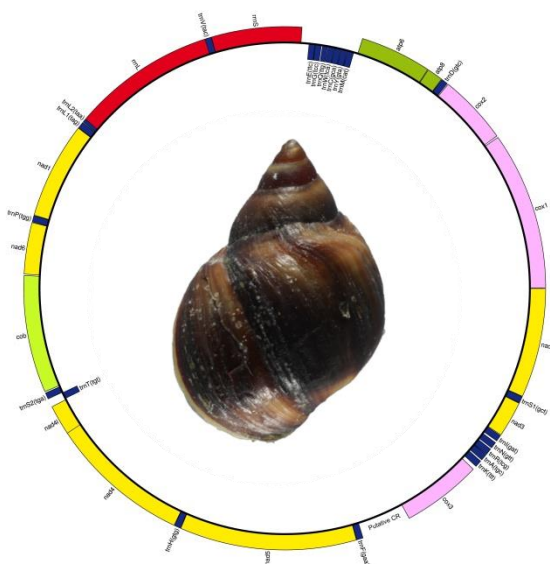# Mitochondrial DNA hyperdiversity and population genetics in the periwinkle *Melarhaphe neritoides* (Mollusca: Gastropoda)

Séverine Fourdrilis

Université Libre de Bruxelles | Faculty of Sciences

Royal Belgian Institute of Natural Sciences | Directorate Taxonomy & Phylogeny

Thesis submitted in fulfilment of the requirements for the degree of Doctor (PhD) in Sciences, Biology

Please refer to this work as:

Fourdrilis S (2017) Mitochondrial DNA hyperdiversity and population genetics in the periwinkle Melarhaphe neritoides *(Linnaeus, 1758) (Mollusca: Gastropoda). PhD thesis, Free University of Brussels.*

PROMOTERS

Prof. Dr. Thierry Backeljau   (90 %, RBINS and University of Antwerp)

Prof. Dr. Patrick Mardulyn   (10 %, Free University of Brussels)

EXAMINATION COMMITTEE

Prof. Dr. Thierry Backeljau   (RBINS and University of Antwerp)

Prof. Dr. Sofie Derycke   (RBINS and Ghent University)

Prof. Dr. Jean-François Flot   (Free University of Brussels)

Prof. Dr. Marc Kochzius   (Vrije Universiteit Brussel)

Prof. Dr. Patrick Mardulyn   (Free University of Brussels)

Prof. Dr. Nausicaa Noret   (Free University of Brussels)

Acknowledgements

Let's be sincere. PhD is like heaven! You savour each morning this taste of paradise, going at work to work on your passion, science. PhD is like heaven, at the BEGINNING. At the end, you simply hate it, you want to give up and burn your articles.

Soyons sincère. Le doctorat c'est le paradis ! On savoure chaque matin ce goût de paradis, à travailler sur notre passion, la science. Le doctorat c'est comme le paradis, au DEBUT ! À la fin, on déteste notre doctorat, on a envie d'abandonner et de brûler nos articles.

But you do not get there easily, and this is the reason why PhD students are grateful to a large number of people! A multitude of unicorns who dropped on your way pieces of rainbows, that you eat, like pieces of sugar feeding your mental energy.

Mais on n'en arrive pas là facilement, et c'est pour cette raison que les doctorants remercient de tout cœur une multitude de personnes ! Une multitude de licornes magiques qui ont déposé sur le chemin des morceaux d'arcs-en-ciel, que tu manges, tels des morceaux de sucre qui boostent ton énergie.

Je remercie très sincèrement toutes les licornes suivantes :

Patrick Mardulyn & **Thierry Backeljau**, mes 2 promoteurs qui se sont engagés avec moi. Merci de m'avoir aidée à chaque fois que j'ai demandé votre aide. Merci Patrick pour les cours de biologie de l'évolution qui m'ont donnés les bases de cette nouvelle discipline dans mon parcours. Une petite chose mais importante : vos emails toujours gentils, c'est super psychologiquement. Tout comme l'enthousiasme incroyable de Thierry. Mais comment faites-vous pour apporter cette énergie positive à tout le monde alors que vous avez 10 fois plus de travail et de stress ? C'est une grande qualité de management incontestablement. Et elle ne sera pas passée inaperçue. The

best supervisor ever ☺! Aah… bien des jours j'ai maudit cette hyperdiversité mitochondriale chez *Melarhaphe*, qui était un problème au début de la thèse et empêchait l'analyse des données ! Et qui d'ailleurs n'avait pas encore été baptisée hyperdiversité ! C'est un chemin scientifique classique, qui part d'une observation inattendue, et aboutit à une exploration en profondeur d'un phénomène. Mais nous, pauvres doctorants trop stressés, on n'ose prendre une voie qui va demander du temps. Car c'est le risque de finir la thèse en retard :p C'est là que votre enthousiasme et votre confiance indéfectible en mon travail m'ont empêchée d'avoir peur de choisir d'explorer cette hyperdiversité. Et maintenant je ne la déteste plus :p

Mes parents les « bénévoles volontaires » ! En fait ils ont été là pour tout dans ma vie c'est simple à résumer. A notre âge, il serait peut-être temps pour Jacques, Stéphane et moi qu'on inverse les rôles et vous rendent la pareille ?!

Ma petite Mamie qui était fière de sa petite fille, enfin c'est ce que j'aime à croire ☺ ;

Mes frères qui n'osent pas dire à leur entourage que leur petite sœur étudie les bigorneaux… C'est tabou en Bretagne. On n'inspecte pas leur anatomie avant de les manger ;

Mes amis Liégeois, qui m'ont gardée en forme pour arriver en un seul morceau jusqu'au début du doctorat. Les **Profondues**, **Doro**, **Bobynou**, **Naud-Naud**, **Canaillotte**, **Colinou**, **Ben-oït**, et tous ceux qui ont pris soin de moi, merci du fond du cœur pour votre soutien guérisseur. Transition à Bruxelles en 2010, futur inconnu et bam je me suis lancée et suis arrivée au bout ! Je ne vous ai pas oublié ces 2 dernières années, non non ;-)

My colleagues, the best cure to prevent your mind of losing its sanity if working at home alone ;

Pandakisses to **Pana**, **CriCri**, **Oihana**, **Nong**, **Maud**, **Silvia**, **Vanya**. You RBINS mates who I shared sweet "international" moments with. That's amazing how the world has no borders in Science :D!

Quentin pour nos discussions sur les F-statistiques et autres épines bioinformatiques, ça a amélioré certaines de mes journées! **Ana** & **Ana** for your

warm welcome in the Azores, **Wiebke** for your sweet company under snow during the intensive workshop in Czech Republic, and the Hong Kong team **Karen**, **Kathy**, **Martin**, **Tommy** and **Terence** for your warm invitation at the Swims research station and your help when I departed in the middle of the night! All of you lightened my steps along the PhD way ;

I am grateful to **Mark Miller**, from the CIPRES Science Gateway team, for replying positively to all my requests for extending my endless MIGRATE-N analyses for days and weeks! Without these computing resources, the Chapter 2 would not be ;

Pandakisses to the members of the Panda Gang who I shared with,

the same PhD universe, funny and funny-haha jokes, the Ostracod song, finding out that lawnmowers are common in Yourope but hovercraft lawnmowas are endemic to Australia, the same sheeps in Lier, the Beer list, do you want some teaaa honeyyy?, seeing the same professional shrimp, Thierry's angels in London, travelling together to Istanbul, blue-biking to Damme after work for a good cause…

Arantza the WinnyPinnyPanda (WPP), **Dinarzarde** the House-Party Panda (HPP), **Marie** the Unipolar Panda (UPii), **Mik** the Bad Idea Panda (BIP-BIP), **Rylan** the Frozen Panda (FP), **Séverine** the White Panda (WiiPii), **Valentina** the Fast Bipolar Panda (FBP)

The thing is: create your parallel crazy universe to be the warrant of your sanity during the PhD. Highly recommended to all PhD students.

I have a special thought for 1 Panda in particular, UPii of course!

Finally someone to eat :D Someone avec qui j'ai pratiqué mon Franglais :D Et discuté en Néerlandais (à partir de 2 phrases c'est une conversation). On s'est senties moins seules dans notre monde bizarre de doctorante :p avec notre angoisse partagée des deadlines, qui apparaissaient et se postponaient inlassablement :p On a bien géré : alternance travail/détente au top. On s'est dégôté un bureau calme et formolé ; on s'est défoulées sur les tables du Corbeau ; on a relargué notre stress à jouer les spéléos dans le noir des grottes Belges ; on s'est rafraîchi le cerveau sous (bien bien dessous) un kayak ; on y a d'ailleurs appris à se désaddicter des smartphones ; Donc UPii, quand on a fini toutes les deux, je t'attends pour le grand feu de joie ! Amène tes allumettes et nos thèses, j'prépare le bûcher :D !

Et ma dernière licorne que je remercie :

To the one who rescued me when I freaked out!

Paper feed fail: call **Super-3D Muisje** ♥

Poor you… bearing a working animal thuis the last 2 years… Ben je beschikbaar het volgende 80 jaren? :s

A Mamie
de la part d'un petit scorpion plus têtu que toi !

A Marine
de la part de ta marraine qui admire ton intelligence et ta créativité

Contents

General Introduction

Nota bene: words in SMALL CAPS are listed in the glossary.

Biodiversity and oceans

Biodiversity, or the contraction of "biological diversity," is the assortment of life components on Earth at all its organizational levels, from genes to ecosystems, and the ecological and evolutionary processes that sustain it. Biodiversity benefits us just by being there, because it is a natural capital that contributes to human well-being, through ECOSYSTEM SERVICES. In 2011, the world's ecosystem services are estimated to circa US$125 trillion per year, against only US$75 trillion per year produced by the entire global economy (Costanza *et al.* 2014).

The usefulness of fundamental research, in the field of biodiversity or any field, lies in the production of useless knowledge to all appearances, because they may or may not lead to new technologies, and because most breakthroughs in research are not and could not be planned but are unexpected (Schwarz 2017).

In this PhD thesis, I am interested in a particular aspect of biodiversity: genetic diversity, which is the variety of DNA makeup in individuals. Genetic diversity varies among individuals, POPULATIONS and species, over a short-term, ecological, time-scale, and over a longer-term, evolutionary, time-scale. Although ecological and evolutionary time-scales overlap, evolution on ecological time-scale is considered as evolutionary changes occurring over tens of generations or fewer, whereas evolution on evolutionary time-scale encompasses millions of years up to hundreds of millions of years (Carroll *et al.* 2007). Associated to physical movement of individuals, genetic diversity varies both over time and space.

Biodiversity is organised as a network of food chains of species that consume, produce and recycle organic matter (Cardinale *et al.* 2012). The richer in species, the greater stability in biomass production and resiliency of ecosystems (Cardinale *et al.* 2012). Likewise, genetic diversity acts as a source of STANDING VARIATION in DNA across a network of individuals and species. The richer in genetic variants this genetic diversity, the higher resistance and adaptive potential of populations in a changing environment, and the lower risk of species extinction (Barrett & Schluter 2008). However, too high genetic diversity might increase the probability of outbreeding depression, because too distant genetic variants produce progeny with intermediate genotypes less adapted to habitat and reduced fitness (Templeton *et al.* 1986).

Marine ecosystems occupy 70.9 % of the surface of the Earth (Eakins & Sharman 2010). They count for less than 2 % of the world's carbon biomass (Groombridge & Jenkins 2002) but they are the cradle of life and encompass 28 animal phyla (13 endemic) (United Nations 2017). In contrast, terrestrial environments contain 11 animal phyla, of which only one is endemic (United Nations 2017). Marine diversity is composed of ~226000 known eukaryotic species, from which nearly 200000 are animal species, and putatively 0.7-1.0 million of estimated species in total when accounting for species that may exist (Appeltans *et al.* 2012). Marine molluscs are the most species diverse phylum in the sea, with 135887 to 164107 estimated species (Appeltans *et al.* 2012). Moreover, marine ecosystems have been evolving for an additional 2.7 billion years compared to terrestrial ecosystems, which resulted in wider phylogenetic diversity of marine organisms than of their terrestrial counterparts. Yet, marine biodiversity has received less attention than its terrestrial counterpart (Danovaro *et al.* 2015).

Littorinidae represent a family of marine gastropod molluscs (Bouchet & Rocroi 2005) commonly called periwinkles and distributed worldwide on seashores. The evolution of Littorinidae has been extensively studied, from a phylogenetic and phylogeographic (e.g. De Wolf *et al.* 2000; McQuaid 1996b; Panova *et al.* 2011; Reid *et al.* 2012; Rolán-Alvarez 2007), to ecological (e.g. De Wolf *et al.* 2004; Diz *et al.* 2017; Johannesson 2003, 2016; Krug 2011; McQuaid 1996a) and population genomic points of view (e.g. Marques *et al.* 2017; Panova *et al.* 2014; Ravinet *et al.* 2016). Nevertheless, some species have been given more attention than others, notably *Littorina saxatilis*. The marine periwinkle *Melarhaphe neritoides* (Linnaeus, 1758) shows intriguing genetic features, but its genetic diversity and evolution have been scarcely studied (Cuña *et al.* 2011; García *et al.* 2013; Johannesson 1992; Libertini *et al.* 2004; Williams *et al.* 2003). *Melarhaphe neritoides* is widely distributed throughout Europe (OBIS data at http://iobis.org/mapper/?taxon=Melarhaphe%20neritoides), where it shows a remarkable homogeneous macrogeographic population genetic structure (inferred from allozyme data) (Johannesson 1992), though locally in Spain it displays huge amounts of mitochondrial (mtDNA) diversity at the COI locus in terms of a large numbers of polymorphic sites ($S = 16$ %), a very high haplotype diversity ($Hd = 0.998$) and 98 % of unique haplotypes, and a very high nucleotide diversity ($\pi = 0.019$) (García *et al.* 2013), suggesting signatures of mitochondrial hyperdiversity. This is usual in fast-evolving prokaryotes and viruses, but unusual in eukaryotes. A study in the periwinkle *Tectarius striatus* reported a high mtDNA variability at the haplotype level ($Hd = 0.934$) but low at the nucleotide level ($\pi = 0.006$), and a complete lack of shared haplotypes and a high genetic differentiation ($F_{ST} = 0.089\text{-}0.139$) between the Cape Verde and the other Macaronesian archipelagos, despite no phylogeographic structure separating the haplotypes of both areas in two reciprocally monophyletic groups (Van den Broeck *et al.* 2008). *Melarhaphe neritoides* is sympatric of *Tectarius striatus* in Macaronesia and shares a similar pattern of very few

shared mtDNA haplotypes, but conversely, shows a very low level of genetic differentiation over thousands of kilometres from Sweden to Greece (Johannesson 1992). These observations gave rise to several questions that will be addressed in the present PhD thesis (see the next section "Objectives and Outline"). The research work starts with investigating why *M. neritoides* does not show genetic structuring like *Tectarius striatus*, and the possibility that it is an artefact of the low variability of allozyme markers. It then explores the particularly high variability in *M. neritoides* mtDNA, what could be the causes generating it during the evolution of *M. neritoides*, and whether such high genetic variability impacts on population genetic inferences.

Laws of evolution

Genetic diversity varies because of four main evolutionary forces acting on DNA makeup of individuals: GENETIC DRIFT, GENE FLOW, MUTATION and NATURAL SELECTION. Mutation generates variation, while genetic drift, gene flow and natural selection change the frequencies of the variants. Variation may subsequently affect all biological levels of organisation, molecules (genes, genomes), populations, species, communities and ecosystems.

The four evolutionary forces act in concert. New alleles arise by mutation and are subsequently under the action of the three other forces. New alleles may be immediately detrimental, such as a misfolded protein involved in vital function, and cause the death of individuals. New alleles may also be neutral, neither selected for nor against, and usually remain in the population with their frequency influenced by genetic drift, and spread thanks to the action of gene flow. Lastly, new alleles may be positively selected because they favour population fitness, and usually remain in the population and increase their frequency over time under the action of genetic drift and gene flow.

Species has been made the biological unit of classification for the purpose of describing biodiversity (Pavlinov 2013; Wilkins 2010). Species are the result of evolution, but the notion of species is much debated and a consensus on its current 26 concepts has still not been reached (Mayden 1997; Wilkins 2011). The complexity resides in the multi-level nature of speciation, a process that unfolds through time and space (Abbott *et al.* 2013). Hence, pinning a name on a biological unit that is the product of an ongoing process is intricate (De Queiroz 2007). The most influential species concept is the biological species concept (BSC) that requires the establishment of reproductive isolation between gene pools (Mayr 1942), essentially meaning the absence of gene exchange between gene pools, followed by the gradual genetic divergence of the pools to form a new taxon i.e. species (Kartavtsev 2011). The BSC implies that allopatric populations have to be evaluated whether they can interbreed when they come into contact, what is constraining and hardly feasible when working with DNA sequences instead of living specimens. The BSC has been revisited later by Coyne & Orr (2004) who added that reproductive isolation between distinct species is substantial but not necessarily complete, because barriers to gene flow are semipermeable and many species from different genera can certainly interbreed but still differentiate. This new definition allows limited gene flow and, consequently, does not enable to recognise cryptic species as distinct species within a gene pool with ongoing speciation. The Phylogenetic Species Concept (PSC) is more appropriate to our study and will be used in this thesis, in which we want to verify the presence/absence of cryptic species in *M. neritoides*. The PSC defines a species as the smallest diagnosable cluster of individual organisms within which there is a parental pattern of ancestry and descent (Cracraft 1983, 1989). The PSC is no more strictly based on the criterion of reproductive isolation to diagnose a taxonomic unit as a species, and as such, even in presence of gene flow and/or ongoing interbreeding, recently diverged lineages can be diagnosed.

Each new genetic data, each new GENE or GENOME sequenced, studied and compared, permit to gain insight on how genetic diversity relates to evolutionary forces and how these forces drive evolution of species. The more we accumulate genetic data, the better we will be able to uncover the full biodiversity in the marine ecosystems, to elucidate the evolutionary relationships among species, and to understand the processes that are the base of marine biodiversity (Ribeiro *et al.* 2017).

Mitochondrial DNA

The mitochondrial genome (mitogenome) is small, typically less than 20 kb, which represents << 0.01 % of a metazoan's total genetic composition. It contains 37 genes in vertebrates and bilaterian invertebrates including molluscs (typically < 0.1 % of the total genes) (Burton & Barreto 2012), or more variable content ranging from 1 to 44 genes in non-bilaterian invertebrates (Lavrov 2014).

Cells contain hundreds of thousands of copies of the mitogenome (Alberts *et al.* 2014), making it easy to extract from tissues.

Mitogenome evolution, which is driven by nonadaptive forces such as genetic drift and mutation pressure (Lynch *et al.* 2006), as well as selection (Castellana *et al.* 2011), and the peculiar properties of mitochondria mentioned hereinafter, can benefit evolutionary studies in many aspects:

Specimen identification and species delimitation

In animals, with only 648 base pairs, a region of the cytochrome *c* oxidase subunit one is able to distinguish between two closely related species and is the ultimate barcode identifier of species, except in Cnidaria, Ctenophora and Porifera for which COI is not a suitable barcode, (Hebert *et al.* 2003a; Hebert *et*

al. 2003b). This method called DNA barcoding is made possible because of the premise that COI sequence divergence is higher among species than within species (Hebert *et al.* 2003b), and can therefore recognise deeply diverged, monophyletic lineages, within a single genetic marker. Initially developed to speed up specimen identification to the level of species by matching COI sequences to previously classified Linnaean species names, DNA barcoding is also used for species discovery i.e. to delineate species that are unknown to science (Hebert *et al.* 2003a). The confusion among DNA barcoding for specimen identification and species discovery generates diffidence among taxonomists with respect to DNA barcoding. Yet, the method remains a powerful tool when used for sequence-based specimen identification in terms of established classification (Collins & Cruickshank 2013). This is why Collins & Cruickshank (2013) suggest the term "specimen identification" in place of "species identification". On the other hand, the use of DNA barcoding for species delimitation is criticised, because based on a single molecular marker as opposed to multiple independent markers (e.g. molecular data, morphological data, ecological data) as in the integrative taxonomy strategy (DeSalle 2006, 2007; Rubinoff 2006a, b; Uiblein 2016; Yeates *et al.* 2011), and because unable to overcome situations where insufficient COI divergence between related species leads to no barcode gap (Wiemers & Fiedler 2007).

Genetic differentiation

Due to its uniparental inheritance and haploidy in the majority of animals (Giles *et al.* 1980; Sato & Sato 2013), mtDNA has an EFFECTIVE POPULATION SIZE four times smaller than nuclear DNA (nDNA) and hence is expected to diverge four times faster than nDNA. The smaller effective population size allows detection of genetic differentiation at a lower – intraspecific – taxonomic level, and at a higher sensitivity in *M. neritoides* presumed to be broadly homogenized.

mtDNA is also a widely used marker to detect recent divergence and infer phylogeographic pattern (Hung *et al.* 2016). Due to its uniparental inheritance, mtDNA only tracks one aspect of the evolutionary history of species, i.e. the history of the organelle inherited from females, which can be misleading in presence of heavy sex-bias in species (Ho & Shapiro 2011). Some exceptions are known, e.g. the doubly uniparental inheritance of mtDNA in some bivalves (Vargas *et al.* 2015), or the haploid males of the two mite species *Tetranychus urticae* and *T. turkestani* which transmit only one copy of the nuclear genome and therefore show a smaller effective population size for the nuclear gene ITS2 than for the mitochondrial gene COI (Navajas & Boursot 2003).

Selection

Selection in mtDNA has been much of debate. Through a comprehensive collection of ~3000 animal species, Bazin *et al.* (2006) showed that positive selection, rather than effective population size or ecological determinants, influences and reduces mtDNA genetic diversity in animals. Berry (2006) and Wares *et al.* (2006) made clear that these selective sweeps occur at deep phylogenetic levels such as phylum or class, over deep evolutionary time of hundreds of millions of years, and that mtDNA evolution is more neutral over recent timescale at which conservation biologists work. Bazin *et al.* (2006) answered that recent populations are also affected by selection, because the observed polymorphism reflects past selective sweeps. Besides, Mulligan *et al.* (2006) questioned the absence of correlation found by Bazin *et al.* (2006) between mtDNA diversity and effective population size, since they found a correlation between mtDNA and allozyme variation in small population of eutherian mammals. There is now abundant evidence that mtDNA is under selection rather than evolving neutrally (Castellana *et al.* 2011; Hershberg & Petrov 2008). Selection, positive or negative, acts on metazoan mitogenomes

and drives adaptation to external environment or maintenance of mitonuclear coadaptation (Hill 2016).

Mitonuclear coevolution

Contemporary mitogenomes arose from the endosymbiosis of a prokaryotic genome 1.45 billion years ago, which progressively lost genes during evolution, genes that were transferred to the nuclear genome (Saccone *et al.* 2006). As a consequence, all functions of mtDNA require interaction with nuclear gene products, i.e. with nuclear-encoded proteins, for miscellaneous functions of mtDNA transcription, mtDNA replication, mtDNA translation and oxidative phosphorylation (cell respiration and energy metabolism) (Fig. 1). Coadaptation of the products encoded in the mitochondrial and nuclear genomes is fundamental to the fitness of species (Bar-Yaacov *et al.* 2012).
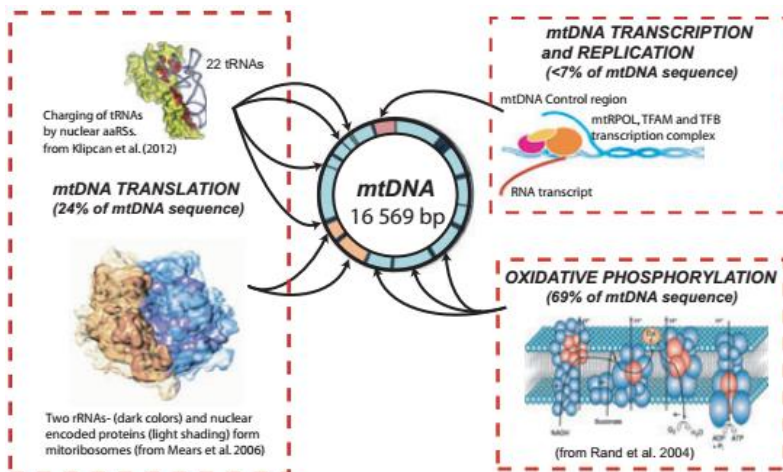
Figure 1. Functions encoded in the mtDNA. Example of human mtDNA. Image from Burton & Barreto (2012).

Mitonuclear coevolution is a non-neutral process and is thought to lead to divergence and speciation after isolation of populations during adequate periods of time (Gershoni *et al.* 2009; Hill 2016).

Cell survival depends on mitochondrial respiration

The mitochondrion is the major source of cellular energy and its function is critical to the life of eukaryotic organisms, with the exception of the microbial eukaryote *Monocercomonoides* sp. that has no mitochondria (Karnkowska *et al.* 2016).

Mitochondria have their own circular DNA molecule, which carries genes coding for proteins involved in cell respiration and energy metabolism (Fig. 2).
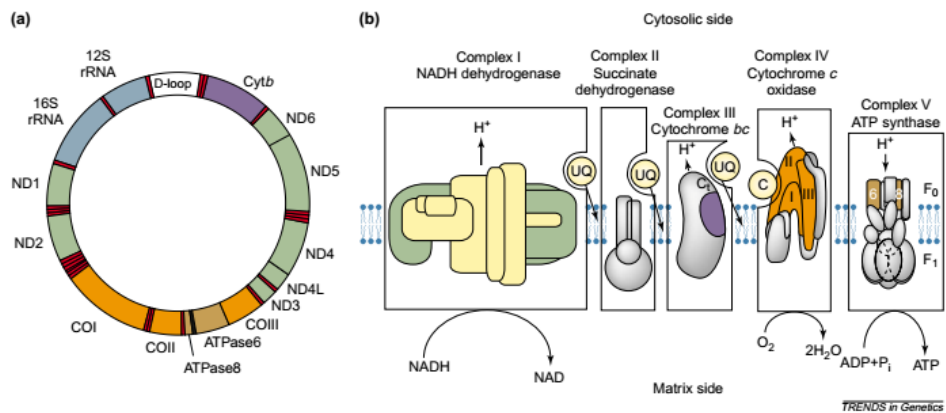
Figure 2. (a) The mitogenome is a circular molecule. (b) The mitochondrial respiratory chain and function of mtDNA-encoded proteins in mitochondrial metabolism (nDNA-encoded proteins are represented in yellow and grey). Image from Blier *et al.* (2001).

During cellular respiration in the mitochondrion, glucose breakdown into pyruvate followed by conversion of pyruvate into acetyl CoA, generate energy in the form of ATP (Fig. 2). Water and carbon dioxide are also produced as by-products. When respiration is aerobic, electrons removed from pyruvate are accepted by oxygen. When oxygen is not available for aerobic respiration, anaerobic respiration (i.e. fermentation) occurs and uses another molecule as terminal acceptor of electrons, such as in oxygen-depleted animal muscle cells, yeasts and some marine invertebrates including mussels and crustaceans, but is less efficient and produces less ATP molecules (Martin & Mentel 2010).

Mutation

Rates of mtDNA evolution are 10–30 times faster than nDNA in bilaterians and drive mitonuclear coevolution and speciation through strong selection pressure (Blier *et al.* 2001; Hill 2016; Lane 2009). A fast mutation rate can quickly produce variants that are suited to changing environment (Lane 2009). Mutations occurring on mtDNA can impact on physiology and fitness of species (Gershoni *et al.* 2014).

Time-scale

At the species level, mtDNA is used to work within a time window from million years ago to thousand years ago.
As an example, in human, mtDNA traced the evolution in recent times and dated the divergence between human and Neandertal around 130,000 years ago, as well as in ancient times and dated the divergence between human and Old World monkeys around 25 million years ago (Soares *et al.* 2009).

Melarhaphe neritoides

Melarhaphe neritoides (Linnaeus, 1758), also called the small periwinkle, is our study system. It is classified as follows:

Mollusca (Phylum)
    Gastropoda (Class)
        Caenogastropoda (Subclass)
            Littorinimorpha (Infraorder)
                Littorinoidea (Superfamily)
                    Littorinidae (Family)
                        Littorininae (Subfamily?)
                          Melarhaphe (Genus)

Other species of the genus *Melarhaphe* are all extinct today.

Within Littorinidae, the phylogenetic relationships of *M. neritoides* with the three subfamilies, Lacuninae, Littorininae and Laevilittorininae, are ambiguous and unresolved. The affiliation of *M. neritoides* to the Littorininae subfamily is supported by the 28S single-gene and the four-gene (28S, 18S, COI, 12S) phylogenies (Williams *et al.* 2003), as well as by morphological evidence of *M. neritoides* showing two important and unique SYNAPOMORPHIES of Littorininae (Reid 1989). The affiliation of *M. neritoides* to the Littorininae is no longer supported in the last up-to-date phylogeny without 18S and with the addition of 110 species (Fig. 3) (Reid *et al.* 2012).
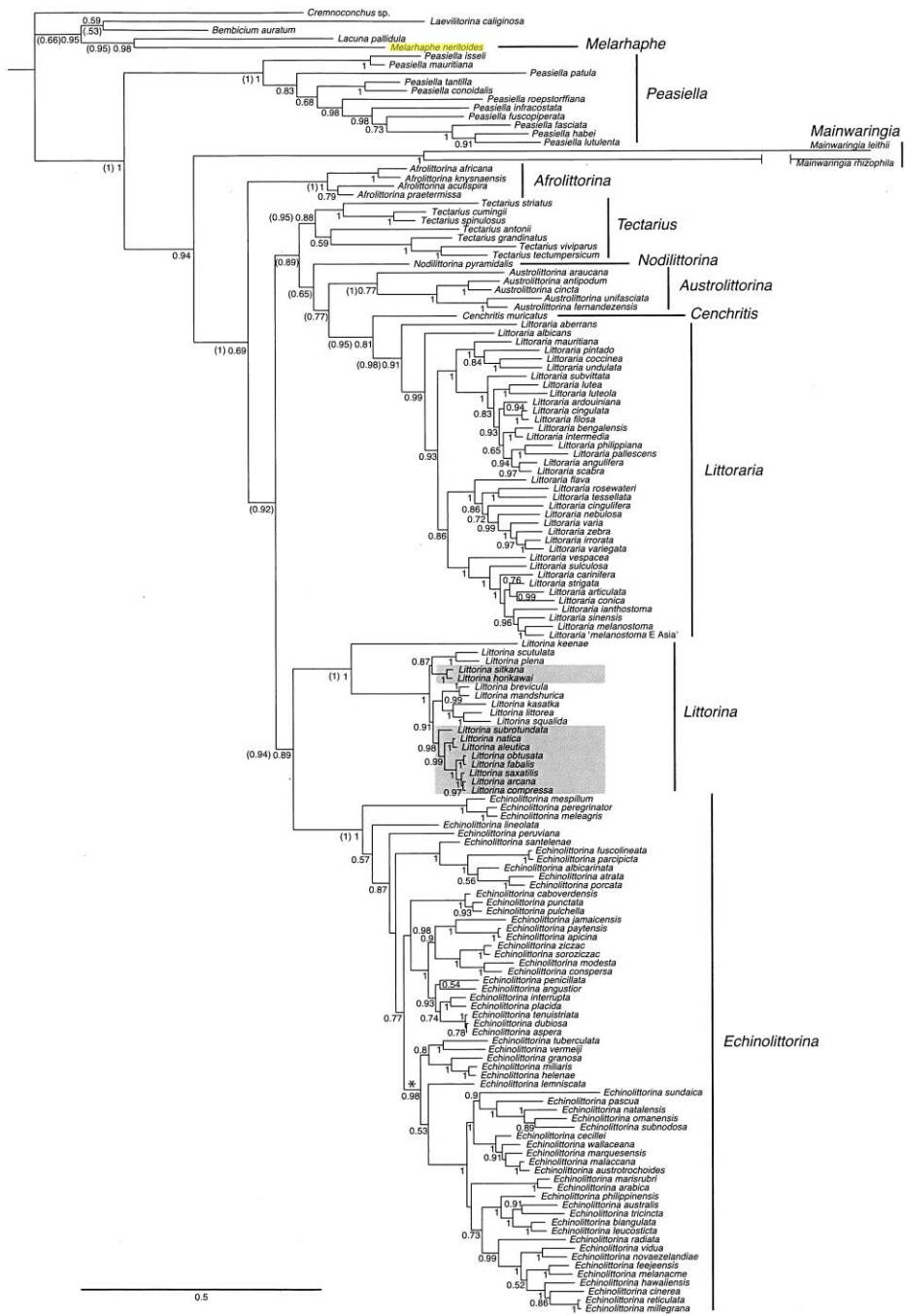
Figure 3. Molecular phylogeny of 147 species of Littorininae generated by MrBayes from concatenated gene sequences from 28S rRNA, 12S rRNA and COI. The full length of the stem leading to *Mainwaringia rhizophila* is not shown; this is 5.2 times the length of the stem of *M. leithii*. The yellow box highlights *M. neritoides*. Grey boxes enclose members of the subgenus *Neritrema*, defined by non-planktotrophic development. Support values are posterior probabilities (PP); values <0.5 are not shown. Support values in parentheses are for generic and deeper clades, from analysis of the same three genes, but excluding the two *Mainwaringia* species. Asterisk indicates clade of the subgenera *Echinolittorina* and *Granulilittorina*. Source: modified from Reid *et al.* (2012).

Melarhaphe neritoides is a small (shell up to 11 mm) marine temperate species (Fig. 4) (Lysaght 1941). The swimming-crawling stage is 0.35-0.42 mm diameter and 0.35-0.38 mm in shell height, and the minimum size at settlement observed in south England is 0.4 mm diameter and 0.37 mm in shell height (Fretter & Manly 1977). In Ireland, recruits are < 2 mm in shell height in their first year, then they grow 0.2 to 0.3 mm in shell height per year, and will take five years to reach 3 mm in height (Myers & McGrath 1993). The growth rate decreases with age, and is very slow in specimens of 6 mm (Lysaght 1941). However, growth rates are variable depending on the season or the location (Hughes & Roberts 1981; Myers & McGrath 1993), and are for instance higher in France than in England (Daguzan 1976).

Figure 4. Shell of *Melarhaphe neritoides*. Drawing from Fretter & Graham (1980).

Melarhaphe neritoides is preferentially found on rocky shores, in the supralittoral zone where adults live in aggregates (Fig. 5), whereas the juvenile stages live further downward in the intertidal zone (Fig. 6) (Lysaght 1941).

Figure 5. Aggregates of *Melarhaphe neritoides* in rock crevices in the Azores.

Figure 6. Intertidal zonation of juvenile (left) and adult (right) *Melarhaphe neritoides* in the Azores.

The resistance of adults to terrestrial conditions is such that they can survive lack of moisture for over 5 months (Patanè 1933). Melarhaphe neritoides is an important intertidal grazer feeding on lichens (Daguzan 1976) and algae, and can control the establishment of ephemeral algae (Silva *et al.* 2015).

Three larval trematodes occur fairly in *M. neritoides*, more heavily in males than females, and the total percentage of parasitism rises with snail size, e.g. 3.3 % of encysted metacercaria in snails averaging 2 mm in height and 87 % in snails of 8.3 mm (Lysaght 1941).

Reproduction is sexual with separate sexes, with internal fertilization. Oogenesis and spermatogenesis occur throughout the year, and the spawning rhythm is fortnightly (Fretter & Graham 1980; Lysaght 1941). However, mature females are found only in some months over the year and the length of the spawning season ranges from 4 to 9 months depending on the latitudinal geographic location (Cronin *et al.* 2000). The smallest mature male and female found in Ireland were 1.20 mm and 1.22 mm respectively (Cronin *et al.* 2000). Spawning has been observed from March to June or from January to June in North Wales, UK, or from September to April in Plymouth, UK (Hughes & Roberts 1981). Population demographic studies report one or two cohorts suggesting that settlement occurs each year but varies in timing and extent (Fretter & Graham 1980; Fretter & Manly 1977; McGrath 1997; Myers & McGrath 1993). Sedentary adults release pelagic egg capsules in the water (Fig. 7) (Lebour 1935).
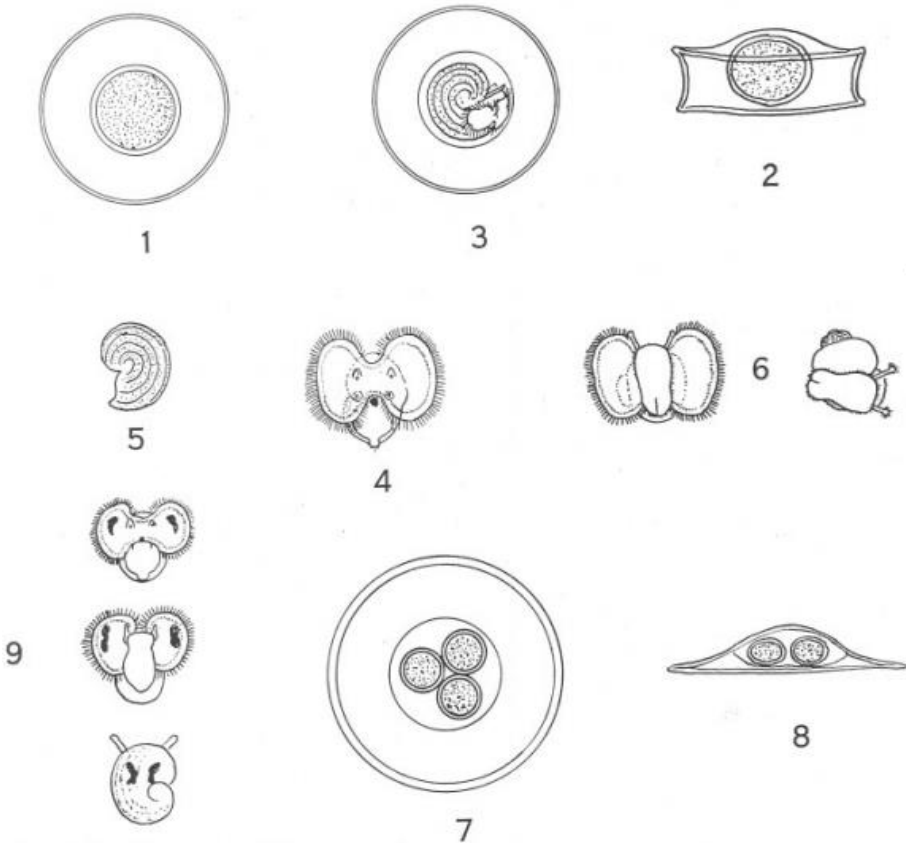
FIG. 1, 2.—Egg capsule of *Littorina neritoides* from plankton, newly laid, 0·18 mm. across.
FIG. 3.—Egg capsule from plankton with veliger nearly ready to hatch.
FIG. 4.—Veliger from plankton, presumably belonging to *L. neritoides*, shell 0·06 mm. across.
FIG. 5.—Shell of same.
FIG. 6.—Late veliger of *Littorina* sp., possibly *L. neritoides*, shell 0·54 mm. high.
FIG. 7, 8.—Egg capsule of *Littorina littorea* 0·96 mm. across.
FIG. 9.—Veligers of *L. littorea*, swimming and crawling, shell 0·48 mm. to 0·5 mm. across.

Figure 7. Morphology of egg capsules and larval stages of *Melarhaphe neritoides*. Drawings from Lebour (1935).

The generation time *G* = 41 months, which is the mean period elapsing between the birth of parents and the birth of their offspring, was calculated from

sequential recaptures of cohorts of winkels marked with Humbrol enamel paint, and following the formula:

$$G = \Sigma l_x m_x x / R_0$$

where l_x is the proportion of winkels surviving from release from the parent to age x, m_x is the age-specific fecundity in terms of the birth rate of female offspring per mother, and R_0 is the net reproductive rate defined by the formula $R_0 = \Sigma l_x m_x$ (Hughes & Roberts 1981).

Marine organisms disperse mostly by ocean currents as larval propagules, and it is commonly thought that the duration of the larval stage is the fundamental determinant of geographic range size, although empirical results have mitigated this paradigm (Luiz *et al.* 2013). We might expect a high potential dispersal of the long-lived planktonic larvae of *M. neritoides* during 4–8 weeks until settlement (Fretter & Manly 1977; Lebour 1935).

The North East Atlantic and the Azores

The North East Atlantic and the Azores archipelago are the two study areas in the PhD.

Quaternary glaciations were prevalent for ca. 80 % of the past two million years, and during the Last Glacial Maximum (LGM), 20-14 thousand years ago glaciers covered most of northern Europe, the Pyrenees and the Alps (Fig. 8) (Weiss & Ferrand 2007). Eustatic sea level in the Mediterranean was 100-120 m lower than present and the Gibraltar Strait was closed. Some rocky habitats remained ice-free in the Azores, the Balkans, the Iberian and Italian peninsula, Ireland, and are assumed to provide glacial refugia for survival of rocky shore species such as *M. neritoides* (Fig. 9) (Ingólfsson 2009; Maggs *et al.* 2008; Weiss & Ferrand 2007).
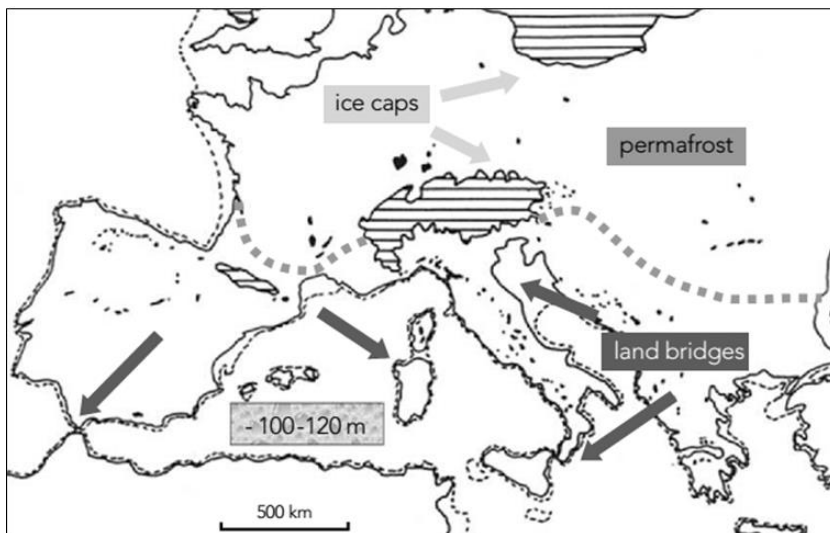
Figure 8. Landscape changes in Southern Europe at LGM (Weiss & Ferrand 2007). Hatched areas represent the ice sheets that covered most of northern Europe, the Pyrenees and the Alps. Permafrost existed over most of Europe. The Mediterranean was 100-120 m lower than present (dashed coastline) resulting in land bridges.
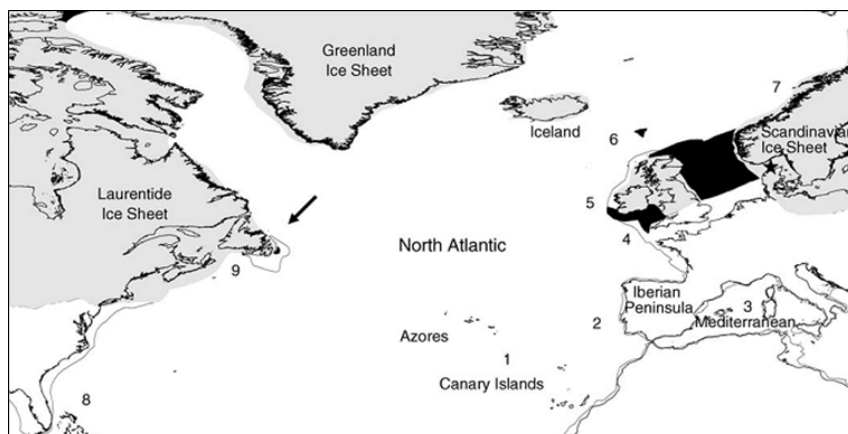
Figure 9. Extent of ice sheets in Atlantic at the LGM around 21000 years ago (Maggs *et al.* 2008), on land (grey) and on sea (solid fill). The finer line delineate the paleocoastline. Potential glacial refugia are numbered.

No signs of mass extinction were found in the molluscan fauna in the Azores during the LGM (Ávila *et al.* 2008), likely because the Azores Plateau was located south of the maximum extension of the polar front which was at 42º N (Dennielou *et al.* 2009).

The Azores archipelago is composed of nine islands about 1500 km west of Portugal, located in the middle of the North Atlantic Ocean between the latitudes 37º N and 40º N and the longitudes 25º W and 31º W. The archipelago has a volcanic origin associated with the triple junction where the American, Eurasian and African plates meet (Fig. 10) (Searle 1980). Despite the volcanic origin, fossiliferous sites from Pleistocene are present in Santa Maria Island and the Formigas' islets in which fossils of *M. neritoides* are dated between 130-120 thousand years ago (Ávila *et al.* 2002). The earliest fossil of the genus *Melarhaphe* is the species *Melarhaphe mausseneti* (Cossman, 1907), now extinct, dated to 55 Ma in France during Upper Paleocene (Reid 1989).
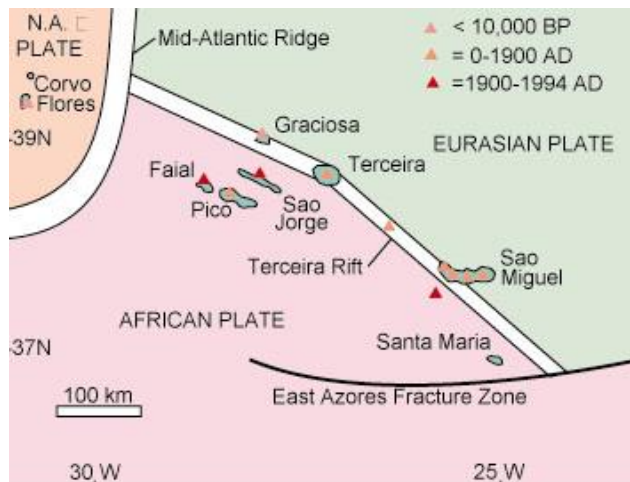
Figure 10. Tectonics of the Azores. From http://volcano.oregonstate.edu/vwdocs/volc_images/africa/azores/tectonics.html.

