 2011. Mitogenome Phylogenetics: The Impact of Using Single Regions and Partitioning Schemes on Topology, Substitution Rate and Divergence Time Estimation. *PLoS One,* 6**,** e27138.

DUPANLOUP, I., SCHNEIDER, S., EXCOFFIER, L. 2002. A simulated annealing approach to define the genetic structure of populations. *Molecular Ecology,* 11**,** 2571-2581.

EDWARDS, S. V. 2009. Is a new and general theory of molecular systematics emerging? *Evolution,* 63**,** 1-19.

ESSELSTYN, J. A., EVANS, B. J., SEDLOCK, J. L., ANWARALI KHAN, F. A., HEANEY, L. R. 2012. Single-locus species delimitation: a test of the mixed Yule–coalescent model, with an empirical application to Philippine round-leaf bats. *Proceedings of the Royal Society B: Biological Sciences*

ETTER W. 2014. A well-preserved isopod from the Middle Jurassic of southern Germany and implications for the fossil record. *Palaeontology,* 57: 931–949.

EZARD, T., FUJISAWA, T., T, B. 2009. SPLITS: SPecies LImits by Threshold Statistics. version 1. R Package.

FAHREIN, K., TALARICO, G., BRABAND, A., PODSIADLOWSKI, L. 2007. The complete mitochondrial genome of Pseudocellus pearsei (Chelicerata: Ricinulei) and a comparison of mitochondrial gene rearrangements in Arachnida. *BMC Genomics,* 8**,** 386.

FAKHER EL ABIARI, A., OULBAZ, Z., MESSOULI, M., COINEAU, N. 1999. A new species of Pseudoniphargus (Crustacea, Amphipoda) from subterranean water of northeastern Morocco: Historical biogeography and evolutionary aspects. *Contributions to Zoology,* 68**,** 161-171.

FEIJÃO, P. C., NEIVA, L. S., AZEREDO–ESPIN, A. M. L. D., LESSINGER, A. C. 2006. AMiGA: the arthropodan mitochondrial genomes accessible database. *Bioinformatics* 22:902–903.

FERNÁNDEZ-PALACIOS, J. M., DE NASCIMENTO, L., OTTO, R., DELGADO, J. D., GARCÍA-DEL-REY, E., ARÉVALO, J. R., WHITTAKER, R. J. 2011. A reconstruction of Palaeo-Macaronesia, with particular reference to the long-term biogeography of the Atlantic island laurel forests. *Journal of Biogeography,* 38**,** 226-246.

FINSTON, T. L., JOHNSON, M. S., HUMPHREYS, W. F., EBERHARD, S. M. & HALSE, S. A. 2007. Cryptic speciation in two widespread subterranean amphipod genera reflects historical drainage patterns in an ancient landscape. *Molecular Ecology,* 16**,** 355-365.

FOLMER, O., BLACK, M., HOEH, W., LUTZ, R., VRIJENHOEK, R. 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology,* 3**,** 294-299.

FONTANETO, D., IAKOVENKO, N., EYRES, I., KAYA, M., WYMAN, M. & BARRACLOUGH, T. G. 2011. Cryptic diversity in the genus Adineta Hudson & Gosse, 1886 (Rotifera: Bdelloidea: Adinetidae): a DNA taxonomy approach. *Hydrobiologia,* 662.

FONTANETO, D., FLOT, J.-F., TANG, C. Q. 2015. Guidelines for DNA taxonomy, with a focus on the meiofauna. *Marine Biodiversity,* 45**,** 433-451.

FOREST, F. 2009. Calibrating the Tree of Life: fossils, molecules and evolutionary timescales. *Annals of Botany,* 104**,** 789-794.

FOURMENT, M., GIBBS, M. J. 2006. PATRISTIC: a program for calculating patristic distances and graphically comparing the components of genetic change. *BMC Evolutionary Biology,* 6**,** 1.

FOULQUIER, A., MALARD, F., LEFEBURE, T., GIBERT, J. & DOUADY, C. J. 2008. The imprint of Quaternary glaciers on the present-day distribution of the obligate groundwater amphipod Niphargus virei (Niphargidae). *Journal of Biogeography,* 35**,** 552-564.

FRASER, B. G., WILLIAMS, D. D., HOWARD, K. W. F. 1996. Monitoring Biotic And Abiotic Processes Across The Hyporheic/Groundwater Interface. *Hydrogeology Journal,* 4**,** 36-50.

FRÉZAL, L., LEBLOIS, R. 2008. Four years of DNA barcoding: Current advances and prospects. *Infection, Genetics and Evolution,* 8**,** 727-736.

FUJISAWA, T., BARRACLOUGH, T. G. 2013. Delimiting Species Using Single-Locus Data and the Generalized Mixed Yule Coalescent Approach: A Revised Method and Evaluation on Simulated Data Sets. *Systematic Biology,* 62**,** 707-724.

FUJITA, M. K., LEACHÉ, A. D., BURBRINK, F. T., MCGUIRE, J. A., MORITZ, C. 2012. Coalescent-based species delimitation in an integrative taxonomy. *Trends in Ecology & Evolution,* 27**,** 480-488.

FUKAMI-KOBAYASHI, K., TATENO, Y. 1991. Robustness of maximum likelihood tree estimation against different patterns of base substitutions. *Journal of Molecular Evolution,* 32**,** 79-91.

GASCA, R., HADDOCK, S. H. D. 2004. Associations between gelatinous zooplankton and hyperiid amphipods (Crustacea: Peracarida) in the Gulf of California. *Hydrobiologia,* 530**,** 529-535.

GIBERT, J., DEHARVENG, L. 2002. Subterranean Ecosystems: A Truncated Functional Biodiversity: This article emphasizes the truncated nature of subterranean biodiversity at both the bottom (no primary producers) and the top (very few strict predators) of food webs and discusses the implications of this truncation both from functional and evolutionary perspectives. *BioScience,* 52**,** 473-481.

GIBERT, J., CULVER, D. C., DOLE-OLIVIER, M.-J., MALARD, F., CHRISTMAN, M. C., DEHARVENG, L. 2009. Assessing and conserving groundwater biodiversity: synthesis and perspectives. *Freshwater Biology,* 54**,** 930-941.

GINÉS, A., GINÉS, J. 2007. Eogenetic karst, glacioeustatic cave pools and anchialine environments on Mallorca Island: a discussion of coastal speleogenesis. *International Journal of Speleology,* 36**,** 57-67.

GOLD, T. 1992. The deep, hot boisphere. *Proceedings of the National Academy of Sciences of the United States of America,* 89**,** 6045-6049.

GRABHERR, M. G., HAAS, B. J., YASSOUR, M., LEVIN, J. Z., THOMPSON, D. A., AMIT, I., ADICONIS, X., FAN, L., RAYCHOWDHURY, R., ZENG, Q., CHEN, Z., MAUCELI, E., HACOHEN, N., GNIRKE, A., RHIND, N., DI PALMA, F., BIRREN, B. W., NUSBAUM, C., LINDBLAD-TOH, K., FRIEDMAN, N., REGEV, A. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology,* 29**,** 644-652.

GRIEBLER, C., AVRAMOV, M. 2015. Groundwater ecosystem services: a review. *Freshwater Science,* 34**,** 355-367.

GRIEBLER, C., MALARD, F., LEFÉBURE, T. 2014. Current developments in groundwater ecology-from biodiversity to ecosystem function and services. *Current Opinion in Biotechnology,* 27**,** 159-167.

GRIEBLER, C., STEIN, H., KELLERMANN, C., BERKHOFF, S., BRIELMANN, H., SCHMIDT, S., SELESI, D., STEUBE, C., FUCHS, A., HAHN, H. J. 2010. Ecological assessment of groundwater ecosystems – Vision or illusion? *Ecological Engineering,* 36**,** 1174-1190.

GUINDON, S., DUFAYARD, J. F., LEFORT, V., ANISIMOVA, M., HORDIJK, W., GASCUEL, O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3–0. *Sysematict Biology,* 59.

GUINDON, S., LETHIEC, F., DUROUX, P., GASCUEL, O. 2005. PHYML Online - a web server for fast maximum likelihood-based phylogenetic inference. *Nucleic Acids Research,* 33**,** W557-W559.

HALIASSOS, A., SIORFANE, N., DRAKOULIS, N. 2001. Use of polymorphic DNA sequences as internal markers for the positive identification of biological samples: The barcode of the future? *Clinical Chemistry,* 47**,** 2079.

HANAGE, W., FRASER, C., SPRATT, B. 2005. Fuzzy species among recombinogenic bacteria. *BMC Biology,* 3**,** 6.

HASSANIN, A. 2006. Phylogeny of Arthropoda inferred from mitochondrial sequences: Strategies for limiting the misleading effects of multiple changes in pattern and rates of substitution. *Molecular Phylogenetics and Evolution,* 38**,** 100-116.

HEBERT, P. D. N., CYWINSKA, A., BALL, S. L., DEWAARD, J. R. 2003. Biological identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences,* 270**,** 313-321.

HEBERT, P. D. N., STOECKLE, M. Y., ZEMLAK, T. S., FRANCIS, C. M. 2004. Identification of Birds through DNA Barcodes. *PLoS Biol,* 2**,** e312.

HELED, J., DRUMMOND, A. J. 2010. Bayesian Inference of Species Trees from Multilocus Data. *Molecular Biology and Evolution,* 27**,** 570-580.

HENNING, W. 1999. *Phylogenetic Systematics,* Urbana And Chicago, University of Illinois Press

HEY, J. 2001. The mind of the species problem. *Trends in Ecology & Evolution,* 16**,** 326-329.

HEY, J. 2006. On the failure of modern species concepts. *Trends in Ecology & Evolution,* 21**,** 447-450.

HEY, J., WAPLES, R. S., ARNOLD, M. L., BUTLIN, R. K., HARRISON, R. G. 2003. Understanding and confronting species uncertainty in biology and conservation. *Trends in Ecology & Evolution,* 18**,** 597-603.

HICKERSON, M. J., CUNNINGHAM, C. W. 2000. Dramatic mitochondrial gene rearrangements in the hermit crab Pagurus longicarpus (Crustacea, Anomura). *Molecular Biology and Evolution* 17:639–644.

HICKERSON, M. J., MEYER, C. P., MORITZ, C. 2006. DNA Barcoding Will Often Fail to Discover New Animal Species over Broad Parameter Space. *Systematic Biology,* 55**,** 729-739.

HIDAKA, H., WATANABE, H., KANO, Y., KOJIMA, S. 2013. Mitochondrial genome rearrangement in a hydrothermal vent–endemic lineage of provannid gastropods provides a new DNA marker for phylogeographical studies. *J Marine Biology Association* UK 93:1053–1058.

HILLIS, D. M. 1987. Molecular Versus Morphological Approaches to Systematics. *Annual Review of Ecology and Systematics,* 18**,** 23-42.

HOLDER, M., LEWIS, P. O. 2003. Phylogeny estimation: traditional and Bayesian approaches. *Nauret Reviews Genetics,* 4**,** 275-284.

HOLSINGER, J. R. (ed.) 2005. *Vicariance and Dispersalist Biogeography,* Amsterdam: Elsevier Academic Press.

HOLSINGER, J. R. 1991. What can vicariance biogeographic models tell us about the distributional history of subterranean amphipods? *Hydrobiologia,* 223**,** 43-45.

HOLSINGER, J. R. 1994. Pattern and process in the biogeography of subterranean amphipods. *Hydrobiologia,* 287**,** 131-145.

HUDSON, R. R. 1991. Gene genealogies and the coalescent process. *Oxford Surveys in Evolutionary Biology,* 7**,** 1-44.

HUELSENBECK, J. P., CRANDALL, K. A. 1997. Phylogeny Estimation and Hypothesis Testing Using Maximum Likelihood. *Annual Review of Ecology and Systematics,* 28**,** 437-466.

HUELSENBECK, J. P., RONQUIST, F. 2001. MrBayes: Bayesian inference of phylogeny. *Bioinformatics,* 17**,** 754-755.

HUGHES, T. J. P. A. J. M. 2011. Neither molecular nor morphological data have all the answers; with an example from Macrobrachium (Decapoda: Palaemonidae) from Australia. *Zootaxa,* 2874**,** 65-68.

HUNTER, R. L., WEBB, M. S., ILIFFE, T. M., BREMER, J. R. A. 2008. Phylogeny and historical biogeography of the cave-adapted shrimp genus Typhlatya (Atyidae) in the Caribbean Sea and western Atlantic. *Journal of Biogeography,* 35**,** 65-75.

HUTCHINSON, G. E. 1959. Homage to Santa Rosalia or why there are so many kinds of animals? *The American Naturalist,* 93**,** 145-159.

IGAWA, T., KURABAYASHI, A., USUKI, C., FUJII, T., SUMIDA, M. 2008. Complete mitochondrial genomes of three neobatrachian anurans: A case study of divergence time estimation using different data and calibration settings. *Gene,* 407**,** 116-129.

ILIFFE,T. M., HART JR , C. W., MANNING R. B. 1983. Biogeography and the caves of Bermuda. *Nature,* 302**,** 141-142.

ILIFFE, T. M., KORNICKER, L. S. 2009. Worldwide Diving Discoveries of Living Fossil Animals from the Depths of Anchialine and Marine Caves. *Smithsonian Contributions to the Marine Sciences***,** 269-280.

ILIFFE, T. M. 2005. *Anchialine caves, biodiversity in,* Elsevier Academic Press, 24-30.

ILIFFE, T. M. 1992. Anchialine cave biology . In Camacho AI, ed. The Natural History of Biospeleology. *Monografias del Museo Nacional de Ciencias Naturales***,** 614-636.

IRISARRI, I., EERNISSE, D. J., ZARDOYA, R. 2014. Molecular phylogeny of Acanthochitonina (Mollusca: Polyplacophora: Chitonida): three new mitochondrial genomes, rearranged gene orders and systematics. *Journal of Natural History* 48:2825–2853.

ITO, A., AOKI, M. N., YOKOBORI, S. I., WADA, H. 2010. The complete mitochondrial genome of Caprella scaura (Crustacea, Amphipoda, Caprellidea), with emphasis on the unique gene order pattern and duplicated control region. *Mitochondrial DNA* 21:183–190.

JAUME, D. 1991. Two new species of the amphipod genus Pseudoniphargus (Crustacea) from Cabrera (Balearic Islands). *Stygologia* 6:177–189.

JAUME, D., HOLSINGER, J., HORTON, T. (2016). PSEUDONIPHARGIDAE KARAMAN, 1993. IN: HORTON, T., LOWRY, J., DE BROYER, C., BELLAN-SANTINI, D., COLEMAN, C. O., DANELIYA, M., DAUVIN, J-C., FIŠER, C., GASCA, R., GRABOWSKI, M., GUERRA-GARCÍA, J. M., HENDRYCKS, E., HOLSINGER, J., HUGHES, L., JAUME, D., JAZDZEWSKI, K., JUST, J., KAMALTYNOV, R. M., KIM, Y.-H., KING, R., KRAPP-SCHICKEL, T., LECROY, S., LÖRZ, A.-N., SENNA, A. R., SEREJO, C., SKET, B.,

TANDBERG, A.H., THOMAS, J., THURSTON, M., VADER, W., VÄINÖLÄ, R., VONK, R., WHITE, K., ZEIDLER, W. (2017). World Amphipoda Database. Accessed through: World Register of Marine Species at http://www.marinespecies.org/aphia.php?p=taxdetails&id=720659 on 2017-05-17

JAŻDŻEWSKI, K., GRABOWSKI, M., KUPRYJANOWICZ, J. 2014. Further records of Amphipoda from Baltic Eocene amber with first evidence of prae-copulatory behaviour in a fossil amphipod and remarks on the taxonomic position of Palaeogammarus Zaddach, 1864. *Zootaxa,* 3765: 401–417.

JOHN, L., PHILIP, S., DAHANUKAR, N., ANVAR ALI, P. H., THARIAN, J., RAGHAVAN, R., ANTUNES, A. 2013. Morphological and Genetic Evidence for Multiple Evolutionary Distinct Lineages in the Endangered and Commercially Exploited Red Lined Torpedo Barbs Endemic to the Western Ghats of India. *PLoS One,* 8**,** e69741.

JONES, G. R. 2014. STACEY: species delimitation and phylogeny estimation under the multispecies coalescent. *bioRxiv*.

JUAN, C., GUZIK, M. T., JAUME, D., COOPER, S. J. B. 2010. Evolution in caves: Darwin's 'wrecks of ancient life' in the molecular era. *Molecular Ecology,* 19**,** 3865-3880.

JÖRGER, K., SCHRODL, M. 2013. How to describe a cryptic species? Practical challenges of molecular taxonomy. *Frontiers in Zoology,* 10**,** 59.

JÖRGER, K., NORENBURG, J., WILSON, N., SCHRODL, M. 2012. Barcoding against a paradox? Combined molecular species delineations reveal multiple cryptic lineages in elusive meiofaunal sea slugs. *BMC Evolutionary Biology,* 12**,** 245.