The formation of the Azores Plateau may have started 36 Ma ago, but the islands are younger (Campan *et al.* 1993). The oldest island, Santa Maria, emerged during the Miocene 8.12 Ma ago, while the youngest, Pico, appeared during the Quaternary 0.25 Ma ago (Fig. 11) (Madeira & Ribeiro 1990).
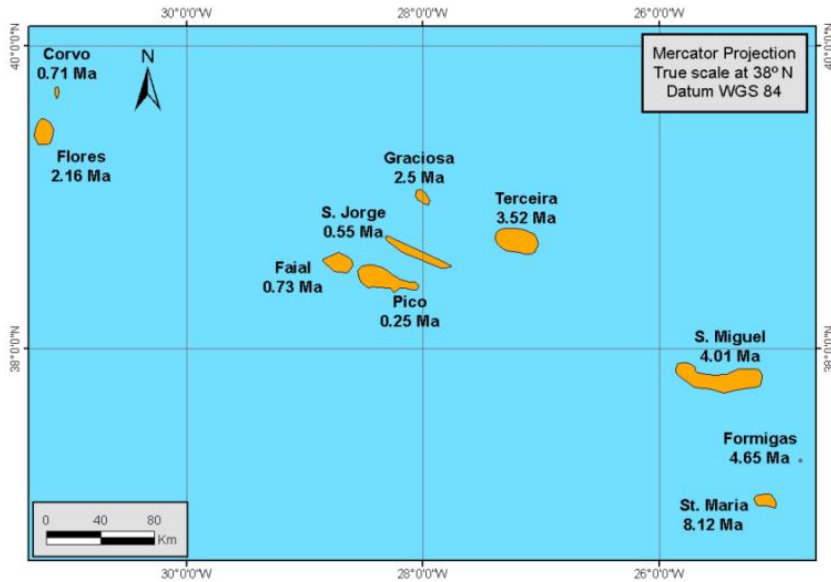
Figure 11. Geological ages of the Azorean islands (Madeira & Ribeiro 1990).

Measuring genetic variation, elucidating evolutionary processes

When studying the evolution of a species, we aim for quantifying the amount of genetic variation in natural populations, and for determining the contribution of each of the four forces in shaping the observed pattern of genetic diversity. This branch of the evolutionary biology is population genetics, and is the subject of this PhD thesis.

Measuring genetic variation and elucidating processes of evolution are paramount to understand boundaries of genetic diversity, connectivity, species delimitation, adaptive potential of species, speciation and taxonomy. Yet it is surprisingly difficult to predict how the processes combine to determine how a population will evolve, or to use our observations of genetic variation in nature to infer how evolution has acted in the past. Because real-time experiments are not possible due to the large time window and population size under study, methods in population genetics are probabilistic and based on models (Méléard 2016), from the simplest Hardy-Weinberg model (Hardy 1908) to the more complex coalescent model (Kingman 1982). Methods test for scenarios of evolution, i.e. the dynamic of genetic diversity among populations through time, by estimating parameters such as, amongst others, allele frequencies expected at equilibrium and those observed, nucleotide substitution rate in DNA, population size, direction and magnitude of selection on DNA, amount of gene copies exchanged over time among populations, and by comparing parameters to models. This results in selecting the scenario for which the estimated parameters approximate the model with the highest confidence. It is therefore important to keep in mind that evolutionary biology attempts to but does not affirm to elucidate the history of species evolution. Evolutionary biology is a dynamic field in which methods and theories coevolve through time.

Some methods are controversial and still debated. In particular, methods for inferring gene flow among populations and consequently assessing GENETIC CONNECTIVITY patterns in species.

Two categories of methods are available for inferring long-term gene flow estimates, namely, F_{ST}-based methods and gene genealogy-based methods:

F_{ST}-based methods use allele frequencies, and gene flow is inferred from F-statistics (Barton & Slatkin 1986; Kelly *et al.* 2010; Neigel *et al.* 1991; Rannala

& Hartigan 1996; Slatkin 1985b; Slatkin & Maddison 1989; Tufto *et al.* 1996; Wright 1931). The classic and usual method infers gene flow N_e*m from the F_{ST} parameter, thanks to the linear relationship between the population genetic differentiation F_{ST}, the migration rate m plus the mutation rate μ, and the effective size N_e of a local population: $F_{ST} = 1/(1+4N_e*(m+\mu))$ (Wright 1949). Derivatives of F_{ST} such as G_{ST} (Nei 1973; Pons & Chaouche 1995), Θ (Weir & Cockerham 1984), N_{ST} (Lynch & Crease 1990; Pons & Petit 1996), φ_{ST} (Excoffier *et al.* 1992) are also commonly used to infer gene flow, as well as the standardized derivatives G'_{ST} (Hedrick 2005), φ'_{ST} (Meirmans 2006), D_{EST} (Chao *et al.* 2008) and G''_{ST} (Meirmans & Hedrick 2011) intended to avoid within-population diversity dependency (Meirmans & Hedrick 2011).

Gene genealogy-based methods estimate genealogies, and convert coalescent times between pairs of alleles into amount of gene flow that would result in a similar distribution of alleles in gene genealogies. Notable methods are the Beerli & Felsenstein method implemented in the MIGRATE-N software (Beerli 2006; Beerli & Felsenstein 1999, 2001; Beerli & Palczewski 2010) and the Nielsen & Wakeley method implemented in the IMA2 software (Hey 2005, 2010; Hey & Nielsen 2004; Hey & Nielsen 2007; Nielsen & Wakeley 2001). From haploid sequence data, the two population parameters θ (the mutation-scaled effective population size, $\theta = 2*N_e*\mu$) and M (the mutation-scaled migration rate, $M = m/\mu$) are also estimated and used to estimate gene flow as the number of immigrants per generation ($N_e*m = 0.5*\theta*M$). But unlike F_{ST}-based methods, sequence data are used in their raw form instead of a summary information, such as allele frequencies, or such as a proportion of rare alleles like in Slatkin's method (1985a). Sequence data are used to build gene genealogies that integrate possible migration (MIGRATE-N), and isolation events and historical changes in population size (IMA2). The genealogy describing the best the migration pattern amongst populations of our sample of

DNA sequences, is selected by either a Maximum Likelihood statistical approach or a Bayesian statistical approach. Then in the selected genealogy, coalescence times – the number of generations that separate alleles from their common ancestor – that are function of migration rate and population size, give estimates of migration rates and population sizes for our sample.

F_{ST}-based methods are popular but rely on too many unrealistic assumptions for natural populations, assuming an equilibrium state known under the Wright's (1931) island model:

- o each population is composed of the same number of individuals, which is a finite size, and constant over time
- o random mating between individuals
- o same reproductive success for all individuals
- o non-overlapping generations
- o mutation at a constant rate over time, and low compared with the migration rates ($\mu << m$) so that mutation can be neglected
- o no selection pressure
- o no recombination
- o migration at a constant rate m over time and equal between each pair of populations, so that a proportion m of individuals are immigrants from other populations at each generation
- o migration is random, so that there is no spatial structure
- o each population has reached an equilibrium between the forces of migration and genetic drift

I am interested in this thesis in the violation of the assumption of a negligible μ, which produces biased and misleading results when applied to highly variable genetic markers whose mutation rates are at the same level or higher than the level of gene flow, i.e. when $N_e\mu \geq N_em$ (Fisher & Bennett 1930; Raybould *et al.* 2002; Whitlock & McCauley 1999; Wright 1931). This results in a F_{ST} index that

reflects the influence of mutation, i.e. $F_{ST} = 1/(1+4N_e*\mu)$, instead of the influence of migration, i.e. $F_{ST} = 1/(1+4N_e*m)$. High μ in DNA generates numerous private alleles with low frequency as well as high within-population genetic diversity, two quantities that may bias genetic differentiation estimates in terms of F_{ST}. On the one hand, F_{ST} is restricted to values much less than 1 (mean maximum $F_{ST} \approx 0.3585$) when the frequency of the most frequent allele is low (near zero) or high (near 1) (Jakobsson *et al.* 2013; Reddy & Rosenberg 2012; Rosenberg & Jakobsson 2008; Rousset 2013). On the other hand, F_{ST} drops to zero when the level of within-population diversity is high (Charlesworth 1998; Gerlach *et al.* 2010; Hedrick 1999, 2005; Heller & Siegismund 2009; Holsinger & Weir 2009; Jost 2008; Leng & Zhang 2011; Long 2009; Meirmans & Hedrick 2011; Nagylaki 1998; Nei 1973; Nei 1987; Neigel 1997; Neigel 2002; Ryman & Leimar 2009; Wang 2012; Whitlock 2011; Wright 1978). This would not be a problem if F_{ST} was strictly used to measure allele fixation as initially intended by Wright (1978), and not to measure differentiation. This bias in genetic differentiation estimates leads to a bias in gene flow estimates inferred from F_{ST}-based methods, and to overestimate connectivity, due to the linear relation between F_{ST} and N_e*m. In contrast, gene genealogy-based methods are suited to highly polymorphic data and produce reliable gene flow estimates over the whole spectrum of mutation rates, because the coalescent model on which these methods rely on is not dependent on the mutation rate (Kingman 1982; Kuhner 2008; Marko & Hart 2011; Wakeley 2001).

Several publications provide detailed discussion about strengths and weaknesses of these two methods (Hart & Marko 2010; Lowe & Allendorf 2010; Marko & Hart 2011; McGovern *et al.* 2010; Pearse & Crandall 2004; Strasburg & Rieseberg 2009).

Objectives and Outline

My aim is to measure and characterise the pattern of mtDNA genetic diversity in the periwinkle *M. neritoides* across its northeastern Atlantic distribution area, a species which shows potential signatures of mtDNA hyperdiversity, and to understand which evolutionary processes shape this genetic diversity, using population genetic tools.

The following questions are addressed:

✓ How much mtDNA genetic diversity does *M. neritoides* harbour?

✓ Which factors are responsible for mtDNA hyperdiversity in *M. neritoides*?

✓ Does mtDNA hyperdiversity reflect population genetic differentiation and structuring in *M. neritoides* despite high dispersal potential in this planktonic-dispersing species?

✓ Are populations of *M. neritoides* genetically connected throughout their distribution area, as one would expect for such a planktonic-dispersing species with long-lived pelagic larvae, and at which rate?

✓ Does mtDNA hyperdiversity induce atypical features in the mitogenome of *M. neritoides*?

In Chapter 1, we **measure mtDNA diversity** in *M. neritoides*, by calculating genetic diversity metrics such as HAPLOTYPE DIVERSITY, haplotype richness, NUCLEOTIDE DIVERSITY and neutral nucleotide diversity, based on a substantial sampling of individuals (610 individuals from six populations over 550 km in the Azores archipelago) and > 10 % of the mitogenome total length (1771 bp for the concatenated 16S-COI-Cyt*b* fragments). Then, we investigate the **factors**

causing this mitochondrial hyperdiversity, by estimating the **mutation rate** at the COI locus, the **effective population size**, and the presence of **selection** on the COI and Cyt*b* genes.

In Chapter 2, we expand the sampling area beyond the Azores archipelago and measure mtDNA diversity in the North East Atlantic, in order to see **whether mtDNA hyperdiversity is prevalent throughout the North East Atlantic**. We calculate indices of population genetic differentiation (F-statistics, AMOVA) to assess the pattern of genetic structure and **test for panmixia** among populations of *M. neritoides*, and highlight the **pitfalls of mtDNA hyperdiversity** in assessing population genetic differentiation.

Furthermore, we quantify **gene flow** using a coalescent-based approach in order to assess the pattern of **genetic connectivity** among populations of *M. neritoides* across its distribution area in the North East Atlantic.

In Chapter 3, we explore the composition and structure of the **whole mitogenome** of *M. neritoides*, and estimate the direction and strength of **selection** on the complete set of protein-coding genes. In comparison with three related species of the same family, Littorinidae, we search for unusual features potentially due to the **influence of mtDNA hyperdiversity**.

Additionally, we carry out a molecular **phylogeny of Littorinimorpha**, the infraorder which *M. neritoides* belongs to, using publicly available whole mitogenome sequence data, in order to test for the robustness of phylomitogenomics for further phylogenomic analyses of Littorinidae.

Chapter 1

Mitochondrial DNA hyperdiversity and its potential causes in the marine periwinkle *Melarhaphe neritoides* (Mollusca: Gastropoda)

Séverine Fourdrilis [1]

Patrick Mardulyn [2]

Olivier J. Hardy [2]

Kurt Jordaens [3]

António Manuel de Frias Martins [4]

Thierry Backeljau [1,5]

[1] Directorate Taxonomy and Phylogeny & JEMU, Royal Belgian Institute of Natural Sciences, Brussels, Belgium

[2] Evolutionary Biology and Ecology, Université Libre de Bruxelles, Brussels, Belgium

[3] Department of Biology, Invertebrate Section, Royal Museum for Central Africa, Tervuren, Belgium

[4] CIBIO, Centro de Investigacão em Biodiversidade e Recursos Genéticos, InBIO Laboratório Associado, Pólo dos Açores, Departamento de Biologia da Universidade dos Açores, University of the Azores, Ponta Delgada, Portugal

[5] Evolutionary Ecology Group, University of Antwerp, Antwerp, Belgium

ABSTRACT

We report the presence of mitochondrial DNA (mtDNA) hyperdiversity in the marine periwinkle *Melarhaphe neritoides* (Linnaeus, 1758), the first such case among marine gastropods. Our dataset consisted of concatenated 16S-COI-Cyt*b* gene fragments. We used Bayesian analyses to investigate three putative causes underlying genetic variation, and estimated the mtDNA mutation rate, selection and the effective population size of the species in the Azores archipelago. The mtDNA hyperdiversity in *M. neritoides* is characterized by extremely high haplotype diversity (Hd = 0.999 ± 0.001), high nucleotide diversity (π = 0.013 ± 0.001), and neutral nucleotide diversity above the threshold of 5 % (π_{syn} = 0.0677). Haplotype richness is very high even at spatial scales as small as 100 m². Yet, mtDNA hyperdiversity does not affect the ability of DNA barcoding to identify *M. neritoides*. The mtDNA hyperdiversity in *M. neritoides* is best explained by the remarkably high mutation rate at the COI locus (μ = 5.82 x 10^{-5} per site per year or μ = 1.99 x 10^{-4} mutations per nucleotide site per generation), whereas the effective population size of this planktonic-dispersing species is surprisingly small (N_e = 5256; CI = 1312-37495) probably due to the putative influence of selection. Comparison with COI nucleotide diversity values in other organisms suggests that mtDNA hyperdiversity may be more frequently linked to high μ values and that mtDNA hyperdiversity may be more common across other phyla than currently appreciated.

INTRODUCTION

The term DNA hyperdiversity is usually applied to populations when neutral nucleotide diversity at selectively unconstrained synonymous sites is ≥ 5 % (Cutter *et al.* 2013), that is when two 100 bp protein-coding DNA sequences

(mitochondrial or nuclear) chosen randomly from a population sample differ on average at five or more synonymous and neutral nucleotide positions. Nucleotide diversity in a sequence alignment is calculated either from pairwise differences at all sites (π) or at SEGREGATING sites only (θ) (Nei 1987; Nei & Miller 1990; Watterson 1975). Yet, π is often preferred because its estimation is less sensitive to sequencing errors and DNA sequence length than θ (Johnson & Slatkin 2008). Nucleotide diversity is also calculated at synonymous sites (π_{syn}) to obtain an estimate of neutral polymorphism reflecting the balance between mutation pressure and genetic drift. This latter measure of π_{syn} is required to observe hyperdiversity. DNA hyperdiversity is usually associated with fast evolving prokaryotes and viruses and less frequently with eukaryotic organisms showing lower rates of evolution (Drake *et al.* 1998). Nevertheless, mitochondrial (mtDNA) or nuclear (nDNA) DNA data retrieved from literature references on 505 animal species, showed signatures of DNA hyperdiversity ($\pi_{syn} \geq 0.05$) in 43 % of the species studied, i.e. 42 % among 394 Chordata, 55 % among 66 Arthropoda, 33 % among 24 Mollusca, 24 % among 17 Echinodermata, and 100 % among 3 Nematoda (Table S1). Although these percentages most probably reflect strong sampling bias, DNA hyperdiversity seems not uncommon in eukaryotes. Rates of mtDNA evolution are 10-30 times faster than nDNA and drive mitonuclear coevolution and speciation through strong selection pressure (Blier *et al.* 2001; Hill 2016; Lane 2009). Hyperdiverse intraspecific mtDNA variation provides a greater density of polymorphic sites for selection to act upon (Cutter *et al.* 2013), and possibly provokes higher speciation rate as observed in birds and reptiles (Eo & DeWoody 2010). Studying mtDNA hyperdiversity is hence interesting to better understand how evolutionary processes such as mutational dynamics and selection that underlie mitonuclear coevolution contribute to speciation (Burton & Barreto 2012).

mtDNA is a popular population genetic marker because of its variability and, as such, is widely used for evolutionary studies at the species level (Féral 2002; Wan *et al.* 2004) and DNA barcoding (Hebert *et al.* 2003b). The main determinants of animal mtDNA diversity are supposed to be mutation rate (μ) and selection, while in contrast to nDNA, effective population size (N_e) and ecology (life history traits) are expected to be less important (Bazin *et al.* 2006; Cutter *et al.* 2006; Dey *et al.* 2013; Lanfear *et al.* 2014; Leffler *et al.* 2012; Nabholz *et al.* 2009; Nabholz *et al.* 2008; Small *et al.* 2007). Indeed, the higher nDNA diversity observed in non-vertebrates vs. vertebrates, in marine vs. non-marine species, and in small vs. large organisms, is not in line with patterns of mtDNA diversity (Bazin *et al.* 2006; Leffler *et al.* 2012). Several determinants shape mtDNA diversity by favoring genetic variation. Three of these are: (1) mutations, i.e. the source of new alleles and increasing genetic variation. (2) diversifying selection and balancing selection that increase genetic variation by favoring extreme or rare phenotypes over intermediate phenotypes (Maruyama & Nei 1981; Mather 1955; Rueffler *et al.* 2006), while other types of selection reduce genetic variation by decreasing the frequency of disadvantageous alleles over time (Anisimova & Liberles 2012). (3) fluctuations in N_e, since more mutations arise in populations with larger N_e (Kimura 1983). As the main determinants of animal mtDNA diversity are supposed to be μ and selection, mtDNA hyperdiversity is more likely to be also explained by high μ or selection on the mitochondrial genome, and we expect that the relationship between N_e and mtDNA diversity may be weakened. Still, at least in eutherian mammals and reptiles mtDNA diversity seems to correlate with N_e (Hague & Routman 2016; Mulligan *et al.* 2006), so that an eventual influence of N_e on mtDNA hyperdiversity cannot a priori be neglected. Yet, far more empirical data are needed to better understand the relative contribution of various determinants of mtDNA hyperdiversity.

In the present work, we investigate three potential determinants of mtDNA hyperdiversity i.e. μ, selection and N_e, in the marine periwinkle *Melarhaphe neritoides* (Linnaeus, 1758) in the Azores archipelago. *Melarhaphe neritoides* is an intertidal gastropod that shows signatures of mtDNA hyperdiversity (see data in García *et al.* 2013). It is a small (shell up to 11 mm) temperate species (Lysaght 1941), in which the sedentary adults produce pelagic egg capsules and long-lived planktonic larvae with high dispersal potential during 4-8 weeks until settlement (Cronin *et al.* 2000; Fretter & Manly 1977; Lebour 1935). *Melarhaphe neritoides* is widely distributed throughout Europe (Fretter & Graham 1980), where it shows a remarkable macrogeographic population genetic homogeneity (inferred from allozyme data) (Johannesson 1992), though locally in Spain it displays huge amounts of mtDNA COI diversity in terms of a large numbers of polymorphic sites (S = 16 %), a very high haplotype diversity (Hd = 0.998) and a very high nucleotide diversity (π = 0.019) (García *et al.* 2013). We studied mtDNA diversity of *M. neritoides* within the archipelago of the Azores because this area provides a vast, though relatively isolated, setting to explore geographic mtDNA variation at different spatial scales.

First, we formally describe and evaluate mtDNA hyperdiversity in *M. neritoides,* by assessing diversity in three mtDNA gene fragments, *viz.* 16S ribosomal RNA (16S), cytochrome oxidase c subunit I (COI) and cytochrome b (Cyt*b*) in substantial numbers of individuals and locations. Second, we survey the literature to compare *M. neritoides* mtDNA hyperdiversity with other littorinids, other planktonic-dispersing gastropods showing high genetic diversity, and other hyperdiverse molluscs in general. Finally, we explore the relationship between mtDNA diversity in *M. neritoides* and (1) μ, (2) selection, (3) N_e, (4) population genetic structuring, and (5) phylogeny.

MATERIALS AND METHODS

Samples and DNA collection

A total of 610 specimens of *M. neritoides* were collected between 1992 and 2012 at six localities in the Azores archipelago, Portugal, *viz.* Varadouro, Faial island (FAI), Fajã Grande, Flores island (FLO), Mosteiros, São Miguel island (MOS), Lajes do Pico, Pico island (PIC), Maia, Santa Maria island (SMA), and Porto Formoso, São Miguel island (SMI) (Fig. 1). These 610 specimens contribute to our analyzed data sets as follows (Table S2): (1) dataset 1: 185 specimens from five islands sequenced for COI (614 bp), 16S (482 bp) and Cyt*b* (675 bp) to investigate mtDNA diversity and demographic history; (2) dataset 2: 223 specimens from one island collected at a single spot of about 100 m² at MOS and sequenced for COI (657 bp) and 213 among these sequenced for 16S (482 bp), to assess microscale mtDNA haplotype richness; (3) dataset 3: 169 specimens from four islands collected between 1992 and 1993, and 175 specimens collected in 2012 at the same four localities, sequenced for COI (578 bp) to generate a temporal series of samples over 20 years for estimating mtDNA μ; (4) dataset 4: 212 specimens from five islands sequenced for COI (605 bp), completed by one COI sequence of *M. neritoides* from the United Kingdom retrieved from GenBank (AJ488608) and 86 COI sequences of seven species from the three littorinid subfamilies Lacuninae, Laevilitorininae and Littorininae (Reid *et al.* 2012) and one species of Pomatiidae available in GenBank, viz. *Bembicium auratum* (Lacuninae) (AJ488606), *Cremnoconchus syhadrensis* (Lacuninae) (AJ488605), *Lacuna pallidula* (Lacuninae) (AJ488604, KT996151), *Laevilitorina caliginosa* (Laevilitorininae) (AJ488607), *Littorina littorea* (Littorininae) (AJ622946, HM884235, HM884236, HM884248, KF643337, KF643416, KF643449, KF643454, KF643456, KF643464, KF643631, KF643658, KF643697,

KF643729, KF643906, KF644042, KF644180, KF644262, KF644330), *Peasiella isseli* (Littorininae) (HE590849), *Pomatias elegans* (Pomatiidae) (JX911283, JQ964789, GQ424199, EU239237-EU239241) and *Tectarius striatus* (Littorininae) (DQ022012-DQ022064), to assess monophyly, possible phylogenetic structuring, and eventual cryptic taxonomic diversity in *M. neritoides*.
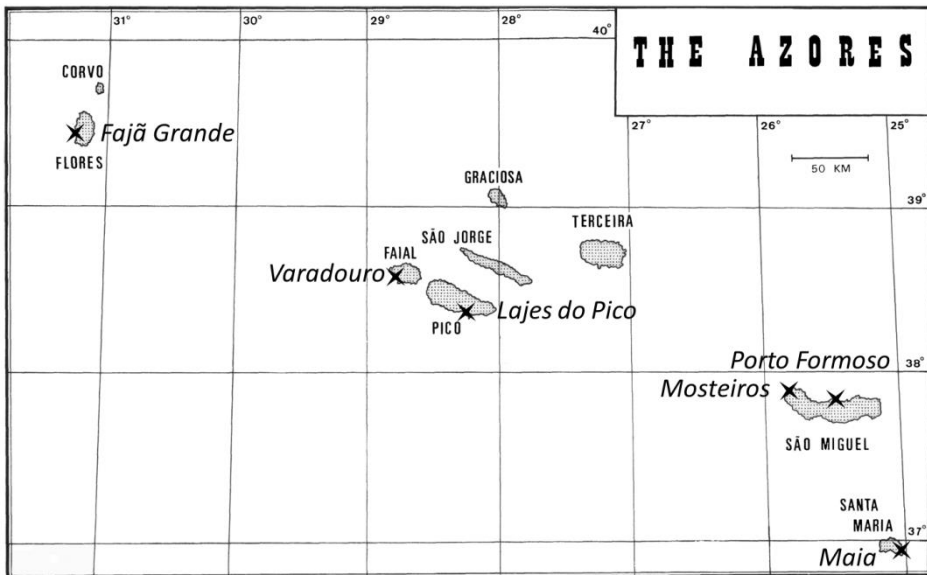
Fig 1. **Sampling sites (cross-shaped symbols) of *M. neritoides* in the Azores archipelago, Portugal.** FAI, Varadouro, Faial island; FLO, Fajã Grande, Flores island; MOS, Mosteiros, São Miguel island; PIC, Lajes do Pico, Pico island; SMA, Maia, Santa Maria island; SMI, Porto Formoso, São Miguel island.

Collected specimens were preserved at -20 °C until DNA analysis. Individual genomic DNA was extracted from foot muscle following the standard protocol of either the NucleoSpin® Tissue kit (Macherey-Nagel GmbH & Co. KG, Düren, Germany) or the DNeasy 96 Blood & Tissue kit (Qiagen GmbH, Hilden,

Germany). Remaining soft body parts and shells have been deposited in the collections of the Royal Belgian Institute of Natural Sciences, Brussels (RBINS) under the general inventory number IG 32962.

16S, COI and Cyt*b* amplification and sequence alignment

PCR amplification was carried out in a 20-µL reaction volume using standard Taq DNA polymerase (Qiagen GmbH, Hilden, Germany) and universal primers LCO1490 (5'-GGTCAACAAATCATAAAGATATTGG-3') and HCO2198 (5'-TAAACTTCAGGGTGACCAAAAAATCA-3') for a 578-to-657 bp region of COI (Folmer *et al.* 1994), universal primers 16Sar (5'-CGCCTGTTTAACAAAAACAT-3') and 16Sbr (5'-CCGGTCTGAACTCAGATCACGT-3') for a 482 bp region of 16S (Simon *et al.* 1994), and littorinid-specific primers 14825 (5'-CCTTCCCGCACCTTCAAATC-3') and 15554 (5'-GCAAATAAAAAGTATCACTCTGG-3') for a 675 bp region of Cyt*b* (Reid *et al.* 1996). The PCR conditions for COI consisted of an initial denaturation at 95 °C for 5 min, 40 cycles of denaturation at 95 °C for 45 s, annealing at 45 °C for 45 s, elongation at 72 °C for 1 min 30 s, and a final elongation at 72 °C for 10 min. The PCR conditions for Cyt*b* were the same except for the annealing step at 48 °C. The PCR conditions for 16S were also the same except for the annealing step at 52 °C, 35 cycles instead of 40, and final elongation for 5 min. PCR products were purified using Exonuclease I and FastAP Thermosensitive Alkaline Phosphatase (Thermo Scientific, Erembodegem-Aalst, Belgium). Sequencing reactions were performed directly on purified PCR products using the BigDye® Terminator v1.1 Cycle Sequencing kit (Life Technologies, Gent, Belgium) and run on an Applied Biosystems 3130xl Genetic Analyser automated capillary sequencer, or outsourced to Macrogen (Rockville, MD, USA). Sample files were assembled, edited and reviewed using ABI Prism® SeqScape® 2.5.0 (Applied Biosystems). The accuracy and reproducibility of the PCR results were

validated by triplicating COI and 16S amplifications on a subset of 20 individuals, using standard Taq DNA polymerase for two replicates and HotStar HiFidelity DNA Polymerase (Qiagen GmbH, Hilden, Germany) for one replicate. Sequence alignments were made with ClustalW (Thompson *et al.* 1994) using default parameters in BioEdit 7.0.9.0 (Hall 1999). All sequences were deposited in GenBank (KT996151-KT997344). The morphology-based identification of *M. neritoides* was used to cross-validate DNA-based identification through DNA barcoding by querying the 185 COI fragments from dataset 1 in the Barcode of Life Datasystem (BOLD) (Ratnasingham & Hebert 2007).

mtDNA diversity

The three gene fragments were concatenated for the 185 specimens of dataset 1, using Geneious 5.3.4 (http://www.geneious.com, Kearse *et al.* 2012). DNA diversity metrics (Tables 1 and 2) were calculated with DnaSP 5.10.1 (Librado & Rozas 2009). Despite the fact that 32-42 specimens were sampled per site, there were no shared haplotypes between sampling sites ($H_s = 0$), i.e. all haplotypes were private (Table 1). Consequently, we examined mtDNA haplotype richness at a microscale, i.e. within a sampling site. Dataset 2, composed of identical fragment lengths across individuals, was used to compute individual-based rarefaction curves for the COI and 16S fragments using EstimateS 8.2.0 (Colwell 2006) in order to assess the relationship between the number of haplotypes observed (H_{obs}) and sample size, and compute the *Chao1* and *Chao2* richness estimators (Chao 1984, 1987). Given that COI and Cyt*b* showed similar diversity levels (Table 1), only COI was used for rarefaction analysis. A logarithmic trendline, best fitting the data, was applied to each rarefaction curve to extrapolate H_{obs} to larger sample sizes.

Table 1. **mtDNA diversity metrics of *Melarhaphe neritoides*.** Statistics describing the number of individuals (N), number of haplotypes (H), number of private haplotypes (H_p), number of shared haplotypes among sampling sites (H_s), number of shared haplotypes within sampling site (H_w), DNA fragment length in base pairs (L), number of segregating sites (S) and its corresponding percentage of the fragment length into brackets, haplotype diversity (Hd) ± standard deviation, Jukes-Cantor corrected nucleotide diversity (π) ± standard deviation, Jukes-Cantor corrected nucleotide diversity at SYNONYMOUS sites (π_{syn}) and Jukes-Cantor corrected nucleotide diversity at NON-SYNONYMOUS sites ($\pi_{non\text{-}syn}$).

	N	H	H_p	H_s	H_w	L	S	Hd	π	π_{syn}	$\pi_{non\text{-}syn}$
16S-COI-Cyt*b*	185	184	184	0	1	1,771	420 (24%)	0.999 ± 0.001	0.013 ± 0.001	0.0677	0.0004
16S	185	77	63	12	2	482	71 (15%)	0.814 ± 0.030	0.004 ± 0.001	-	-
COI	185	156	142	13	1	614	169 (28%)	0.996 ± 0.002	0.018 ± 0.001	0.0736	0.0001
Cyt*b*	185	166	153	9	4	675	180 (27%)	0.998 ± 0.001	0.016 ± 0.001	0.0637	0.0006

Population genetic structure

The monophyly of *M. neritoides* was assessed, and p-distances were compared within and among clades, in order to detect possible cryptic taxa and/or phylogenetic structuring that might contribute to the overall mtDNA hyperdiversity. First, two species trees were produced from dataset 4 using Bayesian inference (BI) and Maximum Likelihood (ML). Seven Littorinidae species were added to the ingroup. The outgroup *Pomatias elegans* belongs to a different family (Pomatiidae), but the same superfamily (Littorinoidea) as *M. neritoides*. Two independent runs of BI were performed using MrBayes 3.2.2 (Ronquist *et al.* 2012) hosted on the CIPRES Science Gateway (Miller *et al.* 2010), under a GTR+G nucleotide substitution model selected according to jModelTest 2.1.4 (Darriba *et al.* 2012), for 4.10^6 generations with a sample frequency of 100 and a 30 % burn-in. Convergence between the two runs onto the stationary distribution was assessed by examining whether the potential scale-reduction factors was close to 1 in the pstat file, standard deviation of split frequencies fell below 0.01 in the log file, and trace plots showed no trend by examining the p files in TRACER 1.6 (Rambaut *et al.* 2014). The final consensus tree was computed from the combination of both runs. ML analysis based on the GTR+G model was conducted in MEGA 6.06 (Tamura *et al.* 2013), with bootstrap consensus trees inferred from 1,000 replicates. Second, three methods of species delimitation were used: (1) the Automatic Barcode Gap Discovery (ABGD, available at http://wwwabi.snv.jussieu.fr/public/abgd/abgdweb.html) method (Puillandre *et al.* 2012), (2) the Bayesian implementation of the Poisson tree Processes (bPTP, available at http://species.h-its.org/ptp/) model (Zhang *et al.* 2013), and (3) the General Mixed Yule Coalescent (GMYC, available at http://species.h-its.org/gmyc/) model (Fujisawa & Barraclough 2013; Pons *et al.* 2006). Finally, sequence divergence within and between clades was assessed by calculating

mean within-group p-distances for the four species comprising more than one sequence (*Littorina littorea*, N = 19; *M. neritoides*, N = 213; *Pomatias elegans*, N = 8; *Tectarius striatus*, N = 53), and mean between groups p-distances for all species pairs (Table S2), using MEGA. Additionally, COI sequence divergence within *M. neritoides* was assessed by generating an intraspecific p-distances distribution from the 185 COI sequences included in dataset 1, using MEGA.

Dataset 1 was subjected to the program ALTER (http://sing.ei.uvigo.es/ALTER/, Glez-Peña *et al.* 2010) to convert the Fasta-formatted sequence alignment to a sequential Nexus-formatted file, which then could be analyzed by NETWORK 4.6.1.2 (www.fluxus-engineering.com, Bandelt *et al.* 1999) to reconstruct a median-joining haplotype network. Population genetic structure in *M. neritoides* was qualitatively investigated with the haplotype network which provides information about phylogeographic structure and gene flow among populations, and quantified by G_{ST} (Pons & Petit 1995), N_{ST} based on a distance matrix of pairwise differences (Pons & Petit 1996) and φ_{ST} (Excoffier *et al.* 1992) using dataset 1 in SPAGEDI 1.4 (Hardy & Vekemans 2002) for G_{ST} and N_{ST} and ARLEQUIN 3.5.1.3 (Excoffier & Lischer 2010) for φ_{ST}.

mtDNA mutation rate

mtDNA evolves fast enough to provide sufficient variation for the estimation of μ over a two-decades period, i.e. the time span of our temporal sampling and corresponding to 5-6 generations of *M. neritoides* (Drummond *et al.* 2003). Dataset 3 comprises different sampling points in time, allowing sequences to be treated as heterochronous data for estimating the number of mutations occurring in the time interval between samples as described in Seo et al. (2002) and Drummond et al. (2002). In this way the mutation rate per nucleotide site per year can be inferred using a Bayesian MCMC method as

implemented in BEAST 2.1.3 (Bouckaert *et al.* 2014) hosted on the CIPRES Science Gateway. The Bayesian MCMC analysis was performed under a HKY substitution model (the closest model to GTR since GTR is not available in BEAST) with empirical base frequencies and a fixed substitution rate of 1.0 and a tree prior set to "*coalescent exponential population*" (chosen after model comparison with the "*coalescent constant population*" and the "*coalescent Bayesian skyline*" priors), a strict clock model assuming a constant substitution rate over time and a prior set to lognormal with M = -5 and S = 1.25. The analysis was run in triplicate for 500 million generations with a sample frequency of 50,000 and 10 % burn-in. Convergence of MCMC chains was assessed by visual examination of the log trace of each posterior distribution showing caterpillar shape in TRACER, and making sure that the ESS value of each statistic was > 200 (Ho & Shapiro 2011). The three runs were combined using LOGCOMBINER 2.1.3 (part of the BEAST package) and the final ESSs were at least 1,100. The estimate of μ was provided under the "*Estimates*" tab in TRACER as the mean of the "*clockRate*" parameter.

Demography, selection and effective population size

Departure from mutation-drift equilibrium indicative of demographic change or SELECTIVE SWEEP was assessed in dataset 1 using Tajima's D (Tajima 1989) and Fu's Fs (Fu 1997) tests implemented in ARLEQUIN and 10,000 coalescent-based simulations were run to calculate p-values. Since Tajima's D and Fu's Fs statistics are sensitive to both demographic change and selection, we also applied Fay & Wu's H statistic (Fay & Wu 2000) to dataset 1 for the single 16S, COI, Cyt*b* fragments and the concatenated 16S-COI-Cyt*b* data using DnaSP to attempt to discriminate between the effects of population size change and selection (Zeng *et al.* 2006). *Tectarius striatus* was the most closely related species to *M. neritoides* (Reid *et al.* 2012) for which the three

same gene fragments of 16S, COI and Cyt*b* were available on Genbank (U46825, AJ488644, U46826), and was therefore used as outgroup for the Fay & Wu test. Confidence intervals were calculated based on 10,000 coalescent-based simulations.

ARLEQUIN was used to construct a distribution of pairwise nucleotide differences between haplotypes (sequence mismatch distribution) and to compare this distribution with the expectations of a sudden expansion model (Harpending 1994; Li 1977; Rogers 1995). Although the analysis complied with the assumption of panmixis, it did not do so with respect to neutrality (see Results), thus limiting the reliability of the results. Three demographic parameters were inferred using a generalized nonlinear least-squares method to determine whether *M. neritoides* has undergone sudden population growth: the rate of population growth $τ = 2\mu t$ (*t* being the time since the expansion), the initial population size before the growth (θ_0), and the final population size after growth (θ_1). The goodness of fit between the observed and expected mismatch distributions was tested by parametric bootstrapping of the sum of squared deviations (Ssd) (Schneider & Excoffier 1999), and by the Harpending Raggedness index (r) (Harpending *et al.* 1998).

A time-calibrated Bayesian skyline plot (BSP) was built for dataset 1 using BEAST, to detect past population dynamics through time and to estimate N_e of *M. neritoides* in the Azores. The coalescent priors used in the skyline plot model assume a random sample of orthologous, non-recombining and neutrally evolving sequences from a panmictic population. The skyline plot model has been shown to be robust to violation of these assumptions and to correctly reconstruct demographic history with mtDNA (Drummond *et al.* 2005). However, recent studies show that violation of these assumptions may still affect the estimated population size variation, and that the BSP is prone to confound the effect of population structure with declines in population size in panmictic populations, or fails to detect population expansion (Grant 2015;

Heller *et al.* 2013). Hence, population structure and selection were assessed beforehand using Tajima's D, Fu's Fs and Fay & Wu's H statistics. The BSP analyses were performed under a HKY substitution model with empirical base frequencies, a fixed substitution rate equal to 1.0, and a piecewise-constant Bayesian skyline model with 10 groups. The prior on the clockRate parameter was set to a log-normal with M = −5 and S = 1.25. Analyses were run in triplicate for 200 million generations with a sample frequency of 20,000 and 10 % burn-in. After combination of the three runs, the final ESSs were at least 1,000. N_e was extracted from the BSP, by dividing the median value of the N_e*τ product in the most recent year 1996 (N_e*τ ≈ 17977) by the generation time τ = 3 years and five months (Hughes & Roberts 1981).

RESULTS

mtDNA diversity of *Melarhaphe neritoides*

Dataset 1, representing the overall population of the Azores archipelago (N = 185 from five localities in the archipelago), contains 184 different and private haplotypes (H = 184; H_p = 184 and hence H_s = 0) (Table 1), except for one haplotype that was found in two individuals from Pico island (H_w = 1). Hence, the frequency of this latter, i.e. the most common, haplotype was 0.0108, while all other haplotypes had a frequency of 0.00541. This remarkable mtDNA diversity is further reflected by a haplotype diversity (based on the concatenated 16S-COI-Cyt*b* data) close to its maximum value 1 (Hd = 0.999 ± 0.001), indicating a probability of less than 0.001 % that two individuals from the same locality share the same haplotype in the overall population of the archipelago. One fourth of the 1771 nucleotide positions are polymorphic (S = 23.7 %) with 167 sites (9.4 %) showing one variant, 225 sites (12.7 %) two variants, 24 sites (1.4 %) three variants, and four sites (0.2 %) four variants.

Moreover, there is on average 1.3 % nucleotide differences per site between two randomly chosen DNA sequences in the overall population (π = 0.013 ± 0.001). More precisely, among the protein-coding COI and Cyt*b* regions (1289 bp), the nucleotide diversity at synonymous sites is π_{syn} = 0.0677 (6.77 %) and at non-synonymous sites is $\pi_{non-syn}$ = 0.0004 (0.04 %).

Repeating COI and 16S PCR amplifications on 20 specimens yielded identical sequence results, confirming that PCR did not generate artificial variation. For 185 query sequences of *M. neritoides* submitted to BOLD, which stores 51 barcodes of *M. neritoides* (data retrieved from BOLD the 13 October 2015), the identification engine returns 100 % correct identifications under one and the same Barcode Index Number (BIN = BOLD:AAG4377), and a similarity percentage ranging 98-100 % for the best match of each queried sequence.

The individual-based rarefaction curves of 16S, COI and 16S-COI (dataset 2) do not reach a plateau, but their steep slopes decrease according to 16S-COI > COI > 16S (Fig. 2). H_{obs} values are close to the maximal sampling size *n* for COI (H_{obs} = 180, *n* = 223) and 16S-COI (H_{obs} = 174, *n* = 197) fragments indicating that a large fraction of the haplotype diversity remains to be discovered, whereas it is further from *n* for 16S fragment (H_{obs} = 71, *n* = 213). The logarithmic trendlines representative of the population growth in the species show inflexion around large sampling sizes (*n* > 500), indicating that additional sampling is likely to yield new haplotypes. Indeed, the *Chao1* (*chao1 mean* = 1596.22, CI = [878.94 ; 3043.34], *n* = 197) and *Chao2* (*chao2 mean* = 1486.20, CI = [842.91 ; 2748.15], *n* = 197) estimators for the concatenated 16S-COI gene fragment suggest that the predicted total haplotype richness of *M. neritoides* would be reached by sampling 1500 individuals per sampling site.
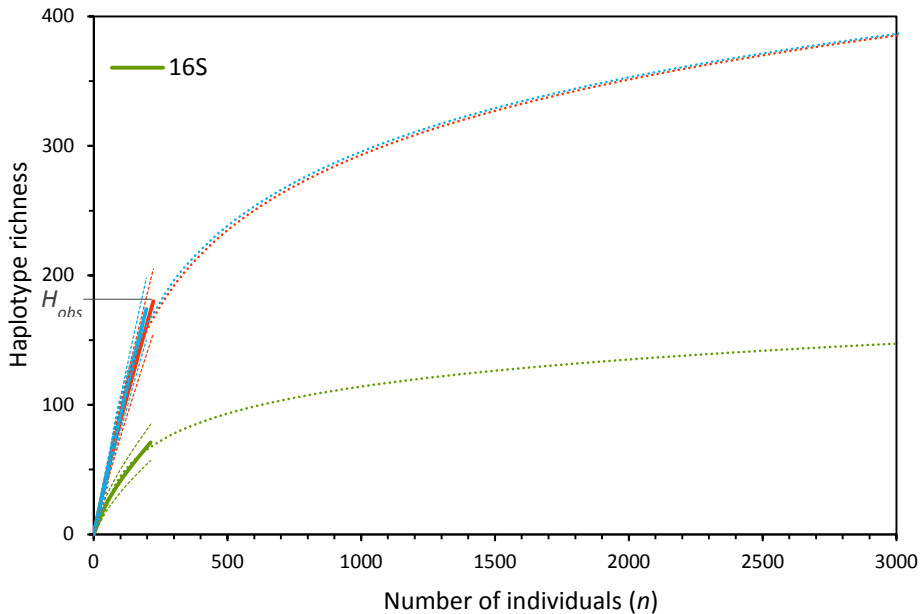
Fig 2. **Individual-based rarefaction curves (solid lines) and 95 % confidence intervals (dashed lines) based on COI, 16S and concatenated 16S-COI data, based on *M. neritoides* specimens sampled in Mosteiros (MOS), São Miguel island.** H_{obs} is the haplotype richness observed in the actual sample (*n*) from MOS. The logarithmic trendlines (dotted lines) show a prediction of the haplotype richness expected for larger sampling size at the MOS sampling site.

Demography, selection and mutation rate

Both Tajima's D and Fu's Fs tests show a significant departure of *M. neritoides* from constant population size or neutrality (D = -2.030, p < 0.01 and Fs = -23.706, p < 0.01), suggesting demographic expansion and/or a potential action of selection. Fay & Wu's H, which is sensitive to positive selection and not to population growth or background selection, shows significant signal of selection for 16S (H = -30.42, CI = [-3.06 ; 1.13]), COI (H = -85.16, CI = [-17.05 ; 6.08]),

Cyt*b* (H = -110.38, CI = [-13.33 ; 5.18]) and the concatenated 16S-COI-Cyt*b* fragment (H = -225.96, CI = [-30.36 ; 11.22]). The unimodal curve of the sequence mismatch distribution (Fig. 3) suggests that population expansion cannot be rejected as $\theta_0 < \theta_1$ ($\tau = 25.543$, $\theta_0 = 4.366$, $\theta_1 = 123.516$). The non-significant values of the sum of squared deviations (Ssd = 0.00118, p = 0.600) and Harpending's Raggedness index (r = 0.0005, p = 0.998) show that the sudden expansion model provides a good fit to the data.
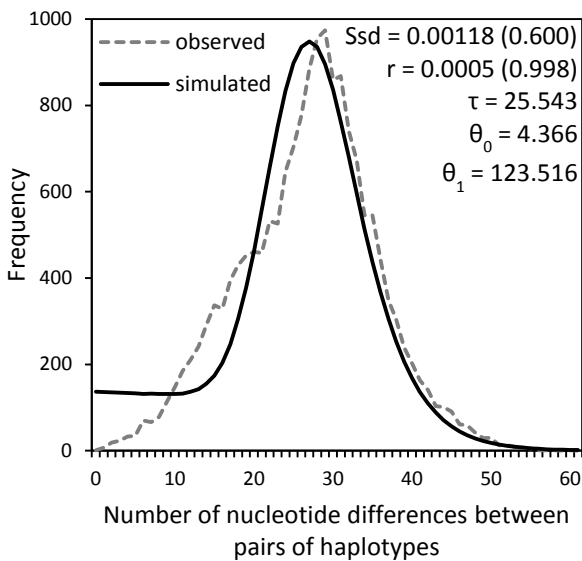
Fig 3. **Mismatch distribution analysis showing the unimodal distribution of the observed number of differences between pairs of haplotypes of *M. neritoides*.** Ssd, sum of squared differences and p-value in parenthesis; r, Harpending's Raggedness index and p-value in parenthesis; τ, time in generations since the last demographic expansion; θ_0, initial population size; θ_1, final population size.

The time-calibrated BSP shows an increase of N_e through time, indicating that *M. neritoides* has been expanding in the Azores archipelago or has undergone selection (Fig. 4). For the year 1996, the BSP gives $N_e{}^*\tau \approx 17977$,

corresponding to N_e ranging from 1312 to 37495 with an average $N_e \approx 5256$ individuals.
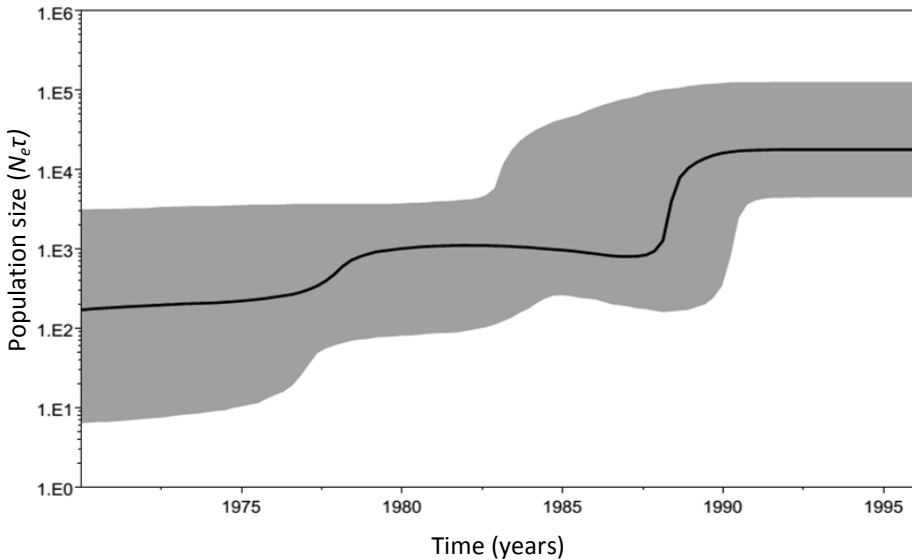
Fig 4. **Historical demographic trends of the median estimate of the maternal effective population size over time (bold line) constructed using a Bayesian skyline plot approach based on concatenated 16S-COI-Cyt*b* haplotypes of *M. neritoides* sampled in 1992, 1993 and 1996.** The y-axis is the product of effective population size (N_e) and generation time (τ) in a log scale, while the x-axis is a linear scale of time in years. The 95% highest probability density (HPD) intervals are shaded in grey and represent both phylogenetic and coalescent uncertainty.

With data sampled in 1992, 1993 and 2012 (i.e. an interval of 20 years), we estimated a mutation rate of $\mu = 5.82 \times 10^{-5}$ per nucleotide site per year at COI. Considering a generation time of τ = 41 months (i.e. 3.42 years), the mutation rate was estimated to be $\mu = 1.99 \times 10^{-4}$ mutations per nucleotide site per generation.

Population genetic structure

All phylogenetic trees provided maximal support for the monophyly of *M. neritoides* (trees not shown). Additionally, the three species delimitation methods, ABGD, bPTP and GMYC, lumped *M. neritoides* as one Molecular Operational Taxonomic Unit (trees not shown). The mean intraspecific p-distance within *M. neritoides* was d = 0.018 ± 0.002, i.e. one order of magnitude greater than the mean intraspecific p-distances of the three other species, *viz. Littorina littorea* (d = 0.004 ± 0.001), *Pomatias elegans* (d = 0.009 ± 0.002) and *Tectarius striatus* (d = 0.006 ± 0.001), but still far below interspecific p-distances ranging from 0.166 to 0.271 for the 36 possible species pairs of Littorinoidea, from 0.166 to 0.246 for the 21 species pairs of Littorinidae, or from 0.187 to 0.225 for the six species pairs of Littorininae (Table S3). The Gaussian distribution of intraspecific COI p-distances in *M. neritoides* (Fig. 5) indicates that the five populations sampled on five different islands of the Azores archipelago form a homogeneous haplotype mixture without any evidence of a DNA barcode gap.

The bush-like pattern of the mtDNA haplotype network (Fig. 6) shows the overwhelming number of unique, private haplotypes represented by single individuals (i.e. singletons), the lack of shared haplotypes between sites, and several homoplastic character states (cycles). The apparent lack of association between genetic variation and geographic location (as revealed by the distribution of colours across the network of Fig. 6) suggests the absence of phylogeographic structure in Azorean *M. neritoides*.

The low and non-significant indices of population genetic differentiation (G_{ST} = 0.0003, p = 0.1676; N_{ST} = 0.0021, p = 0.5346; φ_{ST} = 0.0026, p = 0.2220) make that the hypothesis of panmixis (and hence no population structuring) cannot be rejected.
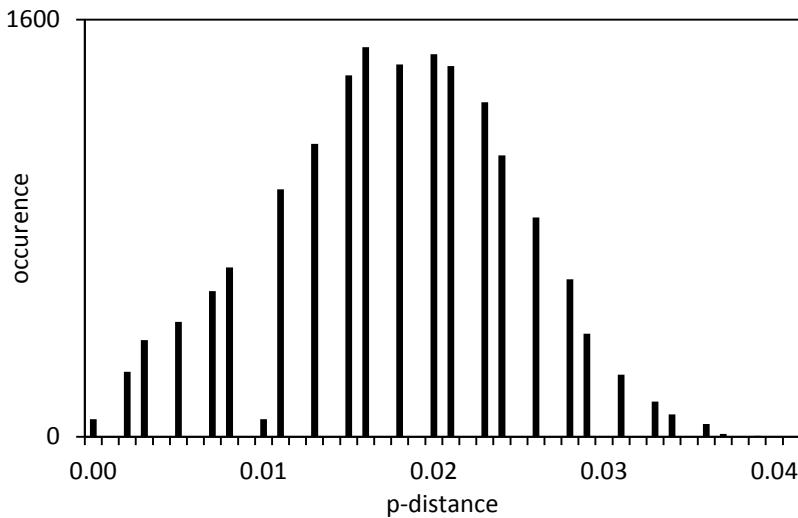
Fig 5. **Distribution of COI pairwise p-distances in *M. neritoides*.**

DISCUSSION

How diverse is the mtDNA of *Melarhaphe neritoides*?

Azorean *M. neritoides* harbours a remarkable amount of intraspecific mtDNA diversity, characterized by very high haplotype diversity and nucleotide diversity with respect to the concatenated 16S-COI-Cyt*b* gene fragments, at the single Cyt*b* gene fragment and at the single COI gene fragment. Moreover, it shows a value of neutral mtDNA nucleotide diversity π_{syn} ≥ the threshold of 5 % for the concatenated 16S-COI-Cyt*b* fragments, and is therefore qualified as hyperpolymorphic. The π_{syn} values for COI and Cyt*b* separately are also ≥ 0.05 and support mtDNA hyperdiversity in *M. neritoides* (Table 1). mtDNA hyperdiversity is also observed in a Spanish population. The COI data retrieved from García et al. (2013) yielded π_{syn} = 0.0762 (7.62 %) and $\pi_{non-syn}$ = 0.0002 (0.02 %) in a local Spanish population of 49 individuals. These values are very

similar to those of COI in the Azorean populations (Table 1). Therefore, mtDNA hyperdiversity is not a local characteristic of *M. neritoides* along the Iberian Atlantic coast, but is shared more broadly in the Azorean populations, and presumably, throughout the species' distribution range.

Fig 6. **Median-joining network of mtDNA in *M. neritoides*.** Branch lengths are proportional to the numbers of mutational steps separating haplotypes. The size of circles is proportional to the number of individuals per haplotype and the sole haplotype shared by two individuals is marked by an arrow. Haplotype origins: Flores island – green; Faial island – blue; Pico island – yellow; São Miguel island – red; Santa Maria island – purple.

The high π values in *M. neritoides* reflect natural variation, not PCR errors, as validated by the identical triplicates of mtDNA sequences and 100 % correct species identification using barcoding. DNA barcoding is based on the premise that COI sequence divergence is higher among species than within species (Hebert *et al.* 2003b), and might be hampered by high mtDNA variation, specifically COI hyperdiversity and high intraspecific sequence divergence in COI. Yet, in spite of the highly variable COI marker in *M. neritoides* (π = 0.018 ± 0.001) and elevated intraspecific p-distance (d = 0.001-0.041), the ability of DNA barcoding to identify *M. neritoides* is not affected by this mtDNA hyperdiversity.

The mtDNA of *M. neritoides* is more diverse than (1) mtDNA of most temperate littorinids and many tropical littorinids, (2) mtDNA of many planktonic-dispersing marine invertebrates, and (3) mtDNA of other hyperdiverse Mollusca (Table 2). More specifically, in comparison with 26 other littorinid species, *M. neritoides* has the highest COI haplotype diversity among temperate species (i.e. *Austrolittorina* spp., *Bembicium vittatum*, *Littorina* spp., *Tectarius striatus*) and the same degree as two tropical species *Echinolittorina reticulata* and *Echinolittorina vidua*. *Melarhaphe neritoides* also has the highest COI nucleotide diversity among temperate species, and shows a higher COI nucleotide diversity than tropical species (i.e. *Bembicium nanum*, *Cenchritis muricatus*, *Echinolittorina* spp., *Littoraria* spp.) except for *Echinolittorina vidua* whose nucleotide diversity (π = 0.041) is about twice that of *M. neritoides* (π = 0.018). In comparison to 15 other non-littorinid marine invertebrates with similar planktonic larval dispersal and high mtDNA variability, *M. neritoides* has the highest COI haplotype diversity. Yet, *M. neritoides* shows the same degree of COI haplotype diversity as the pelagic nudibranch *Glaucus atlanticus* (*Hd* = 0.996) and the annelid *Pygospio elegans* (*Hd* = 0.996). Regarding COI nucleotide diversity, *M. neritoides* has the highest value among annelids, arthropods, cnidarians, echinoderms, other gastropods, and some bivalves (but

not all). Two bivalves, viz. *Brachidontes pharaonis* and *Tridacna maxima*, show very high COI nucleotide diversities that probably reflect ongoing speciation in the three lineages of the *Brachidontes spp.* complex (Terranova *et al.* 2007) and in the four lineages in *Tridacna maxima* (Nuryanto & Kochzius 2009). The literature data in Table 2 suggest that there is no obvious correlation between π and *Hd*. However, more data are needed to corroborate this observation.

We estimated the neutral component of the COI nucleotide hyperdiversity in *M. neritoides*, i.e. π_{syn} = 0.074, on which the diagnosis of hyperdiversity is based. In comparison to eight other mollusc species with hyperdiverse mtDNA (Table 2), *M. neritoides* is situated in the lower part of the neutral nucleotide diversity range (π_{syn} = [0.066-0.256]).

Table 2. **Overview of mtDNA diversity in other Littorinidae, various highly diverse planktonic-dispersers and hyperdiverse mollusc species.** Taxa are listed by decreasing value of haplotype diversity. An, Annelida; Ar, Arthropoda; Ch, Chordata; Cn, Cnidaria; Ec, Echinodermata; Mo, Mollusca; Ne, Nematoda; Po, Porifera; d, direct larval development; p, planktonic larval development (pelagic larval duration given in parenthesis); n/a, not applicable; *N*, number of individuals; *L*, locus length in base pairs; *Hd*, haplotype diversity; *π*, nucleotide diversity; - missing data; * *π* calculated at synonymous sites only (π_{syn}).

Species	larval development	sampling area	*N*	locus	*L*	*Hd*	*π*	Reference
Other Littorinidae								
temperate species								
Littorina saxatilis	Mo	d	North Atlantic	453	ND1-tRNApro-ND6-Cyt*b*	1154	0.940	0.005	(Doellman *et al.* 2011)
			North Atlantic	778	Cyt*b*	607	0.905	0.009	(Panova *et al.* 2011)
Tectarius striatus	Mo	p (unknown)	Macaronesia	109	COI-Cyt*b*	993	0.934	0.006	(Van den Broeck *et al.* 2008)
Littorina keenae	Mo	p (unknown)	North Pacific	584	ND6-Cyt*b*	762	0.815	0.003	(Lee & Boulding 2007)
Littorina littorea	Mo	p (28-42 days)	North Atlantic	488	COI	424	0.810	0.004	calculated from data in Wares et al. (2002), Williams et al. (2003), Williams & Reid (2004), Giribet et al. (2006), Blakeslee et al. (2008), Layton et al. (2014)
Littorina plena	Mo	p (64 days)	NE Pacific	135	Cyt*b*	414	0.775	0.006	(Lee & Boulding 2009)

Littorina obtusata	Mo	d	North Atlantic	46	COI	582	0.762	0.006	calculated from data in Wares & Cunningham (2001)
			NW Atlantic	31	COI	574	0.127	0.001	calculated from data in Layton et al. (2014)
Bembicium vittatum	Mo	d	Indian Ocean	40	12S	324	0.730	-	(Kennington *et al.* 2012)
Austrolittorina unifasciata	Mo	p (4 weeks)	Australia	102	COI	658	0.541	0.002	calculated from data in Colgan et al. (2003), Williams et al. (2003), Waters et al. (2007)
Littorina scutulata	Mo	p (37-70 days)	NE Pacific	265	Cyt*b*	414	0.389	0.003	(Lee & Boulding 2009)
Littorina subrotundata	Mo	d	NE Pacific	229	Cyt*b*	414	0.297	0.001	(Lee & Boulding 2009)
Austrolittorina antipodum	Mo	p (4 weeks)	New Zealand	40	COI	658	0.146	0.001	calculated from data in Williams et al. (2003), Waters et al. (2007)
Littorina sitkana	Mo	d	NE Pacific	146	Cyt*b*	414	0.093	0.001	(Lee & Boulding 2009)
tropical species									
Echinolittorina reticulata	Mo	p (3-4 weeks)	Indo-Pacific	37	COI	1251	1.000	0.009	(Reid *et al.* 2006)
Echinolittorina vidua	Mo	p (3-4 weeks)	Indo-Pacific	92	COI	1217	0.996	0.041	(Reid *et al.* 2006)
Echinolittorina trochoides C	Mo	p (3-4 weeks)	Indo-Pacific	14	COI	1251	0.989	0.006	(Reid *et al.* 2006)
Littoraria coccinea glabrata	Mo	p (unknown)	Indian Ocean	45	COI	451	0.954	0.006	(Silva *et al.* 2013)
Echinolittorina trochoides A	Mo	p (3-4 weeks)	Indo-Pacific	46	COI	1251	0.943	0.009	(Reid *et al.* 2006)

Echinolittorina trochoides B	Mo	p (3-4 weeks)	Indo-Pacific	18	COI	1251	0.935	0.004	(Reid *et al.* 2006)
Bembicium nanum	Mo	p (weeks)	Australia	54	COI	806	0.920	0.006	(Ayre *et al.* 2009)
Echinolittorina trochoides E	Mo	p (3-4 weeks)	Indo-Pacific	21	COI	1251	0.900	0.003	(Reid *et al.* 2006)
Echinolittorina trochoides D	Mo	p (3-4 weeks)	Indo-Pacific	20	COI	1251	0.884	0.003	(Reid *et al.* 2006)
Cenchritis muricatus	Mo	p (4 weeks)	Caribbean	77	COI	282	0.850	0.008	(Díaz-Ferguson *et al.* 2012)
Echinolittorina ziczac	Mo	p (3-4 weeks)	Caribbean Sea	31	COI	431	0.750	0.004	(Díaz-Ferguson *et al.* 2012)
Echinolittorina lineolata	Mo	p (3-4 weeks)	South Atlantic	496	COI	441	0.704	0.003	calculated from Genbank data KJ857561-KJ858054 and Williams & Reid
			South Atlantic	442	Cyt*b*	203	0.284	0.002	calculated from Genbank data KM210838-KM211279
Littoraria scabra	Mo	p (unknown)	Indo- Pacific	50	COI	527	0.690	0.003	(Silva *et al.* 2013)
Littoraria irrorata	Mo	p (4 weeks)	NE Atlantic	238	COI	682	0.546	0.004	calculated from data in Díaz-Ferguson et al. (2010), Robinson et al. (2010), Reid et al. (2010)
Other highly diverse planktonic-dispersing marine invertebrates									
Glaucus atlanticus	Mo	p (lifelong)	Worldwide	112	COI	658	0.996	0.014	calculated from data in Churchill et al. (2013), Churchill et al. (2014), Wecker et al. (2015)

Species									
Pygospio elegans	An	p (4-5 weeks)	North Sea	23	COI	600	0.996	0.014	(Kesäniemi *et al.* 2012)
Argopecten irradians concentricus	Mo	p (5-19 days)	NW Atlantic	219	mtDNA	1025	0.982	0.008	(Marko & Barr 2007)
Brachidontes pharaonis	Mo	p (weeks)	Mediterranean-Red Sea	34	COI	618	0.973	0.039	(Terranova *et al.* 2007)
Ruditapes philippinarum	Mo	p (2-3 weeks)	NW Pacific	170	COI	644	0.960	0.010	(Mao *et al.* 2011)
Cellana sandwicensis	Mo	p (4 days)	Hawaii	109	COI	612	0.960	0.006	(Bird *et al.* 2007)
Holothuria nobilis	Ec	p (13-26 days)	Indo-Pacific	360	COI	559	0.942	0.008	(Uthicke & Benzie 2003)
Tridacna maxima	Mo	p (9 days)	Indo-Pacific	211	COI	484	0.940	0.023	(Nuryanto & Kochzius 2009)
Tridacna crocea	Mo	p (1 week)	Indo-Malaysia	300	COI	456	0.930	0.015	(Kochzius & Nuryanto 2008)
Pachygrapsus crassipes	Ar	p (95 days)	NE Pacific	346	COI	710	0.923	0.009	(Cassone & Boulding 2006)
Tripneustes gratilla	Ec	p (18 days)	Indo-Pacific	83	COI	573	0.902	0.004	calculated from data in Lessios et al. (2003)
Holothuria polii	Ec	p (13-26 days)	Mediterranean Sea	158	COI	484	0.873	0.005	(Vergara-Chen *et al.* 2010)
Nacella magellanica	Mo	p (unknown)	SW Atlantic	171	COI	573-650	0.868	0.004	(Aranzamendi *et al.* 2011)
Bursa fijiensis	Mo	p (8 weeks)	SW Pacific	59	COI	566	0.848	0.003	(Castelin *et al.* 2012)
Acropora cervicornis	Cn	p (4 days)	Caribbean	160	mtCR	941	0.847	0.006	(Vollmer & Palumbi 2007)

Other hyperdiverse mollusc species

Pliocardia kuroshimana	Mo	p	NW Pacific	3	mtDNA	513	1.000	0.256*	(James *et al.* 2016)
Bulinus forskalii	Mo	-	-	12	mtDNA	339	1.000	0.167*	(James *et al.* 2016)
Pyrgulopsis intermedia	Mo	d	-	15	mtDNA	528	0.924	0.148*	(James *et al.* 2016)
Euhadra brandtii	Mo	n/a	-	14	mtDNA	558	0.989	0.098*	(James *et al.* 2016)
Biomphalaria glabrata	Mo	d	-	7	mtDNA	579	0.714	0.092*	(James *et al.* 2016)
Achatinella mustelina	Mo	n/a	-	69	mtDNA	675	0.992	0.078*	(James *et al.* 2016)
Quincuncina infucata	Mo	d	-	5	mtDNA	453	1.000	0.067*	(James *et al.* 2016)
Pyrgulopsis thompsoni	Mo	d	-	7	mtDNA	657	0.952	0.066*	(James *et al.* 2016)

mtDNA divergence and population structuring in *Melarhaphe neritoides*

We investigated whether population genetic structure through time and space, and cryptic taxa, could contribute to the mtDNA hyperdiversity in *M. neritoides*. The monophyly of *M. neritoides* and the Gaussian distribution of its intraspecific p-distances, suggest that *M. neritoides* does not conceal cryptic taxa in the Azores. Conversely, the overwhelming number of private haplotypes (Fig. 6) at first glance suggests that populations are strongly differentiated because of the apparent lack of shared haplotypes. Yet, the bush-like mtDNA haplotype network (Fig. 6) is suggestive of complete population mixing (Nielsen & Slatkin 2013). Indeed, recurrent long-term gene flow homogenising the gene pool of *M. neritoides* over the 600 km between the Azorean islands implies an absence of population genetic structure (differentiation), as is reflected in the G_{ST}, N_{ST} and φ_{ST} values that are not significantly different from zero. This is congruent with the low level of differentiation and high potential for long range gene flow between Swedish and Cretan populations of *M. neritoides* (Johannesson 1992). Currently no other data on population genetic differentiation and gene flow in *M. neritoides* are available. The possibility of long-distance gene flow may suggest that the mtDNA diversity of *M. neritoides* in the Azores is the result of larval influx from European populations. Yet, while short-lived Pleistocene westward-flowing sea surface currents allowed the colonization of the Azores from Eastern Atlantic areas (Ávila *et al.* 2009), the eastward-flowing Azores Current nowadays (Barton 2001) suggests that larval transport predominantly occurs from the Azores towards the North East Atlantic coasts and the Mediterranean Sea, and that the Azores rather may act as a source of new, dispersing, haplotypes than as a sink receiving new haplotypes. Another hypothesis is that the large mtDNA diversity in the Azores is explained by the old age of the Azorean populations, which were located south of the maximum extension of the polar front that was at 42° N during the LGM and were

therefore not reduced by the glaciation (Dennielou *et al.* 2009). This hypothesis could be supported if a higher mtDNA diversity is observed in the Azores than in other European areas that were impacted by the LGM and are consequently younger populations (see Fig. 9 and references therein in the General Introduction of the thesis for the localisation of glacial refugia in Europe). However, no data are currently available to date to verify this hypothesis. Hence, all current evidence suggests that mtDNA hyperdiversity in *M. neritoides* is not due to (1) population structuring, (2) admixture of divergent local populations, (3) lumping of cryptic taxa, or (4) influx of new haplotypes from distant European populations.

mtDNA mutation rate in *Melarhaphe neritoides*

We investigated whether mtDNA mutation rate explains mtDNA hyperdiversity. The mutation rate is the rate at which new mutations arise in each generation of a species and accumulate per DNA sequence, and differs from the substitution rate that accounts for the fraction of new mutations that do not persist in the face of evolutionary forces (Barrick & Lenski 2013). Accordingly, neutral synonymous mutations reflect the mutation rate (Barrick & Lenski 2013). Mutation rates in most nuclear eukaryotic genomes are generally extremely low because elaborate molecular mechanisms correct errors in DNA replication and repair DNA damage, whereas viral and animal mitochondrial genomes have no, or far less efficient, repair mechanisms and thus have much higher mutation rates (Ballard & Whitlock 2004; Drake *et al.* 1998). Overall, synonymous mutations become fixed at a rate that appears to be uniform across various taxa (Kondrashov 2008), and mtDNA mutation rates lie in a narrow range of 10^{-8}-10^{-7} mutations per nucleotide site per generation across e.g. arthropods, echinoderms, chordates, molluscs and nematodes (Table 3). Surprisingly, our estimate of the mtDNA COI mutation rate in *M. neritoides* ($\mu =$

1.99 x 10^{-4} per site per generation) is 1000 to 10000-fold higher than commonly estimated for the mtDNA mutation rates in metazoans from these phyla. So, if our inference is correct, it seems likely that this high mtDNA mutation rate substantially contributed to generating the mtDNA hyperdiversity in *M. neritoides*. Our mutation rate estimate was obtained from mtDNA sequence data of *M. neritoides* itself, not from closely related species, and is therefore expected to be more accurate and species-specific. Bayesian MCMC estimates of substitution rates based on heterochronous mtDNA samples may be susceptible to an upward bias when populations have a complex demographic history (e.g. bottleneck) or pronounced population structure. Hence, such a biased estimates may reflect other processes like migration, selection and genetic drift rather than mutation (Navascués & Emerson 2009). However, this study did not provide evidence of population structure in *M. neritoides*, reducing therefore the risk of bias in the estimate of *μ*. Bayesian MCMC inferences based on heterochronous mtDNA samples over short timescales may also overestimate generational mutation rates by an order of magnitude in comparison to phylogenetically derived mutation rates, because they may account for short-lived, slightly deleterious mutations at non-synonymous sites (Ho *et al.* 2005; Penny 2005; Subramanian & Lambert 2011). Since *μ* in *M. neritoides* was estimated over a short period of 20 years, it may be subject to such a bias. However, while this bias could have generated an order of magnitude overestimation of *μ*, it cannot entirely account for the extreme value inferred, which is 10^3 to 10^4 fold higher than usually estimated for other organisms (Subramanian & Lambert 2011).

Invertebrates with shorter generation times have higher mtDNA mutation rates, as their mitochondrial genomes are copied more frequently (Thomas *et al.* 2010). In comparison to the generation times of invertebrates analyzed by Thomas et al. (2010), ranging from 8 days in the hydrozoan *Hydra magnipapillata* to 1825 days in the coral *Montastraea annularis* and the seastar

Pisaster ochraceus, the generation time of *M. neritoides* (τ ≈ 1250 days) is not particularly short and therefore its mtDNA mutation rate would be expected to be at the lower side. Yet, *M. neritoides* has a high mtDNA mutation rate ($\mu = 5.82 \times 10^{-5}$ per site per year) that does not fall within the range of mutation rates of these invertebrates with longer generation times than *M. neritoides*, i.e. from $\mu = 3 \times 10^{-10}$ per site per year in *Montastraea annularis* (Fukami & Knowlton 2005) to $\mu = 2.81 \times 10^{-6}$ per gene per year in *Pisaster ochraceus* (Popovic *et al.* 2014).

High mtDNA mutation rates may be more frequently linked to hyperdiversity than previously thought in the widely used COI marker. Indeed, neutral nucleotide diversities of ≥ 0.05 have been reported in 222 other species among Arthropoda, Chordata, Echinodermata, Mollusca and Nematoda (Table S1), suggesting the possibility of underlying high mtDNA mutation rates.

Table 3. **mtDNA mutation rates per site per generation in various metazoans ranked according to decreasing *μ*.** Ar, Arthropoda; Ch, Chordata; Ec, Echinodermata; Mo, Mollusca; Ne, Nematoda.

Species		μ	locus	Reference
Melarhaphe neritoides	Mo	1.99×10^{-4}	COI	this study
Homo sapiens sapiens	Ch	6.00×10^{-7}	mt genome	(Kivisild 2015)
Caenorhabditis elegans	Ne	1.60×10^{-7}	mt genome	(Denver *et al.* 2000)
Mytilus edulis	Mo	9.51×10^{-8}	COI	(Wares & Cunningham 2001)
Drosophila melanogaster	Ar	6.20×10^{-8}	mt genome	(Haag-Liautard *et al.* 2008)
Asteria rubens	Ec	4.84×10^{-8}	COI	(Wares & Cunningham 2001)
Nucella lapillus	Mo	4.43×10^{-8}	COI	(Wares & Cunningham 2001)
Euraphia spp.	Ar	3.80×10^{-8}	COI	(Wares & Cunningham 2001)
Idotea balthica	Ar	3.60×10^{-8}	COI	(Wares & Cunningham 2001)
Semibalanus balanoides	Ar	2.76×10^{-8}	COI	(Wares & Cunningham 2001)
Littorina obtusata	Mo	2.49×10^{-8}	COI	(Wares & Cunningham 2001)
Sesarma spp.	Ar	2.10×10^{-8}	COI	(Wares & Cunningham 2001)
Alpheus spp.	Ar	1.90×10^{-8}	COI	(Wares & Cunningham 2001)
Prochilodus spp.	Ch	0.27×10^{-8}	COI	(Turner *et al.* 2004)

Demography and selection in *Melarhaphe neritoides*

We investigated whether selection, mtDNA demographic history and N_e explain mtDNA hyperdiversity. Equilibrium between variation gained by mutations and variation lost by genetic drift should be reached if the effective population size has been stable over time and in absence of population structure or selection (Kimura 1983). According to the negative Tajima's D, Fu's Fs and Fay & Wu's H, the unimodal sequence mismatch distribution and the BSP trend, the phylogeny of Azorean *M. neritoides* has been shaped either by demographic expansion or selection, or a combination of both.

The effective mtDNA population size of *M. neritoides* estimated in this paper is $N_e \approx 5256$ (CI = 1312-37495) for the concatenated 16S-COI-Cyt*b* gene fragments. This is relatively small in comparison to mtDNA N_e of other littorinids with planktonic larval stages and high dispersal potential like *Littorina plena* (N_e = 160526 – 33728571) and *Littorina scutulata* (N_e = 90790 – 3814286) (Table 4), except for the mtDNA N_e in the planktonic dispersing *Littorina keenae* (N_e = 135) (Lee & Boulding 2007). Yet, this latter value refers to one sampling site only, whereas another sampling site of *Littorina keenae* showed a much larger mtDNA N_e (N_e = 31797). Surprisingly, and somewhat counterintuitively, the mtDNA of *M. neritoides* is also smaller than that of periwinkles without planktonic larval stages, such as *Littorina sitkana* (N_e = 105263 – 1400000) and *Littorina subrotundata* (N_e = 25000 – 1942857) (Lee & Boulding 2009). However, past putative selection in *M. neritoides* likely confounds the BSP inference by reducing the overall mtDNA diversity and thus the mtDNA N_e estimate. As such, mtDNA variation in *M. neritoides* is still remarkably high, despite this signal of a reduction of its diversity by selection. This strengthens the hypothesis that the mtDNA hyperdiversity in *M. neritoides* is best explained by a high μ of mtDNA.

mtDNA N_e and mtDNA hyperdiversity may be positively correlated such as in the lined shore crab *Pachygrapsus crassipes* (N_e = 167000 to 1020000; COI *Hd* = 0.923; *π* = 0.009) (Cassone & Boulding 2006). Yet, this relationship has been questioned (Bazin *et al.* 2006; Piganeau & Eyre-Walker 2009), because Bazin et al. (2006) showed that mtDNA diversity is not linked to mtDNA N_e, but rather to *µ* and selection. Conversely, Nabholz et al. (2009; 2008) found no link between selection and mtDNA N_e, but confirmed that mtDNA diversity is strongly linked to *µ*. Our present work shows a link between mtDNA hyperdiversity and high mtDNA *µ*, and the putative influence of selection on N_e estimation making mtDNA N_e a poor indicator of mtDNA hyperdiversity.

Table 4. **mtDNA effective population sizes (N_e) for various taxa.** The 95% confidence interval is given in parenthesis when available. Ar, Arthropoda; Ch, Chordata; Mo, Mollusca.

Taxon		N_e	locus	Reference
Littorina keenae	Mo	135 (42-2490)	ND6-Cyt*b*	(Lee & Boulding 2007)
Melarhaphe neritoides	Mo	5256 (1312-37495)	COI-16S-Cyt*b*	this study
Homo & *Pan*	Ch	5900 – 10 000	mt genome	(Piganeau & Eyre-Walker 2009)
Felidae & Canidae	Ch	130 000 – 430 000	mt genome	(Piganeau & Eyre-Walker 2009)
Pachygrapsus crassipes	Ar	167 000 – 1 020 000	COI	(Cassone & Boulding 2006)
Cardinalis cardinalis	Ch	193 000 (4000-701000)	ND2-Cyt*b*	(Smith & Klicka 2013)
Murinae	Ch	230 000 – 730 000	mt genome	(Piganeau & Eyre-Walker 2009)
Littorina sitkana	Mo	105 263 – 1 400 000	Cyt*b*	(Lee & Boulding 2009)
Littorina subrotundata	Mo	25 000 – 1 942 857	Cyt*b*	(Lee & Boulding 2009)
Littorina scutulata	Mo	90 790 – 3 814 286	Cyt*b*	(Lee & Boulding 2009)
Littorina plena	Mo	160 526 – 33 728 571	Cyt*b*	(Lee & Boulding 2009)

CONCLUSIONS

The mtDNA hyperdiversity of *M. neritoides* is characterized by a high haplotype diversity (Hd = 0.999 ± 0.001), a high nucleotide diversity (π = 0.013 ± 0.001) and a high neutral nucleotide diversity (π_{neu} = 0.0678) for the concatenated 16S-COI-Cyt*b* gene fragments. The mutation rate at the COI locus is μ = 1.99 x 10^{-4} mutations per nucleotide site per generation, which is a very high value. Demographic analyses revealed that *M. neritoides* in the Azores underwent a population expansion, but the effective population size N_e was surprisingly small for a planktonic-developing species (N_e = 5256; CI = 1312-37495) probably due to the putative influence of selection on *M. neritoides* mtDNA. As a result, N_e is not linked to mtDNA hyperdiversity and is a poor indicator of this latter. Mitochondrial DNA hyperdiversity is best explained by a high mtDNA μ in *M. neritoides*. Mitochondrial DNA hyperdiversity may be more common across eukaryotes than currently known.

DNA Deposition

The following information was supplied regarding the deposition of DNA

sequences: new sequences are available from Genbank with accession numbers from KT996151 to KT997344.

Data Availability

The following information was supplied regarding data availability:
Dataset 1: https://dx.doi.org/10.6084/m9.figshare.3437333.v4;
Dataset 2 (16S): https://dx.doi.org/10.6084/m9.figshare.3437390.v2;
Dataset 2 (COI): https://dx.doi.org/10.6084/m9.figshare.3437474.v2;
Dataset 2 (16S_COI): https://dx.doi.org/10.6084/m9.figshare.3437489.v2;
Dataset 3: https://dx.doi.org/10.6084/m9.figshare.3437498.v2;
Dataset 4: https://dx.doi.org/10.6084/m9.figshare.3437531.v2.

Supplemental Information

Table S1. **List of 215 animal species with hyperdiverse DNA ($\pi_{syn} > 0.05$).**

Genus	species	phylum	DNA	π_{syn}	references
Martes	*flavigula*	Ch	mitochondrial	0.3527	James et al. 2016
Petaurista	*alborufus*	Ch	mitochondrial	0.3226	James et al. 2016
Etheostoma	*virgatum*	Ch	mitochondrial	0.2859	James et al. 2016
Liparthrum	*pilosum*	Ar	mitochondrial	0.2633	James et al. 2016
Andropadus	*tephrolaemus*	Ch	mitochondrial	0.2611	James et al. 2016
Vesicomya	*kuroshimana*	Mo	mitochondrial	0.2562	James et al. 2016
Podarcis	*hispanica*	Ch	mitochondrial	0.2429	James et al. 2016
Nectarinia	*mediocris*	Ch	mitochondrial	0.2392	James et al. 2016
Notropis	*sabinae*	Ch	mitochondrial	0.2266	James et al. 2016
Eurycea	*multiplicata*	Ch	mitochondrial	0.2242	James et al. 2016
Eliurus	*webbi*	Ch	mitochondrial	0.2215	James et al. 2016
Lacerta	*tangitana*	Ch	mitochondrial	0.2189	James et al. 2016
Lefua	*echigonia*	Ch	mitochondrial	0.2176	James et al. 2016
Spermophilus	*citellus*	Ch	mitochondrial	0.2143	James et al. 2016
Scartomyzon	*congestus*	Ch	mitochondrial	0.2133	James et al. 2016
Anolis	*marmoratus*	Ch	mitochondrial	0.2076	James et al. 2016
Tarentola	*boettgeri*	Ch	mitochondrial	0.2055	James et al. 2016

Pogonomyrmex	*rugosus*	Ar	mitochondrial	0.2051	James et al. 2016
Proctoporus	*unsaacae*	Ch	mitochondrial	0.204	James et al. 2016
Gymnogobius	*castaneus*	Ch	mitochondrial	0.2036	James et al. 2016
Carlia	*rubrigularis*	Ch	mitochondrial	0.2011	James et al. 2016
Podarcis	*vaucheri*	Ch	mitochondrial	0.2001	James et al. 2016
Spermophilus	*townsendi*	Ch	mitochondrial	0.1992	James et al. 2016
Lonchophylla	*thomasi*	Ch	mitochondrial	0.1944	James et al. 2016
Carollia	*castanea*	Ch	mitochondrial	0.189	James et al. 2016
Eurycea	*spelaeus*	Ch	mitochondrial	0.1885	James et al. 2016
Neotoma	*albigula*	Ch	mitochondrial	0.1871	James et al. 2016
Cratogeomys	*tylorhinus*	Ch	mitochondrial	0.1855	James et al. 2016
Semnopithecus	*entellus*	Ch	mitochondrial	0.1816	James et al. 2016
Cratogeomys	*gymnurus*	Ch	mitochondrial	0.1812	James et al. 2016
Astyanax	*fasciatus*	Ch	mitochondrial	0.179	James et al. 2016
Tangara	*gyrola*	Ch	mitochondrial	0.178	James et al. 2016
Cratogeomys	*fumosus*	Ch	mitochondrial	0.1764	James et al. 2016
Pogonomyrmex	*barbatus*	Ar	mitochondrial	0.1761	James et al. 2016
Lepidiota	*negatoria*	Ar	mitochondrial	0.1742	James et al. 2016
Philoria	*loveridgei*	Ch	mitochondrial	0.1742	James et al. 2016
Procavia	*capensis*	Ch	mitochondrial	0.1739	James et al. 2016
Myotis	*blythii*	Ch	mitochondrial	0.1736	James et al. 2016
Mabuya	*fogoensis*	Ch	mitochondrial	0.1711	James et al. 2016
Icterus	*dominicensis*	Ch	mitochondrial	0.17	James et al. 2016
Cherax	*quinquecarinatus*	Ar	mitochondrial	0.1697	James et al. 2016
Eliurus	*minor*	Ch	mitochondrial	0.1689	James et al. 2016
Bufo	*punctatus*	Ch	mitochondrial	0.1681	James et al. 2016
Bulinus	*forskalii*	Mo	mitochondrial	0.1668	James et al. 2016
Lepus	*oiostolus*	Ch	mitochondrial	0.1666	James et al. 2016
Anolis	*punctatus*	Ch	mitochondrial	0.1631	James et al. 2016
Cratogeomys	*merriami*	Ch	mitochondrial	0.1623	James et al. 2016
Celatoblatta	*montana*	Ar	mitochondrial	0.1615	James et al. 2016
Aphanarthrum	*bicolor*	Ar	mitochondrial	0.1612	James et al. 2016
Hypseleotris	*klunzingeri*	Ch	mitochondrial	0.1572	James et al. 2016
Proctoporus	*guentheri*	Ch	mitochondrial	0.1548	James et al. 2016
Hypocnemis	*cantator*	Ch	mitochondrial	0.1537	James et al. 2016
Glossophaga	*soricina*	Ch	mitochondrial	0.1517	James et al. 2016
Phrynocephalus	*vlangalii*	Ch	mitochondrial	0.1509	James et al. 2016
Solea	*senegalensis*	Ch	mitochondrial	0.15	James et al. 2016
Pyrgulopsis	*intermedia*	Mo	mitochondrial	0.1481	James et al. 2016
Tarentola	*darwini*	Ch	mitochondrial	0.1462	James et al. 2016
Xiphorhynchus	*ocellatus*	Ch	mitochondrial	0.1459	James et al. 2016
Baiomys	*musculus*	Ch	mitochondrial	0.1457	James et al. 2016
Cacicus	*uropygialis*	Ch	mitochondrial	0.1446	James et al. 2016
Lachesis	*muta*	Ch	mitochondrial	0.1441	James et al. 2016

Neochlamisus	*bebbianae*	Ar	mitochondrial	0.1435	James et al. 2016
Mabuya	*spinalis*	Ch	mitochondrial	0.1415	James et al. 2016
Caenorhabditis	*brenneri*	Ne	nuclear	0.141	Dey et al. 2013
Tarentola	*caboverdianus*	Ch	mitochondrial	0.1395	James et al. 2016
Noturus	*gyrinus*	Ch	mitochondrial	0.1392	James et al. 2016
Maoricicada	*mangu*	Ar	mitochondrial	0.1391	James et al. 2016
Thamnophis	*elegans*	Ch	mitochondrial	0.1389	James et al. 2016
Neochlamisus	*platani*	Ar	mitochondrial	0.1385	James et al. 2016
Eurycea	*tynerensis*	Ch	mitochondrial	0.1383	James et al. 2016
Darevskia	*raddei*	Ch	mitochondrial	0.1382	James et al. 2016
Caenorhabditis	*sp 5*	Ne	mitochondrial	0.138	Cutter et al. 2012
Spermophilus	*erythrogenys*	Ch	mitochondrial	0.1373	James et al. 2016
Timema	*podura*	Ar	mitochondrial	0.1367	James et al. 2016
Aegotheles	*bennettii*	Ch	mitochondrial	0.1366	James et al. 2016
Coscinasterias	*muricata*	Ec	mitochondrial	0.1362	James et al. 2016
Gazella	*subgutturosa*	Ch	mitochondrial	0.1346	James et al. 2016
Roeboides	*occidentalis*	Ch	mitochondrial	0.1338	James et al. 2016
Ascaphus	*truei*	Ch	mitochondrial	0.1323	James et al. 2016
Bacillus	*grandii*	Ar	mitochondrial	0.1319	James et al. 2016
Aphanarthrum	*piscatorium*	Ar	mitochondrial	0.1318	James et al. 2016
Gymnogobius	*taranetzi*	Ch	mitochondrial	0.1287	James et al. 2016
Mogera	*wogura*	Ch	mitochondrial	0.1266	James et al. 2016
Ciona	*savignyi*	Ch	nuclear	0.126	Small et al. 2007
Nothopsyche	*ruficollis*	Ar	mitochondrial	0.1252	James et al. 2016
Microtus	*guentheri*	Ch	mitochondrial	0.1247	James et al. 2016
Pongo	*pygmaeus*	Ch	mitochondrial	0.121	James et al. 2016
Andropadus	*masukuensis*	Ch	mitochondrial	0.1204	James et al. 2016
Tanakia	*lanceolata*	Ch	mitochondrial	0.1204	James et al. 2016
Lacerta	*lepida*	Ch	mitochondrial	0.1188	James et al. 2016
Microtus	*savii*	Ch	mitochondrial	0.1176	James et al. 2016
Glyphorynchus	*spirurus*	Ch	mitochondrial	0.1164	James et al. 2016
Trinomys	*gratiosus*	Ch	mitochondrial	0.1162	James et al. 2016
Tanakia	*limbata*	Ch	mitochondrial	0.1161	James et al. 2016
Pseudotylosurus	*angusticeps*	Ch	mitochondrial	0.1157	James et al. 2016
Eothenomys	*smithii*	Ch	mitochondrial	0.1152	James et al. 2016
Plethodon	*elongatus*	Ch	mitochondrial	0.1149	James et al. 2016
Salamandra	*algira*	Ch	mitochondrial	0.114	James et al. 2016
Pituophis	*catenifer*	Ch	mitochondrial	0.1128	James et al. 2016
Palpopleura	*portia*	Ar	mitochondrial	0.112	James et al. 2016
Microlophus	*albemarlensis*	Ch	mitochondrial	0.111	James et al. 2016
Anolis	*oculatus*	Ch	mitochondrial	0.1087	James et al. 2016
Charina	*bottae*	Ch	mitochondrial	0.1078	James et al. 2016
Aerodramus	*spodiopygius*	Ch	mitochondrial	0.1069	James et al. 2016
Troglodytes	*troglodytes*	Ch	mitochondrial	0.1057	James et al. 2016

Apodemus	*peninsulae*	Ch	mitochondrial	0.1056	James et al. 2016
Peromyscus	*merriami*	Ch	mitochondrial	0.1056	James et al. 2016
Sabanejewia	*aurata*	Ch	mitochondrial	0.1045	James et al. 2016
Crotalus	*oreganus*	Ch	mitochondrial	0.1039	James et al. 2016
Microtus	*agrestis*	Ch	mitochondrial	0.1039	James et al. 2016
Podarcis	*lilfordi*	Ch	mitochondrial	0.1008	James et al. 2016
Microcebus	*rufus*	Ch	mitochondrial	0.1002	James et al. 2016
Cherax	*tenuimanus*	Ar	mitochondrial	0.0996	James et al. 2016
Xiphorhynchus	*spixii*	Ch	mitochondrial	0.0996	James et al. 2016
Anolis	*fuscoauratus*	Ch	mitochondrial	0.0994	James et al. 2016
Holbrookia	*maculata*	Ch	mitochondrial	0.0992	James et al. 2016
Euhadra	*brandtii*	Mo	mitochondrial	0.098	James et al. 2016
Cherax	*destructor*	Ar	mitochondrial	0.0975	James et al. 2016
Cottus	*bairdii*	Ch	mitochondrial	0.0969	James et al. 2016
Thrichomys	*apereoides*	Ch	mitochondrial	0.0959	James et al. 2016
Lepus	*mandshuricus*	Ch	mitochondrial	0.0952	James et al. 2016
Trichosurus	*caninus*	Ch	mitochondrial	0.0928	James et al. 2016
Lacerta	*schreiberi*	Ch	mitochondrial	0.0925	James et al. 2016
Peromyscus	*boylii*	Ch	mitochondrial	0.0924	James et al. 2016
Leptynia	*attenuata*	Ar	mitochondrial	0.0922	James et al. 2016
Petrochromis	*polyodon*	Ch	mitochondrial	0.0919	James et al. 2016
Biomphalaria	*glabrata*	Mo	mitochondrial	0.0917	James et al. 2016
Nectarinia	*notata*	Ch	mitochondrial	0.0891	James et al. 2016
Blarina	*carolinensis*	Ch	mitochondrial	0.089	James et al. 2016
Crematogaster	*captiosa*	Ar	mitochondrial	0.0887	James et al. 2016
Plethodon	*stormi*	Ch	mitochondrial	0.0887	James et al. 2016
Xiphorhynchus	*guttatus*	Ch	mitochondrial	0.0883	James et al. 2016
Crocidura	*dsinezumi*	Ch	mitochondrial	0.088	James et al. 2016
Ctenomys	*boliviensis*	Ch	mitochondrial	0.0866	James et al. 2016
Lacerta	*nairensis*	Ch	mitochondrial	0.0861	James et al. 2016
Gorilla	*gorilla*	Ch	mitochondrial	0.0858	James et al. 2016
Reticulitermes	*lucifugus*	Ar	mitochondrial	0.0857	James et al. 2016
Carollia	*brevicauda*	Ch	mitochondrial	0.0854	James et al. 2016
Peromyscus	*eremicus*	Ch	mitochondrial	0.0852	James et al. 2016
Mantella	*aurantiaca*	Ch	mitochondrial	0.0851	James et al. 2016
Pelodytes	*punctatus*	Ch	mitochondrial	0.0849	James et al. 2016
Rana	*virgatipes*	Ch	mitochondrial	0.0847	James et al. 2016
Bufo	*rangeri*	Ch	mitochondrial	0.0835	James et al. 2016
Cacicus	*cela*	Ch	mitochondrial	0.0835	James et al. 2016
Sigmodon	*toltecus*	Ch	mitochondrial	0.0832	James et al. 2016
Bufo	*terrestris*	Ch	mitochondrial	0.0829	James et al. 2016
Todus	*angustirostris*	Ch	mitochondrial	0.0829	James et al. 2016
Trinomys	*dimidiatus*	Ch	mitochondrial	0.0823	James et al. 2016
Strongylocentrotus	*droebachiensis*	Ec	mitochondrial	0.082	James et al. 2016