KARAMAN, G. S., RUFFO, S. 1989. Tyrrhenogammarus sardous, new genus and species with a description of several new taxa of the genus Pseudoniphargus Chevreux, 1901 from Sicily (Amphipoda, Gammaridea). *Animalia (Catania),* 16**,** 161-192.

KARAMAN, G. S. 1993. Crustacea Amphipoda di acqua dolce. In Fauna d`Itàlia Vol. 31. Bologna: Edizione Calderini, pp 337.

KARAMAN, G. S. 1978. Revision of the genus Pseudoniphargus family Gammaridae. *Bollettino del Museo Civico di Storia Naturale di Verona,* 5**,** 239-258.

KATOH, S. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution***,** 772-780.

KELLY, D. W., MACISAAC, H. J., HEATH, D. D. 2006. Vicariance and dispersal effects on phylogeographic structure and speciation in a widespread estuarine invertebrate. *Evolution,* 60**,** 257-267.

KEKKONEN, M., MUTANEN, M., KAILA, L., NIEMINEN, M., HEBERT, P. D. N. 2015. Delineating Species with DNA Barcodes: A Case of Taxon Dependent Method Performance in Moths. *PLoS One,* 10**,** e0122481.

KI, J., HOP, H., KIM, S., KIM, I., PARK, H., LEE, J. 2010. Complete mitochondrial genome sequence of the arctic gammarid, Onisimus nanseni (Crustacea; Amphipoda): novel gene structures and unusual control region features. *Comparative Biochemistry Physiology Part D Genomics Proteomics* 5:105–115.

KILPERT, F., PODSIADLOWSKI, L. 2006. The complete mitochondrial genome of the common sea slater, Ligia oceanica (Crustacea, Isopoda) bears a novel gene order and unusual control region features. *BMC Genomics,* 7.

KILPERT, F., PODSIADLOWSKI, L. 2010a. The mitochondrial genome of the Japanese skeleton shrimp Caprella mutica (Amphipoda: Caprellidea) reveals a unique gene order and shared apomorphic translocations with Gammaridea. *Mitochondrial DNA,* 21**,** 77-86.

KILPERT, F., PODSIADLOWSKI, L. 2010b. The Australian fresh water isopod (Phreatoicidea: Isopoda) allows insights into the early mitogenomic evolution of isopods. *Comparative Biochemistry Physiology Part D Genomics Proteomics* 5:36–44.

KIMURA, M. 1983. *The Neutral Theory of Molecular Evolution*, Cambridge University Press.

KING, R. A., BRADFORD, T., AUSTIN, A. D., HUMPHREYS, W. F., COOPER, S. J. B. 2012. Divergent molecular lineages and not-so-cryptic species: The first descriptions of stygobitic chiltoniid amphipods (*Talitroidea: Chiltoniidae*) from Western Australia. *Journal of Crustacean Biology,* 32**,** 465-488.

KNOWLES, L. L., CARSTENS, B. C. 2007. Delimiting Species without Monophyletic Gene Trees. *Systematic Biology,* 56**,** 887-895.

KNOWLES, L. L., MADDISON, W. P. 2002. Statistical Phylogeography. *Molecular Ecology,* 11**,** 2623-2635.

KRAYTSBERG, Y., SCHWARTZ, M., BROWN, T. A., EBRALIDSE, K., KUNZ, W. S., CLAYTON, D. A., VISSING, J., KHRAPKO, K. 2004. Recombination of human mitochondrial DNA. *Science* 304:981.

KREBES, L., BASTROP, R. 2012. The mitogenome of Gammarus duebeni (Crustacea Amphipoda): a new gene order and non–neutral sequence evolution of tandem repeats in the control region. *Comparative Biochemistry Physiology Part D Genomics Proteomics* 7:201–211.

KURABAYASHI, A., SUMIDA, M., YONEKAWA, H., GLAW, F., VENCES, M., HASEGAWA, M. 2008. Phylogeny, recombination, and mechanisms of stepwise mitochondrial genome reorganization in mantellid frogs from Madagascar. *Molecular Biology and Evolution* 25:874–891.

KURABAYASHI, A., YOSHIKAWA, N., SATO, N., HAYASHI, Y., OUMI, S., FUJII, T., SUMIDA, M. 2010. Complete mitochondrial DNA sequence of the endangered frog Odorrana ishikawae (family Ranidae) and unexpected diversity of mt gene arrangements in ranids. *Molecular Phylogenetics and Evolution* 56:543–553.

LANFEAR, R., CALCOTT, B., HO, S., GUINDON, S. 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution,* 29**,** 1695 - 1701.

LANGMEAD, B., TRAPNELL, C., POP, M., SALZBERG, S. L. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology,* 10**,** R25.

LARSON, E. R., CASTELIN, M., WILLIAMS, B. W., OLDEN, J. D., ABBOTT, C. L. 2016. Phylogenetic species delimitation for crayfishes of the genus Pacifastacus. *PeerJ,* 4**,** e1915.

LARTILLOT, N., PHILIPPE, H. 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Molecular Biology and Evolution,* 21**,** 1095-1109.

LASLETT, D., CANBÄCK, B. 2008. ARWEN: a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. *Bioinformatics* 24:172–175.

LAVAL, P. 1980. Hyperiid Amphipods as Crustacean Parasitoids Associated With Gelatinous Zooplankton. *Oceanography and Marine Biology / An annual Review* 18**,** 11-56.

LEAVITT, S. D., MOREAU, C. S., THORSTEN LUMBSCH, H. 2015. The Dynamic Discipline of Species Delimitation: Progress Toward Effectively Recognizing Species Boundaries in Natural Populations. *In:* UPRETI, K. D., DIVAKAR, K. P., SHUKLA, V. & BAJPAI, R. (eds.) *Recent Advances in Lichenology: Modern Methods and Approaches in Lichen Systematics and Culture Techniques, Volume 2.* New Delhi: Springer India.

LEFEBURE, T., DOUADY, C. J., GOUY, M., TRONTELJ, P., BRIOLAY, J., GIBERT, J. 2006. Phylogeography of a subterranean amphipod reveals cryptic diversity and dynamic evolution in extreme environments. *Molecular Ecology,* 15**,** 1797-1806.

LEFEBURE, T., DOUADY, C. J., MALARD, F., GIBERT, J. 2007. Testing dispersal and cryptic diversity in a widely distributed groundwater amphipod (*Niphargus rhenorhodanensis*). *Molecular Phylogenetics and Evolution,* 42**,** 676-686.

LEJEUSNE, C., CHEVALDONNE, P. 2006. Brooding crustaceans in a highly fragmented habitat: the genetic structure of Mediterranean marine cave-dwelling mysid populations. *Molecular Ecology,* 15**,** 4123-4140.

LI, H., LIU, J., XIONG, L., ZHANG, H., ZHOU, H., YIN, H., JING, W., LI, J., SHI, Q., WANG, Y., NIE, L. 2017. Phylogenetic relationships and divergence dates of softshell turtles (Testudines: Trionychidae) inferred from complete mitochondrial genomes. *Journal of Evolutionary Biology*.

LINDER, H. P., HARDY, C. R., RUTSCHMANN, F. 2005. Taxon sampling effects in molecular clock dating: An example from the African Restionaceae. *Molecular Phylogenetics and Evolution,* 35**,** 569-582.

LIU, M., ZHANG, Z., PENG, Z. 2014. The mitochondrial genome of the water spider Argyroneta aquatica (Araneae: Cybaeidae). *Zoologica Scripta* 44:179–190.

LOWRY, J. K., MYERS, A. A. 2013. A Phylogeny and Classification of the Senticaudata subord. nov. (Crustacea: Amphipoda). *Zootaxa,* 310**,** 001–080.

LOWRY, J. K., MYERS, A. A. 2012. New, mainly southern hemisphere, freshwater families of Amphipoda (Crustacea), together with a description of the first freshwater calliopiid, Lutriwita bradburyi gen. nov. et sp. nov. *Zootaxa* 3499:27–45.

MACEY, J. R., LARSON, A., ANANJEVA, N. B., FANG, Z., PAPENFUSS, T. J. 1997. Two novel gene orders and the role of light–strand replication in rearrangement of the vertebrate mitochondrial genome. *Molecular Biology and Evolution* 14:91–104.

MADDEN, T. L., TATUSOV, R. L., ZHANG, J. 1996. [9] Applications of network BLAST server. *In:* RUSSELL, F. D. (ed.) *Methods in Enzymology.* Academic Press.

MAGNIEZ, G. 1975. Observations sur la biologie de Stenasellus virei (Crustacea Isopoda Asellota des caux souterraines). *International Journal of Speleology,* 7**,** 79-228.

MALLET, J. 1995. A species definition for the modern synthesis. *Trends in Ecology & Evolution,* 10**,** 294-299.

MARGALEF, R. 1970. Anfípodos recolectados en aguas subterráneas ibéricas. *Speleon,* 17**,** 63-65.

MARMONIER, P., VERVIER, P., GIBER, J., DOLE-OLIVIER, M.-J. 1993. Biodiversity in ground waters. *Trends in Ecology & Evolution,* 8**,** 392-395.

MARTÍN, J. A., PUGA-BERNABÉU, A., AGUIRRE, J., BRAGA, J. C. 2014. Miocene Atlantic-Mediterranean seaways in the Betic Cordillera (Southern Spain). *Revista de la Sociedad Geológica de España,* 27**,** 175-186.

MATHIEU, J., ESSAFI, K., CHERGUI, H. 1999. Spatial and temporal variations of stygobite Amphipod populations in interstitial aquatic habitats of karst floodplain interfaces in France and Morocco. *Annales De Limnologie-International Journal of Limnology,* 35**,** 133-139.

MATSUMOTO, Y., YANASE, T., TSUDA, T., NODA, H. 2009. Species–specific mitochondrial gene rearrangements in biting midges and vector species identification. *Medical and Veterinary Entomology,* 23:47–55.

MAYDEN, R. L. 1997. A hierarchy of species concepts: the denouement in the saga of the species problem. *In:* CLARIDGE, M. F., DAWAH, H. A., WILSON, M. R. (eds.) *Species: The units of diversity.* Chapman and Hall.

MEIER, R., SHIYANG, K., VAIDYA, G., NG, P. K. L. 2006. DNA Barcoding and Taxonomy in Diptera: A Tale of High Intraspecific Variability and Low Identification Success. *Systematic Biology,* 55**,** 715-728.

MELEG, I. N., ZAKŠEK, V., FIŠER, C., KELEMEN, B. S., MOLDOVAN, O. T. 2013. Can Environment Predict Cryptic Diversity? The Case of *Niphargus* Inhabiting Western Carpathian Groundwater. *PLoS One,* 8**,** e76760.

MESSOULI, M., MESSANA, G., YACOUBI-KHEBIZA, M. 2006. Three new species of Pseudoniphargus (Amphipoda), from the groundwater of three Mediterranean islands, with notes on the Ps. adriaticus. *Subterranean Biology,* 4**,** 79-101.

MESSOULI, M., MESSANA, G., YACOUBI–KHEBIZA, M., COINEAU, N. 2001. Pseudoniphargus (subterranean crustacean amphipod) from Morocco: systematics, phylogeny and ecological and biogeographic aspects In: 13th International Congress of Speleology. Brazilia: SBE, 391–394.

MEYER, C. P., PAULAY, G. 2005. DNA Barcoding: Error Rates Based on Comprehensive Sampling. *PLoS Biol,* 3**,** e422.

MOLINARI, J. Ã., GUTIÈRREZ, E. E., DE ASCENÇÃO, A. A., NASAR, J. M., ARENDS, A., MÀRQUEZ, R. J. 2005. Predation by giant centipedes, Scolopendra gigantea, on three species of bats in a Venezuelan cave. *Caribbean Journal of Science,* 41**,** 340-346.

MONAGHAN, M. T., WILD, R., ELLIOT, M., FUJISAWA, T., BALKE, M., INWARD, D. J. G., LEES, D. C., RANAIVOSOLO, R., EGGLETON, P., BARRACLOUGH, T. G., VOGLER, A. P. 2009. Accelerated Species Inventory on Madagascar Using Coalescent-Based Models of Species Delineation. *Systematic Biology*.

MORA, J. A., VAN HEININGEN, M., BERMÚDEZ, R., RUIZ-RUANO, F., ALCALÁ, A. 2011. *Cueva del Yeso de Baena,* Córdoba, Ayuntamiento de Baena.

MORITZ, C., BROWN, W. 1987. Tandem duplications in animal mitochondrial DNAs: variation in incidence and gene content among lizards. *Proceedings of the National Academy of Sciences* USA 84:7183–7187.

MORITZ, C., PATTON, J. L., SCHNEIDER, C. J., SMITH, T. B. 2000. Diversification of Rainforest Faunas: An Integrated Molecular Approach. *Annual Review of Ecology and Systematics,* 31**,** 533-563.

MORRISON, D. A. 2011. Species: A History of the Idea. *Systematic Biology,* 60**,** 239-241.

NADLER, S. A., DE LEÓN, G. P.-P. 2011. Integrating molecular and morphological approaches for characterizing parasite cryptic species: implications for parasitology. *Parasitology,* 138**,** 1688-1709.

NASH III, T. H. 1996. *Lichen Biology,* Cambrige, Cambridge University Press.

NEAR, T. J., MEYLAN, P. A., SHAFFER, H. B. 2005. Assessing concordance of fossil calibration points in molecular clock studies: an example using turtles. *The American Naturalist,* 165.

NEE, S., MAY, R. M., HARVEY, P. H. 1994. The Reconstructed Evolutionary Process. *Philosophical Transactions of the Royal Society of London B: Biological Sciences,* 344**,** 305-311.

NEGRISOLO, E., BABBUCCI, M., PATARNELLO, T. 2011. The mitochondrial genome of the ascalaphid owlfly Libelloides macaronius and comparative evolutionary mitochondriomics of neuropterid insects. *BMC Genomics* 12:221.

NGUYEN, L. T., SCHMIDT, H. A., VON HAESELER, A., MINH, B.Q. (2015) IQ-TREE: A fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. *Molecular Biology and Evolution*, 32:268-274.

NOTENBOOM, J. 1986. The species of the genus Pseudoniphargus Chevreux, 1901 (Amphipoda) from northern Spain. *Contrib Zool* 56:75–122.

NOTENBOOM, J. 1987a. Lusitanian species of the amphipod Pseudoniphargus Chevreux, 1901 with a key to all the known Iberian species. *Contributions to Zoology* 57:191–206.

NOTENBOOM, J. 1987b. Species of the genus Pseudoniphargus Chevreux, 1901 (Amphipoda) from the Betic Cordillera of southern Spain. *Contributions to Zoology* 57:87–150.

NOTENBOOM, J. 1988a. Parapseudoniphargus baetis, new genus, new species, a stygobiont amphipod crustacean from the Guadaquivir river basin (Southern Spain), with phylogenetic implications. *Journal of Crustacean Biology* 8:110–121.

NOTENBOOM, J. 1991. Marine Regressions and the evolution of groundwater dwelling amphipods (Crustacea). *Journal of Biogeography* 18:437–454.

NOTENBOOM, J. 1988b. Phylogenetic relationships and biogeography of the groundwater-dwelling amphipod genus Pseudoniphargus (Crustacea), with emphasis on the Iberian species. *Contributions to Zoology,* 58**,** 159-204.

NYGREN, A. 2014. Cryptic polychaete diversity: a review. *Zoologica Scripta,* 43**,** 172-183.

NYLANDER, J. A. A. 2004. MrAIC.pl v1.4.6. Evolutionary Biology Centre, Uppsala University

OJALA, D., MERKEL, C., GELFAND, R., ATTARDI, G. 1980. The tRNA genes punctuate the reading of genetic information in human mitochondrial DNA. *Cell* 22:393–403.

OJALA, D., MONTOYA, J., ATTARDI, G. 1981. tRNA punctuation model of RNA processing in human mitochondria. *Nature* 290:470–474.

PAGE, T. J., HUGHES, J. M., REAL, K. M., STEVENS, M. I., KING, R. A., HUMPHREYS, W. F. 2016. Allegory of a cave crustacean: systematic and biogeographic reality of Halosbaena (Peracarida: Thermosbaenacea) sought with molecular data and multiple scales. *Marine Biodiversity* DOI 10.1007/s12526-016-0565-3.

PAGE, R. D. M., HOLMES, E. C. 1998. *Molecular Evolution: A Phylogenetic Approach,* New York, Wiley-Blackwell.

PAPADOPOULOU, A., MONAGHAN, M. T., BARRACLOUGH, T. G., VOGLER, A. P. 2009. Sampling Error Does Not Invalidate the Yule-Coalescent Model for Species Delimitation. A Response to Lohse 2009. *Systematic Biology,* 58**,** 442-444.

PAUL, C. R. C. 1992 The recognition of ancestors. *Historical Biology* 6, 239–250. (doi:10.1080/10292389209380433)

PEARLMAN, H. 2016 U. S. GEOLOGICAL SURVEY. 2014. The World's Water [Online]. Available: http://water.usgs.gov/edu/earthwherewater.html. Accessed on 2017-05-08

PENG, Y., LEUNG, H. C. M., YIU, S. M., CHIN, L. Y. 2012. IDBA-UD: a *de novo* assembler for single-cell and metagenomic sequencing data with highly uneven depth *Bioinformatics,* 28**,** 1420-1428.