Stenonema	*vicarium*	Ar	mitochondrial	0.0816	James et al. 2016
Parus	*caeruleus*	Ch	mitochondrial	0.0815	James et al. 2016
Phrynocephalus	*frontalis*	Ch	mitochondrial	0.0804	James et al. 2016
Micrurus	*altirostris*	Ch	mitochondrial	0.0802	James et al. 2016
Caenorhabditis	*remanei*	Ne	nuclear	0.078	Cutter et al. 2006
					Jovelin et al. 2009
					Dey et al. 2012
Achatinella	*mustelina*	Mo	mitochondrial	0.0778	James et al. 2016
Nectarinia	*humbloti*	Ch	mitochondrial	0.0766	James et al. 2016
Xiphorhynchus	*picus*	Ch	mitochondrial	0.0756	James et al. 2016
Maoricicada	*campbelli*	Ar	mitochondrial	0.0754	James et al. 2016
Rana	*chensinensis*	Ch	mitochondrial	0.0746	James et al. 2016
Leptynia	*caprai*	Ar	mitochondrial	0.0743	James et al. 2016
Leptura	*modicenotata*	Ar	mitochondrial	0.0738	James et al. 2016
Neochlamisus	*bimaculatus*	Ar	mitochondrial	0.0732	James et al. 2016
Ameiva	*chrysolaema*	Ch	mitochondrial	0.0732	James et al. 2016
Poeciliopsis	*infans*	Ch	mitochondrial	0.073	James et al. 2016
Aegla	*jarai*	Ar	mitochondrial	0.0729	James et al. 2016
Uma	*notata*	Ch	mitochondrial	0.0719	James et al. 2016
Aerodramus	*salangana*	Ch	mitochondrial	0.0717	James et al. 2016
Varroa	*jacobsoni*	Ar	mitochondrial	0.0713	James et al. 2016
Aegla	*grisella*	Ar	mitochondrial	0.071	James et al. 2016
Myrmecocystus	*mimicus*	Ar	mitochondrial	0.0709	James et al. 2016
Macaca	*tonkeana*	Ch	mitochondrial	0.0708	James et al. 2016
Drosophila	*innubila*	Ar	nuclear	0.0699	Dyer & Jaenike 2004
Chelonibia	*testudinaria*	Ar	mitochondrial	0.0699	James et al. 2016
Nectarinia	*sovimanga*	Ch	mitochondrial	0.0699	James et al. 2016
Diadema	*antillarum*	Ec	mitochondrial	0.0699	James et al. 2016
Auxis	*rochei*	Ch	mitochondrial	0.069	James et al. 2016
Hemideina	*maori*	Ar	mitochondrial	0.0686	James et al. 2016
Diadema	*paucispinum*	Ec	mitochondrial	0.0679	James et al. 2016
Neomys	*anomalus*	Ch	mitochondrial	0.0677	James et al. 2016
Phrynocephalus	*theobaldi*	Ch	mitochondrial	0.0672	James et al. 2016
Quincuncina	*infucata*	Mo	mitochondrial	0.0667	James et al. 2016
Pyrgulopsis	*thompsoni*	Mo	mitochondrial	0.0664	James et al. 2016
Sorex	*preblei*	Ch	mitochondrial	0.0656	James et al. 2016
Sarotherodon	*melanotheron*	Ch	mitochondrial	0.0644	James et al. 2016
Spermophilus	*parryii*	Ch	mitochondrial	0.0644	James et al. 2016
Oxymycterus	*dasytrichus*	Ch	mitochondrial	0.0643	James et al. 2016
Ochotona	*cansus*	Ch	mitochondrial	0.064	James et al. 2016
Aphelocoma	*californica*	Ch	mitochondrial	0.0636	James et al. 2016
Psittacus	*erithacus*	Ch	mitochondrial	0.0632	James et al. 2016
Euxinia	*maeoticus*	Ar	mitochondrial	0.0629	James et al. 2016
Timema	*cristinae*	Ar	mitochondrial	0.0627	James et al. 2016

Gazella	*gazella*	Ch	mitochondrial	0.0625	James et al. 2016
Glossophaga	*commissarisi*	Ch	mitochondrial	0.0618	James et al. 2016
Aegotheles	*wallacii*	Ch	mitochondrial	0.0617	James et al. 2016
Laudakia	*caucasia*	Ch	mitochondrial	0.0614	James et al. 2016
Garrulax	*canorus*	Ch	mitochondrial	0.0602	James et al. 2016
Dendroica	*adelaidae*	Ch	mitochondrial	0.0593	James et al. 2016
Baiomys	*taylori*	Ch	mitochondrial	0.0591	James et al. 2016
Drosophila	*recens*	Ar	nuclear	0.059	Dyer et al. 2007
Plethodon	*richmondi*	Ch	mitochondrial	0.0583	James et al. 2016
Alytes	*obstetricans*	Ch	mitochondrial	0.0577	James et al. 2016
Proechimys	*cuvieri*	Ch	mitochondrial	0.0572	James et al. 2016
Cyclura	*nubila*	Ch	mitochondrial	0.0571	James et al. 2016
Sorex	*minutissimus*	Ch	mitochondrial	0.0568	James et al. 2016
Panthera	*pardus*	Ch	mitochondrial	0.0549	James et al. 2016
Microcebus	*murinus*	Ch	mitochondrial	0.0541	James et al. 2016
Neomicropteryx	*matsumurana*	Ar	mitochondrial	0.054	James et al. 2016
Artibeus	*jamaicensis*	Ch	mitochondrial	0.0532	James et al. 2016
Carollia	*sowelli*	Ch	mitochondrial	0.0531	James et al. 2016
Carollia	*perspicillata*	Ch	mitochondrial	0.053	James et al. 2016
Sorex	*thompsoni*	Ch	mitochondrial	0.0518	James et al. 2016
Calomys	*musculinus*	Ch	mitochondrial	0.0512	James et al. 2016
Uma	*scoparia*	Ch	mitochondrial	0.0507	James et al. 2016
Blarina	*brevicauda*	Ch	mitochondrial	0.0506	James et al. 2016
Moschus	*chrysogaster*	Ch	mitochondrial	0.05	James et al. 2016

Ar, Arthropoda; Ch, Chordata; Ec, Echinodermata; Mo, Mollusca; Ne, Nematoda.

Table S2. **Specimens samples and datasets used in this study.**

Locality	N	sampling date	N1	N2	N3	N4
FAI	43	06/28/1993	42		43	43
FAI	46	07/06/2012			46	
FLO	42	1992	39		42	42
FLO	45	07/10/2012			45	
MOS	223	06/29/2012		223		
PIC	45	10/14/1993	37		45	45
PIC	43	07/04/2012			43	
SMA	43	04/17/1996	32			43
SMI	39	07/31/1993	35		39	39
SMI	41	06/28/2012			41	
Total	610		185	223	344	212

N, number of individuals; $N1$, $N2$, $N3$, $N4$, number of individuals used in dataset 1, dataset 2, dataset 3 and dataset 4 respectively.

Table S3. **Estimates of average evolutionary divergence over COI sequence pairs within and between groups.**

Species	n	d_1	d_2 [1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
[1] *Bembicium auratum*	1	n/c		0.017	0.015	0.016	0.018	0.017	0.018	0.016	0.017
[2] *Cremnoconchus syhadrensis*	1	n/c	0.196		0.017	0.017	0.018	0.017	0.018	0.016	0.018
[3] *Lacuna pallidula*	2	0.000 ± 0.000	0.166	0.212		0.016	0.017	0.017	0.017	0.016	0.017
[4] *Laevilitorina caliginosa*	1	n/c	0.166	0.190	0.174		0.018	0.017	0.018	0.016	0.017
[5] *Littorina littorea*	19	0.004 ± 0.001	0.224	0.228	0.212	0.219		0.017	0.016	0.017	0.019
[7] *Peasiella isseli*	1	n/c	0.212	0.230	0.200	0.198	0.214		0.018	0.017	0.018
[9] *Tectarius striatus*	53	0.006 ± 0.001	0.202	0.246	0.185	0.222	0.187	0.225		0.017	0.019
[6] *Melarhaphe neritoides*	213	0.018 ± 0.002	0.181	0.201	0.172	0.176	0.223	0.204	0.212		0.017
[8] *Pomatias elegans*	8	0.009 ± 0.002	0.209	0.244	0.199	0.203	0.254	0.231	0.271	0.228	

n, number of sequences used; d_1, number of base differences per site (p-distance) from averaging over all sequence pairs within each group ± standard error; n/c, cases in which it was not possible to estimate evolutionary distances; d_2, number of base differences per site (p-distance) from averaging over all sequence pairs between groups (under diagonal) and standard error estimates (above diagonal).

Chapter 2

Implications of mtDNA hyperdiversity for assessing population genetic differentiation and connectivity: deceiving haplotypic differentiation in a panmictic periwinkle in the North East Atlantic

Séverine Fourdrilis [1]
Patrick Mardulyn [2]
Thierry Backeljau [1,3]

[1] Directorate Taxonomy and Phylogeny & JEMU, Royal Belgian Institute of Natural Sciences, Brussels, Belgium

[2] Evolutionary Biology and Ecology, Université Libre de Bruxelles, Brussels, Belgium

[3] Evolutionary Ecology Group, University of Antwerp, Antwerp, Belgium

ABSTRACT

Mitochondrial DNA usually shows moderate to substantial amounts of variation in natural populations. Yet, recent studies increasingly uncover cases of hyperdiverse mtDNA with $\pi_{syn} \geq 5$ %, suggesting that this phenomenon is more common than currently appreciated. At first sight, such mtDNA hyperdiversity can easily lead to erroneous interpretations with respect to population genetic differentiation and misuse of D_{EST} to assess fixation in place of haplotypic differentiation. We illustrate this by using hyperdiverse mtDNA markers to assess population genetic differentiation and connectivity in *Melarhaphe neritoides*, a gastropod with high dispersal potential, in the NE Atlantic (NEA). We surveyed mtDNA variation at the combined 16S-COI-Cyt*b* loci in 399 specimens throughout the NEA. Nearly all specimens (except 13) had haplotypes private to populations. Most populations did not share any haplotype, suggesting at first glance a lack of gene flow and thus a strong population genetic differentiation. Yet, the haplotype network showed no signs of phylogeographic or other haplotype structuring, a pattern typical of high ongoing gene flow. Hence, no significant genetic structure was observed over the NEA. Gene flow estimates revealed high rates of genetic connectivity, predominantly eastward, throughout the NEA. As such, *M. neritoides* seems to be panmictic over the entire NEA and the apparent lack of shared mtDNA haplotypes among populations is not due to a lack of gene flow, but is caused by (1) a very high mutation rate that conceals the signal of gene flow and/or (2) a too low sampling effort to detect shared haplotypes.

INTRODUCTION

Larval dispersal refers to the physical intergenerational spread of larvae away from a source to a destination or settlement site, and is characterized by rates

of exchange of migrants called connectivity (Pineda *et al.* 2007). Connectivity plays a fundamental role in population dynamics, genetic structure and diversity, and shapes demographic stability and resilience of populations over time. The study of connectivity addresses questions like "how far do larvae disperse ?" and "how many larvae disperse ?". Estimating connectivity therefore provides important information about the spatial scale at which populations are connected via dispersal of early life stages (Cowen *et al.* 2007). Direct measurement of larval dispersal (e.g. tracking of tagged organisms) is limited to ecological time-scales and constrained by the difficulty of monitoring larvae – tiny propagules spreading in an open and wide marine medium (Kool *et al.* 2013). An alternative approach is the indirect measurement of larval dispersal via genetic markers (Hedgecock *et al.* 2007). Genetic markers provide a powerful tool to assess connectivity, by sampling adult specimens instead of directly monitoring larvae, and by using individual genotypic data to estimate population genetic parameters allowing to estimate genetic exchange between populations (Kool *et al.* 2013). Thus, genetic markers provide estimates of gene flow among populations, beyond ecological time-scales, by integrating gene flow over many generations at evolutionary time-scales. Hence, indirect gene flow estimates are more than an average picture of current gene flow, they are the result of the cumulative effect of gene flow over spatiotemporal scales.

Planktonic dispersers are marine organisms that essentially disperse as planktonic larvae during early life stages and that subsequently become sedentary after settlement. In this way they can be contrasted with organisms that have no dispersing planktonic larval stage (direct developers) and organisms that are capable of lifelong dispersal. If such planktonic dispersers have a long pelagic larval duration (PLD), then they are expected to have the potential for long-distance dispersal and high rates of gene flow, and consequently to show little, if any, population genetic differentiation (F_{ST}) even

over thousands of kilometres (Kyle & Boulding 2000; Shanks 2009). Yet, at least three issues may blur these paradigmatic expectations: (1) The correlation between PLD and F_{ST} may be poor in reef species ($r^2 = 0.29$) due to larval self-recruitment and oceanographic barriers (Selkoe & Toonen 2011; Treml *et al.* 2012; Weersing & Toonen 2009), but also and more surprisingly, in non-reef species from open coastal habitats owing to the coastal boundary layer retaining larvae near the shoreline (Hameed *et al.* 2016; Nickols *et al.* 2015); (2) The correlation between PLD and F_{ST} may be biased by errors in the estimation of F_{ST}, non-equilibrium F_{ST} values, and variation in effective population size (N_e) (Faurby & Barber 2012); (3) Although genetic markers must be sufficiently variable to detect population genetic structuring and to estimate gene flow, too variable genetic markers may conceal gene flow by deceivingly suggesting population genetic differentiation, even in the presence of long PLD and high dispersal potential, because high mutation rates may provoke a shortfall of shared haplotypes among populations and/or require unrealistic sample sizes to detect shared haplotypes (Fourdrilis *et al.* 2016). Hence the expected relationship between long PLD and the capacity to maintain high connectivity among distant populations is not straightforward. Accordingly, more quantitative studies of population genetic connectivity are needed for planktonic-dispersing species, and such studies should rely on gene genealogy-based methods rather than on F_{ST}-based methods that may be biased if markers are too polymorphic (Charlesworth 1998; Jost 2008; Kuhner 2008; Wang 2012; Whitlock 2011; Whitlock & McCauley 1999; Wright 1978).

In the North East Atlantic (NEA), including the Mediterranean Sea, genetic connectivity in planktonic dispersers is still poorly documented (Marti-Puig *et al.* 2013). Very few studies provide estimates of genetic connectivity of planktonic dispersers over their entire distribution in the NEA (exceptions are: Fratini *et al.* 2016; Penant *et al.* 2013; Quesada *et al.* 1998; Santos *et al.* 2012; So *et al.*

2011; Van den Broeck *et al.* 2008; Wilke & Pfenninger 2002), and particularly rocky intertidal communities are understudied (Thompson *et al.* 2002). Quantitative studies of genetic connectivity are therefore needed to assess the amount of gene flow and the spatial range of larval dispersal for planktonic dispersers throughout NEA. Therefore, in this paper, we investigate macrogeographic population genetic connectivity in the intertidal periwinkle *Melarhaphe neritoides* (Linnaeus, 1758) (Gastropoda: Littorinidae), a rocky shore species in the NEA with a long pelagic larval stage lasting 4 to 8 weeks (Fretter & Manly 1977). As such, *M. neritoides* (1) has a potential for long-distance dispersal and high rate of gene flow (Johannesson 1992), and (2) is likely to show no or very little genetic differentiation among geographically distant populations (Johannesson 1992). The species is distributed in the NEA from South Norway to the Canary Islands i.e. over approx. 4000 km, and even up to 5500 km if the Cape Verde Islands are included (Lewis & Tambs-Lyche 1962; Rolán & Groh 2005; Rosewater 1981), and over a West-East beeline distance of 6000 km from the Azores in the Atlantic to Lebanon in the eastern Mediterranean and into the Black Sea (Cordeiro *et al.* 2015; Öztürk *et al.* 2014; Ramos-Esplá *et al.* 2014). Distribution data can be obtained from the Ocean Biogeographic Information System (OBIS, Grassle 2000) at http://iobis.org/mapper/?taxon=Melarhaphe%20neritoides. *Melarhaphe neritoides* shows hyperdiverse mtDNA with an extremely high haplotype diversity (Hd = 0.999 ± 0.001) and a high neutral nucleotide diversity (π_{syn} = 6.8 %) for 16S, COI and Cyt*b* in the Azores (Fourdrilis *et al.* 2016), and for COI (Hd = 0.998; π_{syn} = 7.6 %) at the Galician coast (García *et al.* 2013).

The present study asks whether a marine gastropod with long-lived planktonic-dispersing stage like *M. neritoides* maintains, as expected, long-distance genetic connectivity throughout the NEA and at which rate. It does so by: (1) assessing differentiation among populations using mtDNA markers to explore whether Johannesson's (1992) allozyme data reflect a real pattern of

homogeneity and long-distance gene flow, (2) assessing mtDNA differentiation among populations at several spatial scales within the range 1-6000 km to test for panmixis throughout the NEA, (3) comparing several scenarios of gene flow among three oceanographic areas in the distribution range of *M. neritoides*, *viz.* the Azores, the NEA coast and the Mediterranean Sea, and quantifying gene flow among the three oceanographic areas. As the mtDNA of *M. neritoides* is hyperdiverse (Fourdrilis *et al.* 2016), we also use this study to provide a real-life illustration of how mtDNA hyperdiversity can deceivingly suggest strong population genetic differentiation in the presence of intense gene flow.

MATERIALS AND METHODS

Sample collection and DNA sequencing

We obtained 407 specimens of *M. neritoides* from 12 localities (hereafter referred to as "populations") throughout the species' distribution range in the NEA (Fig. 1, Table S1). Figure 1 was created using the open source geographic information system QGIS 2.8.8 (QGIS Development Team 2004-2014) and shoreline data from the "Global Self-consistent Hierarchical High-resolution Geography" database (Wessel & Smith 1996). All specimens were preserved at -20 °C until DNA analysis, then preserved in ethanol and deposited in the collections of the Royal Belgian Institute of Natural Sciences, Brussels (RBINS) under the general inventory number IG 32962. Genomic DNA extraction, amplification and sequencing of the 16S (482 bp), COI (614 bp) and Cyt*b* (675 bp) mtDNA gene fragments, sequence assembly and alignment, were performed as described in Fourdrilis *et al.* (2016), in 399 specimens. The mtDNA dataset comprises 1197 sequences of 16S, COI and Cyt*b* gene fragments, 555 of which were previously published in Fourdrilis *et al.* (2016) (GenBank: KT996152-KT997344), and 642 were obtained from 214

newly sequenced specimens (GenBank: KX537775-KX538416). The three gene fragments were concatenated using GENEIOUS 5.3.4 (http://www.geneious.cm, Kearse *et al.* 2012) and as such combined 16S-COI-Cyt*b* haplotypes (1771 bp) were obtained for 399 specimens, or when mentioned, were used separately. A nuclear gene fragment (714 bp) of the Internal Transcribed Spacer 2 (ITS2) region was sequenced in 18 specimens, 10 of them are among the 399 specimens mentioned above and 8 of them are new specimens (Table S1), using the primers LSU-1 (5'-CTAGCTGCGAGAATTAATGTGA-3') and LSU-3 (5'-ACTTTCCCTCACGGTACTTG-3') (Wade *et al.* 2006) with similar conditions than COI amplification except for the annealing step at 50 °C. Mixed chromatograms of heterozygous individuals were deconvoluted using CHAMPURU 1.0 (Flot 2007). In total, the nuclear DNA (nDNA) dataset is composed of 15 homozygous and 3 heterozygous individuals and comprises 21 sequences.

Population genetic diversity and differentiation analyses

Diversity metrics were computed using DNASP 5.10.1 (Librado & Rozas 2009) to characterise mtDNA diversity in 11 sampled populations taken separately and pooled (hereafter referred to as "total population"), and nDNA diversity in 5 sampled populations.

Diploidy in nDNA data was also exploited to search for the presence of potential cryptic species in *M. neritoides*, as a complement to the analyses based on mtDNA in Fourdrilis *et al.* (2016). A network-based haploweb was reconstructed using HAPLOWEBMAKER (unpublished program). Second, a Maximum Likelihood tree-based haploweb was reconstructed using RAxML 8.2.10 (Stamatakis 2014) hosted on the CIPRES Science Gateway (Miller *et al.* 2010), based on the HKY+I model and with bootstrap consensus trees inferred

from 1,000 replicates. INKSCAPE (https://inkscape.org) was used for colouring nodes in the network, and for drawing curves between haplotypes found co-occurring in heterozygous individuals in the tree. Such haplowebs are a network or a phylogenetic tree, which provide additional connections between haplotypes co-occurring in heterozygous individuals. It allows to delimitate closely-related species that reached mutual allelic exclusivity (i.e. groups of individuals that share no allele), even when they have not yet reached reciprocal allelic monophyly and that the tree looks monophyletic (Flot *et al.* 2010).

Population genetic differentiation was assessed in the overall population by calculating G_{ST} (Pons & Petit 1995) and N_{ST} based on a distance matrix of pairwise differences (Pons & Petit 1996) using SPAGEDI 1.4 (Hardy & Vekemans 2002), φ_{ST} based on a distance matrix of pairwise differences (Excoffier *et al.* 1992) using ARLEQUIN 3.5.1.3 (Excoffier & Lischer 2010), and D_{EST} using SPADE (Chao & Shen 2010). Population genetic differentiation was also assessed among pairs of populations by calculating pairwise φ_{ST} using ARLEQUIN. The significance of pairwise φ_{ST} was corrected for multiple test biases using the Sequential Bonferroni procedure (Rice 1989) and only p-values that remained significant after these corrections were considered to be meaningful.
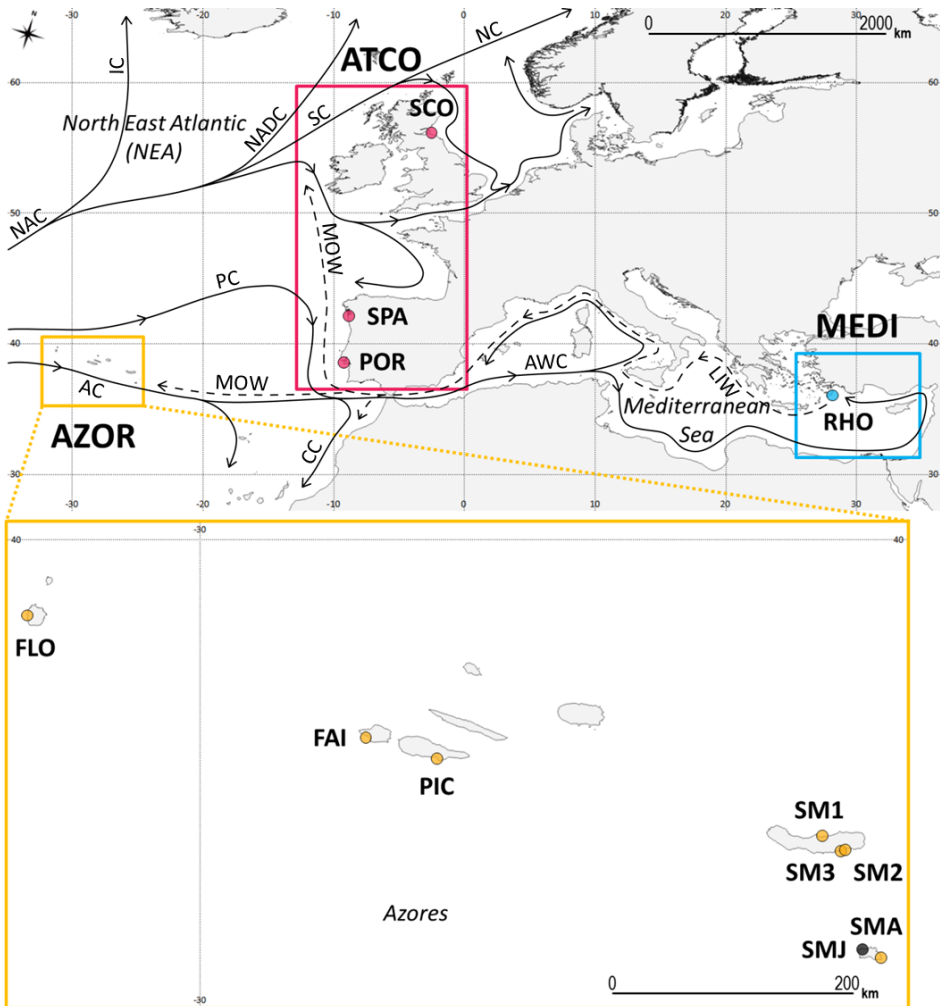
Fig 1. **Distribution range of *Melarhaphe neritoides*** (ETRS89 Lambert azimuthal equal-area projection, EPSG:3035) and 12 sites sampled: Fajã Grande, Flores island, Azores, Portugal (FLO); Varadouro, Faial island, Azores, Portugal (FAI); Lajes do Pico, Pico island, Azores, Portugal (PIC); Porto Formoso, São Miguel island, Azores, Portugal (SM1); port of Ribeira Quente, São Miguel island, Azores, Portugal (SM2); shore of Ribeira Quente, São Miguel island, Azores, Portugal (SM3); Maia, Santa Maria island, Azores, Portugal (SMA); Anjos, Santa Maria island, Azores, Portugal (SMJ); North

Berwick, Scotland, United Kingdom (SCO); Lisbon, Portugal (POR); Vigo, Spain (SPA); Kamiros Skala, Rhodes island, Greece (RHO). The site SMJ in grey was used in the population genetic analyses based on nDNA only. The arrows represent the major surface (solid line) and deep (dashed line) sea currents: Azores Current (AC); Atlantic Water Current (AWC); Canary Current (CC); Irminger Current (IC); Levantine Intermediate Water (LIW); Mediterranean Outflow Water (MOW); North Atlantic Current (NAC); North Atlantic Drift Current (NADC); Norwegian Current (NC); Portugal Current (PC); Slope/Shelf Edge Current (SC).

Hierarchical analyses of molecular variance (AMOVA, Excoffier *et al.* 1992) of Tamura-Nei distances among haplotypes were performed using ARLEQUIN, in order to quantify population genetic differentiation among groups (φ_{CT}), among populations within groups (φ_{SC}) and within populations (φ_{IS}) at several geographic scales and test for panmixis. The significance of φ-statistics was assessed using 90000 permutations of individuals among populations, and of populations among geographic groupings. Three groupings were defined to represent the three oceanographic areas of interest (Fig. 1), i.e. the North East Atlantic coast (ATCO, $N = 95$), the remote Azores archipelago at the southwesternmost border of the distribution area (AZOR, $N = 265$), and the Mediterranean (MEDI, $N = 39$). The AMOVA with three groups contains nine populations following a sampling scheme $k=5,3,1$ (i.e. first group including 5 populations, second group including 3 populations and third group including 1 population) and hence provides adequate statistical power (i.e. p-value ≤ 0.05 and at least 20 unique permutations) at this level (Fitzpatrick 2009).

Population genetic connectivity analyses

Population genetic connectivity in *M. neritoides* was qualitatively investigated by reconstructing a median-joining haplotype network (Bandelt *et al.* 1999) using POPART 1.7 (http://popart.otago.ac.nz). A network such as this, provides

information about phylogeographic structure and gene flow among populations. Population genetic connectivity was then assessed by quantifying long-term gene flow, or immigration rate (i.e. $N_e m$ the effective number of immigrants per generation), among the three oceanographic areas AZOR, ATCO and MEDI in the NEA basin. First, using Slatkin's private allele method (Slatkin 1985b) implemented in GENEPOP 4.5.1 (Rousset 2008). Second, using the Bayesian MCMC method implemented in MIGRATE-N 3.6.11 (Beerli 2006) and hosted on the CIPRES Science Gateway (Miller *et al.* 2010). This latter estimates the mutation-scaled population size ($\theta = 2N_e\mu$ for haploid mtDNA) for each area and the mutation-scaled immigration rate ($M = m/\mu$). Subsampling the three oceanographic groups to get equal sample sizes is not necessary as the difference between the largest (AZOR, $N = 265$) and the smallest (MEDI, $N = 39$) sample sizes is less than ten-fold. Five models of dispersal were evaluated (Fig. 2): (M1) a full migration model with three population sizes and six immigration rates, (M2) an island model where all areas share a single mean estimate of θ and exchange genes with all other areas at the same mean rate, (M3) a source-sink model with three population sizes and three directional West-to-East immigration rates, where the main sink is MEDI receiving immigrants from the sources AZOR and ATCO, and the second sink is ATCO receiving immigrants from AZOR, (M4) a source-sink model with three population sizes and three directional East-to-West immigration rates, where the main sink is AZOR receiving immigrants from the sources MEDI and ATCO, and the second sink is ATCO receiving immigrants from MEDI, and (M5) a panmictic model with one population size parameter. We ran MIGRATE-N analyses under a F84 mutational model, with a windowed uniform prior for θ and M, the bounds of which are (0; 2) and (0; 9500) respectively. For each model, we ran three replicates using four MCMC chains with relative temperatures of 1.0, 1.5, 3.0 and 100000, and of 500 million generations, which sampled 1 of every 100 iterations. The first 30 % of generations were discarded

from each run as burn-in. The five analyses were computationally expensive and required 4-9 weeks depending on the model, although replicates were run simultaneously for some models using the message passing interface version of MIGRATE-N.
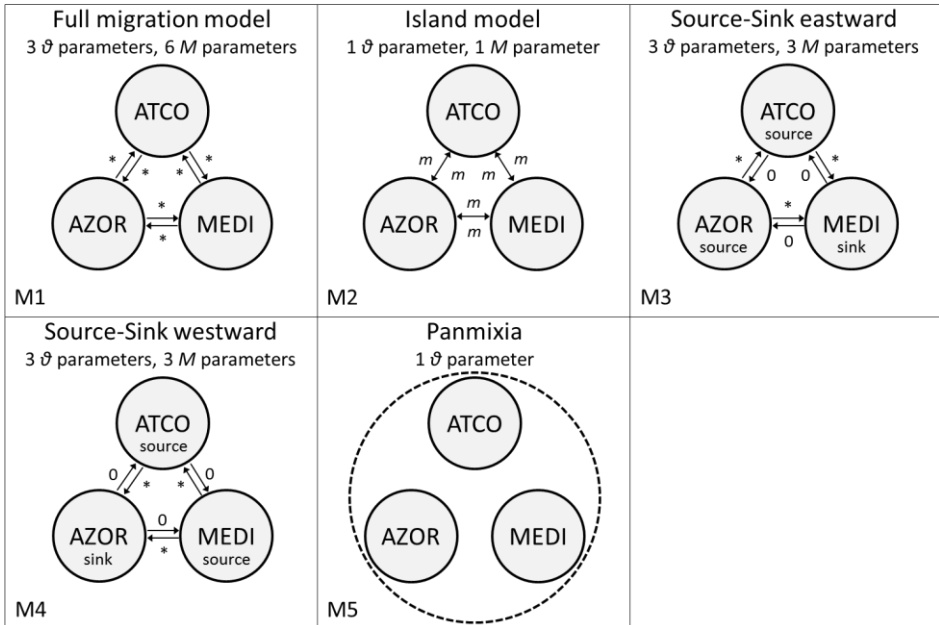
Fig 2. **Diagrams of migration models tested in MIGRATE-N** for Bayesian inferences of mutation-scaled immigration rate (M) and mutation-scaled population sizes (θ): full migration model (M1), island model (M2), source-sink "eastward" model (M3), source-sink "westward" model (M4), and panmixia (M5). Arrows represent directions of gene flow among the three oceanographic groups AZOR (Azores), ATCO (North East Atlantic coast) and MEDI (Mediterranean Sea). *, variable migration rate parameter; m, symmetrical migration rate parameter; 0, migration rate parameter not estimated.

Convergence of MCMC chains was assessed by visual examination of the log trace of each posterior distribution showing caterpillar shape, and making sure

that the Effective Sampling Size value of each statistic was > 200 (Ho & Shapiro 2011), using the 'coda' package (Plummer *et al.* 2006) in R 3.0.2 (R Development Core Team 2011). The R script is available upon request. The models were ranked using log Bayes factors (LBF) and probabilities (p), that compare the marginal likelihood of each model calculated using the thermodynamic integration method implemented in MIGRATE-N (Beerli & Palczewski 2010). The ranking tells how useful a model is to infer a relationship between the pattern of connectivity hypothesised and the biology of *M. neritoides*. The most useful information is found in the model ranked first. The effective number of immigrants per generation was calculated for haploid data with the equation $N_e m = 0.5 * \theta_{recipient} * M$. The effective population size was calculated with the equation $N_e = \theta / 2\mu$ where $\mu = 1.99 \times 10^{-4}$ mutations per nucleotide site per generation (Fourdrilis *et al.* 2016).

Testing of mtDNA hyperdiversity impact

The impact of mtDNA hyperdiversity on population genetic differentiation and connectivity assessment was investigated, using two datasets – with and without hyperdiversity – to compare outcomes of differentiation statistics, haplotype networks, and gene flow estimates. The dataset A contains the original 16S-COI-Cyt*b* hyperdiverse data. The dataset B has a reduced hyperdiversity and contains low-polymorphism data, resulting from the modification of the original dataset A in which hypervariable nucleotide sites, likely homoplasic, were removed. To this end, original data were imported to NETWORK 5.0.0.1 (Bandelt *et al.* 1999) and hypervariable nucleotide sites were identified in the *.sta* outfile as the characters showing a weight > 1, which correspond to fast-mutating nucleotide sites and/or sites segregating for 2 or more nucleotides i.e. showing three or more alleles. Among a total of 540 variable sites, 346 hypervariable nucleotide sites were deleted from the

sequence alignment, representing respectively 10 %, 24 % and 23 % from the length of 16S, COI and Cyt*b* gene fragments. This procedure permits to preserve the high genetic diversity (*Hd* moved from 0.999 ± 0.001 to 0.822 ± 0.001) while concurrently lowering polymorphism (*S* decreased from 30 to 13 % and π from 0.013 ± 0.001 to 0.001 ± 0.001) (Table 1).

RESULTS

mtDNA diversity

With on average 30 % polymorphic sites, the mtDNA in *M. neritoides* is highly polymorphic (Table 1). Haplotype and nucleotide diversities are very high when the 11 populations are pooled (*Hd* = 0.999 ± 0.001; π = 0.013 ± 0.001), but also in individual populations (*Hd* = 0.993 ± 0.021 to 1.000 ± 0.005-0.008; π = 0.012 to 0.014 ± 0.001). Hyperdiversity, i.e. nucleotide diversity at synonymous sites, which reflects neutral polymorphism shaped by the balance between mutation pressure and genetic drift, is observed when the 11 populations are pooled (π_{syn} = 6.86 %), and in individual populations (π_{syn} = 6.16 % to 7.49 %). In contrast, non-neutral polymorphism is low (π_{nonsyn} = 0.05 % maximum). Diversity metrics in single genes show that hyperdiversity is more pronounced in COI (π_{syn} = 7.25 %) than Cyt*b* (π_{syn} = 6.57 %), but overall, both genes show a similar level of variability in terms of haplotype diversity (*Hd* $_{COI}$ = 0.995 ± 0.001; *Hd* $_{Cytb}$ = 0.998 ± 0.001), proportion of polymorphic sites (*S* $_{COI}$ = 33 %; *S* $_{Cytb}$ = 34 %) and of private haplotypes (89.6 % in COI; 89.3 % in Cyt*b*). The estimation of π_{syn} is not applicable to 16S and ITS2 because they are not protein-coding and have consequently no synonymous and non-synonymous sites. The variability of mtDNA in 16S is high, distributed over one fifth of the sites (*S* = 22 %) and generating a large number of haplotypes (*Hd* = 0.842 ± 0.001), but these haplotypes differ from each other by a nucleotide differences

(π = 0.004 ± 0.001). Likewise, the variability of nDNA in ITS2 is high (Hd = 0.995 ± 0.001) with low nucleotide diversity (π = 0.007 ± 0.001), but only a small proportion of sites located between the nucleotide positions 139-153, 230-240 and 329-353 is responsible for this variability (S = 1.5 %). Hyperdiversity in *M. neritoides* seems therefore a mitochondrial phenomenon, which prevails in protein-coding genes, not a nuclear phenomenon, and not expanded to the entire species.

The ITS2 tree-based haploweb shows two well-supported clades (bootstrap value > 90 %), which each contain an haplotype from the same heterozygous individual SF7 (Fig. 3). Thus, these two clades form one allele pool, called single-locus field for recombination (sl-FFR) (Doyle 1995; Flot *et al.* 2010), and are considered as belonging to the same species. This confirms the monophyly of *M. neritoides* inferred from mtDNA data in Fourdrilis *et al.* (2016). In the ITS2 network-based haploweb (Fig. 3), the three haplotypes connected through heterozygous individuals form three sl-FFRs composed of one individual each. The remaining 14 haplotypes, which include 13 private haplotypes and one haplotype shared by two homozygous individuals, yield 14 additional sl-FFRs according to the criterion of mutual allelic exclusivity that considers two gene pools as distinct species when they share no alleles. In total, the ITS2 network-based haploweb yields 17 sl-FFRs and hence 17 potential cryptic species in *M. neritoides*. However, the high haplotype diversity in ITS2 leads to largely undersampled haplotype richness and heterozygous individuals in a small dataset of 18 individuals here. With such a small sample size in terms of number of heterozygous individuals, the network is for the most part composed of haplotypes from homozygous individuals, which are not connected to other haplotypes but are each assigned to a sl-FFR following the mutual allelic exclusivity criterion, what inflates the number of potential cryptic species. A larger sample of heterozygotes is necessary to reliably use the network-based haploweb as a tool for delimiting cryptic species inside *M. neritoides*. Tree-

based haplowebs have been shown to perform better with small datasets than network-based haplowebs (Flot *et al.* 2011), because additional information can be drawn from the presence of clades and from shared haplotypes that are not pooled but displayed on separate branches.
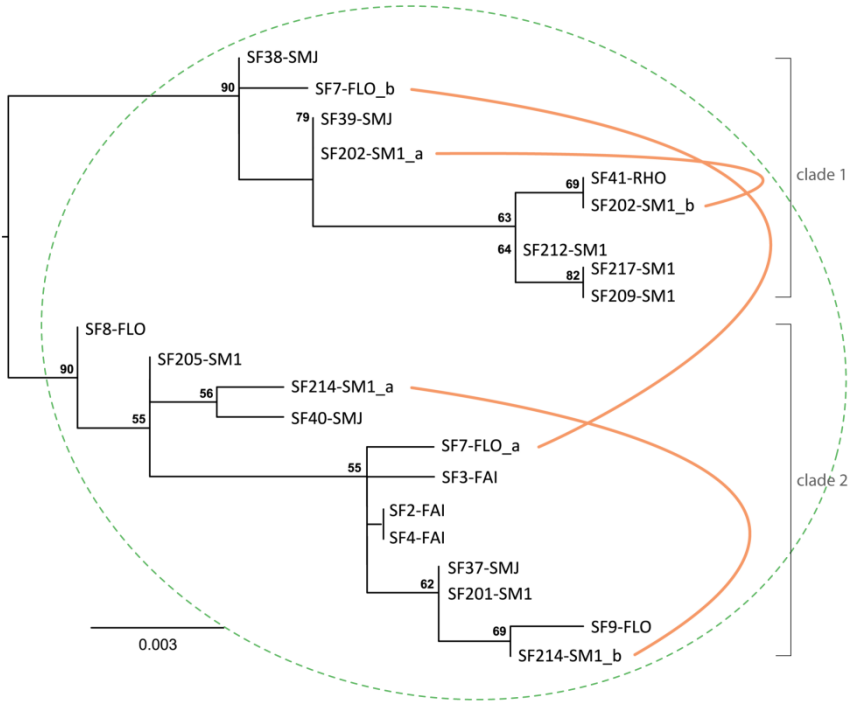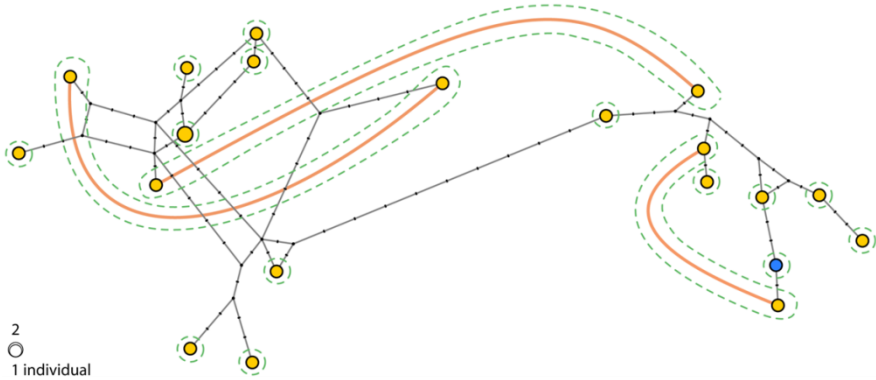
Fig 3. **ITS2 network-based (top) and Maximum Likelihood tree-based (bottom) haplowebs in *M. neritoides*.** Orange curves connect haplotypes co-occurring in heterozygous individuals. Green dashed lines delineate sl-FFRs. Haplotype origins in the network: AZOR, Azores archipelago – yellow; MEDI, Mediterranean Sea – blue. Bootstrap values inferior to 50 in the tree are not indicated.

Table 1. **Genetic diversity in *Melarhaphe neritoides* in the North East Atlantic**. mtDNA dataset A is the original hyperdiverse dataset. mtDNA dataset B has reduced polymorphism. N, number of individuals; H, number of haplotypes; H_p, number of private haplotypes; H_s, number of haplotypes shared with other populations; H_w, number of haplotypes shared within population; L, DNA fragment length in base pair; S, number of segregating sites and its corresponding percentage of the fragment length into brackets; Hd, haplotype diversity and its standard deviation; π, Jukes-Cantor corrected nucleotide diversity and its standard deviation; π_{syn}, Jukes-Cantor corrected nucleotide diversity at synonymous sites; π_{nonsyn}, Jukes-Cantor corrected nucleotide diversity at non-synonymous sites; n/a, not applicable. For the abbreviation of population names, see Fig. 1.

mtDNA	N	H	H_p	H_s	H_w	L	S	$Hd \pm$ SD	$\pi \pm$ SD	π_{syn}	π_{nonsyn}
16S	399	145	112	33	3	486	106 (22%)	0.842 ± 0.001	0.004 ± 0.001	n/a	n/a
COI	399	309	277	32	3	614	200 (33%)	0.995 ± 0.001	0.018 ± 0.001	0.0725	0.0001
Cyt*b*	399	328	293	35	6	675	230 (34%)	0.998 ± 0.001	0.016 ± 0.001	0.0657	0.0005
16S-COI-Cyt*b* (dataset A)											
Total population	399	390	386	4	1	1775	536 (30%)	0.999 ± 0.001	0.013 ± 0.001	0.0686	0.0003
FAI	42	42	42	0	0	1775	205 (12%)	1.000 ± 0.005	0.013 ± 0.001	0.0749	0.0003
FLO	39	39	37	2	0	1775	183 (10%)	1.000 ± 0.006	0.012 ± 0.001	0.0620	0.0005
PIC	37	36	34	3	1	1775	185 (10%)	0.998 ± 0.007	0.012 ± 0.001	0.0616	0.0003
POR	38	38	37	2	0	1775	173 (10%)	1.000 ± 0.006	0.012 ± 0.001	0.0636	0.0003
RHO	39	39	38	1	0	1775	187 (11%)	1.000 ± 0.006	0.013 ± 0.001	0.0661	0.0002
SCO	18	17	15	4	0	1775	120 (7%)	0.993 ± 0.021	0.013 ± 0.001	0.0706	0.0000
SM1	35	35	35	0	0	1775	210 (12%)	1.000 ± 0.007	0.013 ± 0.001	0.0673	0.0004
SM2	37	37	36	1	0	1775	195 (11%)	1.000 ± 0.006	0.013 ± 0.001	0.0692	0.0001
SM3	43	43	43	0	0	1775	239 (14%)	1.000 ± 0.005	0.014 ± 0.001	0.0728	0.0003

SMA	32	32	32	0	0	1775	217 (12%)	1.000 ± 0.008	0.014 ± 0.001	0.0708	0.0005
SPA	39	38	36	4	0	1775	183 (10%)	0.999 ± 0.006	0.014 ± 0.001	0.0729	0.0002
16S-COI-Cyt*b* (dataset B)										
Total population	399	161	134	27	3	1429	191 (13%)	0.822 ± 0.001	0.001 ± 0.001	n/a	n/a
nDNA	18	20	19	1	1	714	11 (1.5 %)	0.995 ± 0.001	0.007 ± 0.001	n/a	n/a

Table 2. **Pairwise mtDNA differentiation (φ$_{ST}$) among 11 populations of *Melarhaphe neritoides* in the North East Atlantic and associated probabilities of significance (in parenthesis) (below diagonal) and haplotypes shared between pairs of populations (above diagonal).** Significant φ$_{ST}$ values before correction for multiple test biases (α = 0.050) are in bold. No values remained significant after sequential Bonferroni correction (α = 0.001). For the abbreviation of population names, see Fig. 1.