PERSEKE, M., FRITZSCH, G., RAMSCH, K., BERNT, M., MERKLE, D., MIDDENDORF, M., BERNHARD, D., STADLER, P. F., SCHLEGEL, M. 2008. Evolution of mitochondrial gene orders in echinoderms. *Molecular Phylogenetics and Evolution* 47:855–864.

PHILLIPS, M. J. 2009. Branch-length estimation bias misleads molecular dating for a vertebrate molecular phylogeny. *Gene* 441:132–140.

PONS, J., BARRACLOUGH, T. G., GOMEZ-ZURITA, J., CARDOSO, A., DURAN, D. P., HAZELL, S., KAMOUN, S., SUMLIN, W. D., VOGLER, A. P. 2006. Sequence-Based Species Delimitation for the DNA Taxonomy of Undescribed Insects. *Systematic Biology,* 55**,** 595-609.

PONS, J., BAUZA–RIBOT, M., JAUME, D., JUAN, C. 2014. Next–generation sequencing, phylogenetic signal and comparative mitogenomic analyses in Metacrangonyctidae (Amphipoda: Crustacea). *BMC Genomics* 15:566.

PONS, J., RIBERA, I., BERTRANPETIT, J., BALKE, M. 2010. Nucleotide substitution rates for the full set of mitochondrial protein–coding genes in Coleoptera. *Molecular Phylogenetics and Evolution* 56:796–807.

PRENDINI, L., FRANCKE, O. F., VIGNOLI, V. 2010. Troglomorphism, trichobothriotaxy and typhlochactid phylogeny (Scorpiones, Chactoidea): more evidence that troglobitism is not an evolutionary dead-end. *Cladistics,* 26**,** 117-142.PORTER, M. L. 2007. Subterranean biogeography: What have we learned from molecular techniques? *Journal of Cave and Karst Studies,* 69**,** 179-186.

POSADA, D., CRANDALL, K. A. 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics,* 14**,** 817-818.

POSADA, D., CRANDALL, K. A. 2001. Intraspecific gene genealogies: trees grafting innto networks *Trends in Ecology & Evolution,* 16**,** 37-45.

POULSON, T. L. 1963. Cave Adaptation in Amblyopsid Fishes. *The American Midland Naturalist,* 70**,** 257-290.

POWELL, J. R. 2012. Accounting for uncertainty in species delineation during the analysis of environmental DNA sequence data. *Methods in Ecology and Evolution,* 3**,** 1-11.

PRETUS, J. L. 1988. A new stygobiont amphipod, Pseudoniphargus mercadali N. sp., from the Island of Minorca (Balearic Archipelago) *Stygologia,* 4**,** 229-241.

PRETUS, J. L. 1990. Three New species of the genus Pseudoniphargus (Crustacea, Amphipoda) in Balearic ground waters. *Stygologia,* 5**,** 101-118.

PROUDLOVE, G., WOOD, P. J. 2003. The blind leading the blind: cryptic subterranean species and DNA taxonomy. *Trends in Ecology & Evolution,* 18**,** 272-273.

PUILLANDRE, N., LAMBERT, A., BROUILLET, S., ACHAZ, G. 2012. ABGD, Automatic Barcode Gap Discovery for primary species delimitation. *Molecular Ecology,* 21**,** 1864-1877.

PYBUS, O. G. 2006. Model Selection and the Molecular Clock. *PLoS Biology,* 4**,** e151.

R DEVELOPMENT CORE TEAM 2008. R: A Language and Environment for Statistical Computing. Version 2.12 ed. Vienna, Austria: R Foundation for Statistical Computing.

RAMBAUT, A., SUCHARD, M. A., XIE, D., DRUMMOND, A. J. 2015. Tracer v1.6.

RAWLINGS, T. A., COLLINS, T. M., BIELER, R. 2001. A Major Mitochondrial Gene Rearrangement Among Closely Related Species. *Molecular Biology and Evolution,* 18**,** 1604-1609.

RENNER, S. S. 2004. Bayesian analysis of combined chloroplast loci, using multiple calibrations, supports the recent arrival of Melastomataceae in Africa and Madagascar. *American Journal of Botany,* 91**,** 1427-1435.

RENNER, S. S. 2016. A Return to Linnaeus's Focus on Diagnosis, Not Description: The Use of DNA Characters in the Formal Naming of Species. *Systematic Biology,* 65**,** 1085-1095.

REYDON, T. A. C. 2005. On the nature of the species problem and the four meanings of 'species'. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences,* 36**,** 135-158.

RICE, P., LONGDEN, I., BLEASBY, A. 2000. EMBOSS: the European molecular biology open software suite. *Trends in Genetics* 16:276–277.

RÖGL, F. 1998. Palaeogeographic considerations for Mediterranean and Parathetys eaways (Oligocene to Miocene). *Annalen des Naturhistorischen Museums in Wien 99A*: 279–310.

ROUCH, R., DANIELOPOL, D. L. 1987. L'origine de la faune aquatique souterraine entre le paradigme du re´fuge et la mode`le de la colonization active. *Stygologia,* 3**,** 345-372.

ROUSSEAU, F., BURROWES, R., PETERS, A. F., KUHLENKAMP, R., DE REVIERS, B. 2001. A comprehensive phylogeny of the Phaeophyceae based on nrDNA sequences resolves the earliest divergences. *Comptes Rendus de l'Académie des Sciences - Series III - Sciences de la Vie,* 324**,** 305-319.

RUBINOFF, D., CAMERON, S., WILL, K. 2006. A Genomic Perspective on the Shortcomings of Mitochondrial DNA for "Barcoding" Identification. *Journal of Heredity,* 97**,** 581-594.

RUTSCHMANN, F. 2006. Molecular dating of phylogenetic trees: A brief review of current methods that estimate divergence times. *Diversity and Distributions,* 12**,** 35-48.

RUTSCHMANN, F., ERIKSSON, T., SALIM, K. A., CONTI, E. 2007. Assessing calibration uncertainty in molecular dating: the assignment of fossils to alternative calibration points. *Sysematict Biology,* 56.

SAEZ, A. G. & LOZANO, E. 2005. Body doubles. *Nature,* 433**,** 111-111.

SAN MAURO, D., GOWER, D. J., ZARDOYA, R., WILKINSON, M. 2006. A hotspot of gene order rearrangement by tandem duplication and random loss in the vertebrate mitochondrial genome. *Molecular Biology and Evolution* 23:227–234.

SANCHEZ, E. 1989. First records of the genus Pseudoniphargus (Amphipoda) from Gran Canaria, with description of a new species. Stygofauna of the Canary Islands, 17. *Contributions to Zoology,* 59**,** 229-238.

SÁNCHEZ, E. 1990. A new species of *Pseudoniphargus* (Crustacea, Amphipoda) from subterranean waters in Tenerife (Canary Islands). *Hydrobiologia,* 196**,** 51-63.

SANCHEZ, E. 1991. New records of Pseudoniphargus (crustacea, Amphipoda) from ground waters of Tenerife and Hierro, with description of new species. *Stygologia,* 6**,** 53-64.

SANDERSON, M. J. 2003. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics,* 19**,** 301-302.

SANO, N., KURABAYASHI, A., FUJII, T., YONEKAWA, H., SUMIDA, M. 2005. Complete nucleotide sequence of the mitochondrial genome of Schlegel´s tree frog Rhacophorus schlegelii (family Rhacophoridae): duplicated control regions and gene rearrangements. *Genes & Genetic Systems* 80:213–224.

SATLER, J. D., CARSTENS, B. C., HEDIN, M. 2013. Multilocus Species Delimitation in a Complex of Morphologically Conserved Trapdoor Spiders (Mygalomorphae, Antrodiaetidae, Aliatypus). *Systematic Biology,* 62**,** 805-823.

SATO, A., NAKADA, K., AKIMOTO, M., ISHIKAWA, K., ONO, T., SHITARA, H., YONEKAWA, H., HAYASHI, J.-I. 2005. Rare creation of recombinant mtDNA haplotypes in mammalian tissues. *Proceedings of the National Academy of Sciences,* 102**,** 6057-6062.

SCHANDER, C., WILLASSEN, E. 2005. What can biological barcoding do for marine biology? *Marine Biology Research,* 1**,** 79-83.

SCHATTNER, P., BROOKS, A. N., LOWE, T. M. 2005. The tRNAscan–SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Research* 33:W686–W689.

SCHMIDT, S. I., HAHN, H. J. 2012. What is groundwater and what does this mean to fauna? – An opinion. *Limnologica - Ecology and Management of Inland Waters,* 42**,** 1-6.

SCHULTZ, J., WOLF, M. 2009. ITS2 sequence-structure analysis in phylogenetics: A how-to manual for molecular systematics. *Molecular Phylogenetics and Evolution,* 52**,** 520-523.

SHAO, R., DOWTON, M., MURRELL, A., BARKER, S. C. 2003. Rates of Gene Rearrangement and Nucleotide Substitution Are Correlated in the Mitochondrial Genomes of Insects. *Molecular Biology and Evolution,* 20**,** 1612-1619.

SHEFFIELD, N., SONG, H., CAMERON, L., WHITING, M. 2008. A comparative analysis of mitochondrial genomes in Coleoptera (Arthropoda: Insecta) and genome descriptions of six new beetles. *Molecular Biology and Evolution.* 25:2499–2509.

SHEN, H., BRABAND, A., SCHOLTZ, G. 2013. Mitogenomic analysis of decapod crustacean phylogeny corroborates traditional views on their relationships. *Molecular Phylogenetics and Evolution,* 66**,** 776-789.

SHEN, X., TIAN, M., YAN, B., CHU, K. 2015. Phylomitogenomics of Malacostraca (Arthropoda: Crustacea). *Acta Oceanologica Sinica,* 34**,** 84-92.

SHIH, H.-T., ZHOU, X.-M., CHEN, G.-X., CHIEN, I. C., NG, P. K. L. 2011. Recent vicariant and dispersal events affecting the phylogeny and biogeography of East Asian freshwater crab genus Nanhaipotamon (Decapoda: Potamidae). *Molecular Phylogenetics and Evolution,* 58**,** 427-438.

SITES JR, J. W., MARSHALL, J. C. 2003. Delimiting species: a Renaissance issue in systematic biology. *Trends in Ecology & Evolution,* 18**,** 462-470.

SKET, B. (ed.) 2005. *Anchialine caves,* Amsterdam: ElsevierAcademic Press.

SKET, B. 1986. Ecology of the mixohaline hypogean fauna along the Yugoslav coasts. *Stygologia,* 2**,** 317-338.

SOUBRIER, J., STEEL, M., LEE, M. S. Y., SARKISSIAN, C. D., GUINDON, S., HO, S. Y. W., COOPER, A. 2012. The influence of rate heterogeneity among sites on the time dependence of molecular rates. *Molecular Biology and Evolution*.

SPEICH, S. M., JONES, H. L., BENEDICT, E. M. 1986. Review of the Natural Nesting of the Barn Swallow in North America. *The American Midland Naturalist,* 115**,** 248-254.

SPRIBILLE, T., TUOVINEN, V., RESL, P., VANDERPOOL, D., WOLINSKI, H., AIME, M. C., SCHNEIDER, K., STABENTHEINER, E., TOOME-HELLER, M., THOR, G., MAYRHOFER, H., JOHANNESSON, H., MCCUTCHEON, J. P. 2016. Basidiomycete yeasts in the cortex of ascomycete macrolichens. *Science,* 353**,** 488-492.

STAMATAKIS, A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics,* 22**,** 2688-2690.

STARR, H. W., HEGNA, T. A., MCMENAMIN, M. A. S. 2016. Epilogue to the tale of the Triassic amphipod: Rosagammarus McMenamin, Zapata and Hussley, 2013 is a decapod tail (Luning Formation, Nevada, USA). *Journal of Crustacean Biology*, 36: 525–529.

STECH, M., VELDMAN, S., LARRAÍN, J., MUÑOZ, J., QUANDT, D., HASSEL, K., KRUIJER, H. 2013. Molecular Species Delimitation in the *Racomitrium canescens* Complex (Grimmiaceae) and Implications for DNA Barcoding of Species Complexes in Mosses. *PLoS One,* 8**,** e53134.

STOCH, F., GALASSI, D. P. 2010. Stygobiotic crustacean species richness: a question of numbers, a matter of scale. *Hydrobiologia,* 653**,** 217-234.

STOCH, F. 1995. The ecological and historical determinants of crustacean diversity in groundwaters, or: why are there so many species? *Mémories de Biospéologie,* 22**,** 139-160.

STOCK, J. H., ABREU, A. D. 1992. Three new species of Pseudoniphargus (Crustacea: Amphipoda) from the Madeira Archipelago. *Boletim do Museu Municipal do Funchal,* 44**,** 131-155.

STOCK, J. H. 1988. The amphipod genus Pseudoniphargus (Crustacea) in the Canary Islands. *Bijdragen tot de Dierkunde.* 58:47–78.

STOCK, J. H. 1974. The systematics of certain Ponto-Caspian Gammaridae (Crustacea, Amphipoda). *Mitteilungen aus dem Hamburgischen Zoologischen Museum und Institut,* 70**,** 75-95.

STOCK, J. H. 1980. Regression model evolution as exemplified by the genus Pseudoniphargus (Amphipoda). *Bijdragen tot de Dierkunde,* 50**,** 105-144.

STOCK, J. H. 1988. The amphipod genus Pseudoniphargus (Crustacea) in the Canary Islands. *Bijdragen tot de Dierkunde,* 58**,** 47-78.

STOCK, J. H. 1993. Some remarkable distribution patterns in stygobiont Amphipoda. *Journal of Natural History,* 27**,** 807-819.

STOCK, J. H., HOLSINGER, J. R., SKET, B., ILIFFE, T. M. 1986. Two new species of *Pseudoniphargus* (Amphipoda), in Bermudian groundwaters. *Zoologica scripta,* 15**,** 237-249.

STOKKAN, M., JURADO-RIVERA, J. A., JUAN, C., JAUME, D., PONS, J. 2016. Mitochondrial genome rearrangements at low taxonomic levels: three distinct mitogenome gene orders in the genus Pseudoniphargus (Crustacea: Amphipoda). *Mitochondrial DNA***,** 1-11.

STUTZ, H. L., SHIOZAWA, D. K., EVANS, R. P. 2010. Inferring dispersal of aquatic invertebrates from genetic variation: a comparative study of an amphipod and mayfly in Great Basin springs. *Journal of the North American Benthological Society,* 29**,** 1132-1147.

SUKUMARAN, J., KNOWLES, L. L. 2017. Multispecies coalescent delimits structure, not species. *Proceedings of the National Academy of Sciences,* 114**,** 1607-1612.

SUNDBERG, P., THUROCZY VODOTI, E., STRAND, M. 2010. DNA barcoding should accompany taxonomy – the case of Cerebratulus spp (Nemertea). *Molecular Ecology Resources,* 10**,** 274-281.

SUPEK, F., VLAHOVIČEK, K. 2004. INCA: synonymous codon usage analysis and clustering by means of self–organizing map. *Bioinformatics* 20:2329–2330.

SURIC, M., LONCARIC, R., LONCAR, N. 2010. Submerged caves of Croatia: distribution, classification and origin. *Environmental Earth Sciences,* 61.

SWOFFORD, D. L. 2003. PAUP*: Phylogenetic Analysis Using Parsimony (* and Other Methods). *In:* ASSOCIATES, S. (ed.) v 4.0 ed. Sunderland, Massachusetts: Nature Publishing Group.

TALAVERA, G. & VILA, R. 2011. What is the phylogenetic signal limit from mitogenomes? The reconciliation between mitochondrial and nuclear data in the Insecta class phylogeny. *BMC Evolutionary Biology,* 11**,** 315.

TALAVERA, G., DINCĂ, V., VILA, R. 2013. Factors affecting species delimitations with the GMYC model: insights from a butterfly survey. *Methods in Ecology and Evolution,* 4**,** 1101-1110.

TAMURA, K., PETERSON, D., PETERSON, N., STECHER, G., NEI, M. & KUMAR, S. 2011. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution,* 28**,** 2731 - 2739.

TAMURA, K., STECHER, G., PETERSON, D., FILIPSKI, A., KUMAR, S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution,* 30**,** 2725-2729.

TANG, C. Q., HUMPHREYS, A. M., FONTANETO, D., BARRACLOUGH, T. G. 2014. Effects of phylogenetic reconstruction method on the robustness of species delimitation using single-locus data. *Methods in Ecology and Evolution,* 5**,** 1086-1094.

TELETCHEA, F. 2010. After 7 years and 1000 citations: Comparative assessment of the DNA barcoding and the DNA taxonomy proposals for taxonomists and non-taxonomists. *Mitochondrial DNA,* 21**,** 206-226.

TEMPLETON, CRANDALL, K. A., SING, C. F. 1992. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics,* 132**,** 619-33.