	FAI	FLO	PIC	POR	RHO	SCO	SM1	SM2	SM3	SMA	SPA
FAI	0										
FLO	0.006 (0.179)	0			hap 73			hap 49			
PIC	0.012 (0.070)	0.008 (0.145)	0	hap 108		hap 108					hap 108
POR	0.008 (0.122)	0.007 (0.156)	-0.005 (0.666)	0		hap 108					hap 108
RHO	0.009 (0.122)	-0.007 (0.842)	0.010 (0.120)	0.005 (0.207)	0						
SCO	**0.029 (0.006)**	0.006 (0.191)	0.017 (0.061)	**0.025 (0.018)**	0.015 (0.054)	0					hap 108 hap 201
SM1	0.009 (0.123)	-0.005 (0.739)	0.003 (0.300)	0.009 (0.119)	0.003 (0.288)	-0.002 (0.532)	0				
SM2	0.003 (0.325)	-0.011 (0.870)	-0.011 (0.781)	0.000 (0.415)	-0.001 (0.447)	0.009 (0.212)	-0.008 (0.715)	0			
SM3	0.002 (0.321)	-0.011 (0.992)	0.014 (0.063)	0.007 (0.145)	-0.006 (0.839)	0.011 (0.084)	0.001 (0.340)	-0.002 (0.519)	0		
SMA	0.003 (0.287)	0.006 (0.182)	0.010 (0.114)	0.003 (0.269)	0.008 (0.138)	**0.037 (0.003)**	**0.016 (0.042)**	0.001 (0.386)	0.002 (0.329)	0	
SPA	-0.001 (0.513)	-0.009 (0.959)	0.008 (0.134)	0.006 (0.169)	-0.005 (0.762)	0.007 (0.151)	-0.004 (0.697)	-0.011 (0.891)	-0.009 (0.963)	0.002 (0.296)	0

mtDNA population differentiation

In the total population, G_{ST} and φ_{ST} reveal very low, but significant differentiation (G_{ST} = 0.001, p = 0.02; φ_{ST} = 0.005, p = 0.04), whereas N_{ST}, suggests no significant differentiation (N_{ST} = 0.004, p = 1.00), reflecting that haplotype frequencies are for the most part similar among populations (Table 3). The significant unbiased Morisita dissimilarity index (D_{EST} = 0.679, CI = 0.664-0.688) shows strong haplotypic differentiation in the total population, reflecting that haplotypes are for the most part distinct among the 11 populations, up to a complete haplotypic differentiation (D_{EST} = 1) in 47 of the 55 pairs of populations including the two closest populations SM2 and SM3 that are 1.2 km apart. Five out of 390 haplotypes occur in more than one individual (H_s = 4 and H_w = 1), but one haplotype is shared within two individuals of the same population (H_w = 1), so that only four haplotypes are shared among populations (H_s = 4), *viz.* within AZOR (between FLO and SM2), within ATCO (among the three localities), between AZOR and ATCO (among PIC, POR, SCO, SPA), and between AZOR and MEDI (between FLO and RHO) (Table 1). The most common haplotype (hap 108) is shared between AZOR and ATCO, with a low frequency of 0.0125 (Table 2). No haplotypes are shared between ATCO and MEDI. Therefore, the vast majority of haplotypes is private to populations (H_p = 386 out of 390 haplotypes) and represent 96.7 % of the 399 individuals sequenced (Table 1). Four of the 11 populations (FAI, SM1, SM3 and SMA), located in the Azores, share no haplotypes with other populations (H_s = 0).

Genetic differentiation assessment using less polymorphic mtDNA (dataset B) than the hyperdiverse mtDNA (dataset A), leads to the same observation of low but significant (only φ_{ST}) differentiation among populations of *M. neritoides* in the NEA (Table 3). Haplotypic differentiation understandably disappears (D_{EST} = 0.026, not significantly different from zero), in accordance with the reduced

variability of mtDNA in dataset B and the larger proportion of shared haplotypes (17 %).

Table 3. **Population genetic differentiation assessment in *Melarhaphe neritoides* in the North East Atlantic based on DNA data showing two levels of polymorphism**. 16S-COI-Cyt*b* (dataset A) is the original hyperdiverse dataset, 16S-COI-Cyt*b* (dataset B) is the modified dataset with reduced variability. Values significantly different from zero are in bold.

	G_{ST}	*p*	φ_{ST}	*p*	N_{ST}	*p*	D_{EST}	*CI*
16S-COI-Cyt*b*								
dataset A	**0.001**	*0.02*	**0.005**	*0.04*	0.004	*1.00*	**0.679**	*0.664-0.688*
dataset B	0.006	*0.20*	**0.005**	*0.03*	0.004	*0.43*	0.026	*0.000-0.100*

Spatial scale of population genetic differentiation

We assessed population genetic differentiation at several spatial scales within the range 1-6000 km over the NEA basin among the three oceanographic areas ATCO, AZOR and MEDI (Table 4).

At large scale, the AMOVAs show no significant genetic differentiation among groups, covering distances up to 6000 km between ATCO and MEDI, 5000 km between AZOR and MEDI, and 2000 to 4000 km between ATCO and AZOR. The AMOVA at the global scale of the NEA (dataset A) shows very low and significant differentiation at the within-population level (φ_{IS} = 0.007, *p* = 0.03), which reflects high variation among individuals of the same population and not variation among populations of the two groups. This is not an artefact of mtDNA hyperdiversity since it yields identical results using reduced-polymorphism data (dataset B). Globally, all AMOVAs show that > 99 % of the variation is due to within-population variation and not to among-population differentiation (< 1 %) (Table 4). Moreover, none of the pairwise comparisons

of population genetic differentiation (φ_{ST}) show significant differentiation (Table 2). At smaller spatial scales, no population genetic structure is detected, neither between Azorean islands (100-550 km), nor between populations on the same shore (1.2 km). Hence, these data suggest that there is no differentiation among populations over the species' distribution range.

Table 4. **φ-based hierarchical AMOVA results** showing mtDNA genetic differentiation among and within populations of *Melarhaphe neritoides* for several geographical groupings which represent different spatial scales and the three oceanographic areas ATCO, AZOR and MEDI. For each AMOVA are given the spatial scale (in parenthesis), the percentage of among-group variance or within-group variance (σ), the φ-statistic (φ, significant values marked with * for $p < 0.05$) and the associated probability of significance (*p*). For the abbreviation of geographical groupings and population names, see Fig. 1.

Geographical grouping	σ (%)	φ	p	Populations
Shore (1.2 km)				SM1, SM2
φ_{SC} among populations	0.22	0.002	0.31	
φ_{IS} within populations	99.78			
Island (100 km)				SM1, SM2, SM3
φ_{SC} among populations	0	0.000	0.54	
φ_{IS} within populations	100			
Archipelago (550 km)				FAI, FLO, PIC, SM1, SMA
φ_{SC} among populations	0.21	0.002	0.25	
φ_{IS} within populations	99.79			
North East Atlantic dataset A (2000-6000 km)				ATCO (POR, SCO, SPA) vs AZOR (FAI, FLO, PIC, SM1, SMA) vs MEDI (RHO)
φ_{CT} among groups	0.40	0.004	0.08	
φ_{SC} among populations within groups	0.29	0.003	0.15	
φ_{IS} within populations	99.31	0.007*	0.03	
North East Atlantic dataset B (2000-6000 km)				ATCO (POR, SCO, SPA) vs AZOR (FAI, FLO, PIC, SM1, SMA) vs MEDI (RHO)
φ_{CT} among groups	0.54	0.005	0.09	
φ_{SC} among populations within groups	0.17	0.002	0.33	
φ_{IS} within populations	99.30	0.007*	0.03	

Increasing distance

Population genetic connectivity

Gene flow estimates with Slatkin's (1985b) private allele method, using a mean frequency of private alleles of 0.028, yielded N_em = 4.2 for dataset A and N_em = 3.9 for dataset B, without providing further information about how this gene flow (dispersal) is patterned. This number of migrants per generation is greater than 1, but less than 10 and might not be sufficient to counteract genetic drift (Mills & Allendorf 1996). In contrast, gene flow estimation with MIGRATE-N, in presence (dataset A) or absence (dataset B) of mtDNA hyperdiversity, suggests that *M. neritoides* complies with a panmictic model, and hence, rates of gene flow are sufficiently high among localities to make *M. neritoides* behaving as a single panmictic population over its entire distribution range. Indeed, M5 has the lowest log marginal likelihood of the five models of gene flow which were tested, and the highest probability (p = 0.99) (Table 5). The effective population size of *M. neritoides* in the NEA is small (N_e = 1303, CI = 1119–1487; using θ = 0.51865) relatively to the effective population size in the Azores (N_e = 5256, CI = 1312–37495) (Cf. Fourdrilis *et al.* 2016).

Table 5. **Ranking of the models of gene flow tested in MIGRATE-N**, using log Bayes factors (LBF) and probabilities (prob) that are based on the comparison of the log marginal likelihood of each model.

Rank	Model	log marginal likelihood	LBF	prob
mtDNA dataset A				
1	M5 Panmixia	-20642.83381	0	9.99×10^{-1}
2	M3 Source-Sink eastward	-20649.53419	-6.7	1.23×10^{-3}
3	M4 Source-Sink westward	-20933.51334	-290.7	5.74×10^{-127}
4	M1 Full migration model	-21032.13422	-389.3	8.48×10^{-170}
5	M2 Island model	-21655.46617	-1012.6	0.00
mtDNA dataset B				
1	M5 Panmixia	-2692.03118	0	1
2	M3 Source-Sink eastward	-2798.99015	-106.9	3.53×10^{-47}

The panmictic model M5 estimates the parameter θ only, and does not allow us to quantify separately the immigration rates among the three areas, since all localities are pooled into one single population. The full migration model (M1) is similar to M5, assuming gene flow among the three areas and the six directions, but allows us to quantify them separately. Surprisingly, the difference between the log marginal likelihoods of M5 and M1 is very large (LBF = -389.3) and the probability of M1 is near-zero, meaning that M1 is not useful for describing how gene flow is patterned in *M. neritoides*. The model M4 has also a near-zero probability, and M2 has a zero probability. The Source-Sink eastward model (M3) is better ranked than M1 and has a non-zero probability (p = 0.001), and is therefore, useful to describe how gene flow is patterned in *M. neritoides*. Rates of gene flow among the three areas are directional, higher eastward than westward (Fig. 4).
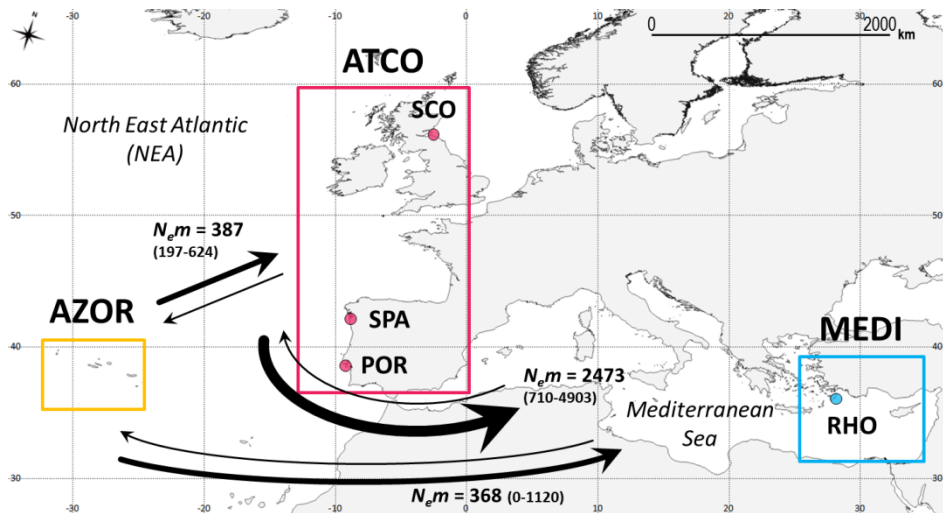

Fig 4. **Connectivity pattern among populations of *Melarhaphe neritoides*.** The arrows represent directions of migration among the three oceanographic groups AZOR (Azores archipelago) in yellow, ATCO (North East Atlantic coast) in pink and MEDI (Mediterranean Sea) in blue. The thicknesses of arrows are proportional to the inferred rates of gene flow.

The Mediterranean sink receives huge numbers of immigrants per generation from the two sources ATCO (N_em = 2473; CI = 710–4903; using $M_{ATCO \rightarrow MEDI}$ = 4587.6) and AZOR (N_em = 368; CI = 0–1120; using $M_{AZOR \rightarrow MEDI}$ = 682.8). The ATCO area also receives huge numbers of immigrants per generation from the Azores (N_em = 387; CI = 197–624; using $M_{AZOR \rightarrow ATCO}$ = 1407.5). MEDI receives more than six times more migrants from ATCO than from AZOR.

Despite high rates of gene flow throughout the distribution area of *M. neritoides*, the mtDNA haplotype network shows a bush-like pattern (Fig. 5) of the private haplotypes represented by single individuals (i.e. singletons) and very few shared haplotypes among sites (sectored circles). Intuitively, such pattern would not be associated with a strong signal of gene flow and population connectivity through time. Yet, it is exactly a pattern one would expect for high gene flow and strong connectivity when using hyperdiverse genetic markers (Nielsen & Slatkin 2013). Moreover, the lack of association between haplotype and geographic location suggests the absence of phylogeographic structure in *M. neritoides* in the NEA, which is also supported by the non-significant difference between N_{ST} and G_{ST} ($N_{ST} - G_{ST}$ = 0.003), indicative of no phylogeographic signal (Pons & Petit 1996).

The hyperdiverse mtDNA data, i.e. the combined 16S-COI-Cyt*b* dataset A and the single COI and Cyt*b* genes, are all associated with a bush-like pattern (Fig. 5 a, b, c). The impact of mtDNA hyperdiversity becomes clear in the haplotype networks of low polymorphism mtDNA data, i.e. the combined 16S-COI-Cyt*b* dataset B and the single 16S gene, showing a more classic star-like pattern (Fig. 5 d, e).
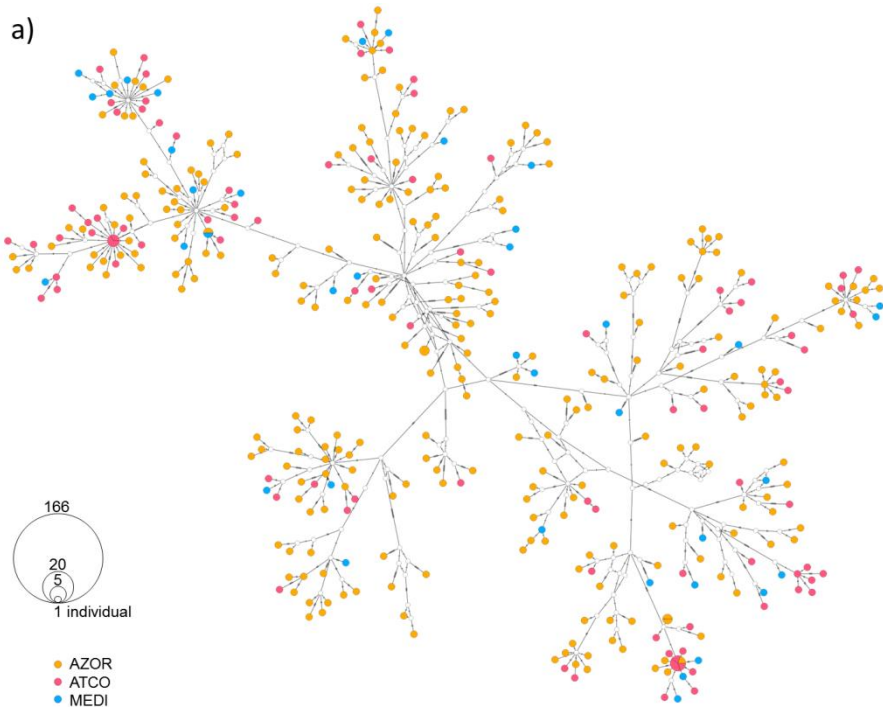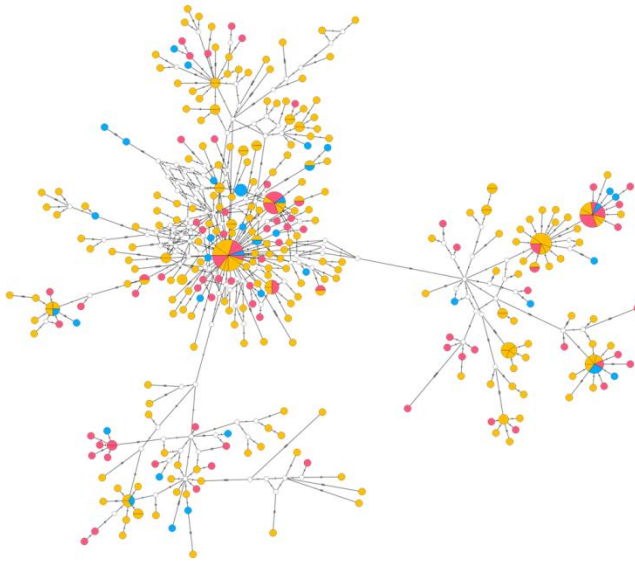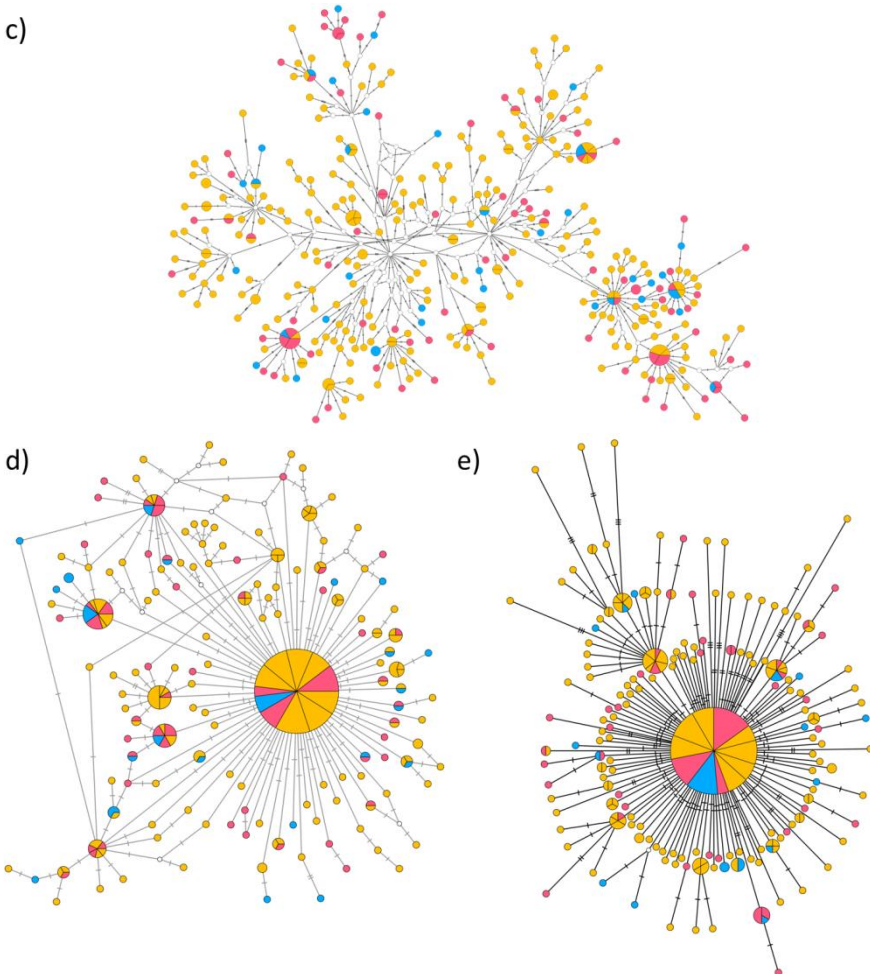
Fig 5. **Median-joining networks of mtDNA (a) concatenated 16S-COI-Cyt*b* (dataset A), (b) COI, (c) Cyt*b*, (d) 16S and (e) concatenated 16S-COI-Cyt*b* (dataset B) in *Melarhaphe neritoides**. The size of circles is proportional to the number of individuals per haplotype. Haplotype origins: AZOR, Azores archipelago – yellow; ATCO, North East Atlantic coast – pink; MEDI, Mediterranean Sea – blue.

DISCUSSION

Assessment of population genetic differentiation based on hyperdiverse mtDNA

The present study confirms that mtDNA in *M. neritoides* is hyperdiverse ($\pi_{syn} \geq$ 5 %), not only in the Azores and Galicia (Fourdrilis *et al.* 2016), but all over the NEA. This mtDNA hyperdiversity results in an overwhelming number of private haplotypes and a paucity of shared haplotypes among the localities sampled throughout the NEA, up to a complete lack of shared haplotypes between localities as close as 1.2 km or as far as 6000 km. Despite nearly complete haplotypic differentiation (D_{EST}) among populations, there is no significant pairwise population genetic differentiation (φ_{ST}). In absence of hyperdiversity, when using low polymorphism mtDNA data, the haplotypic differentiation drops to zero and clearly shows the effect of mtDNA hyperdiversity on D-statistics. Therefore, when using hyperdiverse mtDNA markers, population genetic differentiation as expressed in terms of lack of haplotype sharing may be substantial, but is not indicative of population genetic differentiation in terms of fixation of haplotypes ($D_{EST} \approx 1 \neq \varphi_{ST} \approx 1$). Moreover, from a practical point of view, hyperdiverse mtDNA may need unrealistically high sampling efforts in order to detect single haplotypes more than once and hence reliably assess eventual haplotype sharing among populations (Fourdrilis *et al.* 2016). This phenomenon is explained by the high mutation rates in hyperdiverse mtDNA, which at high speed generate numerous private haplotypes with low frequency that provoke a high within-population genetic diversity (Fourdrilis *et al.* 2016) influencing D_{EST}, but not φ_{ST} (Kronholm *et al.* 2010).

In species like *M. neritoides* mtDNA hyperdiversity represents upper boundary of intra-specific genetic variation, and allowed us to use F_{ST} and D_{EST} at a limit of their applicability for extreme within-population variation. It reveals that F_{ST}

(and relatives) reliably measures differentiation in terms of degree of completion of haplotype fixation and in terms of dissimilarity in haplotype frequencies among populations (G_{ST} = 0.001, p = 0.02; φ_{ST} = 0.005, p = 0.04), and that D_{EST} reliably measures differentiation in terms of lack of haplotype sharing among populations (D_{EST} = 0.679, CI = 0.664-0.688). This is in accordance with the use of F_{ST} initially recommended by Wright (1978, page 82), as well as the use of D_{EST} intended by Jost (2008). The two indices measure two types of population genetic differentiation, and when used for their initial purpose, are complementary.

Is *Melarhaphe neritoides* panmictic?

Our assessment of population genetic differentiation (pairwise φ_{ST}, AMOVA) in *M. neritoides* in the NEA, based on mtDNA markers that are far more variable than Johannesson's (1992) allozyme data, or based on moderately variable mtDNA markers, confirms that the pattern of broad-scale allozyme homogeneity between Cretan and Swedish populations of this species (Johannesson 1992) is not an artefact of the allozyme data.

None of the populations show significant pairwise population genetic differentiation, indicating that there is no mtDNA differentiation in *M. neritoides* throughout the NEA. But weak differentiation is detected at the intra-population level (φ_{IS}), i.e. among individuals within populations, and not among populations within and between groups. Intra-population differentiation in absence of inter-population differentiation reflects very high variability of haplotypes within sampling site, and may be a sampling artefact since the Scottish population in ATCO has a smaller sample size (N = 18) than any other population (N = 32 to 43). As such, its haplotype composition may be more biased than elsewhere due to the extremely high haplotype richness of *M. neritoides* (Fourdrilis *et al.* 2016). At smaller scales (1.2 km, 100 km, 550 km),

our results show no mtDNA differentiation at all. This was also reported either at a very small scale (30 m) between upper and lower shores in Silleiro, Spain (García *et al.* 2013). Thus, *M. neritoides* shows no sign of population genetic structure and therefore, our results suggest that *M. neritoides* is panmictic over the entire NEA basin.

Phylogeographic breaks in the NEA

The Atlantico-Mediterranean transition (defined here as the area encompassing the Gibraltar Strait, the Almeria-Oran Front and the Siculo-Tunisian Strait) and the English Channel potentially form barriers to dispersal, and hence possible phylogeographic breaks for planktonic-dispersing species (Ayata *et al.* 2010; Deli *et al.* 2016; Patarnello *et al.* 2007). Yet, our study did not find any evidence of barriers to gene flow or phylogeographic breaks over the entire NEA basin. Hence, dispersal and mtDNA structuring in *M. neritoides* do not seem to be affected by the phylogeographic breaks of the Atlantico-Mediterranean transition and the English Channel.

Population genetic connectivity in the NEA

To the best of our knowledge, the present work is the first gene flow and genetic connectivity estimation in a marine gastropod over its entire geographic range in the NEA using a coalescent approach. The quantitative assessment of migration rates over evolutionary time-scales in *M. neritoides*, based on gene genealogies using MIGRATE-N, shows substantial gene flow within the whole NEA basin (N_em = 368 to 2473). This high rate of gene flow counteracts genetic drift and ensures homogeneity of the species gene pool over the species' distribution range and over time. Gene flow is directional, towards the Mediterranean Sea, with higher rate eastward than westward, from a main

source area (Atlantic European coasts) and a secondary source area (Azores archipelago). MIGRATE-N gene flow analyses seem not influenced by the amount of mtDNA nucleotide diversity, however, a comparison with more models is necessary to confirm this observation. The quantitative assessment of migration rates over evolutionary time-scales in *M. neritoides*, based on frequency of private alleles using Slatkin's (1985b) method, yields much lower gene flow estimate in the overall population ($N_em \approx 4$ for both dataset A and B) than the estimates obtained in MIGRATE-N. Slatkin's (1985b) method assumes that the private alleles found in samples have reached a quasi-equilibrium distribution and have been all sampled, and therefore the method is insensitive to the number of sampled individuals per population (Barton & Slatkin 1986). Yet, the overwhelming number of private haplotypes in *M. neritoides* (99 % in dataset A and 83 % in dataset B) does not represent a quasi-equilibrium distribution of private alleles in populations but the haplotype richness which is very high and partially sampled (Fourdrilis *et al.* 2016). Slatkin's (1985b) method relies on allele frequency and does not use the information contained in nucleotide variability within alleles, unlike MIGRATE-N, and hence, is not affected by the change in π between dataset A and B. Therefore, when using hyperdiverse mtDNA markers, partial sampling of haplotype richness and non-equilibrium distribution of private alleles may bias gene flow estimation using Slatkin's (1985b) method. It is also noteworthy that selection is potentially acting on *M. neritoides* mtDNA (Fourdrilis *et al.* 2016) and may bias gene flow estimates, by violating the assumption of neutrality which underlies both models of gene flow inference, i.e. the island model in Slatkin's (1985b) method and the coalescent model in MIGRATE-N (Kuhner 2008; Whitlock & McCauley 1999).

A few quantitative estimates of long distance gene flow are available for planktonic-dispersing species in the NEA. Long-distance gene flow has been reported at high rates within the NEA, (1) over 1900 km among Macaronesian

archipelagos (N_em = 18 to 290), but with very limited gene flow over 1500-2500 km between Cape Verde and the three other Macaronesian archipelagos (N_em = 3), in the periwinkle *Tectarius striatus* (Van den Broeck *et al.* 2008), (2) over 3700 km within the Mediterranean (N_em = 60) and over 5000 km from the Atlantic European coasts to the eastern Mediterranean (N_em = 30) in the sea urchin *Paracentrotus lividus* (Penant *et al.* 2013), and (3) over 4500 km along the Atlantic European coasts (N_em = 903) in the bivalve *Scrobicularia plana* (Santos *et al.* 2012). Long-distance gene flow has also been reported at high rate outward the NEA, over 5000 km from Norway to the East coasts of North America (N_em = 80) in the sea cucumber *Cucumaria frondosa* (So *et al.* 2011). In comparison with these estimates, the rate of genetic connectivity in *M. neritoides* is very high, notably from the Atlantic European coasts to the Mediterranean Sea (N_em = 2473). The pelagic larval duration of *M. neritoides* (PLD = 4-8 weeks) is long, and comparable to those of *Paracentrotus lividus* (PLD = 3 weeks) (Gosselin & Jangoux 1998), *Scrobicularia plana* (PLD = 2-4 weeks) (Frenkiel & Mouëza 1979) and *Cucumaria frondosa* (PLD = 6 weeks) (Hamel & Mercier 1996) (the PLD of *Tectarius striatus* is unknown). This suggests that, as expected, planktonic-dispersing species with long-lived larval stage may achieve high levels of gene flow in the NEA basin.

The Atlantic coral-dwelling gall crab *Opecarcinus hypostegus* has a planktonic larval development consisting of five to seven stages (the PLD is unknown). Although this suggests high potential for dispersal, actual gene flow is limited and follows an isolation-by-distance pattern (van Tienderen & van der Meij 2017). Like *M. neritoides*, *O. hypostegus* shows an extreme degree of mtDNA COI variation (Hd = 0.999; π = 0.026; 22 % polymorphic sites; H_p = 187 out of 195 specimens) (van Tienderen & van der Meij 2017). This mtDNA hyperdiversity was interpreted as an early sign of speciation resulting from adaptive genetic divergence over the coral host species which maintains extreme mtDNA variation. Yet, Fu and Li's F and Tajima's D for the COI data of

O. hypostegus were non-significant (van Tienderen & van der Meij 2017) and hence do not point to effects of selection and/or demographic expansion. Moreover, the nucleotide diversity at synonymous sites in *O. hypostegus* well-above the threshold of 5 % (π_{syn} = 10.2 %) (calculated from the sequence data of van Tienderen & van der Meij 2017) suggests a high degree of neutral polymorphism. This is in accordance with the bush-like pattern of the mtDNA haplotype network (Fig. 3 in van Tienderen & van der Meij 2017) typical of mtDNA hyperdiversity and too low sampling effort, and the absence of cryptic species. Therefore, the mtDNA hyperdiversity in *O. hypostegus* may result from an elevated mutation rate, like demonstrated in *M. neritoides* (Fourdrilis *et al.* 2016), and may be maintained on account of limited gene flow rather than of selection suggested by the authors, unlike *M. neritoides*.

The pattern of gene flow in *M. neritoides* is congruent with the history of the sea currents in the NEA (Fig. 1). Short-lived Pleistocene sea surface currents allowed the colonization of Macaronesia from Eastern Atlantic areas (Ávila *et al.* 2009). However, now the Azores Current flows eastward to Gibraltar where its surface water enters and disperses within the Mediterranean water through the Atlantic Water Current (El-Geziry & Bryden 2010; Johnson & Stevens 2000), suggesting that larval transport predominantly occurs from Macaronesia towards the Mediterranean Sea. Originating from the Gulf Stream, the North Atlantic Current (Rowe *et al.* 2013) branches into the Irminger Current (Gyory *et al.* 2013), the North Atlantic Drift Current (Bischof *et al.* 2003a) and the Slope/Shelf Edge Current (Gyory *et al.* 2003), which flow northeastward through the NEA and likely transport larvae from the Azores to the Atlantic European coasts above 50°N to Iceland, the British Isles and France. The average flow of the Portugal Current is southward to Africa (Bischof *et al.* 2003b), feeding the Canary Current and also entering the Mediterranean in a shallow surface layer (Barton 2001), suggesting that larval transport predominantly occurs from the Atlantic European coasts to the Mediterranean

Sea. In the opposite directions, gene flow appears weaker from MEDI to ATCO and AZOR, and from ATCO to AZOR, as it goes against mainstream currents and rather follows the Levantine Intermediate Water and the Mediterranean Outflow Water that flow below 500 m deep westward to Macaronesia and northward to Ireland (Bozec *et al.* 2011; El-Geziry & Bryden 2010), as well as the seasonal northward flow of the Portugal Current in winter. Therefore, the Atlantic European coasts and Macaronesia are most probably a source of new, dispersing, haplotypes supplying the Mediterranean rather than a sink receiving new haplotypes. Besides, genetic exchange within ATCO from the populations POR and SPA to SCO might likely be reduced since the surface circulation along the Atlantic European coasts is southward, and since the Atlantic water that enters the North Sea via the English Channel leaves the North Sea along the Norwegian coast without reaching Scotland (Ecomare).

The glacial history of the NEA provides no evidence of impact on the current distribution of genetic diversity in *M. neritoides*. Rocky shore species are believed to have become fragmented during the Last Glacial Maximum (LGM) when the littoral was mostly covered by ice. They have survived the LGM in glacial refugia, i.e. ice-free rocky shore habitats in the Azores, the Balkans, the Iberian Peninsula, the Italian Peninsula, the English Channel and Southwest Ireland (Ingólfsson 2009; Maggs *et al.* 2008; Weiss & Ferrand 2007). The present day distribution of the periwinkle *Littorina saxatilis*, which lives in sympatry with *M. neritoides*, but which has no planktonic larval stage, is the outcome of expansion from glacial refugia after the LGM amongst other factors that resulted in strong population structure (Panova *et al.* 2011). In *M. neritoides*, the absence of genetic and phylogeographic structuring throughout the NEA and the pervasive mtDNA hyperdiversity do not allow to detect signs of eventual influences of quaternary ice ages on the present day distribution of the species.

CONCLUSIONS

Melarhaphe neritoides shows no genetic structure and is panmictic over its entire distribution range, though with a predominantly eastward gene flow. The Mediterranean acts as a sink receiving large numbers of immigrants per generation from primarily the NEA coasts (N_em = 2473, CI = 710–4903), and secondarily from the Azores (N_em = 368, CI = 0–1120). The mtDNA hyperdiversity (π_{syn} ≥ 5 %) of *M. neritoides* results in a lack of shared haplotypes among the localities sampled throughout the NEA, up to a complete haplotypic differentiation between localities as close as 1.2 km or as far as 6000 km, which contrasts with the absence of population genetic differentiation and the high gene flow. Thus, the deceiving (nearly) complete haplotypic mtDNA differentiation among populations, is not reflecting a lack of gene flow, but results from the concealed signal of gene flow by the high mutation rate and/or from a too low sampling effort to detect shared haplotypes. When using such mtDNA hyperdiverse markers, population genetic differentiation as expressed in terms of lack of haplotype sharing may be substantial, but is not indicative of population genetic differentiation in terms of fixation of haplotypes (D_{EST} ≈ 1 ≠ φ_{ST} ≈ 1). Because F_{ST} (and relatives) and D_{EST} measure two types of population genetic differentiation, misuse of these indices can lead to erroneous interpretations of population genetic differentiation. However, when using F_{ST} accordingly to the original recommendation of Wright (1978, page 82) and D_{EST} as intended by Jost (2008), in presence of mtDNA hyperdiversity, F_{ST} (and relatives) reliably measures fixation and D_{EST} reliably measures haplotypic differentiation, and are complementary indices.

ACKNOWLEDGEMENTS

Supplemental Information

Table S1. **Locations of collection sites and numbers of specimens for *Melarhaphe neritoides* samples used in the present study**.

Sampling site	N	$N1$	$N2$	Sampling date	WGS84 coordinates	
					Latitude	Longitude
FAI	44	42	3	06/28/1993	N 38.56632	W 28.77069
FLO	40	39	3	1992	N 39.45817	W 31.26401
PIC	37	37	0	10/14/1993	N 38.39633	W 28.25684
POR	38	38	0	08/07/2013	N 38.70514	W 9.14312
RHO	40	39	1	10/11/2011	N 36.27311	E 27.82419
SCO	18	18	0	05/28/1997	N 56.06206	W 2.71623
SM1	35	35	7	07/31/1993	N 37.82305	W 25.42695
SM2 (port)	37	37	0	06/30/2012	N 37.7350	W 25.29717
SM3 (praia)	43	43	0	06/30/2012	N 37.7295	W 25.30801
SMA	32	32	0	04/17/1996	N 36.94016	W 25.01322
SMJ	4	0	4	04/1996	N 37.00472	W 25.15831
SPA	39	39	0	08/06/1995	N 42.22458	W 8.76987
Total	407	399	18			

N, total number of sampled individuals in the present study; *N1*, number of sampled individuals for analyses based on mitochondrial DNA; *N2*, number of sampled individuals for analyses based on nuclear DNA.

Chapter 3

Comparative mitogenomics of Littorinidae (Mollusca: Gastropoda) and molecular phylogeny of Littorinimorpha

Séverine Fourdrilis [1]
Thierry Backeljau [1,2]

[1] Directorate Taxonomy and Phylogeny & JEMU, Royal Belgian Institute of Natural Sciences, Brussels, Belgium

[2] Evolutionary Ecology Group, University of Antwerp, Antwerp, Belgium

ABSTRACT

The mitochondrial genome of *Melarhaphe neritoides* (Gastropoda: Littorinidae) has a total length of 15,676 bp and consists of the conventional set of 13 protein-coding genes, 22 tRNA genes, 2 rRNA genes and a partial putative control region (474 bp) flanked by *trnF(gaa)* and *cox3* genes. Mitogenome characteristics are comparable with that of three other Littorinidae. The usage of synonymous codons is not random and shows an over-usage of A and T at the third codon positions, and reflects the overall negative AT skew pattern in the protein-coding genes. Major differences are: 1) ATT as a start codon for *atp8* in *Melarhaphe* instead of ATG in *Littorina*, 2) a positive GC skew, and 3) the cloverleaf structure for *trnM* without a TΨC-loop in *Melarhaphe*. As expected, purifying selection is a dominant force driving non-synonymous polymorphisms of protein-coding genes and their functions, while mutation and drift drive synonymous polymorphisms. Two genes appear positively selected in the mitogenome of *L. saxatilis* and the clade *fabalis*/*obtusata*, respectively *nad4* and *nad5*, and may contribute to adaptive divergence in these two lineages. Among Littorinimorpha, mtDNA gene order is rearranged among families, but conserved within families and congruent with phylogenetic relationships.

INTRODUCTION

Evolutionary biology is extensively studied in Littorinidae, from a phylogenetic and phylogeographic (De Wolf *et al.* 2000; McQuaid 1996a; Panova *et al.* 2011; Reid *et al.* 2012; Rolán-Alvarez 2007), to ecological (De Wolf *et al.* 2004; Johannesson 2003, 2016; McQuaid 1996b) and population genomic points of view (Marques *et al.* 2017; Panova *et al.* 2014; Ravinet *et al.* 2016). *Melarhaphe neritoides* (Linnaeus, 1758) is a littorinid periwinkle that has been

more scarcely studied (Cuña *et al.* 2011; García *et al.* 2013; Johannesson 1992; Libertini *et al.* 2004; Williams *et al.* 2003). The species shows mitochondrial DNA (mtDNA) hyperdiversity, i.e. its selectively neutral nucleotide diversity is above the threshold of 5 % for the 16S ribosomal RNA (*rrnL*), cytochrome oxidase c subunit I (*cox1*) and cytochrome b (*cob*) genes (π_{syn} = 6.9 %), which is generated by an extremely high mtDNA mutation rate (μ = 5.82 × 10^{-5} per site per year) (Fourdrilis *et al.* 2016). Elevated mutational pressure on synonymous variation in mitogenomes promotes adaptive mutations and mitonuclear coevolution (Castellana *et al.* 2011) and is thus a fundamental force for evolution (Lynch *et al.* 2006). This suggests that mtDNA hyperdiversity may induce changes in base composition, gene order and tRNA structure. Recombination in mtDNA can induce gene rearrangements and increases genetic diversity, hence mtDNA hyperdiversity might be a sign of recombination and change in gene order (Chen 2013; Ma & O'Farrell 2015). Besides, gene order arrangements are unusually frequent in Mollusca (Boore & Brown 1998; Grande *et al.* 2008), up to the family level (Rawlings *et al.* 2010). On the other hand, the crucial metabolic functions of mitochondrial protein-coding genes constraint mtDNA variation (Blier *et al.* 2001; Castellana *et al.* 2011) and may limit the impact of mtDNA hyperdiversity.

We report here the nearly complete mitogenome of *M. neritoides*, the only extant species of the genus *Melarhaphe*. We investigate mitogenome composition and structure, and putative signatures of natural selection on protein-coding genes (PCGs). We carry out a comparative mitogenomic analysis of *M. neritoides* and the three other species from the family Littorinidae, viz. *Littorina fabalis*, *Littorina obtusata* and *Littorina saxatilis* (hereafter referred to as "*Littorina* sp."), whose the mitogenomes have been sequenced (Marques *et al.* 2017). These three *Littorina* species and *M. neritoides* are sympatric in some parts of their distribution areas. Furthermore, we provide the first phylomitogenomic analysis of the Littorinimorpha.

MATERIALS AND METHODS

Specimen collection and DNA extraction

We collected one specimen of *M. neritoides* on 6 July 2012 in the port of Varadouro, Faial island, Azores, Portugal (N 38.56633, W 28.77068), and preserved it at -20 °C until DNA analysis. We extracted genomic DNA from foot muscle using the NucleoSpin® Tissue kit (Macherey-Nagel GmbH & Co. KG, Germany). All remaining body parts and the shell were deposited in the collections of the Royal Belgian Institute of Natural Sciences, Brussels (RBINS) under the general inventory number IG 32962 and specimen voucher INV.134051.

Mitogenome sequencing and annotation

Sequencing and assembly of the mitogenome were performed by the Beijing Genomics Institute (Hong Kong) on an Illumina HiSeq4000 platform following manufacturer's instructions and Tang *et. al.* 's pipeline (Blaxter *et al.* 2005). The mitogenome was annotated with the MITOS WebServer (Bernt *et al.* 2013), followed by manual curation. Tandem repeats in the Control Region (CR) were identified using the RepeatMasker Web Server (http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker). We compared the boundaries of predicted PCGs and rRNAs with the three other mitogenomes of littorinids published to date (Marques *et al.* 2017), i.e *Littorina fabalis*, *Littorina obtusata* and *Littorina saxatilis* (Table 1), using Geneious 6.1.8 (Kearse *et al.* 2012). We followed Boore (2006) for naming conventions and Cameron (2014) for annotation recommendations. The graphical representation of the *M. neritoides* mitogenome was drawn with OGDRAW (Lohse *et al.* 2013).

Table 1. Summary of sample information of species used in this study**.**

Infraorder	Family	Species	Accession Number	Reference
Littorinimorpha	Cassidae	*Galeodea echiniphora*	NC_028003	Osca et al. 2015
	Hydrobiidae	*Potamopyrgus antipodarum*	NC_020790	Neiman et al. 2010
		Potamopyrgus estuarinus	NC_021595	Neiman et al. 2010
	Littorinidae	*Littorina fabalis*	KU952092	Marques et al. 2016
		Littorina obtusata	KU952093	Marques et al. 2016
		Littorina saxatilis	KU952094	Marques et al. 2016
		Melarhaphe neritoides	unpublished	this study
	Naticidae	*Naticarius hebraeus*	NC_028002	Osca et al. 2015
	Pomatiopsidae	*Oncomelania hupensis*	NC_013073	unpublished
		Oncomelania hupensis hupensis	NC_012899	unpublished
		Oncomelania hupensis robertsoni	NC_013187	unpublished
		Tricula hortensis	NC_013833	unpublished
	Ranellidae	*Monoplex parthenopeus* [a]	NC_013247	Cunha et al. 2009
	Strombidae	*Lobatus gigas* [b]	NC_024932	Márquez et al, 2014
	Vermetidae	*Ceraesignum maximum* [c]	NC_014583	Rawlings et al. 2010
		Dendropoma gregarium	NC_014580	Rawlings et al. 2010
		Eualetes tulipa	NC_014585	Rawlings et al. 2010
		Thylacodes squamigerus	NC_014588	Rawlings et al. 2010
Neogastropoda	Buccinidae	*Buccinum pemphigus*	NC_029373	unpublished
	Conidae	*Conus striatus*	NC_030536	unpublished
	Nassariidae	*Nassarius reticulatus*	NC_013248	Cunha et al. 2009

[a, b, c] Names in the original publication were respectively: *Cymatium parthenopeum*, *Strombus gigas* and *Dendropoma maximum* but are currently unaccepted names and have been updated in the present publication.

Mitogenome composition and organization

We conducted analyses of nucleotide composition and relative synonymous codon usage (RSCU) using MEGA 7.0 (Kumar *et al.* 2016). We calculated nucleotide skew statistics using the formulas: AT skew = [A – T]/[A + T] and GC skew = [G – C]/[G + C] (Perna & Kocher 1995). We predicted and compared secondary structures of tRNAs among the four littorinids using MITOS WebServer. We used mitogenome sequences (17 taxa) that were available in Genbank (Table 1) in order to examine gene order rearrangements at the family level within Littorinidae, and at the infraorder level within Littorinimorpha.

Divergence and selection

We estimated sequence divergence (p-distance) among *M. neritoides* and three *Littorina* species, excluding the CR, using MEGA 7.0. We performed a maximum likelihood estimation of the ratio (ω) of non-synonymous (d_N) to synonymous (d_S) substitution rates (Angelis *et al.* 2014) to measure the direction and magnitude of natural selection acting on PCGs in the four littorinids, using branch models which allow ω to vary among branches in the phylogeny (Yang 1998; Yang & Nielsen 1998) and which are implemented in CODEML in the PAMLX 1.3.1 package (Xu & Yang 2013). We compared two branch models, viz. the free-ratios model which assumes one ω ratio for each branch in the tree, and the two-ratios model which assumes one ω ratio for the foreground branch (specified *a priori*, one lineage at a time) putatively under positive selection and one ω ratio for the remaining background branches, to the null model which yields an averaged ω_0 for the whole tree. Significance was assessed by a likelihood ratio test.

Littorinimorpha phylogeny

We employed Bayesian (BI) and Maximum Likelihood (ML) approaches, implemented respectively in MrBayes 3.2.6 (Ronquist *et al.* 2012) and RAxML 8.2.9 (Stamatakis 2014) both hosted on the CIPRES Science Gateway (Miller *et al.* 2010), to carry out a phylomitogenomic analysis of 18 Littorinimorpha taxa (Table 1) based on their concatenated PCGs. Three species from the Neogastropoda were used as outgroup (Table 1). Sequence data were aligned using the MAFFT online server (http://mafft.cbrc.jp/alignment/server/). The concatenated dataset was divided into 39 data blocks (for the first, second and third codon positions of the 13 PCGs). The optimal partition strategy of each block (Table 2), restricted to GTR+G model of sequence evolution as recommended by Stamatakis (Stamatakis 2016), was selected by PartitionFinder 2.1.1 (Lanfear *et al.* 2016).

Table 2. **The best partitioning scheme selected by PartitionFinder** for the BI and ML phylogenetic analyses.

Subset	Best model	# sites	Regions
1	GTR+G	516	*cox1*_pos1
2	GTR+G	516	*cox1*_pos2
3	GTR+G	516	*cox1*_pos3
4	GTR+G	1115	*atp6*_pos1, *cox2*_pos1, *cob*_pos1, *cox3*_pos1
5	GTR+G	681	*atp6*_pos2, *nad3*_pos2, *nad4l*_pos2, *cox2*_pos2
6	GTR+G	231	*cox2*_pos3
7	GTR+G	1103	*nad5*_pos1, *nad4*_pos1, *atp8*_pos1
8	GTR+G	1102	*nad4*_pos2, *nad5*_pos2, *atp8*_pos2
9	GTR+G	231	*atp8*_pos3, *nad6*_pos3
10	GTR+G	233	*atp6*_pos3
11	GTR+G	532	*nad3*_pos1, *nad4l*_pos1, *nad1*_pos1
12	GTR+G	966	*cox3*_pos2, *nad1*_pos2, *cob*_pos2
13	GTR+G	583	*nad1*_pos3, *cox3*_pos3
14	GTR+G	177	*nad6*_pos1
15	GTR+G	534	*nad2*_pos3, *nad6*_pos2
16	GTR+G	383	*cob*_pos3
17	GTR+G	576	*nad4l*_pos3, *nad3*_pos3, *nad2*_pos1
18	GTR+G	462	*nad4*_pos3
19	GTR+G	585	*nad5*_pos3
20	GTR+G	358	*nad2*_pos2

For the BI analysis, the final consensus tree was computed from the combination of two independent MCMC runs of 10,000,000 generations each, sampling every 100 generations and discarding the first 2,500,000 generations. Convergence was assessed in TRACER. 1.6 (Rambaut *et al.* 2014). For ML analysis, the bootstrap consensus tree was inferred from 1000 replicates.

RESULTS AND DISCUSSION

Mitogenome organisation and composition

The near complete mitogenome of *M. neritoides* is 15,676 bp long and comprises 37 genes including 13 PCGs, 2 rRNAs genes, 22 tRNAs genes, and a putative non-coding Control Region, as is typical for animal mitogenomes (Table 3, Fig. 1). The CR is partial (474 bp) and is flanked by *trnF(gaa)* and *cox3*. No repetitive sequences were detected in the CR. All genes are encoded on the heavy (H) strand, except eight tRNAs located on the light strand (L). All PCGs start with the canonical ATG codon, like in the *Littorina* species, except for *atp8* that starts with an ATT codon in *M. neritoides*. Stop codons are mainly TAA, or TAG, like in the *Littorina* species, but they are not assigned to the same genes. Intergenic sequences in *M. neritoides* (from 1 to 78 bp) and *Littorina* sp. (from 1 to 72 bp), have similar lengths, the longest being that between *trnE(ttc)* and *rrnS* in all four species. Still, *M. neritoides* has fewer intergenic sequences than *Littorina* sp. (23 vs 28) and hence, their total length across the mitogenome is shorter (250 vs 328-335 bp). Overlapping adjacent genes are slightly more common in *M. neritoides* (*rrnS* and *trnV(tac)*, *trnV(tac)* and *rrnL*, *rrnL* and *trnL2(taa)*, *nad4l* and *nad4*, *nad5* and *trnF(gaa)*) than *Littorina* sp. (*trnG(tcc)* and *trnE(ttc)*, *trnV(tac)* and *rrnL*, *rrnL* and *trnL2(taa)*, *nad3* and *trnS1(gct)*).

Table 3. **Organization of the mitochondrial genome of *Melarhaphe neritoides* and comparison with *Littorina fabalis*, *Littorina obtusata* and *Littorina saxatilis* (Marques *et al.* 2017).** Differences in start and stop codon, length and intergenic nucleotides between *M. neritoides* and the three other species are shaded in grey. * Numbers of intergenic nucleotides separating a gene from the next one; negative values represent overlapping nucleotides in adjacent genes. ** Number (and corresponding percentage) of residues in the amino acid sequence of *Melarhaphe neritoides* that differ with any of the three other species. [a] In *Littorina fabalis*. [b] In *Littorina obtusata*. [c] In *Littorina saxatilis*.

Gene	Strand	Location	Start codon	Stop codon	Length (bp)	Intergenic nucleotides*	Amino acid changes**
cox1	H	1-1536	ATG	TAA	1536	11 30	10/511 (2 %)
cox2	H	1548-2234	ATG	TAA	687	5 2	11/228 (5 %)
trnD(gtc)	H	2240-2307			68 69	1	
atp8	H	2309-2467	ATG	TAA TAG	159	2 13	12/52 (23 %)
atp6	H	2470-3165	ATT ATG	TAA TAG	696	38 31	35/231 (15 %)
trnM(cat)	L	3204-3271			68	1	
trnY(gta)	L	3273-3340			68	1 11	
trnC(gca)	L	3342-3407			66 65	1	
trnW(tca)	L	3409-3474			66	2 1	
trnQ(ttg)	L	3477-3533			57 58	7 11	
trnG(tcc)	L	3541-3607			67	0 -1	
trnE(ttc)	L	3608-3672			65 71	78 72	
rrnS	H	3751-4632			882 894[a,b] 895[c]	-3	
trnV(tac)	H	4630-4696			67 68	-23 -22	
rrnL	H	4674-6087			1414 1415	-36 -10	
trnL2(taa)	H	6052-6121			70 67	2 8	

|							
trnL1(tag)	H	6124-6191			68 67	0	
nad1	H	6192-7133	ATG	TAG TAA	942 939	0 7	42/313 (13 %)
trnP(tgg)	H	7134-7200			67 68	1 2	
nad6	H	7202-7705	ATG	TAA TAG	504 513	8 9	61/167 (37 %)
cob	H	7714-8853	ATG	TAG TAA	1140	9 18[a,b] 17[c]	38/379 (10 %)
trnS2(tga)	H	8863-8929			67 68	0 5	
trnT(tgt)	L	8930-8999			70 71[a]	8	
nad4l	H	9008-9304	ATG	TAA TAG	297	-7	14/98 (14 %)
nad4	H	9298-10668	ATG	TAA TAG[c]	1371	0 8[a] 9[b,c]	108/456 (24 %)
trnH(gtg)	H	10669-10732			64 66	0 1	
nad5	H	10733-12454	ATG	TAA	1722 1719	-1 21[a,b] 23[c]	129/573 (23 %)
trnF(gaa)	H	12454-12520			67 69	0	
CR (partial)		12521-12994			474	0	
cox3	H	12995-13774	ATG	TAA	780	32 33	19/259 (7 %)
trnK(ttt)	H	13807-13878			72 73	11 6[a] 5[b]	
trnA(tgc)	H	13890-13957			68 67	1	
trnR(tcg)	H	13959-14027			69	10 5	
trnN(gtt)	H	14038-14104			67	15 13[a,c] 14[b]	
trnI(gat)	H	14120-14186			67 69	3 4	
nad3	H	14190-14543	ATG	TAA	354	0 -1	22/117 (19 %)
trnS1(gct)	H	14544-14611			68 67	0	
nad2	H	14612-15673	ATG	TAG TAA	1062 1059	3 5	112/353 (32 %)
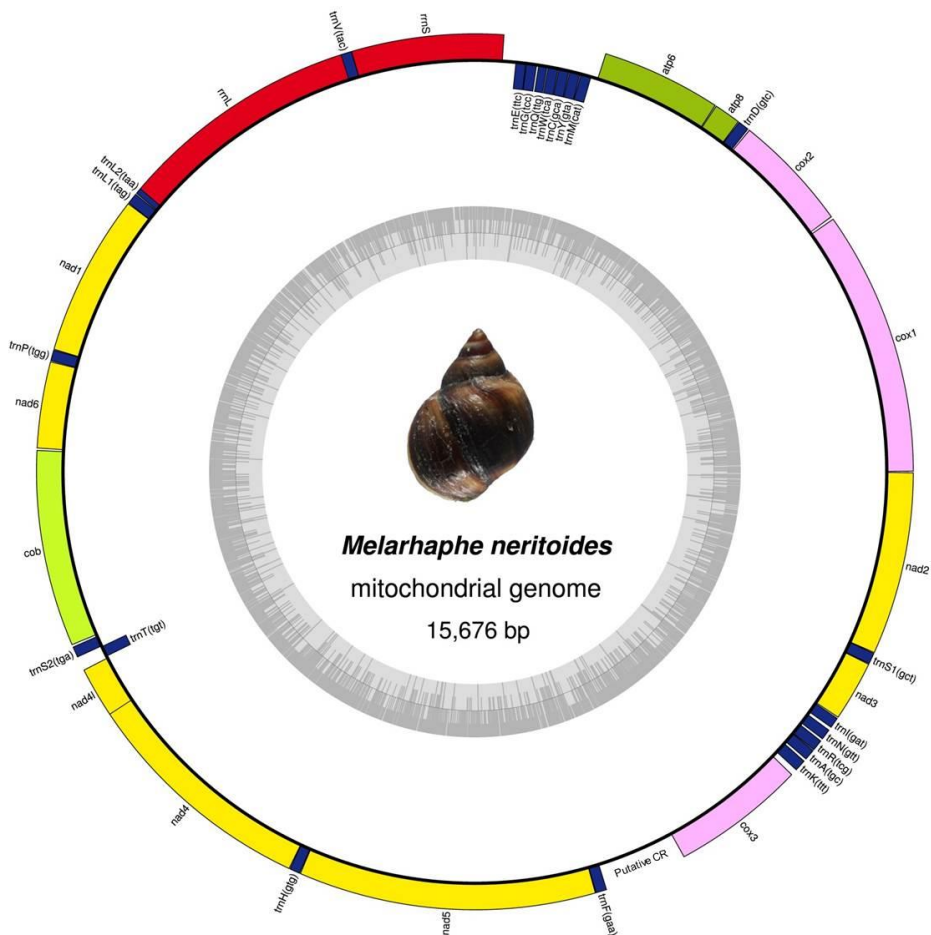
Fig 1. **Gene map of the *Melarhaphe neritoides* mitogenome.** Genes encoded on the H strand are mapped outside the outer circle and are transcribed counterclockwise. Genes encoded on the L strand are mapped inside the outer circle and are transcribed clockwise. The inner circle plot represents G + C% content; the darker lines are, the higher their G + C% is. Photo credit: Yves Barette (RBINS).

The overall nucleotide composition of the mitogenome of *M. neritoides* is AT-rich, significantly biased toward A and T, with A = 28.9 %, C = 16.6 %, G = 17.1 % and T = 37.4 % (Table 4). All regions of the genome are AT-rich with an overall AT content of 66.3 %, but with the lowest values in *rrnS* (48.6 %) and *rrnL* (51.1 %) and at highest in *atp8* (69.2 %). This is similar to the AT content in *Littorina* sp. (66.2-66.9 %), but more than in other Littorinimorpha from the family Vermetidae (59-63 %), and similar to less than in other caenogastropods (65.2-70.1 %) (Rawlings *et al.* 2010). The overall GC content in *M. neritoides* is 33.7 % (33.1 to 33.8 % in *Littorina* sp.).

Table 4. **Nucleotide composition of the mitochondrial genome of *Melarhaphe neritoides* and *Littorina* sp..**

Region	A%	C%	G%	T%	A + T%	G + C%	AT skew	GC skew
mitogenome *M. neritoides*	28.9	16.6	17.1	37.4	66.3	33.7	-0.129	0.012
cox1	26.0	17.3	19.3	37.4	63.3	36.7	-0.180	0.055
cox2	29.1	17.5	18.5	34.9	64.0	36.0	-0.091	0.028
atp8	30.2	15.7	15.1	39.0	69.2	30.8	-0.127	-0.019
atp6	24.9	17.1	15.5	42.5	67.4	32.6	-0.261	-0.049
rrnS	30.7	13.8	37.5	17.9	48.6	51.3	0.263	0.462
rrnL	36.1	12.5	36.4	15.0	51.1	48.9	0.413	0.489
nad1	25.7	16.8	17.1	40.4	66.1	33.9	-0.222	0.009
nad6	26.8	15.7	16.1	41.5	68.3	31.8	-0.215	0.013
cob	24.6	18.7	17.7	39.0	63.6	36.4	-0.226	-0.027
nad4l	27.9	15.8	17.2	39.1	67.0	33.0	-0.167	0.042
nad4	26.4	19.0	15.2	39.4	65.8	34.2	-0.198	-0.111
nad5	25.7	20.1	15.9	38.3	64.0	36.0	-0.197	-0.117
CR (partial)	32.7	19.8	15.8	31.6	64.3	35.7	0.016	-0.112
cox3	24.4	18.6	22.2	34.9	59.3	40.8	-0.177	0.088
nad3	28.2	15.0	19.5	37.3	65.5	34.5	-0.139	0.130
nad2	27.7	13.2	17.5	41.6	69.3	30.7	-0.201	0.140
mitogenome *L. fabalis*	29.9	18.9	14.9	36.3	66.2	33.8	-0.097	-0.119
mitogenome *L. obtusata*	29.9	19.1	14.7	36.4	66.2	33.8	-0.098	-0.129
mitogenome *L. saxatilis*	30.4	18.9	14.1	36.5	66.9	33.1	-0.091	-0.145

The negative value of the AT skewness for the whole mitogenome (-0.129) indicates a bias toward the use of more As over Ts, except for the two rRNA genes and the CR in which T is more common. The positive value of the GC skewness in the whole mitogenome (+0.012) indicates a bias toward the use of more Cs over Gs, except for genes *atp8*, *atp6*, *cob*, *nad4*, *nad5* and the CR.

We analysed codon usage in PCGs to determine preferentially used synonymous codons (Fig. 2).
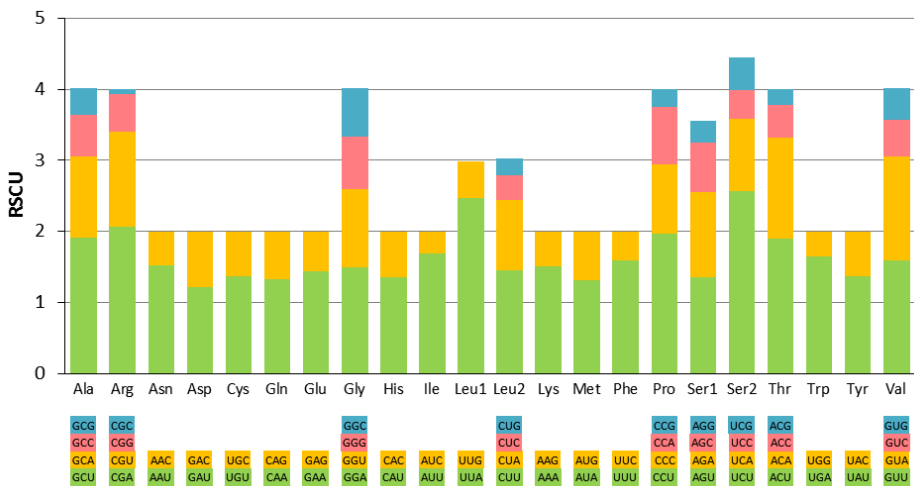

Fig 2. **Relative synonymous codon usage (RSCU) of the mitochondrial genome of *Melarhaphe neritoides*.** The 22 codon families consisting of a total of 62 two- and four-fold degenerate synonymous codons are plotted on the x-axis. The label for the 2 or 4 codons within each family is shown below the x-axis, and the colours correspond to the colours in the stacked columns. The most used synonymous codon in each family is in green. The RSCU values are shown on the y-axis.

Table 5. Relative synonymous codon usage (RSCU) of each amino acid in the mitochondrial genome of *Melarhaphe neritoides***.**

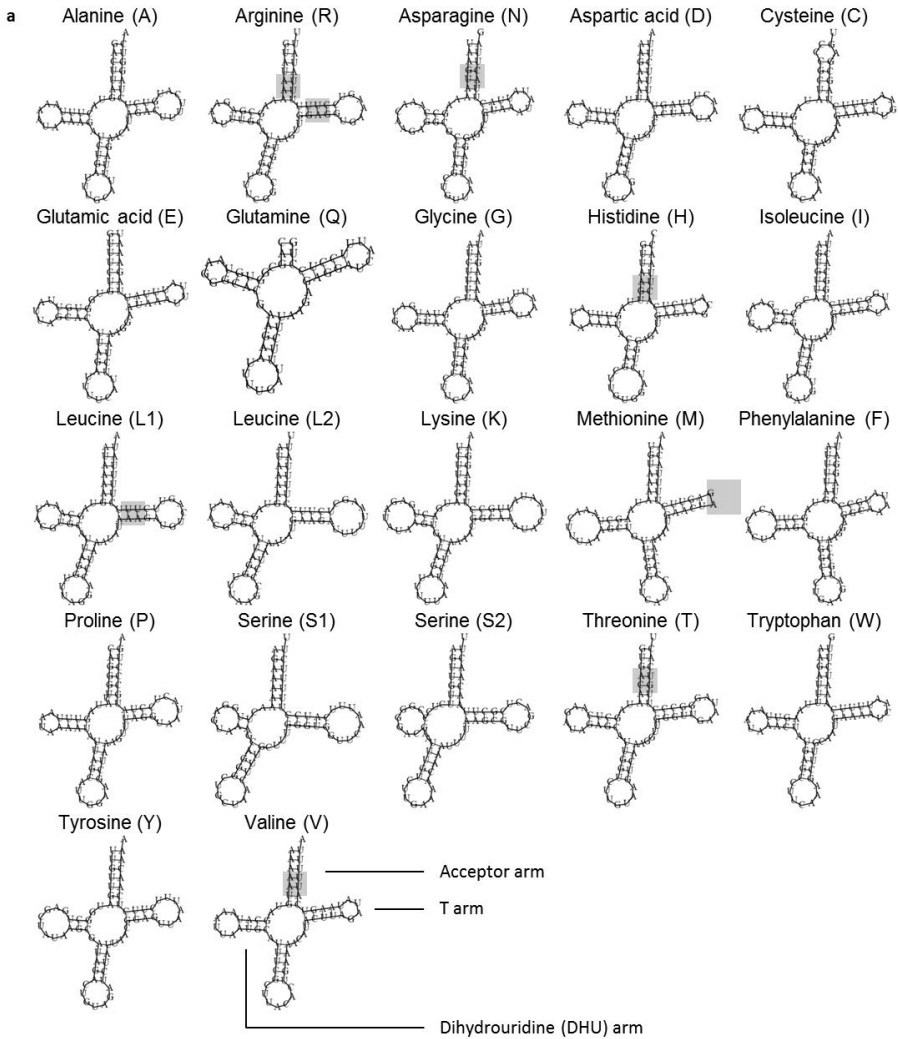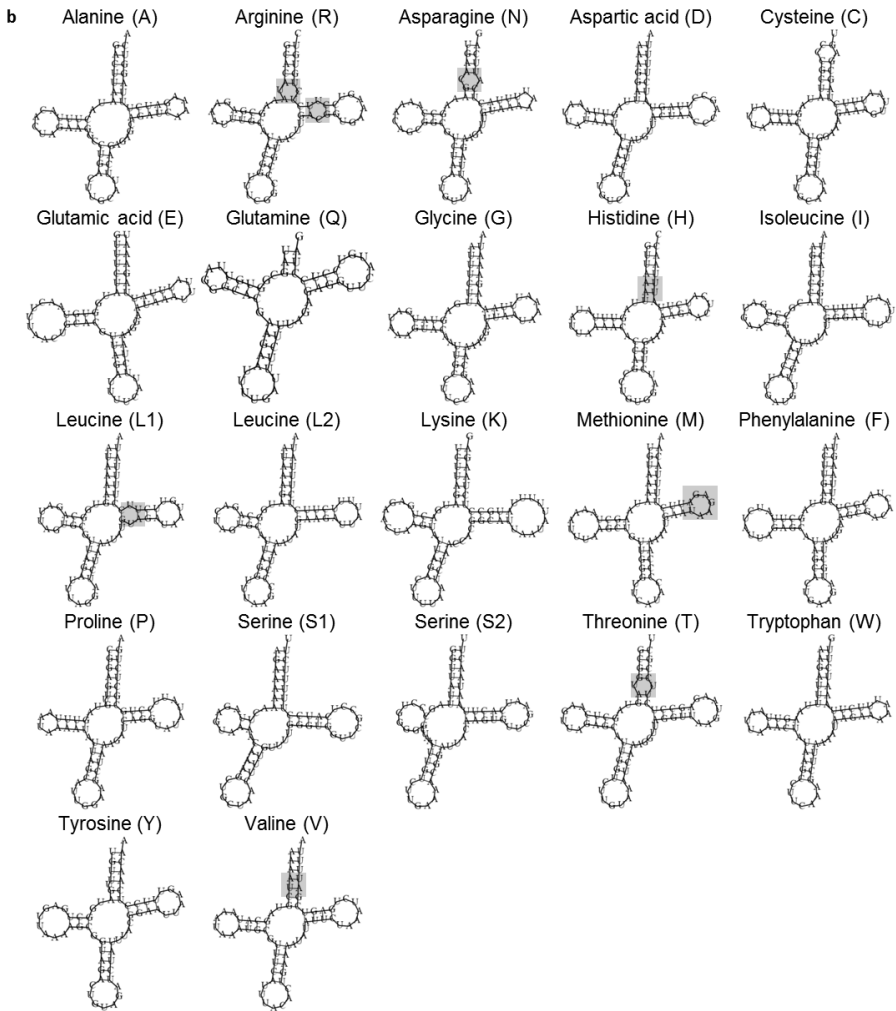| Amino acid	Codon	Count	RSCU	%	Amino acid	Codon	Count	RSCU	%
Ala	GCU	131	1.92	3.49	Lys	AAA	68	1.51	1.81
	GCA	77	1.13	2.05		AAG	22	0.49	0.59
	GCC	40	0.59	1.07	Met	AUA	123	1.31	3.28
	GCG	25	0.37	0.67		AUG	65	0.69	1.73
Arg	CGA	31	2.07	0.83	Phe	UUU	253	1.59	6.74
	CGU	20	1.33	0.53		UUC	66	0.41	1.76
	CGG	8	0.53	0.21	Pro	CCU	71	1.97	1.89
	CGC	1	0.07	0.03		CCC	35	0.97	0.93
Asn	AAU	97	1.52	2.58		CCA	29	0.81	0.77
	AAC	31	0.48	0.83		CCG	9	0.25	0.24
Asp	GAU	47	1.22	1.25	Ser1	AGU	66	1.36	1.76
	GAC	30	0.78	0.80		AGA	58	1.19	1.55
Cys	UGU	26	1.37	0.69		AGC	34	0.7	0.91
	UGC	12	0.63	0.32		AGG	15	0.31	0.40
Gln	CAA	50	1.33	1.33	Ser2	UCU	125	2.57	3.33
	CAG	25	0.67	0.67		UCA	49	1.01	1.31
Glu	GAA	64	1.44	1.70		UCG	22	0.45	0.59
	GAG	25	0.56	0.67		UCC	20	0.41	0.53
Gly	GGA	95	1.5	2.53	Thr	ACU	79	1.9	2.10
	GGU	69	1.09	1.84		ACA	59	1.42	1.57
	GGG	47	0.74	1.25		ACC	19	0.46	0.51
	GGC	43	0.68	1.15		ACG	9	0.22	0.24
His	CAU	55	1.36	1.47	Trp	UGA	90	1.65	2.40
	CAC	26	0.64	0.69		UGG	19	0.35	0.51
Ile	AUU	227	1.69	6.05	Tyr	UAU	98	1.37	2.61
	AUC	42	0.31	1.12		UAC	45	0.63	1.20
Leu1	UUA	248	2.47	6.61	Val	GUU	99	1.6	2.64
	UUG	51	0.51	1.36		GUA	90	1.45	2.40
Leu2	CUU	145	1.45	3.86		GUC	32	0.52	0.85
	CUA	99	0.99	2.64		GUG	27	0.44	0.72
	CUC	35	0.35	0.93					
	CUG	24	0.24	0.64	**Total**		**3754**	**62**	**100**

Note: % = Count/3754.

The codon family encoding the amino acid Ser2 is the most used, followed by equally used codon families encoding Ala, Arg, Gly, Pro, Thr and Val. The five most prevalent codons representing one fourth of the 62 synonymous codons are Phe (TTT) (6.74 %), Leu1 (TTA) (6.61 %), Ile (ATT) (6.05 %), Leu2 (CTT) (3.86 %) and Ala (GCT) (3.49 %), while Arg (CGC) (0.03 %) is the rarest codon (Table 5). We observe an over-usage of two-fold and four-fold degenerate synonymous codons with A or T in the third position in comparison to other synonymous codons. Therefore, the usage of synonymous codons is not random and is AT biased in *M. neritoides*, and reflects the AT skew pattern in the whole mitogenome.
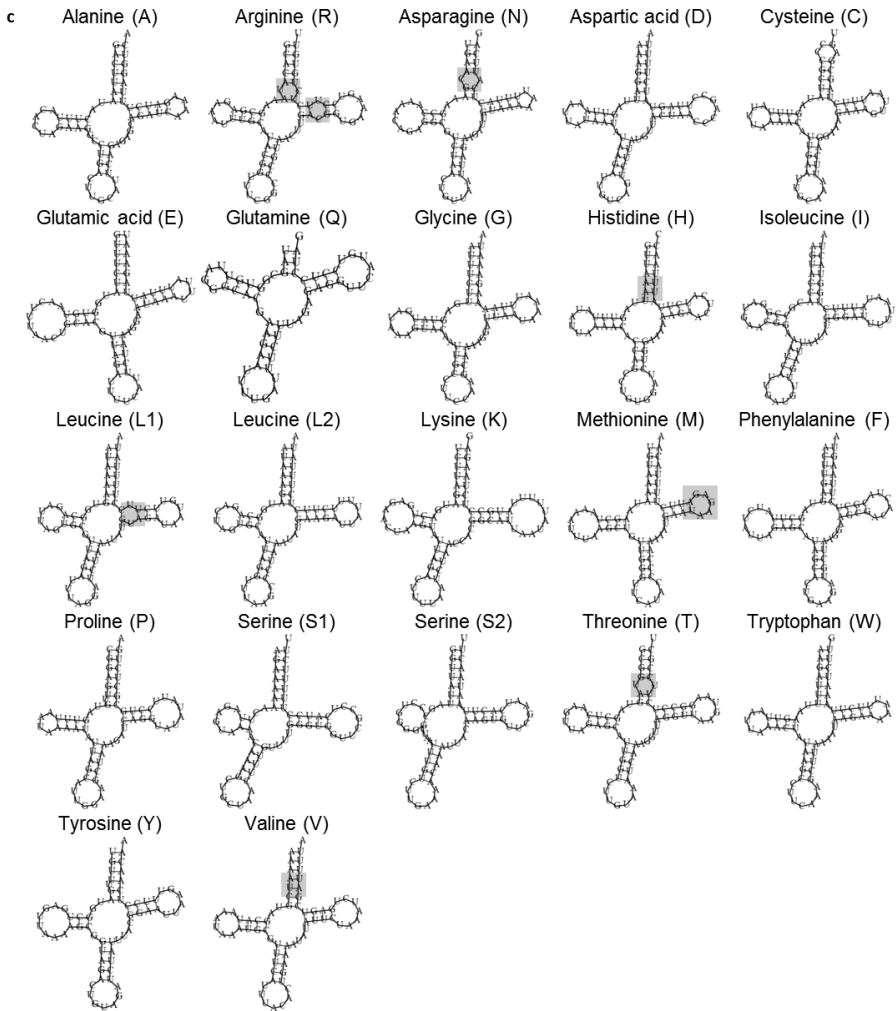
The overall composition of the mitogenome of *M. neritoides* in terms of the 37 genes, anticodons of the 22 tRNA genes, strand-specific distribution, start codons (except for one), nucleotide composition and AT skew, is similar to *Littorina* sp. (Marques *et al.* 2017), and is in line with the trend in Mollusca (Castellana *et al.* 2011). As such, *M. neritoides* shows a conspicuous positive GC skew (+0.04), like in other Mollusca (Castellana *et al.* 2011), but in strong contrast with the negative GC skew (mean -0.13) in *Littorina* sp..

tRNA secondary structure

All 22 tRNA genes are present in *M. neritoides*. They range in length from 57 to 72 bp. Only two tRNA genes do not fold into the typical cloverleaf secondary structure: *trnM(cat)* lacks the loop in the T arm and *trnS2(tga)* has a dihydrouridine (DHU) arm that forms a loop without a stem (Fig. 3). The loss of complete DHU arm and/or T arm in *trnS2* is an occasional event, that in Mollusca occurs far less frequently than in other metazoan phyla (Jühling *et al.* 2012). Conversely, *trnS1(gct)* often lacks the D arm in metazoans, but features the classical cloverleaf in *M. neritoides* and *Littorina* sp. in line with the trend in Mollusca (Jühling *et al.* 2012). Additionally, minor differences are observed.
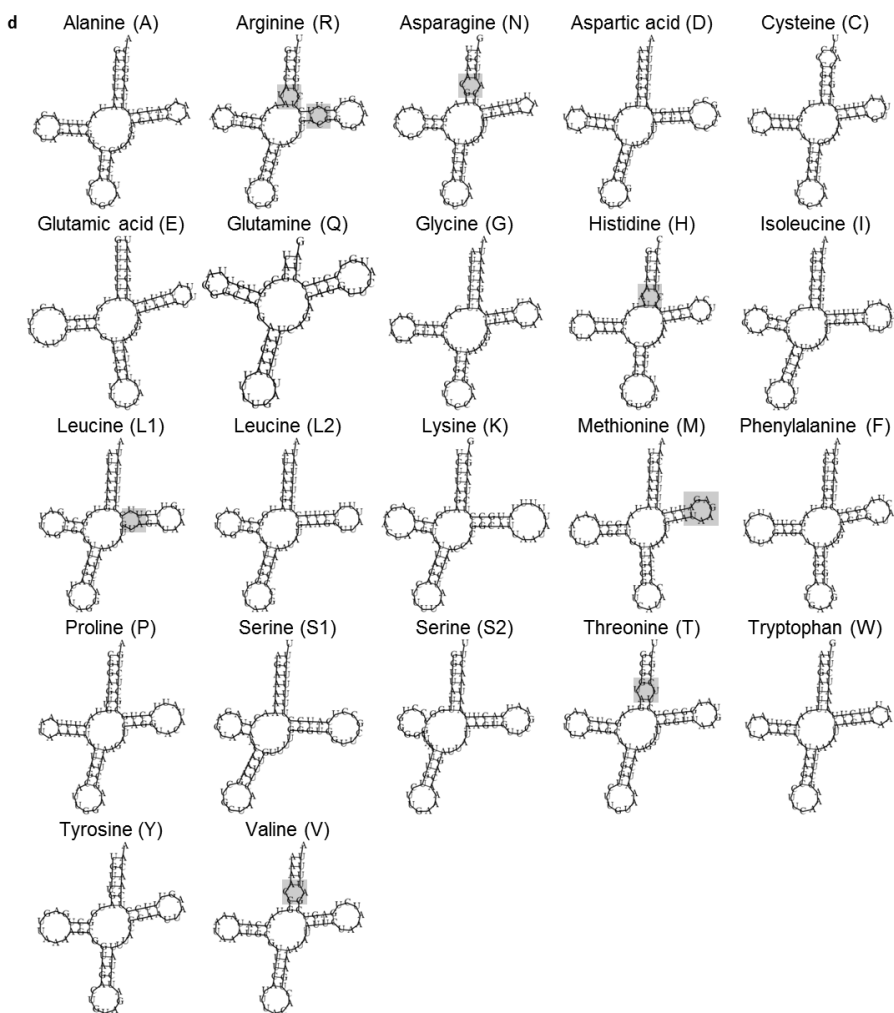
a

Alanine (A)

Arginine (R)

Asparagine (N)

Aspartic acid (D)

Cysteine (C)

Glutamic acid (E)

Glutamine (Q)

Glycine (G)

Histidine (H)

Isoleucine (I)

Leucine (L1)

Leucine (L2)

Lysine (K)

Methionine (M)

Phenylalanine (F)

Proline (P)

Serine (S1)

Serine (S2)

Threonine (T)

Tryptophan (W)

Tyrosine (Y)

Valine (V)

Acceptor arm

T arm

Dihydrouridine (DHU) arm

b

Alanine (A)   Arginine (R)   Asparagine (N)   Aspartic acid (D)   Cysteine (C)

Glutamic acid (E)   Glutamine (Q)   Glycine (G)   Histidine (H)   Isoleucine (I)

Leucine (L1)   Leucine (L2)   Lysine (K)   Methionine (M)   Phenylalanine (F)

Proline (P)   Serine (S1)   Serine (S2)   Threonine (T)   Tryptophan (W)

Tyrosine (Y)   Valine (V)

Fig 3. **Cloverleaf structures of identified tRNA genes in the mitogenome of (a)** *Melarhaphe neritoides,* **(b)** *Littorina fabalis,* **(c)** *Littorina obtusata* **and (d)** *Littorina saxatilis.* Grey boxes mark regions showing differences between *Melarhaphe neritoides* and *Littorina* sp..

Other minor differences are observed in *M. neritoides* with *L. saxatilis*, such as the cloverleaf structure of *trnR(tcg)*, *trnN(gtt)*, *trnH(gtg)*, *trnT(tgt)* and *trnV(tac)* that does not show a loop within the stem of the Acceptor arm, and that of *trnR(tcg)* and *trnL1(tag)* that does not show a loop within the stem of the T arm (Fig. 3). These loops are also absent in the six mollusc species of Vermetidae for which mtDNA genomes are available (Rawlings *et al.* 2010). We observe fewer differences among the three closely related *Littorina* species than between *Littorina* and *Melarhaphe*, as only the cloverleaf structures of *trnH(gtg)* and *trnV(tac)* in *L. fabalis* and *L. obtusata* do not show the loop in the Acceptor arm of *L. saxatilis*. The 22 cloverleaf structures are identical between *L. fabalis* and *L. obtusata*.

Sequence divergence, evolutionary rates and selection

The overall nucleotide divergence between mitogenomes is 67 % between *M. neritoides* and *L. saxatilis*, and 67.5 % between *M. neritoides* and both *L. fabalis* and *L. obtusata*. The genus *Melarhaphe* is clearly more distantly related to the genus *Littorina*, than are the *Littorina* species from each other (3.3 %) (Marques *et al.* 2017). The divergence between *M. neritoides* and *Littorina* sp. is nevertheless lower for single PCGs, which show the following divergence rank order from the fastest to the slowest evolving gene: *nad6* (36.1 %) > *nad2* (33.6 %) > *nad5* (30.9 %) > *nad4* (30.3 %) > *nad3* (29.9 %) > *atp8* (26.2 %) > *atp6* (25.9 %) > *nad1* (25.2 %) > *nad4l* (23.7 %) > *cox3* (22.6 %) > *cob* (22 %) > *cox2* (19.4 %) > *cox1* (19.2 %).

At the protein level, the proportion of amino acids differing between *M. neritoides* and *Littorina* sp. varies from 2 % in *cox1* to 37 % in *nad6* (Table 3). The most conservative genes between *Melarhaphe* and *Littorina* are *cox1* > *cox2* > *cox3* > *cob*, with ≤ 10 % of amino acid differences, followed by *nad1* > *nad4l* > *atp6* > *nad3* > *atp8* > *nad5* > *nad4* > *nad2* showing more than 10 %

amino acid differences, to the least conservative *nad6* gene. In mammals, rates of amino acid substitution have been linked to the stringency of structural and functional constraints in proteins (Tourasse & Li 2000).

The significant non-synonymous/synonymous substitution ratios for the concatenated PCGs reveal signatures of purifying selection (ω < 1) on the mitogenome of all five lineages *L. fabalis*, *L. obtusata*, *L. saxatilis*, the branch upstream the clade formed by *L. fabalis* and *L. obtusata*, and *M. neritoides* (Fig. 4, Table 6).
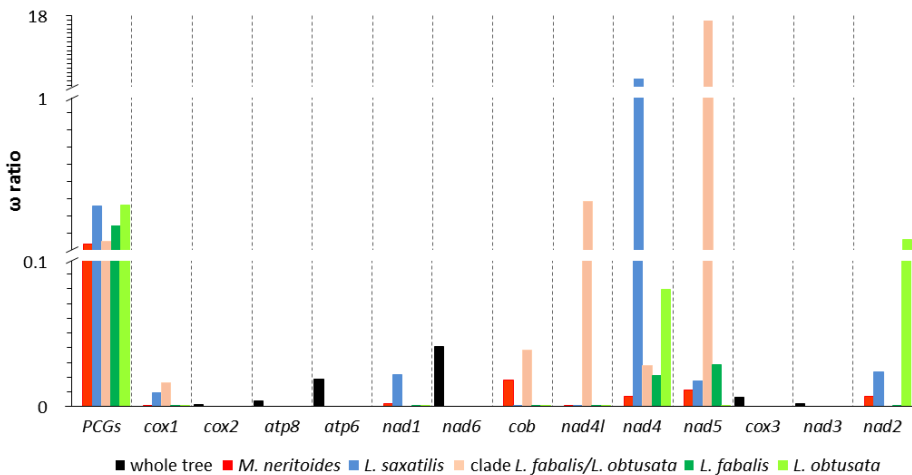
Fig 4. Evolutionary rates (ω) for each PCG among four littorinid mitogenomes.

The ratios are variable among lineages, indicating stronger purifying selection on the mitogenome of *M. neritoides* (ω = 0.1361) and on the branch leading to the clade *fabalis*/*obtusata* (ω = 0.1534) than on the mitogenome of *L. fabalis* (ω = 0.2438), *L. obtusata* (ω = 0.3667) and *L. saxatilis* (ω = 0.3595).

Table 6. **Results from the PAMLX analyses**. Models selected to best fit the data and best describe natural selection acting on protein-coding genes (PCGs) in *Melarhaphe neritoides*, *Littorina fabalis*, *Littorina obtusata* and *Littorina saxatilis*. Values marked in grey indicate positive selection.

Region	PCGs	cox1	cox2	atp8	atp6	nad1	nad6	cob	nad4l	nad4	nad5	cox3	nad3	nad2
ML tree	A	A	A	A	A	B	A	A	A	A	A	A	A	B
Model	1	1	0	0	0		0	1	1	1	1	0	0	1
LRT	34.75	5.55				3.87		4.14	6.91	7.17	11.10			8.56
ω_0			0.0018	0.0041	0.0185		0.0408					0.0062	0.0021	
$\omega_{L.\ fabalis}$	0.2438	0.0001				0.0001		0.0001	0.0001	0.0212	0.0287			0.0001
$\omega_{L.\ obtusata}$	0.3667	0.0001				0.0001		0.0001	0.0001	0.0798	0.0001			0.1638
$\omega_{clade\ fabalis/obtusata}$	0.1534	0.0164						0.0386	0.3827	0.0278	16.891			
$\omega_{M.\ neritoides}$	0.1361	0.0010				0.0019		0.0184	0.0010	0.0072	0.0115			0.0068
$\omega_{L.\ saxatilis}$	0.3595	0.0096				0.0218		0.0001	0.0001	3.5994	0.0177			0.0238

ML tree: A = ((*Littorina fabalis*,*Littorina obtusata*),*Melarhaphe neritoides*,*Littorina saxatilis*); B = (*Littorina fabalis*,*Littorina obtusata*,*Melarhaphe neritoides*,*Littorina saxatilis*).

Model: 0 = null model, 1 = branch model 1 (free-ratios model); the branch model 2 (two-ratios model) was never selected.

LRT: value of the likelihood ratio test for the comparison of the free-ratios model against the null model, significantly greater than the critical chi-square value for 1 degree of freedom with a significance level of 0.05 (χ^2 = 3.84).

ω: d_N/d_S ratio, for the null model (ω_0), or for the free-ratios model for each branch in the tree.

Yet, analyses conducted on single PCGs reveal disparity in terms of selection among genes and among lineages. On the one hand, most PCGs show ω values below 0.1, ranging from 0.0001 to 0.0798, thus indicating strong purifying selection on most PCGs in all five lineages, while more relaxed selective constraints act on *nad2* in *L. obtusata* (ω = 0.1638) and on *nad4l* in the clade *fabalis/obtusata* (ω = 0.3827). In *M. neritoides*, genes under purifying selection are ranked as follows, by decreasing strength of selection (increasing ω): *cox1*, *nad4l* > *cox2* > *nad1* > *nad3* > *atp8* > *cox3* > *nad2* > *nad4* > *nad5* > *cob* > *atp6* > *nad6*. Although this rank order differs from that in *Littorina* sp. and other organisms (Castellana *et al.* 2011), notably *cob* usually along the three *cox* genes for which purifying selection is the most efficient, the strength of purifying selection on *cob* in *M. neritoides* and the *Littorina* species lies in the range reported in other organisms such as chaetognaths (Marletaz *et al.* 2015), fishes (Sun *et al.* 2011), insects and some vertebrates (Castellana *et al.* 2011). On the other hand, *nad4* in *L. saxatilis* and *nad5* in the clade *fabalis/obtusata* show exceptionally high d_N/d_S ratios, which is significantly higher (ω = 3.5994 and ω = 16.8910 respectively) than that in other littorinids, and is indicative of strong positive selection on *nad4* and *nad5* in these two lineages. Positive selection has been reported on PCGs both in marine invertebrates and vertebrates (Bazin *et al.* 2006; da Fonseca *et al.* 2008; Longo *et al.* 2016; Ma *et al.* 2015), probably as adaptation to environment (i.e. thermal adaption, hypoxia tolerance) or to maintain mitonuclear coadaptation (Gershoni *et al.* 2009; Hill 2016). No positive selection is detected in *M. neritoides* lineage.

Purifying selection is a dominant evolutionary force acting on the mitogenomes of *M. neritoides*, *L. fabalis*, *L. obtusata* and *L. saxatilis*, and is expected to maintain crucial mitochondrial gene functions (Castellana *et al.* 2011), since mtDNA-encoded proteins are responsible for the oxidative phosphorylation. Additionally, strong positive selection on *nad4* and *nad5* suggests that these genes play an important role in mitogenome evolution in *L. saxatilis* and in the

clade *fabalis*/*obtusata*, and may promote adaptive divergence in ecotypes of *L. saxatilis* and between *L. saxatilis* and the clade *fabalis*/*obtusata*. Whether environmental or genetic factors contribute equally to adaptive divergence in the two lineages need to be addressed in further studies. In *M. neritoides*, our results from a previous study (Fourdrilis *et al.* 2016) consolidate a scenario in which mutation and genetic drift drive synonymous polymorphism in mtDNA while purifying selection drives non-synonymous polymorphism and preserves protein-coding gene functions.

Phylogeny of Littorinimorpha

The phylogeny, based on the complete set of mitochondrial protein-coding gene sequence data in 18 Littorinimorpha taxa, recovered the 8 families used in this study as distinct lineages within Littorinimorpha (Fig. 5). Hydrobiidae, Littorinidae, Naticidae, Pomatiopsidae and Vermetidae are more closely related to each other, than they are to Cassidae and Ranellidae or to Strombidae.