TIMMERMANS, M., DODSWORTH, S., CULVERWELL, C., BOCAK, L., AHRENS, D., LITTLEWOOD, D., PONS, J., VOGLER, A. 2010. Why barcode? High-throughput multiplex sequencing of mitochondrial genomes for molecular systematics. *Nucleic Acids Research,* 38**,** e197.

TOUSSAINT, E. F. A., MORINIÈRE, J., MÜLLER, C. J., KUNTE, K., TURLIN, B., HAUSMANN, A. BALKE, M. 2015. Comparative molecular species delimitation in the charismatic Nawab butterflies (Nymphalidae, Charaxinae, Polyura). *Molecular Phylogenetics and Evolution,* 91**,** 194-209.

TRONTELJ, P., FISER, C. 2009. Cryptic species diversity should not be trivialised. *Systematics and Biodiversity,* 7.

TRONTELJ, P., DOUADY, C. J., FIŠER, C., GIBERT, J., GORIČKI, Š., LEFÉBURE, T., SKET, B., ZAKŠEK, V. 2009. A molecular test for cryptic diversity in ground water: how large are the ranges of macro-stygobionts? *Freshwater Biology,* 54**,** 727-744.

TRONTELJ, P., GORICKI, S., POLAK, S., VEROVNIK, R., ZAKŠEK, V., SKET, B. 2007. Age estimates for some subterranean taxa and lineages in the Dinaric Karst. *Acta Carsologica,* 36**,** 183-189.

TSAOUSIS, A. D., MARTIN, D. P., LADOUKAKIS ED, POSADA D, ZOUROS E. 2005. Widespread recombination in published animal mtDNA sequences. *Molecular Biology and Evolution* 22:925–933.

TUINSTRA, J. & VAN WENSEM, J. 2014. Ecosystem services in sustainable groundwater management. *Science of The Total Environment,* 485–486**,** 798-803.

VILSTRUP, J. T., HO, S. Y., FOOTE, A. D., MORIN, P. A., KREB, D., KRÜTZEN, M., PARRA, G. J., ROBERTSON, K. M., STEPHANIS, R., VERBORGH, P., WILLERSLEV, E., ORLANDO, L., GILBERT, M. T. P. 2011. Mitogenomic phylogenetic analyses of the Delphinidae with an emphasis on the Globicephalinae. *BMC Evolutionary Biology,* 11**,** 1-10.

WANG, Y., HUANG, X., QIAO, G. 2013. Comparative analysis of mitochondrial genomes of five aphid species (Hemiptera: Aphididae) and phylogenetic implications. *PLoS One* 8:e77511.

WAUGH, J. 2007. DNA barcoding in animal species: progress, potential and pitfalls. *Bioessays,* 29**,** 188-197.

WERTHEIM, J.O., FOURMENT, M., POND, S. L. K. 2012. Inconsistencies in estimating the age of HIV-1 subtypes due to heterotachy. *Molecular Biology and Evolution*. 29:451–456.

WHITE, W. B. 2005. Hydrogeology of Karst Aquifers. *In:* CULVER, D. C. & WHITE, W. B. (eds.) *Encyclopedia of caves.* Burlington, Mass. : Elsevier Academic Press, 680.

WIELSTRA, B., ARNTZEN, J. 2011. Unraveling the rapid radiation of crested newts (Triturus cristatus superspecies) using complete mitogenomic sequences. *BMC Evoutionaryl Biology,* 11**,** 162.

WIENS, J. J. 2007. Species Delimitation: New Approaches for Discovering Diversity. *Systematic Biology,* 56**,** 875-878.

WIENS, J. J., CHIPPINDALE, P. T., HILLIS, D. M. 2003. When Are Phylogenetic Analyses Misled by Convergence? A Case Study in Texas Cave Salamanders. *Systematic Biology,* 52**,** 501-514.

WILKINS, J. S. 2011. Philosophically speaking, how many species concepts are there? *Zootaxa***,** 58-60.

WILSON, G. D. F., EDGECOMBE, G. D. 2003. The Triassic isopod Protamphisopus wianamattensis (Chilton) and comparison with extant taxa (Crustacea, Phreatoicidea). *Journal of Paleontology,* 77: 454–470.

WRIGHT, F. 1990. The 'effective number of codons' used in a gene. *Gene* 87:23–29.

WU, C.-I. 2001. The genic view of the process of speciation. *Journal of Evolutionary Biology,* 14**,** 851-865.

WYMAN, S. K., JANSEN, R. K., BOORE, J. L. 2004. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20:3252–3255.

XIA, X., XIE, Z., SALEMI, M., CHEN, L., WANG, Y. 2003. An index of substitution saturation and its application. *Molecular Phylogenetics and Evolution,* 26**,** 1-7.

XIA, X., LEMEY, P. 2009. Assessing substitution saturation with DAMBE. Pp. 615-630 in Philippe Lemey. Marco Salemi and Anne-Mieke Vandamme. eds. The Phylogenetic Handbook: A Practical Approach to DNA and Protein Phylogeny. 2nd edition Cambridge University Press.

YANG, J. S., YANG, W. J. 2008. The complete mitochondrial genome sequence of the hydrothermal vent galatheid crab Shinkaia crosnieri (Crustacea: Decapoda: Anomura): A novel arrangement and incomplete tRNA suite. *BMC Genomics* 9:257.

YANG, Z. & RANNALA, B. 2010. Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences,* 107**,** 9264-9269.

YANG, Z. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *Journal of Molecular Evolution,* 39**,** 306 - 314.

YANG, Z. 2004. A heuristic rate smoothing procedure for maximum likelihood estimation of species divergence times. *Acta Zoologica Sinica,* 50**,** 645-656.

YANG, Z. 2006. *Computational molecular evolution,* USA, Oxford University Press.

YANG, Z. 2008. Empirical evaluation of a prior for Bayesian phylogenetic inference. *Royal Society Philosophical Transactions Biological Sciences* 363:4031–4039.

ZAKŠEK, V., SKET, B. & TRONTELJ, P. 2007. Phylogeny of the cave shrimp Troglocaris: Evidence of a young connection between Balkans and Caucasus. *Molecular Phylogenetics and Evolution,* 42**,** 223-235.

ZAKŠEK, V., SKET, B., GOTTSTEIN, S., FRANJEVIĆ, D., TRONTELJ, P. 2009. The limits of cryptic diversity in groundwater: phylogeography of the cave shrimp Troglocaris anophthalmus (Crustacea: Decapoda: Atyidae). *Molecular Ecology,* 18**,** 931-946.

ZHANG, C., ZHANG, D.-X., ZHU, T., YANG, Z. 2011. Evaluation of a Bayesian Coalescent Method of Species Delimitation. *Systematic Biology,* 60**,** 747-761.

ZHANG, J., KAPLI, P., PAVLIDIS, P., STAMATAKIS, A. 2013. A general species delimitation method with applications to phylogenetic placements. *Bioinformatics,* 29**,** 2869-2876.

ZUKER, M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research* 31:3406–3415.

Appendix

Appendix 1 List of all known members of Pseudoniphargidae prior to this study

Species	Descriptor	Distribution	Habitat
Parapseudoniphargus baetis	Notenboom, 1988	Spain (Jaén; Sevilla)	fresh
Pseudoniphargus adriaticus	S. Karaman, 1955	Western Mediterranean & Adriatic Sea	brackish; marine
Pseudoniphargus affinis	Notenboom, 1987	Spain (Granada)	fresh
Pseudoniphargus africanus	Chevreux, 1901	Algeria	fresh
Pseudoniphargus associatus	Sánchez, 1991	Canary Islands (Tenerife)	fresh
Pseudoniphargus branchiatus	Stock, 1980	Spain (Albacete; Alicante; Valencia)	fresh
Pseudoniphargus brevipedunculatus	Stock, 1980	Açores (Faial; Pico)	fresh
Pseudoniphargus burgensis	Notenboom, 1986	Spain (Burgos)	fresh
Pseudoniphargus callaicus	Notenboom, 1987	Spain (La Coruña)	fresh
Pseudoniphargus candelariae	Sánchez, 1990	Canary Islands (Tenerife)	fresh
Pseudoniphargus carpalis	Stock, Holsinger, Sket & Iliffe, 1986	Bermuda	fresh
Pseudoniphargus cazorlae	Notenboom, 1987	Spain (Jaén)	fresh
Pseudoniphargus cupicola	Stock, 1988	Canary Islands (La Palma)	brackish
Pseudoniphargus daviui	Jaume, 1991	Balearic Islands (Cabrera)	fresh
Pseudoniphargus duplus	Messouli, Messana & Yacoubi-Khebiza, 2006	Sicily	fresh?
Pseudoniphargus eborarius	Notenboom, 1986	Spain (Burgos)	fresh
Pseudoniphargus elongatus	Stock, 1980	Spain (Burgos; Santander)	fresh
Pseudoniphargus fontinalis	Stock, 1988	Canary Islands (Gran Canaria; Tenerife)	fresh; brackish
Pseudoniphargus fragilis	Notenboom, 1987	Spain (Granada; Málaga)	fresh
Pseudoniphargus gibraltaricus	Notenboom, 1987	Spain (Cádiz)	fresh
Pseudoniphargus gomerae	Stock, 1988	Canary Islands (La Gomera)	fresh
Pseudoniphargus gorbeanus	Notenboom, 1986	Spain (Alava)	fresh
Pseudoniphargus gracilis	Notenboom, 1987	Spain (Almería)	fresh
Pseudoniphargus granadensis	Notenboom, 1987	Spain (Granada)	fresh
Pseudoniphargus grandimanus	Stock, Holsinger, Sket & Iliffe, 1986	Bermuda	fresh; brackish; marine
Pseudoniphargus grandis	Notenboom, 1987	Spain (Granada; Málaga)	fresh

Pseudoniphargus guernicae	Notenboom, 1986	Spain (Vizcaya)	fresh
Pseudoniphargus illustris	Notenboom, 1987	Spain (Jaén)	fresh
Pseudoniphargus incantatus	Notenboom, 1986	Spain (Guipúzcoa; Navarra)	fresh
Pseudoniphargus inconditus	Karaman & Ruffo, 1989	Sicily	fresh
Pseudoniphargus italicus	Karaman & Ruffo, 1989	Sicily	fresh
Pseudoniphargus jereanus	Notenboom, 1986	Spain (Burgos)	fresh
Pseudoniphargus latipes	Notenboom, 1987	Spain (Albacete; Jaén; Sevilla)	fresh
Pseudoniphargus leucatensis	Bréhier & Jaume, 2009	France	brackish
Pseudoniphargus littoralis	Stock & Abreu, 1993	Madeira islands (Madeira)	brackish; marine?
Pseudoniphargus longicarpus	Notenboom, 1986	Spain (Oviedo)	fresh
Pseudoniphargus longicauda	Stock, 1988	Canary Islands (Tenerife)	fresh
Pseudoniphargus longiflagellum	Fakher-el-Abiari, Oulbaz, Messouli & Coineau, 1999	Morocco	freh
Pseudoniphargus longipes	Coineau & Boutin, 1996	Morocco	fresh
Pseudoniphargus longispinum	Stock, 1980	Portugal	fresh
Pseudoniphargus macrotelsonis	Stock, 1980	Algeria	fresh
Pseudoniphargus macrurus	Stock & Abreu, 1993	Madeira Islands (Madeira)	fresh
Pseudoniphargus margalefi	Notenboom, 1987	Spain (Alicante)	fresh
Pseudoniphargus maroccanus	Boutin & Coineau, 1988	Morocco	fresh
Pseudoniphargus mateusorum	Stock, 1980	Portugal	brackish?
Pseudoniphargus mercadali	Pretus, 1988	Balearic Islands (Mallorca; Menorca)	fresh
Pseudoniphargus montanus	Notenboom, 1986	Spain (León; Oviedo)	fresh
Pseudoniphargus multidens	Stock, 1988	Canary Islands (La Palma)	fresh; brackish
Pseudoniphargus nevadensis	Notenboom, 1987	Spain (Almería; Granada)	fresh
Pseudoniphargus obritus	Messouli, Messana & Yacoubi-Khebiza, 2006	Corsica	brackish?
Pseudoniphargus pedrerae	Pretus, 1990	Balearic Islands (Ibiza; Formentera)	fresh; brackish
Pseudoniphargus pedunculatus	Sánchez, 1989	Canary Islands (Gran Canaria)	fresh
Pseudoniphargus pityusensis	Pretus, 1990	Balearic Islands (Ibiza; Formentera)	fresh; brackish
Pseudoniphargus planasiae	Messouli, Messana & Yacoubi-Khebiza, 2006	Tuscan Archipelago (Pianosa)	fresh
Pseudoniphargus porticola	Stock, 1988	Canary Islands (Tenerife)	brackish
Pseudoniphargus portosancti	Stock & Abreu, 1993	Madeira Islands (Porto Santo)	fresh
Pseudoniphargus racovitzai	Pretus, 1990	Balearic Islands (Mallorca)	fresh
Pseudoniphargus romanorum	Coineau & Boutin, 1996	Morocco	fresh
Pseudoniphargus ruffoi	Coineau & Boutin, 1996	Morocco	fresh

Pseudoniphargus salinus	Stock, 1988	Canary Islands (El Hierro)	brackish
Pseudoniphargus semielongatus	Notenboom, 1986	Spain (Oviedo; Santander)	fresh
Pseudoniphargus sodalis	Karaman & Ruffo, 1989	Sicily	fresh
Pseudoniphargus sorbasiensis	Notenboom, 1987	Spain (Almería)	fresh
Pseudoniphargus spiniferus	Notenboom, 1986	Spain (Navarra)	fresh
Pseudoniphargus stocki	Notenboom, 1987	Spain (Cádiz; Málaga)	fresh
Pseudoniphargus triasi	Jaume, 1991	Balearic Islands (Cabrera)	fresh
Pseudoniphargus unisexualis	Stock, 1980	Spain (Guipúzcoa)	fresh
Pseudoniphargus unispinosus	Stock, 1988	Canary Islands (Tenerife)	fresh
Pseudoniphargus vasconiensis	Notenboom, 1986	Spain (Alava; Guipúzcoa)	fresh
Pseudoniphargus vomeratus	Notenboom, 1987	Spain (Jaén)	fresh

Appendix 2: Publication: see pdf document mspd

Mitochondrial genome rearrangements at low taxonomic levels three distinct mitogenome gene orders in the genus Pseudoniphargus Crustacea Amphipoda

Taylor & Francis
Taylor & Francis Group

Mitochondrial DNA Part A

DNA Mapping, Sequencing, and Analysis

ISSN: 2470-1394 (Print) 2470-1408 (Online) Journal homepage: http://www.tandfonline.com/loi/imdn21

Mitochondrial genome rearrangements at low taxonomic levels: three distinct mitogenome gene orders in the genus Pseudoniphargus (Crustacea: Amphipoda)

Morten Stokkan, Jose A. Jurado-Rivera, Carlos Juan, Damià Jaume & Joan Pons

To cite this article: Morten Stokkan, Jose A. Jurado-Rivera, Carlos Juan, Damià Jaume & Joan Pons (2016) Mitochondrial genome rearrangements at low taxonomic levels: three distinct mitogenome gene orders in the genus Pseudoniphargus (Crustacea: Amphipoda), Mitochondrial DNA Part A, 27:5, 3579-3589, DOI: 10.3109/19401736.2015.1079821

To link to this article: http://dx.doi.org/10.3109/19401736.2015.1079821

Published online: 02 Sep 2015.

Submit your article to this journal

Article views: 125

View related articles

View Crossmark data

Citing articles: 1 View citing articles

Full Terms & Conditions of access and use can be found at
http://www.tandfonline.com/action/journalInformation?journalCode=imdn21

Download by: [Inst Medit Estudios Avanzados] **Date:** 18 May 2017, At: 09:06

Mitochondrial DNA

http://informahealthcare.com/mdn
ISSN: 2470-1394 (print), 2470-1408 (electronic)

Mitochondrial DNA Part A, 2016; 27(5): 3579–3589
© 2015 Taylor & Francis. DOI: 10.3109/19401736.2015.1079821

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

Mitochondrial genome rearrangements at low taxonomic levels: three distinct mitogenome gene orders in the genus *Pseudoniphargus* (Crustacea: Amphipoda)

Morten Stokkan[1], Jose A. Jurado-Rivera[1,2], Carlos Juan[1,2], Damià Jaume[1], and Joan Pons[1]

[1]*Department of Biodiversity and Conservation, Instituto Mediterraneo de Estudios Avanzados (IMEDEA, CSIC-UIB), Esporles, Spain and* [2]*Departament de Biologia, Universitat de les Illes Balears, Palma, Spain*

Abstract

A comparison of mitochondrial genomes of three species of the amphipod *Pseudoniphargus* revealed the occurrence of a surprisingly high level of gene rearrangement involving protein-coding genes that is a rare phenomenon at low taxonomic levels. The three *Pseudoniphargus* mitogenomes also display a unique gene arrangement with respect to either the presumed Pancrustacean order or those known for other amphipods. Relative long non-coding sequences appear adjacent to the putative breakage points involved in gene rearrangements of protein coding genes. Other details of the newly obtained mitochondrial genomes – e.g., gene content, nucleotide composition and codon usage – are similar to those found in the mitogenomes of other amphipod species studied. They all contain the typical mitochondrial genome set consisting of 13 protein-coding genes, 22 tRNAs, and two rRNAS, as well as a large control region. The secondary structures and characteristics of tRNA and ribosomal mitochondrial genes of these three species are also discussed.

Keywords

Pseudoniphargidae, non-coding region, RNA structure, transposition and reversal events

History

Received 24 April 2015
Revised 26 July 2015
Accepted 29 July 2015
Published online 1 September 2015

Introduction

A typical metazoan mitochondrial genome (mitogenome) contains 13 protein-coding genes (PCGs), two ribosomal RNA genes (rRNA), 22 transfer RNA genes (tRNA), and a non-coding control region (CR) also called the D-loop that includes the origin of replication (Boore, 1999). However, gene order differs remarkably across higher taxonomic ranks (family and above) due to the occurrence of gene rearrangements such as transpositions and reversals (Arndt & Smith, 1998; Bauza-Ribot et al., 2009; Bensch & Harlid, 2000; Dowton & Austin, 1999; Hickerson & Cunningham, 2000; Irisarri et al., 2014). Rearrangements of this kind, especially those affecting tRNA genes, are frequently detected when comparing animal mitogenomes belonging to different taxonomic families or even between genera within the same family (Arndt & Smith, 1998; Dowton et al., 2009; Kurabayashi et al., 2008, 2010). But rearrangements within the same genus, especially those involving PCGs, are rather rare (Matsumoto et al., 2009; Rawlings et al., 2001). Mitochondrial gene rearrangements can occur as a product of one of these four major events: reversals, transpositions, reverse transpositions, or tandem duplications with subsequent random loss (TDRL) (Chaudhuri et al., 2006; Moritz & Brown, 1987; San Mauro et al., 2006). Reversals consist of one or several genes that switch from one DNA strand to the other, while transpositions involve a shift to another location without changing the sense of the strand. Both processes are deduced to have operated in reverse transpositions. TDRLs are more complex processes where a segment including more than one gene undergoes duplication and subsequently, suffers the deletion of particular gene copies at random, producing novel gene arrangements that can differ considerably from the ancestral gene order. The latter is an interesting mechanism from a phylogenetic point of view, since it is the only shift that can be considered with confidence to be irreversible (Chaudhuri et al., 2006). The rearrangement events that took place in a particular mitogenome with respect to an ancestral gene order can be heuristically determined using strong interval trees (Berard et al., 2007; Bernt et al., 2007; Perseke et al., 2008). The ultimate causes of rearrangement events in the mitochondrial genome remain unclear, although nuclear substitution rate has been shown to correlate with rearrangement frequency (Shao et al., 2003). Furthermore, the origin of replication – that is located in the control region – also has been postulated to be a hot-spot for gene rearrangements (Arunkumar & Nagaraju, 2006; Kraytsberg et al., 2004; Macey et al., 1997; San Mauro et al., 2006). Besides, life history and ecology traits such as founder effect and parasitism have been postulated to play a role in the gene rearrangement rate in some animal lineages (Tsaousis et al., 2005). Despite gene content of mitogenomes is mostly conserved across metazoans, nucleotide and aminoacid composition, as well as codon usage, vary largely across major taxonomic groups. Thus, insects and other arthropods display very AT-rich mitogenomes with PCGs showing heavily biased codon usages (Sheffield et al., 2008; Yang, 2008), while vertebrates show lower AT-richness (Sano et al., 2005).

The amphipod crustacean family Pseudoniphargidae (Karaman, 1993) contains only two genera, *Pseudoniphargus* (Chevreux, 1901) and *Parapseudoniphargus* (Notenboom, 1988),

Correspondence: Joan Pons, Department of Biodiversity and Conservation, Instituto Mediterraneo de Estudios Avanzados (IMEDEA), Miquel Marquès 21, 07190 Esporles, Spain. Tel: +34 971611332. E-mail: jpons@imedea.uib-csic.es

whereupon the latter is represented by only one species. *Pseudoniphargus* has 68 species currently described, all being obligate stygobionts (occurring exclusively in groundwaters), and displays a broad disjunct geographic distribution despite its members presumably having a poor dispersal ability (Notenboom, 1991). Endemic island *Pseudoniphargus* species are found in Bermuda as well as in the Atlantic archipelagoes of the Canaries, Madeira, and Azores. Species of the genus are also found on the southern margin of the Mediterranean (e.g., Algeria, Morocco, and southern Spain including the Balearic Islands) as well as in northern Spain and the south coast of France (Messouli et al., 2001; Notenboom, 1986, 1987a, b; Stock, 1980, 1988; Stock & Abreu, 1992; Stock et al., 1986). The mitogenome of *P. daviui* (Jaume, 1991), of around 15 kb, was sequenced and used as an outgroup in a broad scope biogeographic study (Bauzà-Ribot et al., 2012) although its gene order, composition, and secondary structures were not addressed. Here we analyze in greater detail the characteristics of this mitogenome along with newly obtained mitochondrial genomes of two additional congeneric species, namely *P. gorbeanus* (Notenboom, 1986) and *P. sorbasiensis* (Notenboom, 1987b), which display a different gene order, particularly focusing on gene rearrangements and secondary structures.

Methods

Sampling

Specimens of *P. gorbeanus* were collected at Artzegi'ko Koba (Zigoitia, Araba, Northern Spain), whereas those of *P. daviuvi* came from the island of Cabrera (Balearic Islands, Spain). Individuals of *P. sorbasiensis* were collected at Cueva del Agua (Sorbas, Almería; Southern Spain). All samples were kept in 96% ethanol immediately after collection and stored at −20 °C.

DNA extraction, amplification, and sequencing

DNA was extracted using Qiagen DNeasy Blood & Tissue kit (Qiagen, Hilden, Germany) from whole specimens following the protocol of the manufacturer. The entire mitogenomes were amplified in two large amplicons using species-specific primers designed to match the nucleotide sequences of cytochrome oxidase subunit 1 (*cox1*) and cytochrome oxidase b (*cob*), which were amplified using universal primers (Supplementary Table 1). The amplicon sizes were about 9 and 6 kb. Amplifications were performed using Herculase™ II Fusion DNA polymerase (Agilent Technologies, Santa Clara, CA) following the recommended protocol except for the addition of 10 µl of 5 M betaine per sample in *P. sorbasiensis* as an amplification enhancer. Large amplicons were purified using Invitek MSB® Spin PCRapace purification kit (Invitek GmbH, Berlin, Germany), and their concentration estimated using fluorometric quantification (Qubit Thermo Fisher Scientific Inc., Waltham, MA). The amplicons obtained for *P. sorbasiensis* and *P. gorbeanus*, in addition to four other taxa (12 amplicons in total), were pooled in equimolar concentration of 100 ng/µl and the batch sequenced using Junior NGS 454 sequencing as one single library.

Gene annotation and sequence analysis

Reads were assembled with CodonCode Aligner v5.0.1 (Codoncode Corporation, Dedham, MA), allowing a 10% mismatch to acquire the entire mitogenome, and using species-specific *cox1* and *cob* sequences as baits (Bauzà-Ribot et al., 2012). Gene annotation was performed with DOGMA web server (Wyman et al., 2004) and the 5′ and 3′ gene ends manually refined by comparing with mitogenome annotations of other crustaceans. Secondary structures of tRNAs were corroborated using Arwen (Laslett & Canbäck, 2008) and tRNAscan-SE (Schattner et al., 2005) search servers in parallel. Finally, annotations were checked with MITOS web server (Bernt et al., 2013), particularly to estimate the secondary structures of rRNA sequences. These were manually refined with Mfold web server (Zuker, 2003) and graphically visualized using VARNA v3.9 (Darty et al., 2009). Palindromes in non-coding spacer regions and the control region were identified using the Mfold. Nucleotide and aminoacid composition, as well as relative synonymous codon usage (RSCU), were calculated using MEGA v5.05 (Tamura et al., 2011). The effective number of codons (ENC) (Wright, 1990) was estimated with INCA V2.1 (Supek & Vlahoviček, 2004). AT and GT skews were estimated as described in Pons et al. (2014), and ggskew options within the EMBOSS v6.6.0 package (Rice et al., 2000).

Mitochondrial genome rearrangements

Gene rearrangements with respect to the putative ancestral Pancrustacean and other known amphipod gene orders were investigated using CREx (Bernt et al., 2007). This software calculates and creates ''strong interval trees'', to heuristically disentangle the possible processes involved (Berard et al., 2007; Bernt et al., 2007; Perseke et al., 2008). A ''common interval'' is a subset of genes that appears successively in two or more input orders. A ''strong interval'' is defined by Berard et al. (2007) as ''A common interval I of a permutation P is a strong interval of P if it commutes with every common interval of P.'' This means both original and derived mitogenomes share a particular subset of consecutive genes or any rearrangement of the original one (strong common intervals), although they must not overlap, and be either disjoint or completely contained within the new arrangement. Their analysis produces a strong interval tree whose nodes can be defined as linear increasing if they appear as in the original gene order, or linear decreasing if they appear in reverse position to the original gene order (Berard et al., 2007). Otherwise, they are defined as prime (shown with rounded shapes, see Supplementary Figures 1–3). These nodes can again be interpreted as produced by different rearrangement events such as those caused by transpositions, reverse transpositions, reversals, or by more complex TDRL events.

Results and discussion

Genome organization, gene order, and rearrangements

The mitochondrial genomes of *P. daviui* (15 155 bp), *P. gorbeanus* (14 190 bp), and *P. sorbasiensis* (15 462 bp) are circular as found in most metazoans and include the canonical 13 PCGs, 22 tRNAs, and two rRNAs (Figure 1). Mitogenome sequences of *P. sorbasiensis* and *P. gorbeanus* are deposited in the EMBL database under accession numbers LN871175 and LN871176, respectively. The sequence coverage was as follows: 73× in *P. daviuvi*, 138× in *P. gorbeanus*, and 270× in *P. sorbasiensis*. The comparison of the mitogenomes of three *Pseudoniphargus* species at both nucleotide and aminoacid levels and taking into account the secondary structure of RNAs allowed a more accurate annotation of gene boundaries and to detect some errors in the published sequence of *P. daviui* (Bauza-Ribot et al., 2009), particularly on the 3′ end of *atp8*, *nad4*, *trnQ*, and *rrnS* genes. These corrections were reported to European Nucleotide Archive (ENA), and they now are included in the FR872383 entry.

The discrepancy in size among the three mitogenomes is largely due to the fact that we were unable to fully recover the control region as well as a small part of *rrnS* (approximately

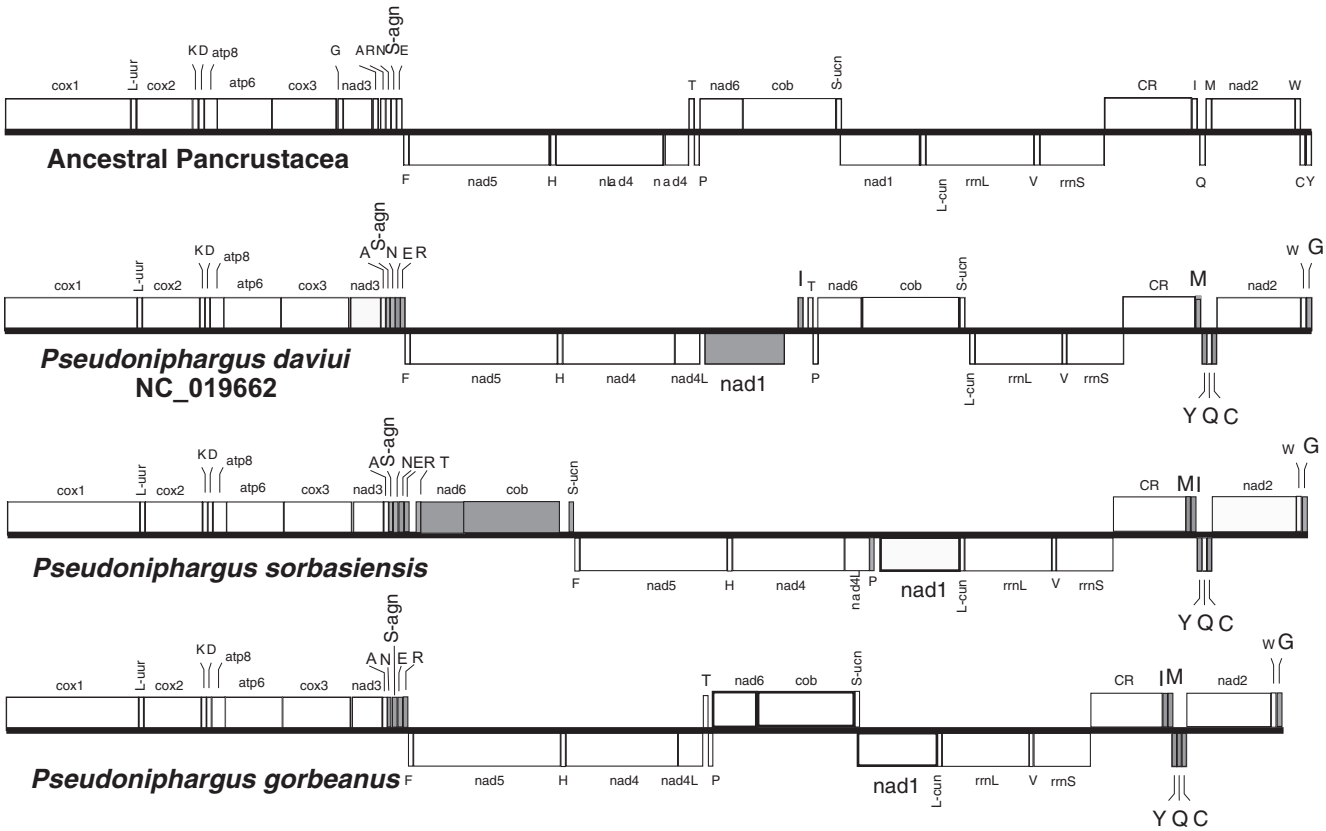


Figure 1. Comparison of mitochondrial gene order in the three *Pseudoniphargus* species studied herein with respect to the presumed Pancrustacean gene order. Genes appearing in a different arrangement than the Pancrustacean order are remarked in grey. Genes placed under the line are coded in the negative strand while genes above are coded in the positive strand.

40 bp) of *P. gorbeanus*, and to the presence of intergenic non-coding regions differing slightly in length. The three mitogenomes show several short non-coding spacers: *P. gorbeanus* (2–5 bp), *P. sorbasiensis* (2–16 bp), and *P. daviui* (2–18 bp). Besides, the last two species display three relatively large non-coding regions of 112 bp, 184 bp, and 450 bp (*P. sorbasiensis*), and 54 bp, 56 bp, and 156 bp (*P. daviui*). These non-coding regions have been compared all along their respective mitogenomes, finding no sequence similarity to other regions.

The three *Pseudoniphargus* mitogenomes show different gene arrangements that in turn differ from both the putative Pancrustacean pattern and the gene orders of amphipods published elsewhere (Bauza-Ribot et al., 2009; Kilpert & Podsiadlowski, 2010; Pons et al., 2014) (Figure 1). It is noteworthy to remark that *Pseudoniphargus* gene orders differ in the position of one or several PCGs while rearrangements involving PCGs are extremely rare at the genus level in other metazoans (Matsumoto et al., 2009; Rawlings et al., 2001). The unexpected length of two PCR amplicons in *P. sorbasiensis* compared with *P. gorbeanus* and *P. daviui* already suggested the occurrence of a putative gene rearrangement (see Supplementary Figure 4). Sequence analyses show that all three genomes have undergone several gene shifts with respect to the putative ancestral Pancrustacean gene order (Boore et al., 1995, 1998; see Figure 1 and Supplementary Figures 1–3). Some of these rearrangements found in *Pseudoniphagus* are shared throughout all currently known amphipod mitogenomes (Pons et al., 2014). For instance, *trnG* gene has undergone a transposition from its ancestral position between *cox3* and *nad3* to occupy an adjacent position to *nad2* and *trnW* genes. The *trnR* gene embedded in the tRNA complex between *nad3* and *nad5* has been transposed to the left of *trnE* in the same complex. Finally, *trnC* gene has relocated from *trnW* to occupy an adjacent position to *nad2* in all amphipods except *Caprella mutica* and *C. scaura* (Ito et al., 2010; Kilpert & Podsiadlowski, 2010). Other rearrangement events appear to be unique to *Pseudoniphargus* and even species specific. In *P. daviui*, the gene coding for *nad1* has relocated downstream of *nad6* and *cob*, possibly through a tandem duplication subsequently affected by a TDRL (Supplementary Figure 1). In addition, in *P. sorbasiensis*, the position of both *nad6* and *cob* has changed through another TDRL event which placed them before *nad5* (Supplementary Figure 2). This rearrangement explains the above-mentioned opposite sizes observed in the agarose gel after electrophoresis of the PCR fragments in *P. sorbasiensis*. *Pseudoniphargus gorbeanus* has apparently undergone one TDRL event shared with the two other *Pseudoniphargus* and other amphipods such as *Bahadzia jaraguensis* (Bauzà-Ribot et al., 2012). *Pseudoniphargus gorbeanus* displays three transpositions and one TDRL event that differ from the presumed ancestral Pancrustaean arrangement (Supplementary Figure 3), although none of them is unique. Moreover, the mitogenome of this species shows no trace of the transposition events involving *trnN* evident in the two other *Pseudoniphargus* species as well as in the rest of amphipods except *Parhyale hawaiiensis*. Finally, *P. daviui* shows four transpositions and two TDRL events differing from the ancestral genome, whereas *P. sorbasiensis* displays five transpositions and two TDRL events.