We have mapped gene orders in Littorinimorpha mitogenomes onto the Littorinimorpha phylogeny to reconstruct the evolutionary history of mitochondrial PCGs gene order rearrangements in Littorinimorpha. Gene order is conserved within Littorinidae. *Melarhaphe neritoides* is grouped with the three *Littorina* sp., and is clearly not affiliated to one of the other families included in the phylogeny. The phylogenetic position of *M. neritoides* within Littorinidae was hitherto not well-resolved or supported (Reid *et al.* 2012). Yet, branch support values in the present reconstruction are strong, so that the mitogenome data provide the first strong evidence for assigning *M. neritoides* to Littorinidae. Gene order is also conserved among Littorinimorpha among nearly all littorinimorph, viz. Cassidae, Hydrobiidae, Littorinidae, Naticidae, Pomatiopsidae, Ranellidae and Strombidae, but is altered in Vermetidae. For example, in the four vermetids, *cox2*, *nad5*, *cox3* and *nad2* have been

translocated while *nad1* and *nad6* have been inverted in *Dendropoma gregarium*. In addition, several tRNAs are also translocated (A, K, N, P, R, T), inverted (L1 and L2, S1 and S2), duplicated (K, L2) or encoded on the opposite strand (T), in all or some of the vermetids. Gene order is not unique to Littorinimorpha and we observe a shared gene order between Littorinimorpha and Neogastropoda (based on three families, viz. Buccinidae, Conidae and Nassariidae), confirming the observations of Rawlings et al. (2010) based on 6 complete plus 4 incomplete Littorinimorpha mitogenomes. Considering gene order and phylogenetic data, only the major rearrangements described by Rawlings et al. (2010), i.e. translocations of tRNAs and PCGs and one gene inversion, has taken place in the evolutionary history of Littorinimorpha, along the lineage leading to the Vermetidae, whereas no such rearrangements occurred in the other Littorinimorpha lineages.
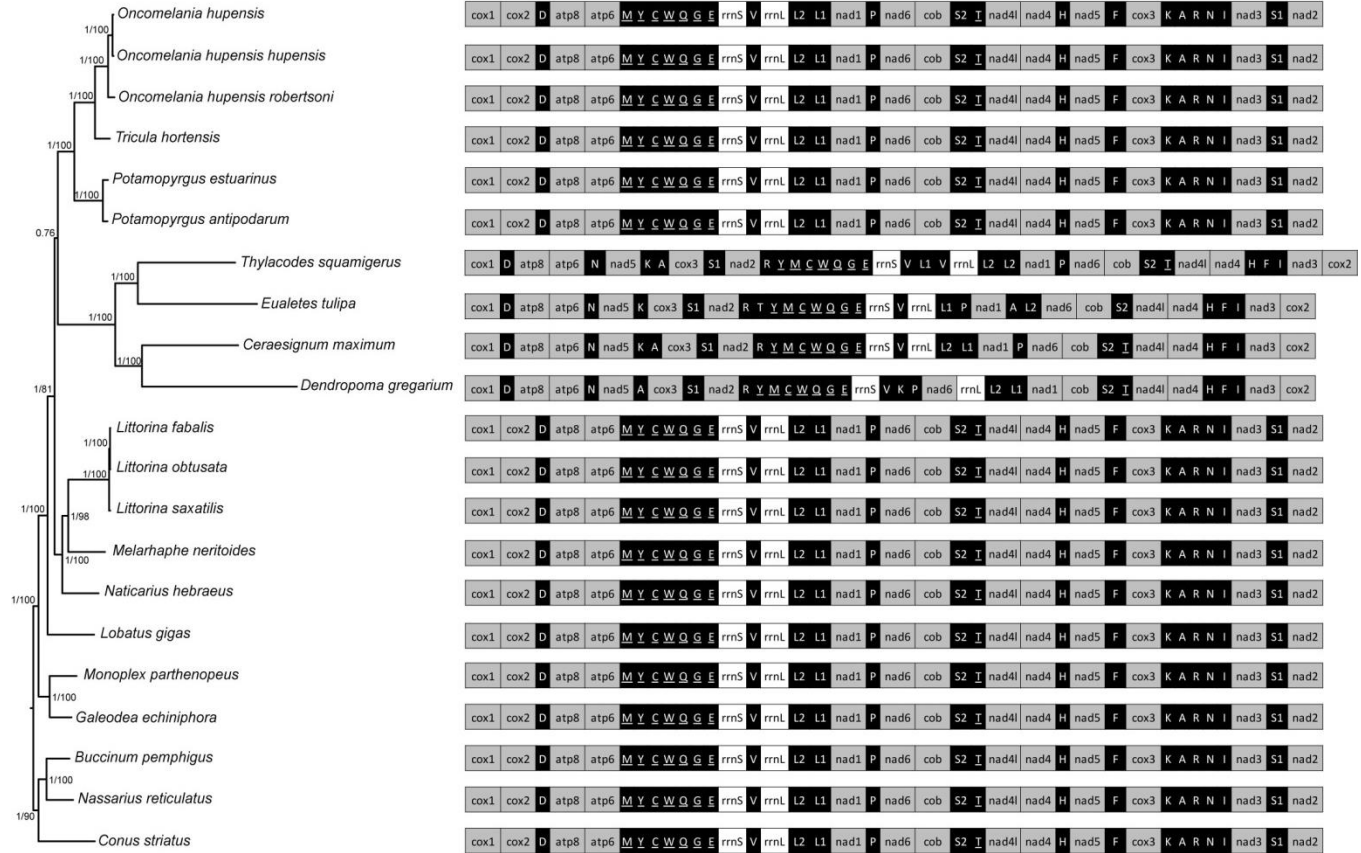
Fig 5. **Phylogenetic tree inferred from the sequences of 13 protein-coding genes of the mitogenomes of 17 Littorinimorpha and 3 outgroups.** Numbers at the nodes are Bayesian posterior probabilities (left) and ML bootstrap values (right). Branches with posterior probability > 0.95 and bootstrap support value > 70 % are considered to be strongly supported. Symbol letters for tRNAs indicate the encoded amino acid and follows the IUPAC-IUB nomenclature for amino acids. Underlined tRNAs are encoded on the L strand.

ACKNOWLEDGEMENTS

General Discussion

My investigation into the characteristics of the mtDNA in *Melarhaphe neritoides* revealed several mechanisms shaping the evolution of this periwinkle species: an elevated mutation rate, positive selection and high rates of migration. It also brought more insights into the evolution of the organelle, driven by strong purifying selection, and whose genome organization is similar to other Littorinimorpha with the exception of Vermetidae and of the *trnM(cat)* secondary structure lacking the TΨC -loop.

✓ **How much mtDNA genetic diversity does *M. neritoides* harbour?**

By measuring genetic diversity in *M. neritoides*, we could reveal the remarkable amount of intraspecific mtDNA diversity, and characterise mtDNA hyperdiversity, with respect to the concatenated 16S-COI-Cyt*b* gene fragment:

- a very high haplotype diversity (Hd = 0.999 ± 0.001) and nucleotide diversity (π = 0.013 ± 0.001) in the Azores archipelago (Chapter 1), observed as well at a wider scale in the North East Atlantic (NEA) (Chapter 2).

- a neutral nucleotide diversity π_{syn} ≥ the threshold of 5 % (π_{syn} = 6.8 %), indicative of **mtDNA hyperdiversity**. This mtDNA hyperdiversity is not a local characteristic but is prevalent over the NEA.

This mtDNA diversity has been validated to **reflect natural variation**, not PCR errors, nor the presence of cryptic taxa or the admixture of divergent and/or subdivided local populations. Deep divergence of about 6 % and mtDNA hyperdiversity (π_{syn} = 18.4 % at COI locus, calculated from data in Chapter 1) have been found in the littorinid *Echinolittorina vidua* without morphological differentiation, providing another case of elevated intraspecific sequence divergence (Williams & Reid 2004). Therefore, mtDNA hyperdiversity reflects

the **upper boundary of the range of values that intraspecific genetic diversity can take**.

The BOLD identification engine yields correct specimen identification to the species level when the percentage of similarity between the queried sequence and the barcodes stored in the database is 99 % or higher, and when all queried sequences are clustered into the same Barcode Index Number (BIN). If the taxon if absent from the database, a record match to a distant species will be found with a low similarity score (80 %), or no record match will be displayed. DNA barcoding returned 100 % correct species-level specimen identification of *M. neritoides* despite an elevated intraspecific COI divergence in this species (1.8 %), two times higher than in other Littorinoidea (e.g. 0.6 % in *Tectarius striatus*). High intraspecific COI divergence may lead to a barcode overlap between intra- and interspecific genetic distances, which results in a lower rate of identification, as reported in northwestern Pacific molluscs showing two times higher intraspecific COI divergence than other marine groups of the same area (Sun *et al.* 2016), in closely related arthropod species (Elias *et al.* 2007; Meier *et al.* 2006), and in too recently diverged lineages (Hickerson *et al.* 2006). Therefore, BOLD identified the 185 sequences (dataset 1, Chapter 1) as *M. neritoides*, but this cannot be taken as a definitive proof of monophyly because the hypothesis of several cryptic species showing intraspecific COI divergences overlapping with the 51 existing barcodes of *M. neritoides* in the reference database cannot be ruled out.

DNA barcoding relies on the premise that COI divergence in metazoans is one order of magnitude higher among species (typically 10-25%) than within species (typically < 3 %) (Hebert *et al.* 2003b). This has been further validated in 20371 vertebrate and invertebrate species by Kartavtsev (2011), estimating interspecific genetic divergence to 3.78-20.57 % (3.78 % among subspecies/sibling species, 11.06 % among species from the same genus, and

20.57 % among species from separate families), and intraspecific genetic divergence (p-distance) to 0.89 % among populations within species. Although the use of a distance threshold is arbitrary, it is necessary when identifying specimens with genetic data. However, assuming taxa are distinct species when the ratio between mean inter- and intraspecific divergence is ≥ 10 (Hebert *et al.* 2004b) at the COI locus does not always hold true (Meier *et al.* 2006; Virgilio *et al.* 2010). Sibling species are, when not yet recognised, cryptic species, which possibly show a ratio < 10 according to the figures in Kartavtsev (2011). Nevertheless, COI has proven efficient in revealing cryptic diversity e.g. in the butterfly species *Astraptes fulgerator* (Hebert *et al.* 2004a) or in a genus of parasitoid flies (Smith *et al.* 2006), and the barcoding gap should not be considered as a predictor of the identification success (Lou & Brian Golding 2010; Meier 2008; Meier *et al.* 2008; Ross *et al.* 2008).

The exhaustiveness of the coverage across different populations and geographic regions of the reference barcode database, to which the query sequence is compared, determines the reliability of DNA barcoding (Barco *et al.* 2016). Hence, the non-exhaustive sampling of barcodes in BOLD is another drawback. Less than 10 % of the estimated number of marine species are represented in BOLD (Barco *et al.* 2016). For marine molluscs, the number of species with available barcodes in 2011 represented only 3-3.5 % of the total number of marine mollusc species (Appeltans *et al.* 2012). The absence of conspecific sequences in the reference database biases DNA barcoding (Virgilio *et al.* 2010), as would do the addition afterwards of new barcodes genetically closer to the query sequences but defined as another species by the BOLD user who uploaded these barcodes. However, no such bias is expected in the barcoding of *M. neritoides* specimens with 51 barcodes of *M. neritoides* (http://www.boldsystems.org/index.php/Taxbrowser_Taxonpage?taxid=81082) and 1781 conspecific barcodes from other genera of the Littorinidae family (http://www.boldsystems.org/index.php/Taxbrowser_Taxonpage?taxid=1767), which includes

the closest genus to *Melarhaphe* (i.e. *Lacuna*), in the BOLD reference database. There are no other barcodes from the genus *Melarhaphe* since *M. neritoides* is the only species in this genus.

Future prospects

A literature survey of COI nucleotide diversity values in Chapter 1 suggests that mtDNA hyperdiversity may be more common across other phyla than currently appreciated.

Investigating potential candidate species for mtDNA hyperdiversity could help assess in which extent mtDNA hyperdiversity is common among eukaryotes.

The highest values of π_{syn} could help define the upper range of values of intraspecific mtDNA genetic diversity, in absence of speciation.

✓ **Are there biological and phylogenetic features correlating with mtDNA hyperdiversity?**

Is there a correlation between species attributes and mtDNA hyperdiversity?

In the list of species in Table 2 in Chapter 1, no attribute in particular (phylum, larval dispersal mode, habitat, climatic zone) seems to be shared exclusively among species putatively showing mtDNA hyperdiversity (π from 10^{-2}) and the mollusc species for which mtDNA hyperdiversity is known ($\pi_{syn} > 5$ %). Hence, **none of the species attribute investigated seems to correlate well with mtDNA hyperdiversity**.

Is there a correlation between mtDNA hyperdiversity and rate of sequence evolution?

Among the 27 species of Littorinidae surveyed in Table 2 in Chapter 1, *Echinolittorina vidua* is the only species with *M. neritoides* showing mtDNA hyperdiversity. Additionally, *Mainwaringia rhizophila* and *Mainwaringia leithii* show an accelerated rate of molecular evolution at two nuclear and three mitochondrial loci relative to other Littorinidae (Reid *et al.* 2012; Reid *et al.* 1996; Williams *et al.* 2003) and could therefore be a potential candidate for hyperdiversity too, although available sequence data do not allow us to calculate π_{syn}.

From a phylogenetic point of view, *M. neritoides* appears to be widely divergent of the Littorininae subfamily (Fig. 3 in General Introduction). The two species *Mainwaringia rhizophila* and *Mainwaringia leithii*, are also widely divergent members of the Littorininae on long branches, but belong to Littorininae (Reid *et al.* 2012; Williams *et al.* 2003). In contrast, *Echinolittorina vidua* and the whole genus *Echinolittorina* show a sister group relationship with the genus *Littorina*, and are not divergent of the Littorininae. Thus, mtDNA hyperdiversity occurs in two Littorinidae species, and potentially in two others, which are divergent or not divergent of other Littorinidae species. Therefore, **phylogenetic divergence does not correlate with mtDNA hyperdiversity**.

<u>Future prospects</u>

Does high *µ* correlate with high substitution rate?
Estimating mutation rates in hyperdiverse mtDNA from several species would allow us to explore the link between mutation rate and substitution rate.

✓ **Which factors are responsible for mtDNA hyperdiversity in *M. neritoides*?**

By evaluating the role of mutation, selection and effective population size in *M. neritoides*, I revealed that **an elevated mutation rate** at the COI locus ($\mu = 5.82 \times 10^{-5}$ per site per year or $\mu = 1.99 \times 10^{-4}$ mutations per nucleotide site per generation), 10^3 to 10^4 fold higher than usually estimated for other organisms (see Table 3 in Chapter 1), is likely the **primary force for generating mtDNA hyperdiversity**.

Selection is a second force, which influences mtDNA hyperdiversity. In Chapter 1, the Fay & Wu's (2000) H statistic shows that positive selection is acting on 16S, COI, and Cyt*b* genes. In Chapter 3, d_N/d_S ratios indicate strong purifying selection on the 13 protein-coding genes of the *M. neritoides* mitogenome. These two tests of neutrality show different things. Fay & Wu's (2000) test deals with overall mtDNA polymorphism, without distinguishing between synonymous and non-synonymous substitutions, and tells us that a significant proportion of nucleotides is variable, and that these polymorphic nucleotides show a frequency higher than expected under neutral evolution and are consequently positively selected. Polymorphic sites comprise synonymous sites and therefore, **some synonymous sites are possibly under positive selection**. The d_N/d_S ratios tell us about the impact of this mtDNA polymorphism at the protein level, and reveal that a very small proportion of this mtDNA polymorphism induces changes in amino acid sequences of the proteins encoded by mitochondrial genes.

Hyperdiversity estimation relies on π_{syn} calculated at synonymous sites, which are assumed to be neutral. However, synonymous sites might not be neutral but under micro-evolutionary selection pressures, because a change of nucleotide base leads to a different codon and some codons are more accurately and/or efficiently translated than others (Cutter & Charlesworth 2006; Hershberg & Petrov 2008). Weak selection for codon usage bias leading to non-neutral synonymous sites has been observed in the nematode

Caenorhabditis remanei (Cutter 2008), as well as strong purifying selection on synonymous sites in the arthropod *Drosophila melanogaster* (Lawrie *et al.* 2013).

Fay & Wu's (2000) test does not inform about the type of positive selection, which can be stabilizing selection or disruptive selection. Stabilizing selection leads to allele fixation and will reduce genetic diversity, whereas disruptive selection favours both types of extreme (rare) phenotypes over intermediate phenotype and has a diversifying effect that will increase genetic diversity (Rueffler *et al.* 2006; Thoday 1959).

To conclude in *M. neritoides*, **positive selection shapes (maintains or reduces) overall polymorphism in mtDNA and thus also neutral polymorphism in mtDNA (mtDNA hyperdiversity)**, and **strong purifying selection reduces non-synonymous polymorphism and therefore maintains a low polymorphism in mtDNA-encoded proteins**.

Investigating taxa with hyperdiverse DNA helps identify micro-evolutionary selective pressures on non-coding elements and the limits of natural selection.

The effective population size was also investigated as potential determinant of mtDNA hyperdiversity. Despite evidence that *M. neritoides* underwent demographic expansion in a recent past (in Chapter 1: significant negative Tajima's *D* and Fu's *Fs*, unimodal curve of the sequence mismatch distribution, non-significant values of the sum of squared deviations and Harpending's Raggedness index, positive slope of the Bayesian Skyline Plot), the effective population size of *M. neritoides*, estimated by means of a Bayesian Skyline Plot, was found to be surprisingly small in the Azores archipelago (N_e = 5256; CI = 1312-37495) for such a planktonic-dispersing organism with high-dispersal potential (see comparison with other planktonic-dispersing organisms in Table 4 of Chapter 1). The effective population size of *M. neritoides*, using the Bayesian approach based on the coalescent implemented in MIGRATE-N in

Chapter 2, was also found to be relatively small in the NEA (N_e = 1303; CI = 1119-1487). Selection likely influences N_e estimation, sweeping mutations to fixation and thereby squeezing ancestry through fewer individuals, which reduces N_e through time. Hence, **N_e is not positively correlated with mtDNA hyperdiversity**. Therefore, **N_e is a poor indicator for inferring the presence of mtDNA hyperdiversity**, as suggested by Bazin *et al.* (2006).

Future prospects

Further studies are needed to investigate the potential origins and causes of the high mtDNA *µ* underlying mtDNA hyperdiversity.

- Is a high mtDNA *µ* specific to the lineage leading to the genus *Melarhaphe* and to the species *M. neritoides*, i.e. after the split between *Melarhaphe* and the other Littorininae genera?

The age of the earliest fossil of *M. neritoides* is recent and dated between 130 and 120 thousand years ago (Ávila *et al.* 2002). But a fossil of another species of the genus *Melarhaphe* is known, *Melarhaphe mausseneti*, now extinct, and dated to 55 Ma (CI = 55.8-58.7) (Reid 1989). Based on this fossil age amongst other time calibrations, Reid *et al.* (2012) calibrated the phylogeny of Littorininae and dated the genus *Melarhaphe* to 130 Ma (CI = 102-161), what corresponds to the split between *Melarhaphe* and the Littorininae subfamily. The genus *Melarhaphe* is older than the oldest fossils of the genera from the Littorininae subfamily (Reid *et al.* 2012), and is the earliest genus to have diverged from the rest of the Littorininae. We may wonder whether a high mtDNA *µ* is a specificity of *M. neritoides* and of the genus *Melarhaphe*.
Estimating the mtDNA *µ* in *Echinolittorina vidua* and in the species from the genus *Mainwaringia* could help addressing this question.

The presence of high mtDNA μ in *E. vidua* and/or in *Mainwaringia* sp., would suggests that an elevated μ is not a specificity of the mtDNA of *M. neritoides* and of the genus *Melarhaphe*, but a trait shared among several Littorinidae. This hypothesis would hold true until an elevated μ is found in species with hyperdiverse mtDNA in other taxonomic levels.

The absence of high mtDNA μ in the other Littorininae would suggest that an elevated μ is a synapomorphy of the genus *Melarhaphe*, or even an AUTAPOMORPHY of *M. neritoides*. However, distinguishing between synapomorphy and autapomorphy is not possible since all other species of *Melarhaphe* are extinct.

- Is a high mtDNA μ caused by a high error rate in mtDNA replication?

DNA repair efficiency, and species generation time, influence the mutation rate (Bromham 2008). Polymerase enzymes that copy DNA can vary greatly in their error rate and species can differ in their repair efficiency (Bromham 2008). Moreover, species with short generation times, such as mice, tend to have faster rates of molecular evolution than species with longer generation times like humans (Bromham 2008). With a generation time of 41 months, *M. neritoides* possibly shows a high number of mtDNA replications per unit time, and therefore a higher chance of acquiring copy errors in mtDNA. By purifying the mitochondrial polymerase γ in *M. neritoides*, assessing its error rate and evaluating it against that of other littorinids, further studies could evaluate the role of the error rate of the polymerase γ in contributing to a high mtDNA μ.

Furthermore, selection may play a role in shaping DNA repair rates, allowing the evolution of higher or lower mutation rates (Bromham 2008). Further studies should assess the direction and magnitude of selection on the *polg* gene encoding the mitochondrial polymerase γ.

- Is a high mtDNA μ caused by mutagenic DNA methylation?

In many eukaryote genomes, a cytosine that is next to a guanine and forms a cytosine-phosphate-guanine (CpG) dinucleotide is far more likely to be methylated than a cytosine not within CpG, for example, 2.59 times more in honeybee (Flores *et al.* 2013). Methylated cytosine spontaneously deaminates to thymine, inducing the replacement of guanine by adenine on the opposite DNA strand and hence the depletion of CpG (Duncan & Miller 1980). For instance, in humans, a C to T (C↔T) transition at methylated cytosines occurs at a rate 10- to 50-fold higher than any other mutation (Flores & Amdam 2011). Methylated genomic regions are therefore subject to an increased mutation rate (Flores *et al.* 2013). Depletion of CpG occurs in genomic regions that are targeted for consistent methylation over several consecutive generations, and hence is a signature that these regions have been methylated over evolutionary time (Flores & Amdam 2011).

Delineating the methylome of *M. neritoides* to assess the pattern of DNA methylation in hyperdiverse mtDNA would help to better understand the contribution of DNA methylation to a high mtDNA μ.

Preliminary data in the overall population of *M. neritoides* in the NEA, using the 399 specimens from Chapter 2, show high C↔T transition rates at the COI locus ($R = 63.07$) and the combined 16S, COI and Cyt*b* loci ($R = 42.96$), which are respectively the highest and the second highest (< A↔G transition rate) of the six rates. However, these data do not allow to distinguish the C↔T transition rate occurring at cytosine nucleotide from that at methylated cytosine.

I propose to acquire whole-genome methylation data at a base-pair resolution, using the commonly used 'bisulfite genomic sequencing' procedure (Frommer *et al.* 1992), to quantify methylated cytosines within CpG dinucleotides in the mitogenome of *M. neritoides* as well as in another species of Littorinidae for comparison. This sequencing procedure enables to calculate the rate of C↔T

transitions at methylated cytosines. An increased transition rate in *M. neritoides* would suggests that DNA methylation contributes to increase *µ* in hyperdiverse mtDNA. Besides, methylome data also allows to locate higher percentages of methylated CpGs within the mitogenome and determine the regions (e.g. to exons, intergenic regions) which have been the most methylated over evolutionary time.

✓ **Does mtDNA hyperdiversity reflect population genetic differentiation and structuring in *M. neritoides* despite high dispersal potential in this planktonic-dispersing species?**

By calculating indices of population genetic differentiation, to assess the influence of genetic drift and the pattern of genetic structure among populations of *M. neritoides*, we observed the absence of genetic differentiation and structure, and revealed **panmixia in *M. neritoides*** throughout the NEA (Chapter 2). Furthermore, quantification of gene flow among populations of *M. neritoides* revealed **high rates of gene flow** across its distribution area in the NEA, and **predominantly eastward** gene flow.

These results highlight an important **pitfall with respect to the use of hyperdiverse mtDNA markers in assessing population genetic differentiation and connectivity**. Hyperdiverse mtDNA markers may easily lead to erroneous interpretations of differentiation statistics and haplotype network bush-like patterns. First, very high mtDNA haplotype richness requires large and unrealistic sampling efforts to sample the total mtDNA haplotype richness, hence sample size is too low to detect shared mtDNA haplotypes. Differentiation statistics such as D_{EST} may reach a maximal value of 1, but are not indicative of population genetic differentiation as expressed in terms of fixation of haplotypes, and only reflect population genetic differentiation as expressed in terms of lack of haplotype sharing. Second, the very high

mutation rate causes an apparent lack of shared mtDNA haplotypes among populations and conceals the signal of gene flow in haplotype network pattern.

Future prospects

- Investigating population genetic connectivity in the nuclear DNA (nDNA) of *M. neritoides*

The nuclear 18S and 28S gene have been used in previous phylogenetic studies and worked for *M. neritoides* (Reid *et al.* 2012; Williams *et al.* 2003). The two nuclear ITS1 and ITS2 genes were sequenced in a few specimens of *M. neritoides* in this PhD thesis (data not shown) but did not provide convenient markers due to the presence of multiple binding. Nevertheless, cloning may successfully separate ITS1 and ITS2 alleles within individual and provide relevant nuclear markers for inferring the genetic connectivity pattern of nDNA in *M. neritoides*. Development of nuclear markers from myoglobin (Mb) genes can also possibly offer relevant nuclear markers although possibly lacking of sufficient variability at the intraspecific level (Brito *et al.* 2001; De Wolf *et al.* 1998; Medeiros *et al.* 1998; Olabarria *et al.* 1998).

Nuclear markers are assumed to be less variable than mitochondrial markers, and should show genetically homogeneous populations and confirm the panmixia pattern found in *M. neritoides* in the NEA based on mtDNA. However, a significant deviation in the sex ratios of *M. neritoides*, with females more frequent than males, was observed on some shores in Ireland (e.g. 1:1.94 in Portnakilly, p < 0.001; 1:1.54 in Lekkycranny, p < 0.01) (Cronin *et al.* 2000), in England (Fretter & Graham 1980) and in France (Daguzan 1976). Nevertheless, approximately equal sex ratios were observed in Ireland (Cronin *et al.* 2000) and Israel (Palant & Fishelson 1968). It is therefore interesting to

examine whether the pattern of connectivity based on nDNA would be similar to the connectivity pattern of the organelle.

- Is there a mirror of mtDNA hyperdiversity in nDNA?

Measuring genetic diversity in protein-coding nuclear genes will allow to see whether π_{syn} is greater than 5 %, the threshold above which nDNA would show hyperdiversity. Amongst the nuclear genes mentioned in the previous paragraph, 18S, 28S, ITS1, ITS2 and Mb, only Mb is protein-coding.

- Is there signs of variability at the phenotypic level?

Sedentary adults of benthic invertebrates, such as rocky-shore periwinkles like *M. neritoides*, are subjected to habitat-specific pressures, and the traits of their shell may respond to environmental factors such as wave action, temperature, salinity, desiccation risk, substrate colour, or predation by crabs (Johannesson 2003). Morphological polymorphism in shell shape has been found in *M. neritoides* among populations of the Iberian Peninsula, likely caused by phenotypic plasticity in absence of mtDNA genetic divergence (Cuña *et al.* 2011; García *et al.* 2013; Queiroga *et al.* 2011).

I observed a shell-colour polymorphism in *M. neritoides* across its distribution area, varying from black to brown or grey (Fig. 1). High rates of population connectivity in *M. neritoides* found in Chapter 2 show that planktonic-dispersing larvae have to cope with multiple adult remote habitats and rock colours over the NEA at the destination site where they settle (Fig. 2).
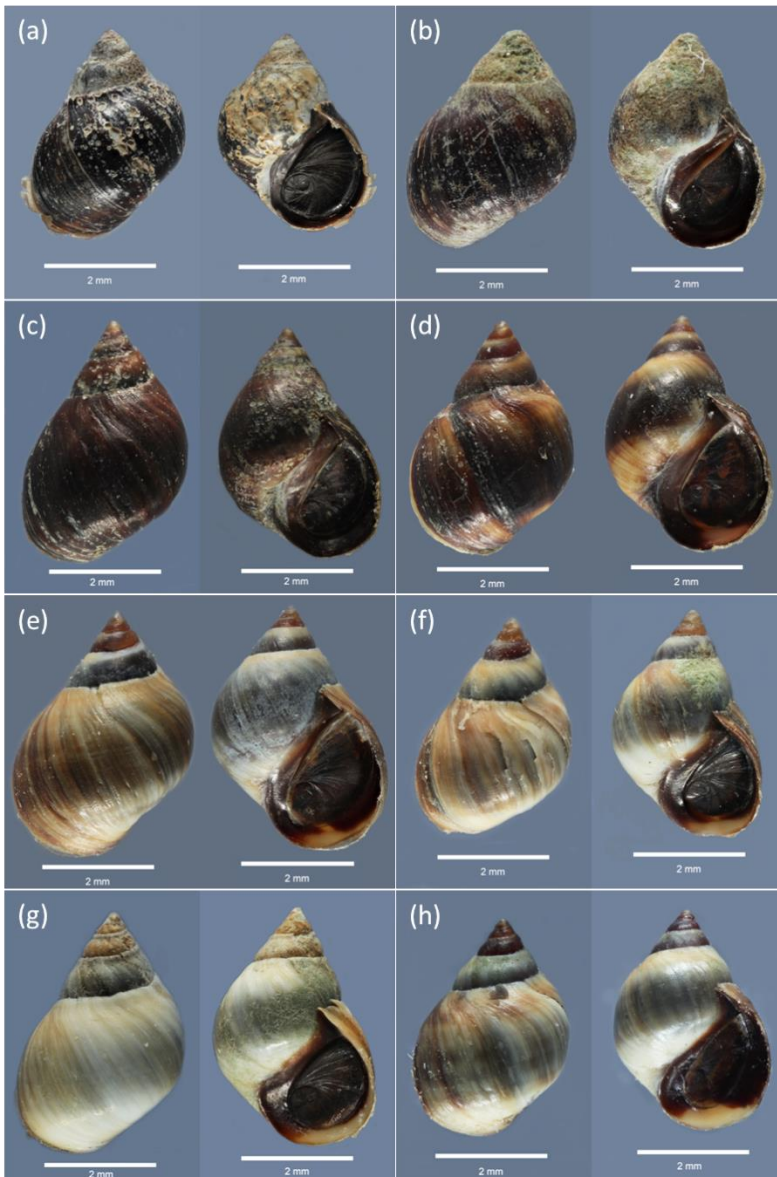
Figure 1. Shell-colour polymorphism in specimens of *M. neritoides* from (a) Sweden; (b) France; (c) (d) and (e) Azores, Pico island (same sampling site); (f) Tunisia; (g) Italy; (h) Greece, Rhodes island.
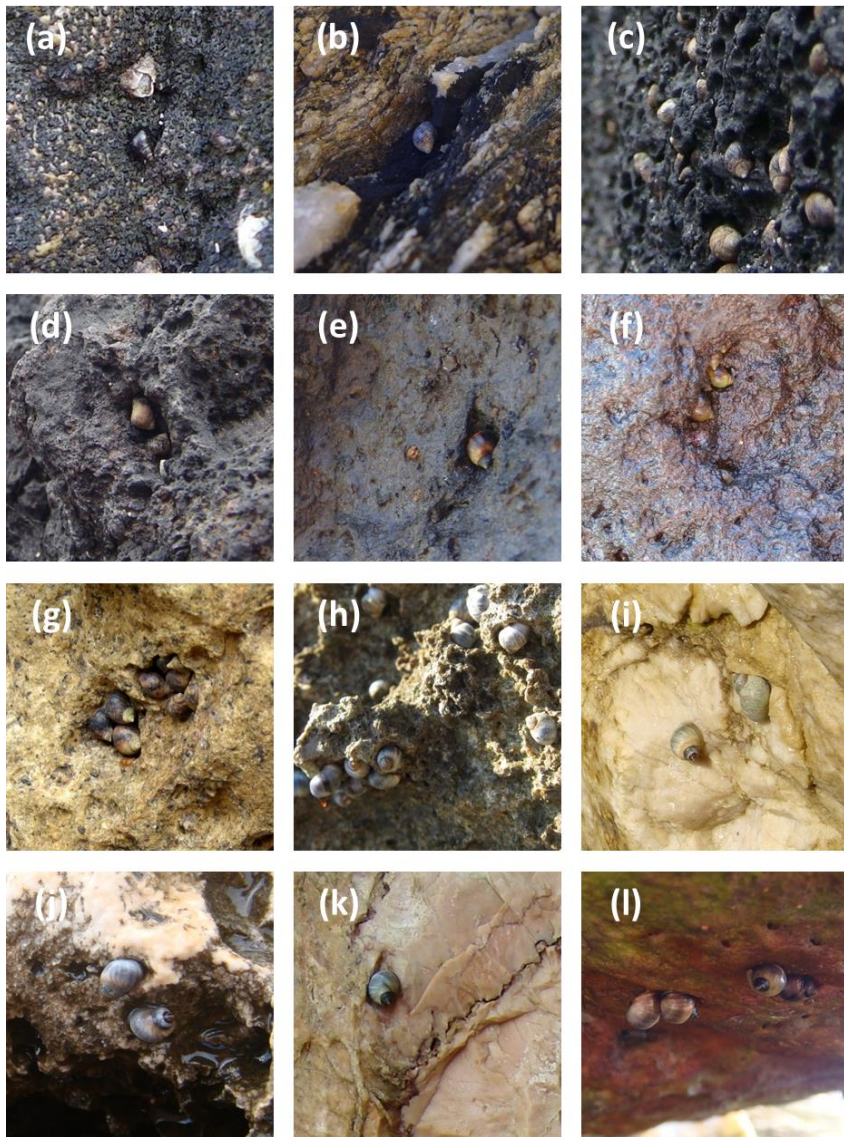
Figure 2. Various colours of rock substrate in habitats of *M. neritoides* in (a) Ireland, Ballyconneely; (b) France, Moëlan-sur-Mer; (c) Azores, Pico island; (d) (e) (f) (g) Azores, São Miguel island; (h) Italy, Tricase Porto; (i) (k) (l) Greece, Rhodes island; (j) Greece, Chalki island.

I hypothesise that this **spatial variation in shell colour might be associated to the colour of the rock substrate in the adult habitat**. Furthermore, high rates of population connectivity imply that the gene pool is homogeneous in *M. neritoides* over its entire distribution, because gene flow mixes alleles and hinders local adaptation based on genetic divergence, suggesting that the **shell-colour polymorphism in *M. neritoides* might be more plastic than genetically-determined**. However, phenotypic plasticity might also have a genetic basis (Agrawal 2001; Fusco & Minelli 2010).

Future studies require the identification of gene(s) involved in metabolic pathways responsible for shell colour and pigmentation (Williams 2016), likely in the nuclear genome, and to carry out comparative analyses at the given locus/loci among habitats differing in their colour substrate, in order to investigate the role of phenotypic plasticity in the shell-colour polymorphism of *M. neritoides*.

✓ **Does mtDNA hyperdiversity induce atypical features in the whole mitogenome of *M. neritoides*?**

By exploring the composition and structure of the entire mitogenome of *M. neritoides* in Chapter 3, and estimating the direction and strength of selection on the complete set of protein-coding genes (PCGs), I identified a major difference in the secondary structure of the transfer RNA *trnM(cat)* which lacks the TΨC-loop, and the action of strong purifying selection as expected globally on the 13 PCGs. Otherwise, the mitogenome of *M. neritoides* shows comparable composition and organisation to the three other Littorinidae species *Littorina fabalis*, *Littorina obtusata* and *Littorina saxatilis*, and to other Littorinimorpha with the exception of Vermetidae.

I concluded that **mtDNA hyperdiversity and the underlying high *μ*, associated to genetic drift and positive selection at the DNA level, drive**

synonymous polymorphism in the mtDNA of *M. neritoides*, while **purifying selection drives non-synonymous polymorphism at the protein level** and preserves protein-coding gene functions.

However, in Chapter 1, μ was estimated at the COI locus only and was not known in the other PCGs. The COI gene undergoes the strongest purifying selection pressure and therefore non-synonymous polymorphism is strongly hindered. Yet, COI shows a very high degree of neutral polymorphism (mtDNA hyperdiversity). This suggests that all other PCGs, which are under weaker purifying selection and carry more non-synonymous polymorphism, might also harbour high neutral polymorphism and possibly high μ.

Future prospects

- Are the other PCGs in *M. neritoides* mitogenome hyperdiverse too?

Interestingly, the H strand remains single-stranded for a long time during mtDNA replication, and is only partially protected by single-stranded-mtDNA-binding proteins (Reyes *et al.* 1998). During this time, the H strand is exposed to nonenzymatic methylation, hydrolytic and oxidative damage, and is thus prone to mutations (Brown & Simpson 1982). The longer the single-stranded state, the higher the probability of mutations. The time is the longest in *cox1*, ca. two hours, and the shortest in Cyt*b*, ca. 80 min (Clayton 1982). In mammals, the time decreases as follows: COI < COII < ATP8 < ATP6 < COIII < NAD3 < NAD4L < NAD4 < NAD1 < NAD5 < NAD2 < Cyt*b*.

I hypothesise that μ would be high in COII in *M. neritoides* and would decrease in the other PCGs following the same order as the duration of single-stranded state. Extracts of genomic DNA from the specimens of *M. neritoides* used for μ estimation in Chapter 1 are available at RBINS, and could be used to sequence the other PCGs and to estimate their μ.

- What are the phylogenetic relationships of the mitogenome of *M. neritoides* with the other members of the Littorinidae family?

In Chapter 3, the mitogenome of *M. neritoides* could be compared to three other species of Littorinidae, viz. *Littorina fabalis*, *Littorina obtusata* and *Littorina saxatilis*, as well as 17 other species of the Littorinimorpha infraorder. Extending the comparative mitogenomic analysis to other Littorinidae would allow us to assess how similar is the mitogenome of *M. neritoides* to closely related species at the family level, and to shed light on the architecture and evolution of mitogenomes in the Littorinidae family.

Mitochondrial gene rearrangements are stable across major taxonomic groups, with the exception of Mollusca showing unusually high numbers of gene order changes (Boore 1999). At the class level, Gastropod mitogenomes show relatively high rates of gene rearrangement between major lineages (Grande *et al.* 2008). This even extends to the family level, such as the Vermetidae showing extensive gene order changes (Rawlings *et al.* 2010).

Mitochondrial gene order is a useful phylogenetic tool (Boore & Brown 1998). A comparison of gene order and of secondary tRNA structures within Littorinidae could bring more insights into possible structural DNA rearrangements among members of the family. For example, the presence or lack of the TΨC-loop in *trnM(cat)*, or a particular gene arrangement in *Mainwaringia rhizophila* possibly correlated with its elevated rate of sequence evolution.

As a by-product of the mitogenomics analysis, the unresolved phylogenetic position of *M. neritoides* within Littorinidae might gain support. Complete mitogenome sequences and mitogenome arrangements can be used to reconstruct robust phylogenies if applied at the proper taxonomic level (Osca *et al.* 2015; Stöger & Schrödl 2013; Williams *et al.* 2014).

In another perspective of reconstructing the phylogeny of Littorinidae by Reid *et al.* (2012), adding nuclear loci to the original dataset, composed of the nuclear 28S gene and the two mitochondrial 12S and COI genes, may improve the resolution of the relationships in the family.

- What are the implications of mtDNA hyperdiversity for mitonuclear coevolution?

Hyperdiverse intraspecific mtDNA variation provides a greater density of polymorphic sites for selection to act upon (Cutter *et al.* 2013), and possibly provokes higher speciation rate as observed in birds and reptiles (Eo & DeWoody 2010). Studying mtDNA hyperdiversity is hence interesting to better understand how evolutionary processes such as mutational dynamics and selection that underlie mitonuclear coevolution contribute to speciation (Burton & Barreto 2012). As rapid evolution of the mtDNA results in significant selection pressure for CO-ADAPTATION amongst nuclear genes that produce proteins that function within the mitochondria, we might expect to see elevated μ amongst these interacting nuclear genes. Investigating the presence of hyperdiversity and of elevated d_N/d_S ratios in *M. neritoides* in nuclear genes coding for mitochondrial proteins involved in the respiratory chain, would help understanding mitonuclear coevolution in invertebrates and evolutionary forces shaping it.

Glossary

AUTAPOMORPHY

An autapomorphy is a derived characteristic unique to a given taxon and not shared with other taxa.

CO‑ADAPTATION

Selection by which harmoniously interacting genes accumulate in the gene pool of a population.

ECOSYSTEM SERVICES

Ecosystems provide a range of services that are of fundamental importance to human well-being, health, livelihoods and survival. Ecosystem services are categorized under 4 types:

- provisioning services that maintain the supply of natural products (e.g. supply of food, timber, water)
- regulatory services that keep different elements of the natural world running smoothly (e.g. filtration of air pollution, moderation of climate, storage of carbon, recycling of dead organic matter)
- supporting services that maintain the provisioning and regulatory services (e.g. photosynthesis, soil formation, provision of healthy habitat for species and genetic diversity)
- cultural services that are obtained from contact with nature (e.g. aesthetic and spiritual benefits) during recreational activities (e.g. hiking, bird watching, gardening)

EFFECTIVE POPULATION SIZE

The theoretical size of an idealized population that has the same genetic properties as those observed in a real population. An example of genetic property is the mean coalescent time.

GENE

A coding sequence of DNA, which determines a function of an organism and exist in several versions named alleles.

GENE FLOW

The exchange of genes between populations, which are usually of the same species.

GENETIC CONNECTIVITY

Refers to the exchange of migrants among geographically separated subpopulations that comprise a metapopulation, encompasses the dispersal phase from reproduction to the completion of the settlement process (including habitat choice and metamorphosis).

GENETIC DRIFT

A change in allele frequencies over time in a population of finite size due to random transmission of parental alleles from parents to offspring and due to the fact that some individuals randomly (irrespective of genotype) produce more offspring than other individuals. Genetic drift can lead to fixation when an allele of a gene attains a frequency of 100 % in the population.

GENOME

The entirety of an organism's hereditary information which is either encoded in the Deoxyribonucleic acid (DNA), or in the Ribonucleic acid (RNA), and includes both the genes and the non-coding regions of DNA or RNA.

HAPLOTYPE DIVERSITY

The probability that two randomly chosen haplotypes (alleles) are different.

$$Hd = n/(n - 1)*(1 - \Sigma x_i^2)$$

where n is the total number of haplotypes,

x_i is the frequency of the haplotype i,

Σx_i is the sum of haplotype frequencies of all haplotypes

LOCUS

Any position in the genome, where there might be one or more alleles segregating.

MUTATION

The process being the ultimate source of genetic variation in the form of new alleles. Mutations are a permanent change to the genome that will be included in any copies made of that genome.

NATURAL SELECTION

The process favouring some traits that make it more likely for an organism to survive and reproduce.

NON-SYNONYMOUS SUBSTITUTION

A nucleotide substitution that alters a codon and causes a change in amino acid residue in the protein sequence (e.g. TGT to TGA will replace Cysteine by Tryptophan in invertebrates).

NUCLEOTIDE DIVERSITY

The average number of nucleotide differences per site between any two DNA sequences chosen randomly from the sampled population.

This is simply the sum of the pairwise differences divided by the number of pairs, and is signified by π.

POPULATION

A reproductive unit of organisms, which share a common gene pool.

SEGREGATING SITES

Positions in a DNA sequence that differ between two or more individuals. Another name for polymorphic sites in a sequence alignment.

SELECTIVE SWEEP

The reduction or elimination of variation at sites that are physically linked to a site under positive selection.

STANDING VARIATION

The presence of more than one allele at a locus in a population at a given time-point.

SYNAPOMORPHY

A synapomorphy is a derived characteristic that is shared by members of a monophyletic group (homologous trait) but not by other clades, and that is used to distinguish a clade from another clade. The character has been inherited from a common ancestor, and is derived from an ancestral character and different from it.

SYNONYMOUS SUBSTITUTION (SILENT)

A nucleotide substitution resulting in a codon specifying the same amino acid as before (e.g. TGT to TGC, both code for Cysteine in invertebrates).

References

Abbott R, Albach D, Ansell S*, et al.* (2013) Hybridization and speciation. *Journal of Evolutionary Biology* **26**, 229-246.

Agrawal AA (2001) Phenotypic plasticity in the interactions and evolution of species. *Science* **294**, 321-326.

Alberts B, Johnson A, Lewis J*, et al.* (2014) *Molecular Biology of the cell*, 6th edn. Taylor & Francis Group, Abingdon, UK.

Angelis K, dos Reis M, Yang Z (2014) Bayesian estimation of nonsynonymous/synonymous rate ratios for pairwise sequence comparisons. *Molecular Biology and Evolution* **31**, 1902-1913.

Anisimova M, Liberles DA (2012) Detecting and understanding natural selection. In: *Codon Evolution: mechanisms and models* (eds. Cannarozzi GM, Schneider A), p. 296. Oxford University Press, New York.

Appeltans W, Ahyong Shane T, Anderson G*, et al.* (2012) The magnitude of global marine species diversity. *Current Biology* **22**, 2189-2202.

Aranzamendi MC, Bastida R, Gardenal CN (2011) Different evolutionary histories in two sympatric limpets of the genus *Nacella* (Patellogastropoda) in the South-western Atlantic coast. *Marine Biology* **158**, 2405-2418.

Ávila SP, Amen RG, Azevedo J, Cachão M, García-Talavera F (2002) Checklist of the Pleistocene marine molluscs of Praínha and Lagoínhas (Santa Maria Island, Azores). *Acoreana* **9**, 343-370.

Ávila SP, Madeira P, Mendes N*, et al.* (2008) Mass extinctions in the Azores during the last glaciation: fact or myth? *Journal of Biogeography* **35**, 1123-1129.

Ávila SP, Marques Da Silva C, Schiebel R*, et al.* (2009) How did they get here? The biogeography of the marine molluscs of the Azores. *Bulletin de la Societe Geologique de France* **180**, 295-307.

Ayata S-D, Lazure P, Thiébaut E (2010) How does the connectivity between populations mediate range limits of marine invertebrates? A case study of larval dispersal between the Bay of Biscay and the English Channel (North-East Atlantic). *Progress in Oceanography* **87**, 18-36.

Ayre DJ, Minchinton TE, Perrin C (2009) Does life history predict past and current connectivity for rocky intertidal invertebrates across a marine biogeographic barrier? *Molecular Ecology* **18**, 1887-1903.

Ballard JWO, Whitlock MC (2004) The incomplete natural history of mitochondria. *Molecular Ecology* **13**, 729-744.

Bandelt HJ, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* **16**, 37-48.

Bar-Yaacov D, Blumberg A, Mishmar D (2012) Mitochondrial-nuclear co-evolution and its effects on OXPHOS activity and regulation. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1819**, 1107-1111.

Barco A, Raupach MJ, Laakmann S, Neumann H, Knebelsberger T (2016) Identification of North Sea molluscs with DNA barcoding. *Molecular Ecology Resources* **16**, 288-297.

Barrett RDH, Schluter D (2008) Adaptation from standing genetic variation. *Trends in Ecology & Evolution* **23**, 38-44.

Barrick JE, Lenski RE (2013) Genome dynamics during experimental evolution. *Nature Reviews Genetics* **14**, 827-839.

Barton ED (2001) Canary and Portugal Currents. In: *Encyclopedia of Ocean Sciences*, pp. 380-389. Academic Press, Oxford.