An additional observation is that the PCGs rearranged by TDRL display large non-coding regions flanking both the novel and prior placements in both *P. daviui and P. sorbasiensis* (see Figure 1). Similar large non-coding areas adjacent to new and old gene placements have been shown to occur in *Culicoides* (Diptera, Ceratopogonidae; Matsumoto et al., 2009), whereas smaller

Table 1. Gene length, AT content, and GC and AT skews for the species *P. daviui*, *P. gorbeanus*, and *P. sorbasiensis*. Asterisk identifies the rRNA of *P. gorbeanus*, where *rrnS* is partial, lacking about 40 bp.

		P. daviui				*P. gorbeanus*				*P. sorbasensis*			
Gene	Strand	Size bp	AT %	AT-skew	GC-skew	Size bp	AT %	AT-skew	GC-skew	Size bp	AT%	AT-skew	GC-skew
Complete	+	15 155	71.2	0.065	−0.292	14 190	68.7	−0.001	−0.314	15 460	69.7	0.016	−0.338
rDNA	−	1717	73.90	−0.077	0.435	1676*	76.9	0.108	−0.357	1705	75.9	−0.065	0.394
tRNA	n.a	1311	70.3	0.036	0.077	1311	71.4	0.022	0.091	1326	70.5	0.041	0.098
PCG	n.a	11 007	66.4	−0.175	0.006	11 001	70.0	−0.149	0.030	11 001	67.5	−0.167	0.012
cob	+	1137	64.8	−0.086	−0.335	1131	69.6	−0.011	−0.283	1137	65.7	−0.102	−0.265
nad2	+	975	72.3	−0.203	−0.307	975	74.0	−0.192	−0.305	975	71.5	−0.180	−0.333
cox1	+	1536	62.4	−0.176	−0.114	1536	64.7	−0.097	−0.020	1536	63.7	−0.149	−0.129
cox2	+	675	62.3	−0.114	−0.284	675	65.9	−0.026	−0.226	675	62.9	−0.107	−0.229
atp8	+	156	66.0	−0.088	−0.322	156	72.4	−0.044	−0.348	156	70.6	−0.091	−0.349
atp6	+	669	67.3	−0.204	−0.323	669	69.6	−0.086	−0.399	669	65.6	−0.143	−0.442
cox3	+	786	63.4	−0.177	−0.153	786	65.5	−0.056	−0.122	786	64.6	−0.130	−0.209
nad3	+	348	67.2	−0.205	−0.104	348	65.4	−0.151	−0.148	348	67.3	−0.180	−0.122
nad5	−	1707	66.7	−0.148	0.441	1707	71.1	−0.221	0.384	1707	68.6	−0.152	0.457
nad4	−	1299	68.1	−0.181	0.420	1299	71.7	−0.250	0.493	1299	70.8	−0.232	0.555
nad4L	−	288	72.9	−0.218	0.535	288	78.4	−0.301	0.546	288	73.9	−0.240	0.571
nad6	+	507	69.8	−0.181	−0.503	507	75.6	−0.130	−0.492	501	70.1	−0.005	−0.453
nad1	−	924	65.9	−0.278	0.431	924	70.2	−0.316	0.470	924	67.6	−0.290	0.463
Control region	+	790	85.5	0.074	0	n.a	n.a	n.a	n.a	572	83.4	0.031175	−0.0964

non-coding spacers (from 12 bp to 60 bp) have been reported in tRNA rearrangements of provannid and vermetid marine gastropods (Hidaka et al., 2013; Rawlings et al., 2001). It is likely that these non-coding spacers represent residual sequence artifacts of prior rearrangement events (Boore et al., 1998).

Sequence composition and codon usage

The AT-content in these mitogenomes is quite high as expected for arthropods: 71.2%, 68.7%, and 69.7% in *P. daviui*, *P. gorbeanus*, and *P. sorbasiensis*, respectively (Table 1). Using metAMiGA (Feijão et al., 2006), we establish that these AT-content values are well within the normal range of crustaceans and in the middle range of peracarids (60.8–76.9%). The highest AT content occurs in the control region (85.5–83.4%) as is common in arthropods (Ki et al., 2010; Liu et al., 2014; Wang et al., 2013; Yang & Yang, 2008), and in a lesser extent in ribosomal and tRNA genes (Table 1 and Supplementary Table 2). PCGs showing the highest AT content were *nad4L* (72.9–78.4%) and *nad2* (71.5–74.0%) in all three species, with minor differences in regard to their coding strand as reported in Pons et al. (2014). Nonetheless, if codon position is taken into account, third positions are considerably AT richer (70.9–77.6%) than first and second sites (63.2–66.4%; see Supplementary Table 2). This composition bias stresses the importance of data partitioning in phylogenetic studies using mitochondrial PCGs (Hassanin, 2006; Pons et al., 2010). There is a significant C over G content, i.e., negative GC-skew, in the PCGs on the positive strand from −0.019 to −0.503, and an opposite trend on the negative strand (0.384–0.555 in *nad1*, *nad5*, *nad4*, and *nad4L*) (Supplementary Table 2). These results are similar to the values reported for most malacostracan mitogenomes (Krebes & Bastrop, 2012; Pons et al., 2014). In addition, GC-skew varies slightly depending on the codon position with first sites being positive (from 0.190 to 0.208), second ones being negative (from −0.113 to −0.095) and third positions being closer to zero (from −0.087 to −0.027) (see Supplementary Table 2). AT-skew is negative for all PCGs and positive in tRNAs, rRNAs, and the control region. AT-skew shows similar trends although it appears to be less prominent (Supplementary Table 2). Accumulative AT-skew plotted for *P. sorbasiensis* indicates that a maximum positive value is reached at around 8000 bp and lowest at around 12 000 bp, clearly indicating the shifts from positive to negative strands in the sequence, where the decline represents the genes on the negative strand (see Supplementary Figure 5). A similar trend is found in *P. daviui*, the maximum positive value starting here around 5500 bp, whereas a maximum negative value is reached around 10 000 bp. In *P. gorbeanus*, the maximum value being reached at around 5200 bp, and experiences a shift in rate from 8800 to 10 800, indicating the falling trend of the curve on the negative-stranded genes. Accumulative GC-skew has been used previously to discover the origin of replication in bacterial circular chromosomes. In our case, it can be identified at the end of the plateau around 14 000 bp in both *P. daviui* and *P. sorbasiensis* (Supplementary Figure 5). Finally, the effective number of codons (ENC) estimated for the mitogenomes of the three species were 48.9, 45.7, and 48.4 for *P. daviui*, *P. gorbeanus*, and *P. sorbasiensis*, respectively. No clear correlation is found between GC content in third positions and the ENC values as reported in other studies (Bauza-Ribot et al., 2009; Kilpert & Podsiadlowski, 2006), perhaps due to the lower A + T bias occurring in *Pseudoniphargus*.

Start and stop codons

Eight out of the 13 protein coding genes (*nad2*, *cox1*, *cox2*, *nad3*, *atp8*, *nad5*, *nad4*, and *nad1*) show unusual start codons (ATY, ATT, TTG, TTG, ATT, TTG, TTT, and GTR, respectively) and in four of them, the stop codons are truncated to either T or TA (*nad2*, *cox1*, *nad5*, and *nad4*). Most of these non-canonical start codons had been already reported in previous studies, including amphipod mitogenomes (Bauza-Ribot et al., 2009; Ito et al., 2010; Ki et al., 2010; Kilpert & Podsiadlowski, 2010). In addition, *P. daviui* and *P. sorbasiensis* display a truncated stop codon in *cox2*. These truncated stop codons are likely to translate to UAA terminal codons that will be completed in later stages by post-transcriptional polyadenylation, as described elsewhere (Ojala et al., 1980, 1981). The large amount of mitogenomes published in recent years corroborates the initial hypothesis that non-canonical start and stop codons are common features in mitochondrial genomes (Boore et al., 2005) which are generally detected due to large overlapping sequence with upstream or downstream genes. Most of these translation exceptions are already codified in the translation tables available at GenBank,

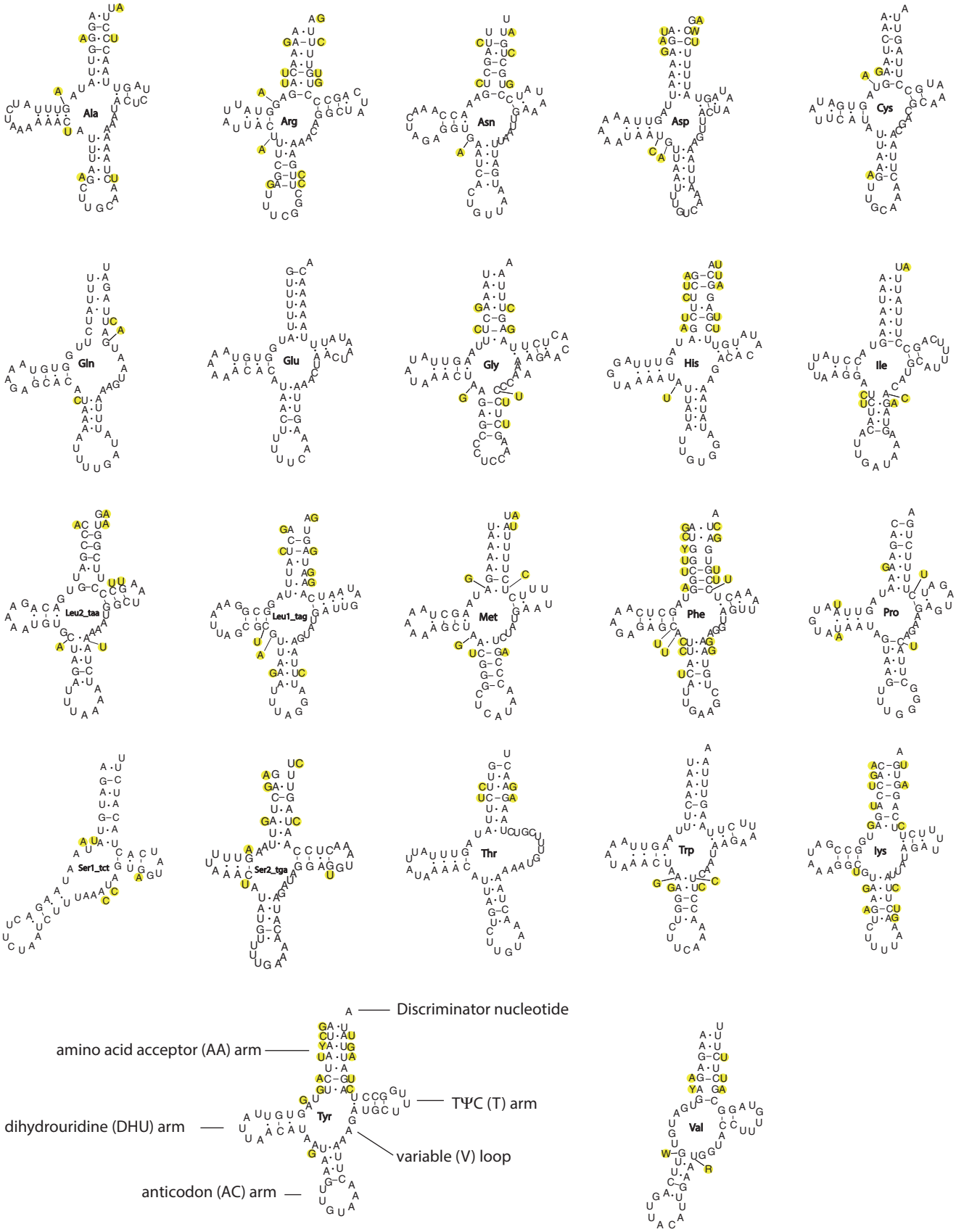


Figure 2. Mitochondrial tRNAs secondary structures in *P. sorbasiensis*. Variable positions are highlighted in grey.

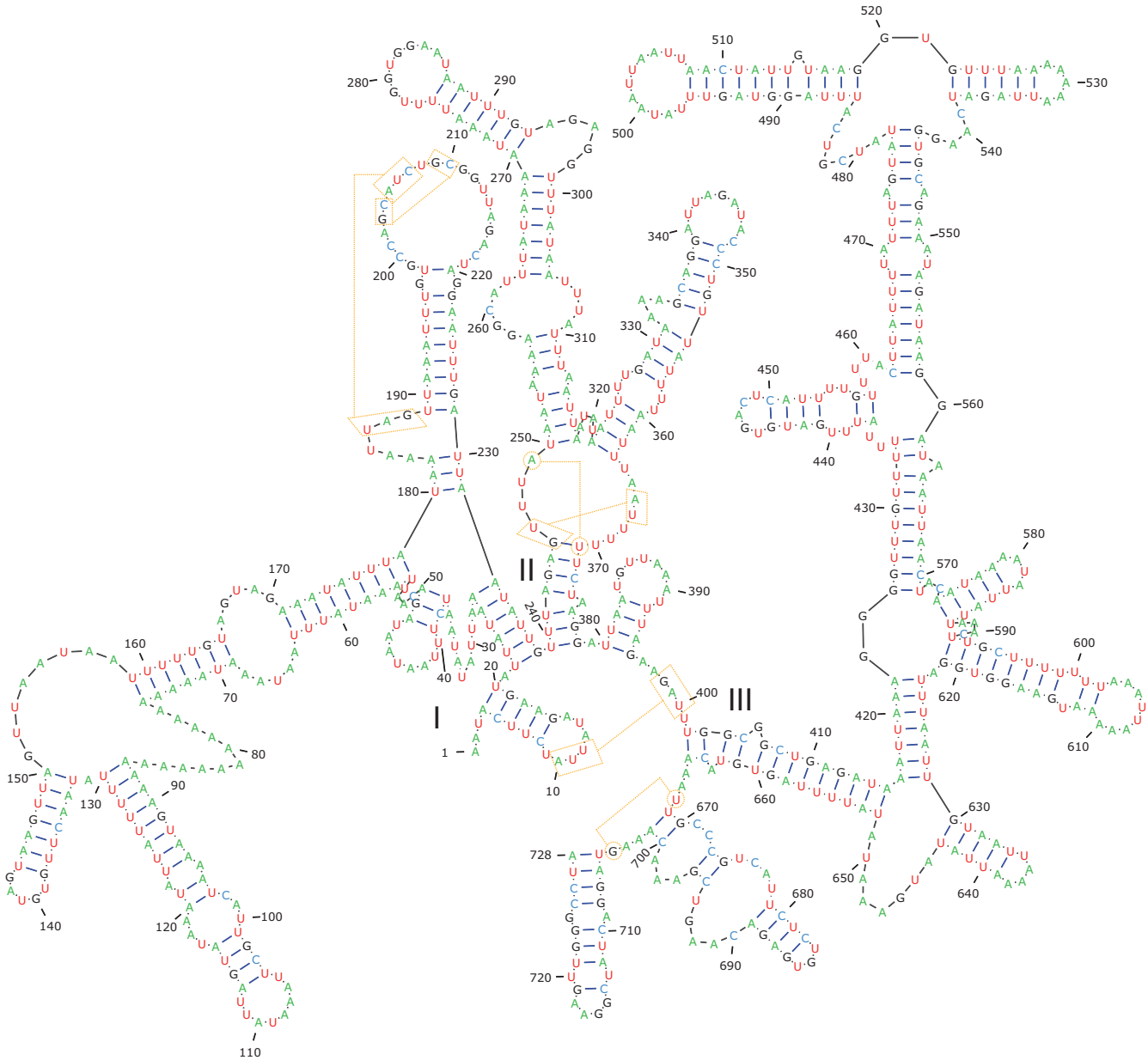


Figure 3. Predicted secondary structure for the small ribosomal mitochondrial RNA (12S) of *P. sorbasiensis*. Different domains are labeled with roman numerals, and sites involved in tertiary folding are remarked with dotted lines.

ENA, and DNA Data Bank of Japan (DDBJ) or can be manually added as translation exception notes.

Transfer RNA genes

We identified all 22 expected tRNA genes in the three mitogenomes, with 14 genes coded on the positive and eight on the negative strand. This strand pattern is the same as in the putative Pancrustacean pattern, although tRNA gene order varies across *Pseudoniphargus* as shown above (Figure 1). Secondary structure is well conserved within all three species, with *P. sorbasiensis* showing the most divergent primary sequences (Figure 2). The tRNAs sequence length varies slightly across the three species as follows: 52–66 bp (*P. sorbasiensis*), 50–64 pb (*P. daviui*), and 51–61 bp (*P. gorbeanus*). The tRNA Threonine lacks the TΨC arm as already reported to be the case in *Metacrangonyx* (Pons et al., 2014), whereas the tRNA genes for Valine and Serine 1 (codon UCN) lack the DUH arm as described for all metazoans (Ki et al., 2010; Kilpert & Podsiadlowski, 2006). The tRNA glutamine lacks the TΨC arm, as reported also in *Caprella mutica* (Kilpert & Podsiadlowski, 2010). The overall structure of the remaining 19 tRNAS is as follows: seven pairs in the aminoacid acceptor (AA) arm, two bases joining AA and dihydrouridine (DUH) arms, 2–4 pairs in the DUH stem, 1–7 nucleotides in the DUH loop, five pairs in the anticodon stem (AC), seven bases in the anticodon loop including the three anticodon nucleotides, 4–5 bases in the variable loop, 2–4 pairs in the pseudo (TΨC) stem, and finally 4–6 nucleotides in the TΨC loop (Figure 2). The presence of two nucleotides between AA and DUH arms, a single nucleotide joining DUH and AC arms, and the absence of nucleotide link between TΨC and AA arms are constant throughout all the tRNA genes. Most of the nucleotide substitutions found in the tRNAs of the three *Pseudoniphargus* mitogenomes are compensatory mutations located on stem regions (AU/GC 35, GU/GC 13, GU/AU 11, and UA/AU three times), although there are also 18 mismatches. These results are

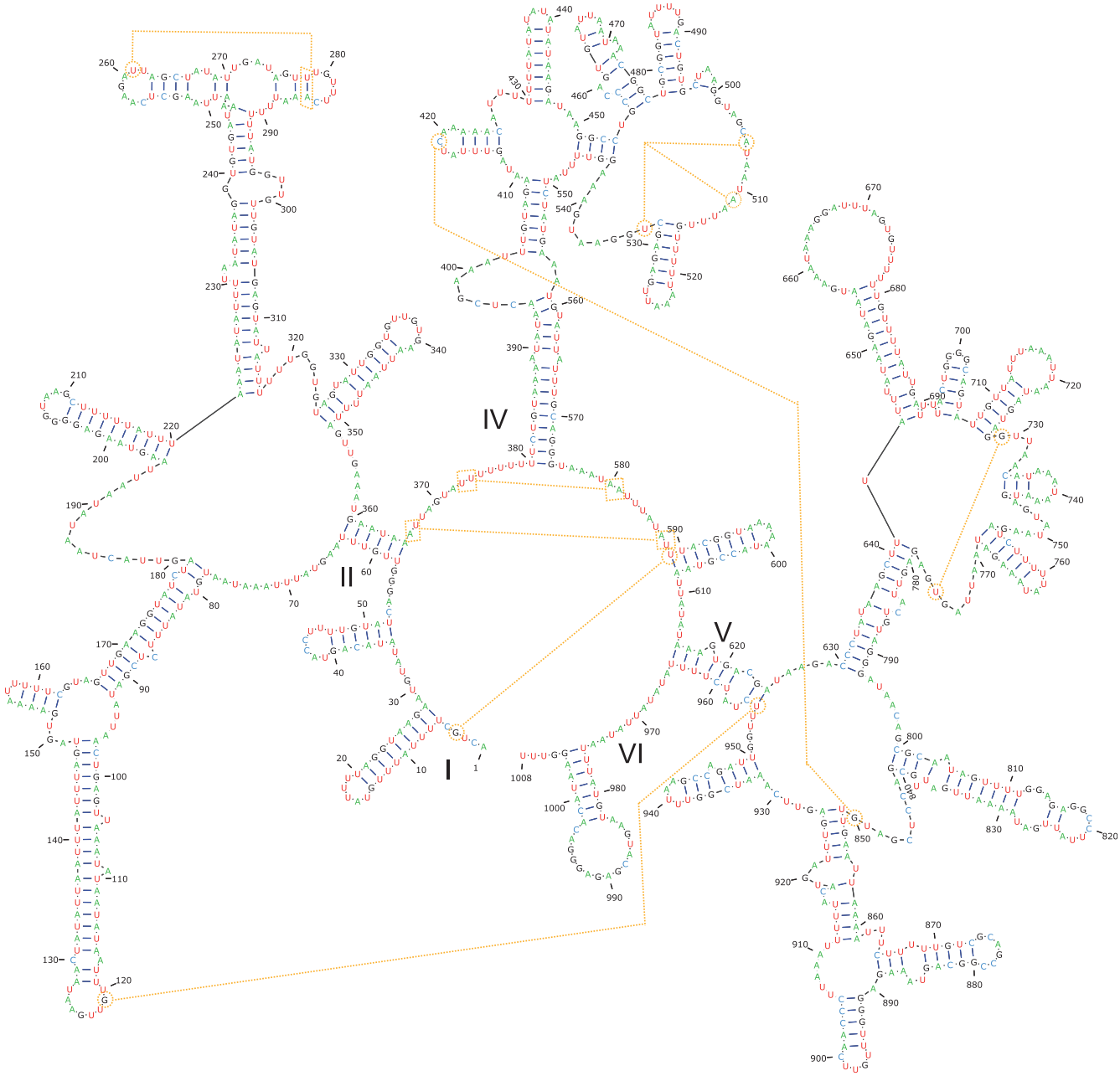


Figure 4. Predicted secondary structure of the large ribosomal mitochondrial RNA (16S) of *P. sorbasiensis*. Details as in Figure 3.

congruent with the pattern found in a previous analysis of 21 *Metacrangonyx* species (Pons et al., 2014), suggesting that compensatory mutations play a key role in the evolution of tRNA sequences.

Ribosomal genes

The *rrnL* and *rrnS* genes show a similar length across the three species, *P. daviui* (*rrnL* 1012 bp; *rrnS* 700 bp), *P. gorbeanus* (*rrnL* 1001 bp; partial *rrnS* 675 bp), and *P. sorbasiensis* (*rrnL* 1008 bp; *rrnS* 698 bp). These are the shortest *rrnL* sequences reported for the Crustacea, although *Metacrangonyx* has a shorter *rrnS* (Pons et al., 2014). The predicted secondary structures are shown in Figures 3 and 4. They were tentatively constructed using MITOS and compared with the already known secondary structures of amphipods and other arthropods (Carapelli et al., 2004; Negrisolo et al., 2011; Pons et al., 2014; Wang et al., 2013). The secondary structure of the small ribosomal unit of *P. sorbasiensis* (Figure 3) was very similar to that described for the amphipod *Metacrangonyx boveii* (Pons et al., 2014), the branchiopod crustacean *Artemia franciscana* (http://www.rna.ccbb.utexas.edu/RNA/Structures/b.16.m.A.franciscana.bpseq) and the neuropterid insect *Libelloides macaronius* (Negrisolo et al., 2011). However, the primary sequence is only conserved in some stretches of domain II and in most of the domain III. In fact, secondary structure of domains I and II had to be reconstructed by finding small anchor regions (i.e., conserved regions on the aligned sequences), and then folding primary sequences in small sections of 30–100 bp using Mfold. Furthermore, the secondary structures shown here, despite being significantly suboptimal in Mfold, were selected because they closely resemble those previously published elsewhere (Carapelli et al., 2004; Negrisolo et al., 2011; Pons et al., 2014; Wang et al., 2013). The first domain of the small subunit of *P. sorbasiensis* (209 bp) is larger than those previously calculated for the crustaceans *M. boveii* (186 bp) and *A. franciscana* (181 bp), but slightly shorter than that of the insect

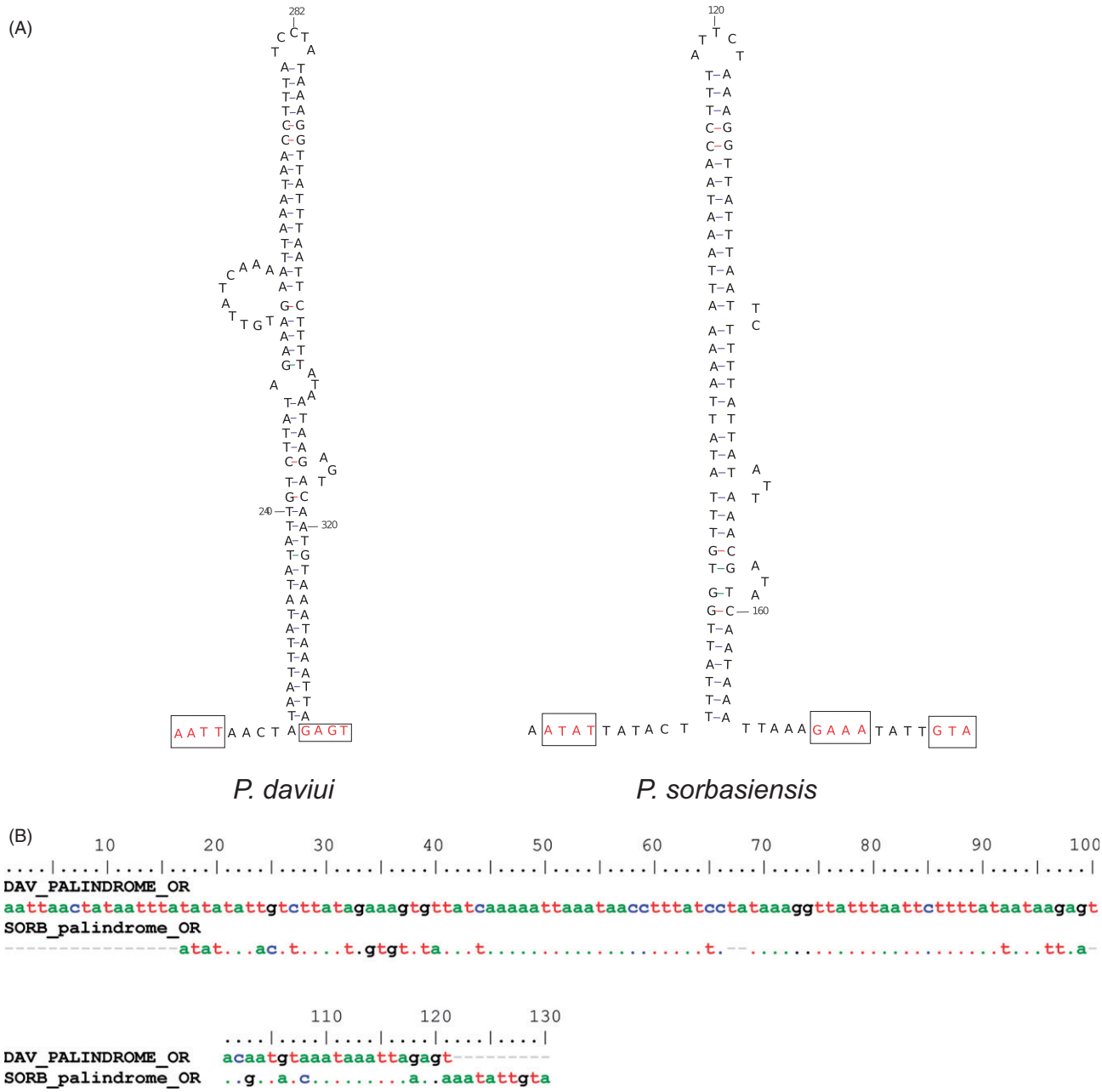


Figure 5. Putative origin of replication for *P. daviui* and *P. sorbasiensis*, with the start and stop motifs indicated in boxes (A). DNA alignment of *P. daviui* and *P. sorbasiensis* palindromes (B).

L. macaronious (217 bp). The large ribosomal unit shows a similar pattern, with the most divergent primary sequences also found in domain I (Figure 4). As in any other arthropods, *P. sorbasiensis* also lacks the domain III, and the last domain (V) is the most conserved (Carapelli et al., 2004; Negrisolo et al., 2011; Pons et al., 2014; Wang et al., 2013). Both secondary structures seem to be folded correctly since all bindings related to tertiary folding are conserved (yellow lines in Figures 3 and 4). Although MITOS is a very useful tool for PCG annotation and building tRNA secondary structures, it appears to be less effective for determining *rrnS* and *rrnL* secondary structures, particularly in the most variable domains. This is probably because it focuses on reconstructing global secondary structures of the ribosomal molecules instead of local motifs that can be better defined throughout conserved stretches from the aligned sequences.

Control region

Both *P. daviui* and *P. sorbasiensis* mitogenomes have a large non-coding region of 790 bp and 572 bp, respectively, situated between *rrnS* and *trnM* genes. These sequences display characteristic features of the control region (CR), such as its high AT-content (85.5% in *P. daviui*; 83.4% in *P. sorbasiensis*) and repetitive motifs (Arunkumar & Nagaraju, 2006). Both control regions were compared using blastn, revealing a 75% sequence identity with 5% gapped positions, suggesting a relative conservation of this fast evolving sequence. In the amphipod genus *Metacrangonyx*, the comparison of 10 control regions found a remarkable conservation only between two very closely related species (Pons et al., 2014). We were unable to detect AT-rich palindromes associated with TATA and GA(N)T motifs in other non-coding areas except at the CR, thus discarding that *Pseudoniphargus* had more than one CR as described in some

amphipods (Ito et al., 2010; Kilpert & Podsiadlowski, 2010). The mitogenomes of *P. daviui* and *P. sorbasiensis* display candidate sequence regions for the origin of replication within the CR (Figure 5). In *P. daviuvi*, we identified the motifs related to the origin of replication (5′ end ATAT and 3′ end GANT motifs) plus a large palindrome within positions 14 501–14 703. We also identified these motifs in *P. sorbasiensis* occurring within positions 8106–8219, but in this species the 3′ end motif differs (GAAA or GTA). Several studies have shown that these motifs appear sometimes modified or truncated (Black & Roehrdanz, 1998; Fahrein et al., 2007). Interestingly, the sequences of both palindromes are relatively conserved (59.23% identity; Figure 5B), suggesting the operation of selective evolutionary constrains in these regions.

Conclusions

The three *Pseudoniphargus* mitochondrial genomes analyzed herein show the occurrence of a remarkable level of gene rearrangement with respect not only to other crustaceans whose mitogenomes are known, but also among them. The latter involve several PCGs, which is extremely rare at low taxonomic levels such as between members of the same genus or family. These rearrangements are flanked by relatively long non-coding sequences that are remnants of presumed breakage points involved in PCGs rearrangements, and likely represent the registry of past TDRL events. In spite of *Pseudoniphargus* displays similar life history traits and habitat to the amphipod genus *Metacrangonyx*, the 21 species studied of the latter genus display an identical mitochondrial gene order, so the unusual environment and geographical isolation of these animals do not necessarily are factors related to variation in gene order and seem to derive from random events (Pons et al., 2014). The high level of mitochondrial gene rearrangements deduced herein in *Pseudoniphargus* would corroborate the family status assigned to this enigmatic group of subterranean amphipods (Karaman, 1993; Lowry & Myers, 2012, 2013). Although gene rearrangements involving PCGs have been shown to occur rarely among closely related species, it is probable that more cases will be reported as next generation sequencing techniques are increasingly applied in comparative mitogenomic studies. Finally, this study shows the importance of reconstructing the secondary structure of ribosomal RNA genes. These structures are crucial to understand the evolutionary pattern of compensatory nucleotide substitutions, their phylogenetic analysis, and to improve the algorithms implemented for the reconstruction of secondary folding, particularly at the most divergent regions.

Acknowledgments

The authors are greatly indebted to Pablo Barranco (University of Almeria, Spain) for the collection of *P. sorbasiensis* specimens used in this study.

Declaration of interest

The authors report no conflicts of interest. The authors alone are responsible for the content and the writing of the paper. This work was supported by Spanish MIMECO grants CGL2009-08256 and CGL2012-33597, partially financed with EU FEDER funds. Morten Stokkan was supported by a MIMECO FPI fellowship.

References

Arndt A, Smith MJ. (1998). Mitochondrial gene rearrangement in the sea cucumber genus *Cucumaria*. Mol Biol Evol 15:1009–16.

Arunkumar KP, Nagaraju J. (2006). Unusually long palindromes are abundant in mitochondrial control regions of insects and nematodes. PLoS One 1:e110.

Bauza-Ribot MM, Jaume D, Juan C, Pons J. (2009). The complete mitochondrial genome of the subterranean crustacean *Metacrangonyx longipes* (Amphipoda): A unique gene order and extremely short control region. Mitochondrial DNA 20:88–99.

Bauzà-Ribot MM, Juan C, Nardi F, Oromí P, Pons J, Jaume D. (2012). Mitogenomic phylogenetic analysis supports continental-scale variance in subterranean thalassoid crustaceans. Curr Biol 22:2069–74.

Bensch S, Harlid A. (2000). Mitochondrial genomic rearrangements in songbirds. Mol Biol Evol 17:107–13.

Berard S, Bergeron A, Chauve C, Paul C. (2007). Perfect sorting by reversals is not always difficult. IEEE/ACM Trans Comput Biol Bioinform 4:4–16.

Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritzsch G, Pütz J, Middendorf M, Stadler PF. (2013). MITOS: Improved de novo metazoan mitochondrial genome annotation. Mol Phylogenet Evol 69:313–19.

Bernt M, Merkle D, Ramsch K, Fritzsch G, Perseke M, Bernhard D, Schlegel M, et al. (2007). CREx: Inferring genomic rearrangements based on common intervals. Bioinformatics 23:2957–8.

Black WC, Roehrdanz RL. (1998). Mitochondrial gene order is not conserved in arthropods: Prostriate and metastriate tick mitochondrial genomes. Mol Biol Evol 15:1772–85.

Boore JL. (1999). Animal mitochondrial genomes. Nucleic Acids Res 27:1767–80.

Boore JL, Collins TM, Stanton D, Daehler LL, Brown WM. (1995). Deducing the pattern of arthropod phylogeny from mitochondrial DNA rearrangements. Nature 376:163–5.

Boore JL, Lavrov D, Brown W. (1998). Gene translocation links insects and crustaceans. Nature 392:667–8.

Boore JL, Macey JR, Medina M. (2005). Sequencing and comparing whole mitochondrial genomes of animals. Meth Enzymol 395:311–48.

Carapelli A, Soto-Adames F, Simon C, Frati F, Nardi F, Dallai R. (2004). Secondary structure, high variability and conserved motifs for domain III of 12S rRNA in the arthropleona (Hexapoda; Collembola). Insect Mol Biol 13:659–70.

Chaudhuri K, Chen K, Mihaescu R, Rao S. (2006). On the tandem duplication-random loss model of genome rearrangement. In: Proceedings of the seventeenth annual ACM-SIAM symposium on discrete algorithms (SODA). New York: ACM. p 564–70.

Chevreux E. (1901). Amphipodes des eaux soutteraines de France et d'Algerie. Bull Soc Zool Fr 26:211–16.

Darty K, Denise A, Ponty Y. (2009). VARNA: Interactive drawing and editing of the RNA secondary structure. Bioinformatics 25:1974–75.

Dowton M, Austin AD. (1999). Evolutionary dynamics of a mitochondrial rearrangement ''hot spot'' in the Hymenoptera. Mol Biol Evol 16:298–309.

Dowton M, Cameron S, Dowavic J, Austin A, Whiting M. (2009). Characterization of 67 mitochondrial tRNA gene rearrangements in the Hymenoptera suggests that mitochondrial tRNA gene position is selectively neutral. Mol Biol Evol 26:1607–17.

Fahrein K, Talarico G, Braband A, Podsiadlowski L. (2007). The complete mitochondrial genome of *Pseudocellus pearsei* (Chelicerata: Ricinulei) and a comparison of mitochondrial gene rearrangements in Arachnida. BMC Genomics 8:386.

Feijão PC, Neiva LS, Azeredo-Espin AMLD, Lessinger AC. (2006). AMiGA: The arthropodan mitochondrial genomes accessible database. Bioinformatics 22:902–3.

Hassanin A. (2006). Phylogeny of Arthropoda inferred from mitochondrial sequences: Strategies for limiting the misleading effects of multiple changes in pattern and rates of substitution. Mol Phylogenet Evol 38:100–16.

Hickerson MJ, Cunningham CW. (2000). Dramatic mitochondrial gene rearrangements in the hermit crab *Pagurus longicarpus* (Crustacea, Anomura). Mol Biol Evol 17:639–44.

Hidaka H, Watanabe H, Kano Y, Kojima S. (2013). Mitochondrial genome rearrangement in a hydrothermal vent-endemic lineage of provannid gastropods provides a new DNA marker for phylogeographical studies. J Marine Biol Assoc UK 93:1053–8.

Irisarri I, Eernisse DJ, Zardoya R. (2014). Molecular phylogeny of Acanthochitonina (Mollusca: Polyplacophora: Chitonida): Three new

mitochondrial genomes, rearranged gene orders and systematics. J Nat Hist 48:2825–53.

Ito A, Aoki MN, Yokobori SI, Wada H. (2010). The complete mitochondrial genome of *Caprella scaura* (Crustacea, Amphipoda, Caprellidea), with emphasis on the unique gene order pattern and duplicated control region. Mitochondrial DNA 21:183–90.

Jaume D. (1991). New species of the amphipod genus *Pseudoniphargus* (Crustacea) from Cabrera (Balearic Islands). Stygologia 6:177–89.

Karaman GS. (1993). Crustacea Amphipoda di acqua dolce. In: Fauna d'Italia. Vol. 31. Bologna: Edizione Calderini. p. 337.

Ki J, Hop H, Kim S, Kim I, Park H, Lee J. (2010). Complete mitochondrial genome sequence of the arctic gammarid, *Onisimus nanseni* (Crustacea; Amphipoda): Novel gene structures and unusual control region features. Comp Biochem Physiol Part D: Genoms Proteom 5:105–15.

Kilpert F, Podsiadlowski L. (2006). The complete mitochondrial genome of the common sea slater, *Ligia oceanica* (Crustacea, Isopoda) bears a novel gene order and unusual control region features. BMC Genom 7:241.

Kilpert F, Podsiadlowski L. (2010). The Australian fresh water isopod (Phreatoicidea: Isopoda) allows insights into the early mitogenomic evolution of isopods. Comp Biochem Physiol Part D: Genom Proteom 5:36–44.

Kraytsberg Y, Schwartz M, Brown TA, Ebralidse K, Kunz WS, Clayton DA, Vissing J, Khrapko K. (2004). Recombination of human mitochondrial DNA. Science 304:981.

Krebes L, Bastrop R. (2012). The mitogenome of *Gammarus duebeni* (Crustacea Amphipoda): A new gene order and non-neutral sequence evolution of tandem repeats in the control region. Comp Biochem Physiol Part D Genom Proteom 7:201–11.

Kurabayashi A, Sumida M, Yonekawa H, Glaw F, Vences M, Hasegawa M. (2008). Phylogeny, recombination, and mechanisms of stepwise mitochondrial genome reorganization in mantellid frogs from Madagascar. Mol Biol Evol 25:874–91.

Kurabayashi A, Yoshikawa N, Sato N, Hayashi Y, Oumi S, Fujii T, Sumida M. (2010). Complete mitochondrial DNA sequence of the endangered frog *Odorrana ishikawae* (family Ranidae) and unexpected diversity of mt gene arrangements in ranids. Mol Phylogenet Evol 56:543–53.

Laslett D, Canbäck B. (2008). ARWEN: A program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. Bioinformatics 24:172–5.

Liu M, Zhang Z, Peng Z. (2014). The mitochondrial genome of the water spider *Argyroneta aquatica* (Araneae: Cybaeidae). Zool Scripta 44:179–90.

Lowry JK, Myers AA. (2012). New, mainly southern hemisphere, freshwater families of Amphipoda (Crustacea), together with a description of the first freshwater calliopiid, *Lutriwita bradburyi* gen. nov. et sp. nov. Zootaxa 3499:27–45.

Lowry JK, Myers AA. (2013). A phylogeny and classification of the Senticaudata subord. nov. (Crustacea: Amphipoda). Zootaxa 310:1–80.

Macey JR, Larson A, Ananjeva NB, Fang Z, Papenfuss TJ. (1997). Two novel gene orders and the role of light-strand replication in rearrangement of the vertebrate mitochondrial genome. Mol Biol Evol 14:91–104.

Matsumoto Y, Yanase T, Tsuda T, Noda H. (2009). Species-specific mitochondrial gene rearrangements in biting midges and vector species identification. Med Vet Entomol 23:47–55.

Messouli M, Messana G, Yacoubi-Khebiza M, Coineau N. (2001). *Pseudoniphargus* (subterranean crustacean amphipod) from Morocco: Systematics, phylogeny and ecological and biogeographic aspects In: 13th international congress of speleology. Brazilia: SBE. p. 391–4.

Moritz C, Brown W. (1987). Tandem duplications in animal mitochondrial DNAs: Variation in incidence and gene content among lizards. Proc Natl Acad Sci USA 84:7183–7.

Negrisolo E, Babbucci M, Patarnello T. (2011). The mitochondrial genome of the ascalaphid owlfly *Libelloides macaronius* and comparative evolutionary mitochondriomics of neuropterid insects. BMC Genomics 12:221.

Notenboom J. (1986). The species of the genus *Pseudoniphargus* Chevreux, 1901 (Amphipoda) from northern Spain. Contrib Zool 56:75–122.

Notenboom J. (1987a). Lusitanian species of the amphipod *Pseudoniphargus* Chevreux, 1901 with a key to all the known Iberian species. Contrib Zool 57:191–206.

Notenboom J. (1987b). Species of the genus *Pseudoniphargus* Chevreux, 1901 (Amphipoda) from the Betic Cordillera of southern Spain. Contrib Zool 57:87–150.

Notenboom J. (1988). *Parapseudoniphargus baetis*, new genus, new species, a stygobiont amphipod crustacean from the Guadaquivir river basin (Southern Spain), with phylogenetic implications. J Crust Biol 8:110–21.

Notenboom J. (1991). Marine regressions and the evolution of groundwater dwelling amphipods (Crustacea). J Biogeogr 18:437–54.

Ojala D, Merkel C, Gelfand R, Attardi G. (1980). The tRNA genes punctuate the reading of genetic information in human mitochondrial DNA. Cell 22:393–403.

Ojala D, Montoya J, Attardi G. (1981). tRNA punctuation model of RNA processing in human mitochondria. Nature 290:470–4.

Perseke M, Fritzsch G, Ramsch K, Bernt M, Merkle D, Middendorf M, Bernhard D, et al. (2008). Evolution of mitochondrial gene orders in echinoderms. Mol Phylogenet Evol 47:855–64.

Pons J, Bauza-Ribot M, Jaume D, Juan C. (2014). Next-generation sequencing, phylogenetic signal and comparative mitogenomic analyses in Metacrangonyctidae (Amphipoda: Crustacea). BMC Genomics 15:566.

Pons J, Ribera I, Bertranpetit J, Balke M. (2010). Nucleotide substitution rates for the full set of mitochondrial protein-coding genes in Coleoptera. Mol Phylogenet Evol 56:796–807.

Rawlings TA, Collins TM, Bieler R. (2001). A major mitochondrial gene rearrangement among closely related species. Mol Biol Evol 18:1604–9.

Rice P, Longden I, Bleasby A. (2000). EMBOSS: The European molecular biology open software suite. Trends Genet 16:276–7.

San Mauro D, Gower DJ, Zardoya R, Wilkinson M. (2006). A hotspot of gene order rearrangement by tandem duplication and random loss in the vertebrate mitochondrial genome. Mol Biol Evol 23:227–34.

Sano N, Kurabayashi A, Fujii T, Yonekawa H, Sumida M. (2005). Complete nucleotide sequence of the mitochondrial genome of Schlegeĺs tree frog *Rhacophorus schlegelii* (family Rhacophoridae): Duplicated control regions and gene rearrangements. Genes Genet Syst 80:213–24.

Schattner P, Brooks AN, Lowe TM. (2005). The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. Nucleic Acids Res 33:W686–9.

Shao R, Dowton M, Murrell A, Barker SC. (2003). Rates of gene rearrangement and nucleotide substitution are correlated in the mitochondrial genomes of insects. Mol Biol Evol 20:1612–19.

Sheffield N, Song H, Cameron L, Whiting M. (2008). A comparative analysis of mitochondrial genomes in Coleoptera (Arthropoda: Insecta) and genome descriptions of six new beetles. Mol Biol Evol 25: 2499–509.

Stock JH. (1980). Regression model evolution as exemplified by the genus *Pseudoniphargus* (Amphipoda). Bijdr Dierkd 50:105–44.

Stock JH. (1988). The amphipod genus *Pseudoniphargus* (Crustacea) in the Canary Islands. Bijdr Dierk 58:47–78.

Stock JH, Abreu AD. (1992). Three new species of *Pseudoniphargus* (Crustacea: Amphipoda) from the Madeira Archipelago. Bol Mus Munic Funchal 44:131–55.

Stock JH, Holsinger JR, Sket B, Iliffe TM. (1986). Two new species of *Pseudoniphargus* (Amphipoda), in Bermudian groundwaters. Zool Scripta 15:237–49.

Supek F, Vlahoviček K. (2004). INCA: Synonymous codon usage analysis and clustering by means of self-organizing map. Bioinformatics 20:2329–30.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. (2011). MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731–9.

Tsaousis AD, Martin DP, Ladoukakis ED, Posada D, Zouros E. (2005). Widespread recombination in published animal mtDNA sequences. Mol Biol Evol 22:925–33.

Wang Y, Huang X, Qiao G. (2013). Comparative analysis of mitochondrial genomes of five aphid species (Hemiptera: Aphididae) and phylogenetic implications. PLoS One 8:e77511.

Wright F. (1990). The 'effective number of codons' used in a gene. Gene 87:23–9.

Wyman SK, Jansen RK, Boore JL. (2004). Automatic annotation of organellar genomes with DOGMA. Bioinformatics 20:3252–5.

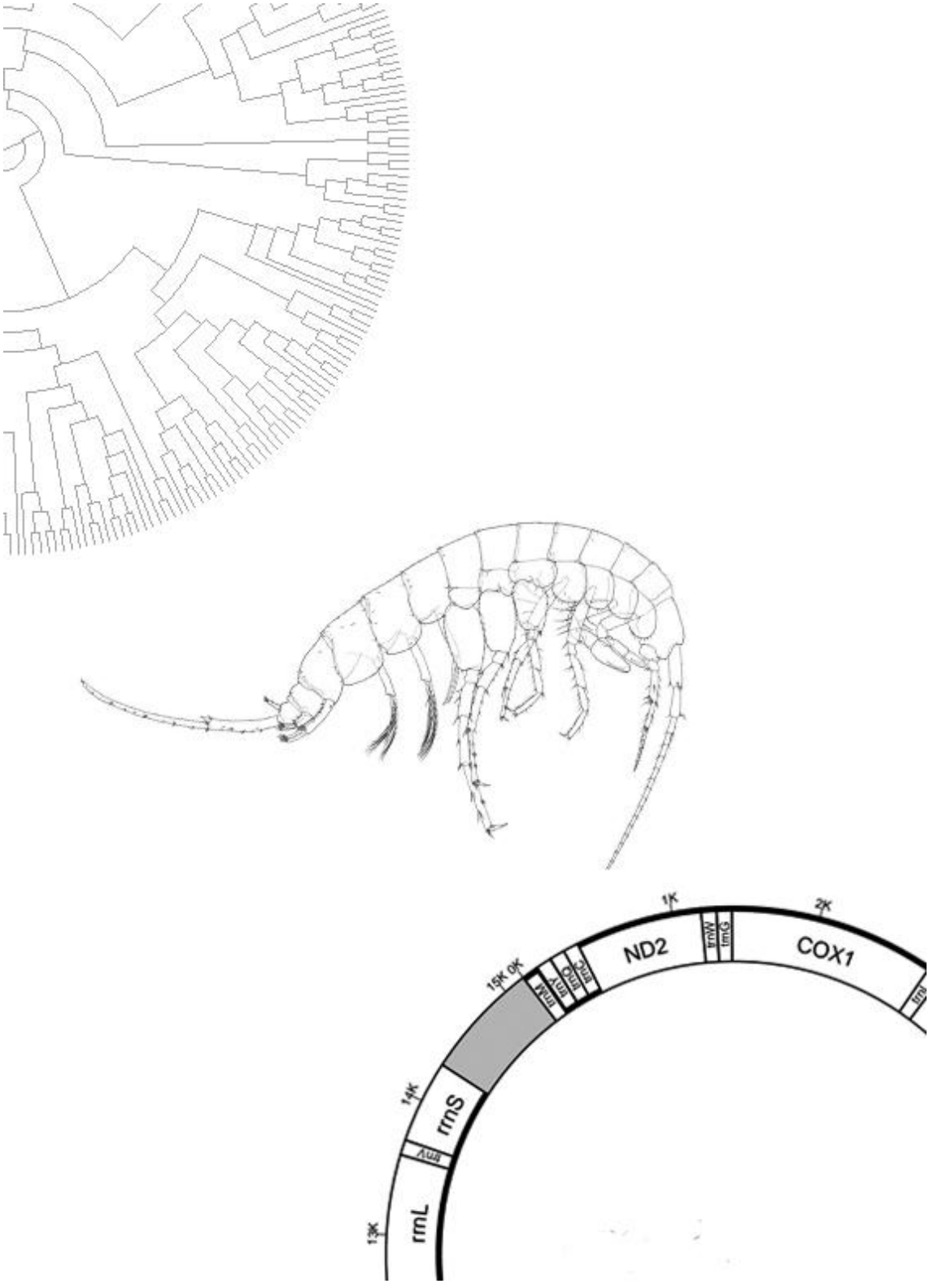
Yang JS, Yang WJ. (2008). The complete mitochondrial genome sequence of the hydrothermal vent galatheid crab *Shinkaia crosnieri* (Crustacea: Decapoda: Anomura): A novel arrangement and incomplete tRNA suite. BMC Genomics 9:257.

Yang Z. (2008). Empirical evaluation of a prior for Bayesian phylogenetic inference. R Soc Phil Trans Biol Sci 363:4031–9.

Zuker M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 31:3406–15.

Supplementary material available online

Supplemental material Table 1 and Figures 1–4.


1K
2K
ND2
COX1
rmS
rmL