Barton NH, Slatkin M (1986) A Quasi-equilibrium theory of the distribution of rare alleles in a subdivided population. *Heredity* **56**, 409-415.

Bazin E, Glémin S, Galtier N (2006) Population size does not influence mitochondrial genetic diversity in animals. *Science* **312**, 570-572.

Beerli P (2006) Comparison of Bayesian and maximum-likelihood inference of population genetic parameters. *Bioinformatics* **22**, 341-345.

Beerli P, Felsenstein J (1999) Maximum-Likelihood Estimation of Migration Rates and Effective Population Numbers in Two Populations Using a Coalescent Approach. *Genetics* **152**, 763-773.

Beerli P, Felsenstein J (2001) Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences* **98**, 4563-4568.

Beerli P, Palczewski M (2010) Unified framework to evaluate panmixia and migration direction among multiple sampling locations. *Genetics* **185**, 313-326.

Bernt M, Donath A, Jühling F*, et al.* (2013) MITOS: Improved de novo metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution* **69**, 313-319.

Berry OF (2006) Mitochondrial DNA and population size. *Science* **314**, 1388-1390.

Bird CE, Holland BS, Bowen BW, Toonen RJ (2007) Contrasting phylogeography in three endemic Hawaiian limpets (*Cellana* spp.) with similar life histories. *Molecular Ecology* **16**, 3173-3186.

Bischof B, Mariano AJ, Ryan EH (2003a) *"The North Atlantic Drift Current." Ocean Surface Currents.* http://oceancurrents.rsmas.miami.edu/atlantic/north-atlantic-drift.html

Bischof B, Mariano AJ, Ryan EH (2003b) *"The Portugal Current System". Ocean Surface Currents.* http://oceancurrents.rsmas.miami.edu/atlantic/portugal.html

Blakeslee AMH, Byers JE, Lesser MP (2008) Solving cryptogenic histories using host and parasite molecular genetics: the resolution of *Littorina littorea*'s North American origin. *Molecular Ecology* **17**, 3684-3696.

Blaxter M, Mann J, Chapman T*, et al.* (2005) Defining operational taxonomic units using DNA barcode data. *Philosophical Transactions of the Royal Society B: Biological Sciences* **360**, 1935-1943.

Blier PU, Dufresne F, Burton RS (2001) Natural selection and the evolution of mtDNA-encoded peptides: evidence for intergenomic co-adaptation. *Trends in Genetics* **17**, 400-406.

Boore JL (1999) Animal mitochondrial genomes. *Nucleic Acids Research* **27**, 1767-1780.

Boore JL (2006) Requirements and standards for organelle genome databases. *OMICS: A Journal of Integrative Biology* **10**, 119-126.

Boore JL, Brown WM (1998) Big trees from little genomes: mitochondrial gene order as a phylogenetic tool. *Current Opinion in Genetics & Development* **8**, 668-674.

Bouchet P, Rocroi JP (2005) Classification and nomenclator of gastropod families. *Malacologia*.

Bouckaert R, Heled J, Kuehnert D*, et al.* (2014) BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* **10**, e1003537.

Bozec A, Lozier MS, Chassignet EP, Halliwell GR (2011) On the variability of the Mediterranean Outflow Water in the North Atlantic from 1948 to 2006. *Journal of Geophysical Research: Oceans* **116**, 1-18.

Brito C, Lourenco P, Medeiros R*, et al.* (2001) Radular myoglobin as a molecular marker in littorinid systematics (Caenogastropoda). *Journal of Shellfish Research* **20**, 411-414.

Bromham L (2008) *Reading the story in DNA: a beginner's guide to molecular evolution* Oxford University Press, New York.

Brown GG, Simpson MV (1982) Novel features of animal mtDNA evolution as shown by sequences of two rat cytochrome oxidase subunit II genes. *Proceedings of the National Academy of Sciences of the United States of America* **79**, 3246-3250.

Burton RS, Barreto FS (2012) A disproportionate role for mtDNA in Dobzhansky–Muller incompatibilities? *Molecular Ecology* **21**, 4942-4957.

Cameron SL (2014) How to sequence and annotate insect mitochondrial genomes for systematic and comparative genomics research. *Systematic Entomology* **39**, 400-411.

Campan A, Royer J, Gente P, Olivet J, Muller R (1993) Evolution of the Azores–Gibraltar plate boundary for the last 36 Ma. *Eos* **74**, 586.

Cardinale BJ, Duffy JE, Gonzalez A*, et al.* (2012) Biodiversity loss and its impact on humanity. *Nature* **486**, 59-67.

Carroll SP, Hendry AP, Reznick DN, Fox CW (2007) Evolution on ecological time-scales. *Functional Ecology* **21**, 387-393.

Cassone BJ, Boulding EG (2006) Genetic structure and phylogeography of the lined shore crab, *Pachygrapsus crassipes*, along the northeastern and western Pacific coasts. *Marine Biology* **149**, 213-226.

Castelin M, Lorion J, Brisset J*, et al.* (2012) Speciation patterns in gastropods with long-lived larvae from deep-sea seamounts. *Molecular Ecology* **21**, 4828-4853.

Castellana S, Vicario S, Saccone C (2011) Evolutionary patterns of the mitochondrial genome in metazoa: exploring the role of mutation and selection in mitochondrial protein–coding genes. *Genome Biology and Evolution* **3**, 1067-1079.

Chao A (1984) Nonparametric estimation of the number of classes in a population. *Scandinavian Journal of Statistics* **11**, 265-270.

Chao A (1987) Estimating the population size for capture-recapture data with unequal catchability. *Biometrics* **43**, 783-791.

Chao A, Jost L, Chiang SC, Jiang YH, Chazdon RL (2008) A two-stage probabilistic approach to multiple-community similarity indices. *Biometrics* **64**, 1178-1186.

Chao A, Shen T-J (2010) Program SPADE (Species Prediction And Diversity Estimation). National Tsing Hua University. URL http://chao.stat.nthu.edu.tw, Hsin-Chu, Taiwan.

Charlesworth B (1998) Measures of divergence between populations and the effect of forces that reduce variability. *Molecular Biology and Evolution* **15**, 538-543.

Chen XJ (2013) Mechanism of homologous recombination and implications for aging-related deletions in mitochondrial DNA. *Microbiology and Molecular Biology Reviews : MMBR* **77**, 476-496.

Churchill CKC, Alejandrino A, Valdés Á, Ó Foighil D (2013) Parallel changes in genital morphology delineate cryptic diversification of planktonic nudibranchs. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **280**, 20131224.

Churchill CKC, Valdés Á, Ó Foighil D (2014) Afro-Eurasia and the Americas present barriers to gene flow for the cosmopolitan neustonic nudibranch *Glaucus atlanticus*. *Marine Biology* **161**, 899-910.

Clayton DA (1982) Replication of animal mitochondrial DNA. *Cell* **28**, 693-705.

Colgan DJ, Ponder WF, Beacham E, Macaranas JM (2003) Gastropod phylogeny based on six segments from four genes representing coding or non-coding and mitochondrial or nuclear DNA. *Molluscan Research* **23**, 123-148.

Collins RA, Cruickshank RH (2013) The seven deadly sins of DNA barcoding. *Molecular Ecology Resources* **13**, 969-975.

Colwell RK (2006) EstimateS: statistical estimation of species richness and shared species from samples. Version 8. Persistent URL <http://purl.oclc.org/estimates>.

Cordeiro R, Borges JP, De Frias Martins AM, Ávila SP (2015) Checklist of the littoral gastropods (Mollusca: Gastropoda) from the Archipelago of the Azores (NE Atlantic). *Biodiversity Journal* **6**, 855-900.

Costanza R, de Groot R, Sutton P*, et al.* (2014) Changes in the global value of ecosystem services. *Global Environmental Change* **26**, 152-158.

Cowen RK, Gawarkiewicz G, Pineda J, Thorrold SR, Werner FE (2007) Population connectivity in marine systems: An overview. *Oceanography and Marine Biology an Annual Review* **20**, 14-21.

Coyne JA, Orr HA (2004) *Speciation* Sinauer Associates, Sunderland, Massachusetts, USA.

Cracraft J (1983) Species concepts and speciation analysis. In: *Current Ornithology* (ed. Johnston RF), pp. 159-187. Plenum Press, New York.

Cracraft J (1989) Speciation and its ontology: the empirical consequences of alternative species concepts for understanding patterns and processes of differentiation. In: *Speciation and its Consequences* (eds. Otte D, Endler JA), pp. 28-59. Sinauer, Sunderland, Massachussets.

Cronin MA, Myers AA, O'Riordan RM (2000) The reproductive cycle of the intertidal gastropod *Melarhaphe neritoides* on the west and south coasts of Ireland. *Proceedings of the Royal Irish Academy. Section B: Biological, Geological and Chemical Science* **100B**, 97-106.

Cuña V, Saura M, Quesada H, Rolan-Alvarez E (2011) Extensive micro-geographical shell polymorphism in a planktotrophic marine intertidal snail. *Marine Ecology Progress Series* **427**, 133-143.

Cutter AD (2008) Multilocus patterns of polymorphism and selection across the X chromosome of *Caenorhabditis remanei*. *Genetics* **178**, 1661-1672.

Cutter AD, Baird SE, Charlesworth D (2006) High nucleotide polymorphism and rapid decay of linkage disequilibrium in wild populations of *Caenorhabditis remanei*. *Genetics* **174**, 901-913.

Cutter AD, Charlesworth B (2006) Selection intensity on preferred codons correlates with overall codon usage bias in *Caenorhabditis remanei*. *Current Biology* **16**, 2053-2057.

Cutter AD, Jovelin R, Dey A (2013) Molecular hyperdiversity and evolution in very large populations. *Molecular Ecology* **22**, 2074-2095.

da Fonseca RR, Johnson WE, O'Brien SJ, Ramos MJ, Antunes A (2008) The adaptive evolution of the mammalian mitochondrial genome. *BMC Genomics* **9**, 119.

Daguzan J (1976) Contribution à l'étude de la croissance et de la longévité de quelques Littorinidae (Mollusques Gastéropodes Prosobranches). *Archives de Zoologie Experimentale et Générale* **117**, 57-80.

Danovaro R, Costantini M, Verde C (2015) The marine genome: structure, regulation and evolution. *Marine Genomics* **24, Part 1**, 1-2.

Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* **9**, 772-772.

De Queiroz K (2007) Species concepts and species delimitation. *Systematic Biology* **56**, 879-886.

De Wolf H, Backeljau T, Verhagen R (1998) Lack of significant esterase and myoglobin differentiation in the periwinkle, *Littorina striata* (Gastropoda, Prosobranchia). *Hydrobiologia* **378**, 27-32.

De Wolf H, Blust R, Backeljau T (2004) The population genetic structure of Littorina littorea (Mollusca: Gastropoda) along a pollution gradient in the Scheldt estuary (The Netherlands) using RAPD analysis. *Science of the Total Environment* **325**, 59-69.

De Wolf H, Verhagen R, Backeljau T (2000) Large scale population structure and gene flow in the planktonic developing periwinkle, Littorina striata, in Macaronesia (Mollusca: Gastropoda). *Journal of Experimental Marine Biology and Ecology* **246**, 69-83.

Deli T, Fratini S, Ragionieri L*, et al.* (2016) Phylogeography of the marbled crab *Pachygrapsus marmoratus* (Decapoda, Grapsidae) along part of the African Mediterranean coast reveals genetic homogeneity across the Siculo-Tunisian Strait versus heterogeneity across the Gibraltar Strait. *Marine Biology Research* **12**, 471-487.

Dennielou B, Jallet L, Sultan N*, et al.* (2009) Post-glacial persistence of turbiditic activity within the Rhône deep-sea turbidite system (Gulf of Lions, Western Mediterranean): Linking the outer shelf and the basin sedimentary records. *Marine Geology* **257**, 65-86.

Denver DR, Morris K, Lynch M, Vassilieva LL, Thomas WK (2000) High direct estimate of the mutation rate in the mitochondrial genome of *Caenorhabditis elegans*. *Science* **289**, 2342-2344.

DeSalle R (2006) Species discovery versus species identification in DNA barcoding efforts: response to Rubinoff. *Conservation Biology* **20**, 1545-1547.

DeSalle R (2007) Phenetic and DNA taxonomy; a comment on Waugh. *Bioessays* **29**, 1289-1290.

Dey A, Chan CKW, Thomas CG, Cutter AD (2013) Molecular hyperdiversity defines populations of the nematode *Caenorhabditis brenneri*. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 11056-11060.

Díaz-Ferguson E, Haney RA, Wares JP, Silliman BR (2012) Genetic structure and connectivity patterns of two Caribbean rocky intertidal gastropods. *Journal of Molluscan Studies* **78**, 112-118.

Díaz-Ferguson E, Robinson JD, Silliman B, Wares JP (2010) Comparative phylogeography of North American Atlantic salt marsh communities. *Estuaries and Coasts* **33**, 828-839.

Diz AP, Álvarez-Rodríguez M, Romero MR, Rolán-Alvarez E, Galindo J (2017) Limited proteomic response in the marine snail *Melarhaphe neritoides* after long-term emersion. *Current Zoology*, 1-7.

Doellman MM, Trussell GC, Grahame JW, Vollmer SV (2011) Phylogeographic analysis reveals a deep lineage split within North Atlantic *Littorina saxatilis*. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **278**, 3175-3183.

Doyle JJ (1995) The irrelevance of allele tree topologies for species delimitation, and a non-topological alternative. *Systematic Botany* **20**, 574-588.

Drake JW, Charlesworth B, Charlesworth D, Crow JF (1998) Rates of spontaneous mutation. *Genetics* **148**, 1667-1686.

Drummond AJ, Nicholls GK, Rodrigo AG, Solomon W (2002) Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics* **161**, 1307-1320.

Drummond AJ, Pybus OG, Rambaut A, Forsberg R, Rodrigo AG (2003) Measurably evolving populations. *Trends in Ecology & Evolution* **18**, 481-488.

Drummond AJ, Rambaut A, Shapiro B, Pybus OG (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology and Evolution* **22**, 1185-1192.

Duncan BK, Miller JH (1980) Mutagenic deamination of cytosine residues in DNA. *Nature* **287**, 560-561.

Eakins BW, Sharman GF (2010) *Volumes of the World's Oceans from ETOPO1.* NOAA National Geophysical Data Center, Boulder, CO, USA. https://ngdc.noaa.gov/mgg/global/etopo1_ocean_volumes.html

Ecomare *Sea currents.* Texel, The Netherlands. http://www.ecomare.nl/en/encyclopedia/natural-environment/water/water-currents/sea-currents/

El-Geziry TM, Bryden IG (2010) The circulation pattern in the Mediterranean Sea: issues for modeller consideration. *Journal of Operational Oceanography* **3**, 39-46.

Elias M, Hill RI, Willmott KR*, et al.* (2007) Limited performance of DNA barcoding in a diverse community of tropical butterflies. *Proceedings of the Royal Society B: Biological Sciences* **274**, 2881-2889.

Eo SH, DeWoody JA (2010) Evolutionary rates of mitochondrial genomes correspond to diversification rates and to contemporary species richness in birds and reptiles. *Proceedings of the Royal Society of London B: Biological Sciences* **277**, 3587-3592.

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10**, 564-567.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**, 479-491.

Faurby S, Barber PH (2012) Theoretical limits to the correlation between pelagic larval duration and population genetic structure. *Molecular Ecology* **21**, 3419-3432.

Fay JC, Wu C-I (2000) Hitchhiking under positive Darwinian selection. *Genetics* **155**, 1405-1413.

Féral J-P (2002) How useful are the genetic markers in attempts to understand and manage marine biodiversity? *Journal of Experimental Marine Biology and Ecology* **268**, 121-145.

Fisher RA, Bennett JH (1930) *The Genetical Theory of Natural Selection* Oxford University Press, Oxford.

Fitzpatrick BM (2009) Power and sample size for nested analysis of molecular variance. *Molecular Ecology* **18**, 3961-3966.

Flores KB, Amdam GV (2011) Deciphering a methylome: what can we read into patterns of DNA methylation? *The Journal of Experimental Biology* **214**, 3155-3163.

Flores KB, Wolschin F, Amdam GV (2013) The role of methylation of DNA in environmental adaptation. *Integrative and Comparative Biology* **53**, 359-372.

Flot J-F (2007) CHAMPURU 1.0: a computer software for unraveling mixtures of two DNA sequences of unequal lengths. *Molecular Ecology Notes* **7**, 974-977.

Flot J-F, Blanchot J, Charpy L*, et al.* (2011) Incongruence between morphotypes and genetically delimited species in the coral genus *Stylophora*: phenotypic plasticity, morphological convergence, morphological stasis or interspecific hybridization? *BMC Ecology* **11**, 22.

Flot J-F, Couloux A, Tillier S (2010) Haplowebs as a graphical tool for delimiting species: a revival of Doyle's "field for recombination" approach and its application to the coral genus *Pocillopora* in Clipperton. *BMC Evolutionary Biology* **10**, 372.

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.

Fourdrilis S, Mardulyn P, Hardy OJ*, et al.* (2016) Mitochondrial DNA hyperdiversity and its potential causes in the marine periwinkle *Melarhaphe neritoides* (Mollusca: Gastropoda). *PeerJ* **4**, e2549.

Fratini S, Ragionieri L, Deli T*, et al.* (2016) Unravelling population genetic structure with mitochondrial DNA in a notional panmictic coastal crab species: sample size makes the difference. *BMC Evolutionary Biology* **16**, 1-15.

Frenkiel L, Mouëza M (1979) Développement larvaire de deux Tellinacea, *Scrobicularia plana* (Semelidae) et *Donax vittatus* (Donacidae). *Marine Biology* **55**, 187-195.

Fretter V, Graham A (1980) The prosobranch molluscs of Britain and Denmark: marine Littorinacea, Part 5. *Journal of Molluscan Studies* **46**, 242-284.

Fretter V, Manly R (1977) The settlement and early benthic life of *Littorina neritoides* (L.) at Wembury, S. Devon. *Journal of Molluscan Studies* **43**, 255-262.

Frommer M, McDonald LE, Millar DS*, et al.* (1992) A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 1827-1831.

Fu Y-X (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**, 915-925.

Fujisawa T, Barraclough TG (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.

Fukami H, Knowlton N (2005) Analysis of complete mitochondrial DNA sequences of three members of the *Montastraea annularis* coral species complex (Cnidaria, Anthozoa, Scleractinia). *Coral Reefs* **24**, 410-417.

Fusco G, Minelli A (2010) Phenotypic plasticity in development and evolution: facts and concepts. *Philosophical Transactions of the Royal Society B: Biological Sciences* **365**, 547-556.

García SD, Diz AP, Sá-Pinto A, Rolán-Alvarez E (2013) Proteomic and morphological divergence in micro-allopatric morphotypes of *Melarhaphe neritoides* in the absence of genetic differentiation. *Marine Ecology Progress Series* **475**, 145-153.

Gerlach G, Jueterbock A, Kraemer P, Deppermann J, Harmand P (2010) Calculations of population differentiation based on GST and D: forget GST but not all of statistics! *Molecular Ecology* **19**, 3845-3852.

Gershoni M, Levin L, Ovadia O*, et al.* (2014) Disrupting mitochondrial–nuclear coevolution affects OXPHOS complex I integrity and impacts human health. *Genome Biology and Evolution* **6**, 2665-2680.

Gershoni M, Templeton AR, Mishmar D (2009) Mitochondrial bioenergetics as a major motive force of speciation. *Bioessays* **31**, 642-650.

Giles RE, Blanc H, Cann HM, Wallace DC (1980) Maternal inheritance of human mitochondrial DNA. *Proceedings of the National Academy of Sciences* **77**, 6715-6719.

Giribet G, Okusu A, Lindgren AR*, et al.* (2006) Evidence for a clade composed of molluscs with serially repeated structures: Monoplacophorans are related to chitons. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 7723-7728.

Glez-Peña D, Gómez-Blanco D, Reboiro-Jato M, Fdez-Riverola F, Posada D (2010) ALTER: program-oriented conversion of DNA and protein alignments. *Nucleic Acids Research* **38**, W14-W18.

Gosselin P, Jangoux M (1998) From competent larva to exotrophic juvenile: a morphofunctional study of the perimetamorphic period of *Paracentrotus lividus* (Echinodermata, Echinoida). *Zoomorphology* **118**, 31-43.

Grande C, Templado J, Zardoya R (2008) Evolution of gastropod mitochondrial genome arrangements. *BMC Evolutionary Biology* **8**, 61.

Grant WS (2015) Problems and cautions with sequence mismatch analysis and Bayesian skyline plots to infer historical demography. *Journal of Heredity* **106**, 333-346.

Grassle JF (2000) The Ocean Biogeographic Information System (OBIS): An on-line, worldwide atlas for accessing, modeling and mapping marine biological data in a multidimensional geographic context. *Oceanography* **13**, 5-7.

Groombridge B, Jenkins M (2002) *World atlas of biodiversity: earth's living resources in the 21st century* University of California Press, Berkeley, USA.

Gyory J, Mariano AJ, Ryan EH (2003) *"The Slope/Shelf Edge Current." Ocean Surface Currents.* http://oceancurrents.rsmas.miami.edu/atlantic/slope.html

Gyory J, Mariano AJ, Ryan EH (2013) *"The Irminger Current." Ocean Surface Currents.* http://oceancurrents.rsmas.miami.edu/atlantic/irminger.html

Haag-Liautard C, Coffey N, Houle D*, et al.* (2008) Direct estimation of the mitochondrial DNA mutation rate in *Drosophila melanogaster*. *PLoS Biology* **6**, e204.

Hague MTJ, Routman EJ (2016) Does population size affect genetic diversity? A test with sympatric lizard species. *Heredity* **116**, 92-98.

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41**, 95-98.

Hameed SO, White JW, Miller SH, Nickols KJ, Morgan SG (2016) Inverse approach to estimating larval dispersal reveals limited population connectivity along 700 km of wave-swept open coast. *Proceedings of the Royal Society B: Biological Sciences* **283**.

Hamel J-F, Mercier A (1996) Early development, settlement, growth, and spatial distribution of the sea cucumber *Cucumaria frondosa* (Echinodermata: Holothuroidea). *Canadian Journal of Fisheries and Aquatic Sciences* **53**, 253-271.

Hardy GH (1908) Mendelian proportions in a mixed population. *Science* **28**, 49-50.

Hardy OJ, Vekemans X (2002) SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes* **2**, 618-620.

Harpending HC (1994) Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Human Biology* **66**, 591-600.

Harpending HC, Batzer MA, Gurven M*, et al.* (1998) Genetic traces of ancient demography. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 1961-1967.

Hart MW, Marko PB (2010) It's about time: divergence, demography, and the evolution of developmental modes in marine invertebrates. *Integrative and Comparative Biology* **50**, 643-661.

Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003a) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London B: Biological Sciences* **270**, 313-321.

Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W (2004a) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 14812-14817.

Hebert PDN, Ratnasingham S, deWaard JR (2003b) Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society of London B: Biological Sciences* **270**, S96-S99.

Hebert PDN, Stoeckle MY, Zemlak TS, Francis CM (2004b) Identification of birds through DNA barcodes. *PLoS Biology* **2**.

Hedgecock D, Barber PH, Edmands S (2007) Genetic approaches to measuring connectivity. *Oceanography* **20**, 70-79.

Hedrick PW (1999) Perspective: highly variable loci and their interpretation in evolution and conservation. *Evolution* **53**, 313-318.

Hedrick PW (2005) A standardized genetic differentiation measure. *Evolution* **59**, 1633-1638.

Heller R, Chikhi L, Siegismund HR (2013) The confounding effect of population structure on Bayesian skyline plot inferences of demographic history. *PLoS ONE* **8**, e62992.

Heller R, Siegismund HR (2009) Relationship between three measures of genetic differentiation GST, DEST and G'ST: how wrong have we been? *Molecular Ecology* **18**, 2080-2083.

Hershberg R, Petrov DA (2008) Selection on codon bias. *Annual Review of Genetics* **42**, 287-299.

Hey J (2005) On the number of new world founders: a population genetic portrait of the peopling of the Americas. *PLoS Biology* **3**, e193.

Hey J (2010) Isolation with migration models for more than two populations. *Molecular Biology and Evolution* **27**, 905-920.

Hey J, Nielsen R (2004) Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics* **167**, 747-760.

Hey J, Nielsen R (2007) Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proceedings of*

the National Academy of Sciences of the United States of America **104**, 2785-2790.

Hickerson MJ, Meyer CP, Moritz C (2006) DNA barcoding will often fail to discover new animal species over broad parameter space. *Systematic Biology* **55**, 729-739.

Hill GE (2016) Mitonuclear coevolution as the genesis of speciation and the mitochondrial DNA barcode gap. *Ecology and Evolution* **6**, 5831-5842.

Ho SYW, Phillips MJ, Cooper A, Drummond AJ (2005) Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Molecular Biology and Evolution* **22**, 1561-1568.

Ho SYW, Shapiro B (2011) Skyline-plot methods for estimating demographic history from nucleotide sequences. *Molecular Ecology Resources* **11**, 423-434.

Holsinger KE, Weir BS (2009) Genetics in geographically structured populations: defining, estimating and interpreting FST. *Nature Reviews Genetics* **10**, 639-650.

Hughes RN, Roberts DJ (1981) Comparative demography of *Littorina rudis*, *L. nigrolineata* and *L. neritoides* on three contrasted shores in North Wales. *Journal of Animal Ecology* **50**, 251-268.

Hung C-M, Drovetski SV, Zink RM (2016) Matching loci surveyed to questions asked in phylogeography. *Proceedings of the Royal Society of London B: Biological Sciences* **283**.

Ingólfsson A (2009) A marine refugium in Iceland during the last glacial maximum: fact or fiction? *Zoologica Scripta* **38**, 663-665.

Jakobsson M, Edge MD, Rosenberg NA (2013) The relationship between F_{ST} and the frequency of the most frequent allele. *Genetics* **193**, 515-528.

James JE, Piganeau G, Eyre-Walker A (2016) The rate of adaptive evolution in animal mitochondria. *Molecular Ecology* **25**, 67-78.

Johannesson K (1992) Genetic variability and large scale differentiation in two species of littorinid gastropods with planktotrophic development, *Littorina littorea* (L.) and *Melarhaphe* (*Littorina*) *neritoides* (L.) (Prosobranchia: Littorinacea), with notes on a mass occurrence of *M. neritoides* in Sweden. *Biological Journal of the Linnean Society* **47**, 285-299.

Johannesson K (2003) Evolution in *Littorina*: ecology matters. *Journal of Sea Research* **49**, 107-117.

Johannesson K (2016) What can be learnt from a snail? *Evolutionary Applications* **9**, 153-165.

Johnson J, Stevens I (2000) A fine resolution model of the eastern North Atlantic between the Azores, the Canary Islands and the Gibraltar Strait. *Deep Sea Research Part I: Oceanographic Research Papers* **47**, 875-899.

Johnson PLF, Slatkin M (2008) Accounting for bias from sequencing error in population genetic estimates. *Molecular Biology and Evolution* **25**, 199-206.

Jost LOU (2008) G_{ST} and its relatives do not measure differentiation. *Molecular Ecology* **17**, 4015-4026.

Jühling F, Pütz J, Bernt M*, et al.* (2012) Improved systematic tRNA gene annotation allows new insights into the evolution of mitochondrial tRNA structures and into the mechanisms of mitochondrial genome rearrangements. *Nucleic Acids Research* **40**, 2833-2845.

Karnkowska A, Vacek V, Zubáčová Z*, et al.* (2016) A eukaryote without a mitochondrial organelle. *Current Biology* **26**, 1274-1284.

Kartavtsev YP (2011) Divergence at Cyt-b and Co-1 mtDNA genes on different taxonomic levels and genetics of speciation in animals. *Mitochondrial DNA* **22**, 55-65.

Kearse M, Moir R, Wilson A*, et al.* (2012) Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647-1649.

Kelly RP, Oliver TA, Sivasundar A, Palumbi SR (2010) A method for detecting population genetic structure in diverse, high gene-flow species. *Journal of Heredity* **101**, 423-436.

Kennington WJ, Hevroy TH, Johnson MS (2012) Long-term genetic monitoring reveals contrasting changes in the genetic composition of newly established populations of the intertidal snail *Bembicium vittatum*. *Molecular Ecology* **21**, 3489-3500.

Kesäniemi JE, Geuverink E, Knott KE (2012) Polymorphism in developmental mode and its effect on population genetic structure of a spionid polychaete, *Pygospio elegans*. *Integrative and Comparative Biology* **52**, 181-196.

Kimura M (1983) *The Neutral Theory of Molecular Evolution*, 1st edn. Cambridge University Press, New York.

Kingman JFC (1982) The coalescent. *Stochastic Processes and their Applications* **13**, 235-248.

Kivisild T (2015) Maternal ancestry and population history from whole mitochondrial genomes. *Investigative Genetics* **6**, 3.

Kochzius M, Nuryanto A (2008) Strong genetic population structure in the boring giant clam, *Tridacna crocea*, across the Indo-Malay archipelago: implications related to evolutionary processes and connectivity. *Molecular Ecology* **17**, 3775-3787.

Kondrashov AS (2008) Another step toward quantifying spontaneous mutation. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 9133-9134.

Kool JT, Moilanen A, Treml EA (2013) Population connectivity: recent advances and new perspectives. *Landscape Ecology* **28**, 165-185.

Kronholm I, Loudet O, de Meaux J (2010) Influence of mutation rate on estimators of genetic differentiation - lessons from *Arabidopsis thaliana*. *BMC Genetics* **11**, 33.

Krug PJ (2011) Patterns of speciation in marine gastropods: a review of the phylogenetic evidence for localized radiations in the sea. *American Malacological Bulletin* **29**, 169-186.

Kuhner MK (2008) Coalescent genealogy samplers: windows into population history. *Trends in Ecology & Evolution* **24**, 86-93.

Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution* **33**, 1870-1874.

Kyle CJ, Boulding EG (2000) Comparative population genetic structure of marine gastropods (*Littorina* spp.) with and without pelagic larval dispersal. *Marine Biology* **137**, 835-845.

Lane N (2009) Biodiversity: On the origin of bar codes. *Nature* **462**, 272-274.

Lanfear R, Frandsen PB, Wright AM, Senfeld T, Calcott B (2016) PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution* **34**, 772-773.

Lanfear R, Kokko H, Eyre-Walker A (2014) Population size and the rate of evolution. *Trends in Ecology & Evolution* **29**, 33-41.

Lavrov DV (2014) Mitochondrial genomes in invertebrate animals. In: *Molecular Life Sciences: An Encyclopedic Reference* (ed. Bell E), pp. 1-8. Springer New York, New York, NY.

Lawrie DS, Messer PW, Hershberg R, Petrov DA (2013) Strong purifying selection at synonymous sites in *D. melanogaster*. *PLoS Genetics* **9**, e1003527.

Layton KKS, Martel AL, Hebert PDN (2014) Patterns of DNA barcode variation in Canadian marine molluscs. *PLoS ONE* **9**, e95003.

Lebour MV (1935) The Breeding of *Littorina neritoides*. *Journal of the Marine Biological Association of the United Kingdom (New Series)* **20**, 373-378.

Lee HJ, Boulding EG (2007) Mitochondrial DNA variation in space and time in the northeastern Pacific gastropod, *Littorina keenae*. *Molecular Ecology* **16**, 3084-3103.

Lee HJ, Boulding EG (2009) Spatial and temporal population genetic structure of four northeastern Pacific littorinid gastropods: the effect of mode of larval development on variation at one mitochondrial and two nuclear DNA markers. *Molecular Ecology* **18**, 2165-2184.

Leffler EM, Bullaughey K, Matute DR*, et al.* (2012) Revisiting an old riddle: what determines genetic diversity levels within species? *PLoS Biology* **10**, e1001388.

Leng L, Zhang D-X (2011) Measuring population differentiation using G_{ST} or D? A simulation study with microsatellite DNA markers under a finite island model and nonequilibrium conditions. *Molecular Ecology* **20**, 2494-2509.

Lessios HA, Kane J, Robertson DR (2003) Phylogeography of the pantropical sea urchin *Tripneustes*: contrasting patterns of population structure between oceans. *Evolution* **57**, 2026-2036.

Lewis JR, Tambs-Lyche H (1962) *Littorina neritoides* in Scandinavia. *Sarsia* **7**, 7-10.

Li W-H (1977) Distribution of nucleotide differences between two randomly chosen cistrons in a finite population. *Genetics* **85**, 331-337.

Libertini A, Trisolini R, Edmands S (2004) A cytogenetic study of the periwinkle *Littorina keenae* Rosewater, 1978 (Gastropoda: Littorinidae). *Journal of Molluscan Studies* **70**, 299-301.

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451-1452.

Lohse M, Drechsel O, Kahlau S, Bock R (2013) OrganellarGenomeDRAW—a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Research*.

Long JC (2009) Update to Long and Kittles's "Human genetic diversity and the nonexistence of biological races" (2003): fixation on an index. *Human Biology* **81**, 799-803.

Longo GC, O'Connell B, Green RE, Bernardi G (2016) The complete mitochondrial genome of the black surfperch, *Embiotoca jacksoni*: Selection and substitution rates among surfperches (Embiotocidae). *Marine Genomics* **28**, 107-112.

Lou M, Brian Golding G (2010) Assigning sequences to species in the absence of large interspecific differences. *Molecular Phylogenetics and Evolution* **56**, 187-194.

Lowe WH, Allendorf FW (2010) What can genetics tell us about population connectivity? *Molecular Ecology* **19**, 3038-3051.

Luiz OJ, Allen AP, Robertson DR*, et al.* (2013) Adult and larval traits as determinants of geographic range size among tropical reef fishes. *Proceedings of the National Academy of Sciences* **110**, 16498-16502.

Lynch M, Crease TJ (1990) The analysis of population survey data on DNA sequence variation. *Molecular Biology and Evolution* **7**, 377-394.

Lynch M, Koskella B, Schaack S (2006) Mutation pressure and the evolution of organelle genomic architecture. *Science* **311**, 1727-1730.

Lysaght AM (1941) The biology and trematode parasites of the gastropod *Littorina neritoides* (L.) on the Plymouth breakwater. *Journal of the Marine Biological Association of the United Kingdom* **25**, 41-67.

Ma H, O'Farrell PH (2015) Selections that isolate recombinant mitochondrial genomes in animals. *eLife* **4**, e07247.

Ma Z, Yang X, Bercsenyi M, *et al.* (2015) Comparative mitogenomics of the genus *Odontobutis* (Perciformes: Gobioidei: Odontobutidae) revealed conserved gene rearrangement and high sequence variations. *International Journal of Molecular Sciences* **16**, 25031-25049.

Madeira J, Ribeiro A (1990) Geodynamic models for the Azores triple junction: a contribution from tectonics. *Tectonophysics* **184**, 405-415.

Maggs CA, Castilho R, Foltz D, *et al.* (2008) Evaluating signatures of glacial refugia for North Atlantic benthic marine taxa. *Ecology* **89**, S108-S122.

Mao Y, Gao T, Yanagimoto T, Xiao Y (2011) Molecular phylogeography of *Ruditapes philippinarum* in the Northwestern Pacific Ocean based on COI gene. *Journal of Experimental Marine Biology and Ecology* **407**, 171-181.

Marko PB, Barr KR (2007) Basin-scale patterns of mtDNA differentiation and gene flow in the bay scallop *Argopecten irradians concentricus*. *Marine Ecology Progress Series* **349**, 139-150.

Marko PB, Hart MW (2011) The complex analytical landscape of gene flow inference. *Trends in Ecology & Evolution* **26**, 448-456.

Marletaz F, Le Parco Y, Liu S, Peijnenburg K (2015) Extreme mitogenomic variation without cryptic speciation in chaetognaths. *bioRxiv* **025957**.

Marques JP, Sotelo G, Larsson T, *et al.* (2017) Comparative mitogenomic analysis of three species of periwinkles: *Littorina fabalis*, *L. obtusata* and *L. saxatilis*. *Marine Genomics* **32**, 41-47.

Marti-Puig P, Costantini F, Rugiu L, Ponti M, Abbiati M (2013) Patterns of genetic connectivity in invertebrates of temperate MPA networks. *Advances in Oceanography and Limnology* **4**, 138-149.

Martin WF, Mentel M (2010) The Origin of Mitochondria. *Nature Education* **3**, 58.

Maruyama T, Nei M (1981) Genetic variability maintained by mutation and overdominant selection in finite populations. *Genetics* **98**, 441-459.

Mather KC (1955) Polymorphism as an outcome of disruptive selection. *Evolution* **9**, 52-61.

Mayden RL (1997) A hierarchy of species concepts: the denouement in the saga of the species problem. In: *Species: The units of diversity* (eds. Claridge MF, Dawah HA, Wilson MR), pp. 381–423. Chapman & Hall.

Mayr E (1942) *Systematics and the origin of species* Columbia University Press, New York.

McGovern TM, Keever CC, Saski CA, Hart MW, Marko PB (2010) Divergence genetics analysis reveals historical population genetic processes leading to contrasting phylogeographic patterns in co-distributed species. *Molecular Ecology* **19**, 5043-5060.

McGrath D (1997) Population structure and recruitment of the intertidal gastropod *Melarhaphe neritoides* (L.) on an exposed rocky shore on the South coast of

Ireland. *Biology and Environment: Proceedings of the Royal Irish Academy* **97B**, 75-79.

McQuaid C (1996a) Biology of the gastropod family Littorinidae. II. Role in the ecology of intertidal and shallow marine ecosystems. *Oceanography and marine biology: an annual review*.

McQuaid C (1996b) Biology of the gastropod family Littornidae. I. Evolutionary aspects. *Oceanography and marine biology: an annual review* **34**, 233-262.

Medeiros R, Serpa L, Brito C*, et al.* (1998) Radular myoglobin and protein variation within and among some littorinid species (Mollusca: Gastropoda). In: *Aspects of Littorinid Biology: Proceedings of the Fifth International Symposium on Littorinid Biology, held in Cork, Ireland, 7–13 September 1996* (eds. O'Riordan RM, Burnell GM, Davies MS, Ramsay NF), pp. 43-51. Springer Netherlands, Dordrecht.

Meier R (2008) DNA sequences in taxonomy, opportunities and challenges. In: *The new taxonomy* (ed. Wheeler QD). CRC Press, Boca Raton.

Meier R, Shiyang K, Vaidya G, Ng PKL (2006) DNA barcoding and taxonomy in diptera: a tale of high intraspecific variability and low identification success. *Systematic Biology* **55**, 715-728.

Meier R, Zhang G, Ali F (2008) The use of mean instead of smallest interspecific distances exaggerates the size of the barcoding gap and leads to misidentification. *Systematic Biology* **57**.

Meirmans PG (2006) Using the amova framework to estimate a standardized genetic differentiation measure. *Evolution* **60**, 2399-2402.

Meirmans PG, Hedrick PW (2011) Assessing population structure: F_{ST} and related measures. *Molecular Ecology Resources* **11**, 5-18.

Méléard S (2016) *Modèles aléatoires en Ecologie et Evolution* Springer-Verlag Berlin Heidelberg.

Miller MA, Pfeiffer W, Schwartz T (2010) Creating the CIPRES Science Gateway for inference of large phylogenetic trees, 1-8.

Mills LS, Allendorf FW (1996) The One-Migrant-per-Generation Rule in Conservation and Management. *Conservation Biology* **10**, 1509-1518.

Mulligan CJ, Kitchen A, Miyamoto MM (2006) Comment on "Population size does not influence mitochondrial genetic diversity in animals". *Science* **314**, 1390-1390.

Myers AA, McGrath D (1993) Settlement and growth of the intertidal gastropod, Melarhaphe neritoides, on the south coast of Ireland. *Journal of the Marine Biological Association of the United Kingdom* **73**, 313-319.

Nabholz B, Glémin S, Galtier N (2009) The erratic mitochondrial clock: variations of mutation rate, not population size, affect mtDNA diversity across birds and mammals. *BMC Evolutionary Biology* **9**, 54-54.

Nabholz B, Mauffrey J-F, Bazin E, Galtier N, Glemin S (2008) Determination of mitochondrial genetic diversity in mammals. *Genetics* **178**, 351-361.

Nagylaki T (1998) Fixation indices in subdivided populations. *Genetics* **148**, 1325-1332.

Navajas M, Boursot P (2003) Nuclear ribosomal DNA monophyly versus mitochondrial DNA polyphyly in two closely related mite species: the influence of life history and molecular drive. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **270**, S124-S127.

Navascués M, Emerson BC (2009) Elevated substitution rate estimates from ancient DNA: model violation and bias of Bayesian methods. *Molecular Ecology* **18**, 4390-4397.

Nei M (1973) Analysis of Gene Diversity in Subdivided Populations. *Proceedings of the National Academy of Sciences* **70**, 3321-3323.

Nei M (1987) *Molecular Evolutionary Genetics* Columbia University Press, New York, NY.

Nei M, Miller JC (1990) A simple method for estimating average number of nucleotide substitutions within and between populations from restriction data. *Genetics* **125**, 873-879.

Neigel JE (1997) A comparison of alternative strategies for estimatinggene flow from genetic markers *Annual Review of Ecology and Systematics* **28**, 105-128.

Neigel JE (2002) Is F_{ST} obsolete? *Conservation Genetics* **3**, 167-173.

Neigel JE, Ball RM, Avise JC (1991) Estimation of single generation migration distances from geographic variation in animal mitochondrial DNA. *Evolution* **45**, 423-432.

Nickols KJ, White JW, Largier JL, Gaylord B (2015) Marine population connectivity: reconciling large-scale dispersal and high self-retention. *The American Naturalist* **185**, 196-211.

Nielsen R, Slatkin M (2013) *An Introduction to Population Genetics: Theory and Applications*, 1st edn. Sinauer Associates Inc., Sunderland, Massachusetts.

Nielsen R, Wakeley J (2001) Distinguishing migration from isolation: A Markov chain Monte Carlo approach. *Genetics* **158**, 885-896.

Nuryanto A, Kochzius M (2009) Highly restricted gene flow and deep evolutionary lineages in the giant clam *Tridacna maxima*. *Coral Reefs* **28**, 607-619.

Olabarria C, Timmermans J-M, Backeljau T (1998) Electrophoretic heterogeneity within and between flat periwinkles (Mollusca: Gastropoda) along an intertidal transect at Ria Ferrol, northwest Spain. *Hydrobiologia* **378**, 11-19.

Osca D, Templado J, Zardoya R (2015) Caenogastropod mitogenomics. *Molecular Phylogenetics and Evolution* **93**, 118-128.

Öztürk B, Dogan A, Bitilis-Bakir B, Salman A (2014) Marine molluscs of the Turkish coasts: an updated checklist. *Turkish Journal of Zoology* **38**, 832-879.

Palant B, Fishelson L (1968) *Littorina punctata* and *M. neritoides* from Israel: ecology and annual cycle of the genital system. *Israel Journal of Zoology* **17**, 145-160.

Panova M, Blakeslee AMH, Miller AW, *et al.* (2011) Glacial history of the North Atlantic marine snail, *Littorina saxatilis*, inferred from distribution of mitochondrial DNA lineages. *PLoS ONE* **6**, e17511.

Panova M, Johansson T, Canbäck B, *et al.* (2014) Species and gene divergence in Littorina snails detected by array comparative genomic hybridization. *BMC Genomics* **15**, 687.

Patanè L (1933) Sul comportamento di *Littorina neritoides* L. mantenuta in ambiente sobaero ed in altro condizioni sperimentali. *Rendiconti della Academia dei Lincei* **6**, 962-967.

Patarnello T, Volckaert FAMJ, Castilho R (2007) Pillars of Hercules: is the Atlantic–Mediterranean transition a phylogeographical break? *Molecular Ecology* **16**, 4426-4444.

Pavlinov IY (2013) The species problem, why again? In: *The Species Problem - Ongoing Issues* (ed. Pavlinov IY), pp. 3-37. InTech, Rijeka, Croatia.

Pearse DE, Crandall KA (2004) Beyond F_{ST}: Analysis of population genetic data for conservation. *Conservation Genetics* **5**, 585-602.

Penant G, Aurelle D, Feral JP, Chenuil A (2013) Planktonic larvae do not ensure gene flow in the edible sea urchin *Paracentrotus lividus*. *Marine Ecology Progress Series* **480**, 155-170.

Penny D (2005) Relativity for molecular clocks. *Nature* **436**, 183-184.

Perna NT, Kocher TD (1995) Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. *Journal of Molecular Evolution* **41**, 353-358.

Piganeau G, Eyre-Walker A (2009) Evidence for variation in the effective population size of animal mitochondrial DNA. *PLoS ONE* **4**, e4396.

Pineda J, Hare JA, Sponaugle S (2007) Larval transport and dispersal in the coastal ocean and consequences for population connectivity. *Oceanography* **20**, 22-39.

Plummer M, Best N, Cowles K, Vines K (2006) CODA: Convergence diagnosis and output analysis for MCMC. *R news* **6**, 7-11.

Pons J, Barraclough TG, Gomez-Zurita J, *et al.* (2006) Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Systematic Biology* **55**, 595-609.

Pons O, Chaouche K (1995) Estimation, variance and optimal sampling of gene diversity. II. Diploid locus. *Theoretical and Applied Genetics* **91**, 122-130.

Pons O, Petit RJ (1995) Estimation, variance and optimal sampling of gene diversity. I. Haploid locus. *Theoretical and Applied Genetics* **90**, 462-470.

Pons O, Petit RJ (1996) Measuring and testing genetic differentiation with ordered versus unordered alleles. *Genetics* **144**, 1237-1245.

Popovic I, Marko PB, Wares JP, Hart MW (2014) Selection and demographic history shape the molecular evolution of the gamete compatibility protein bindin in *Pisaster* sea stars. *Ecology and Evolution* **4**, 1567-1588.

Puillandre N, Lambert A, Brouillet S, Achaz G (2012) ABGD, Automatic Barcode Gap Discovery for primary species delimitation. *Molecular Ecology* **21**, 1864-1877.

QGIS Development Team (2004-2014) QGIS Geographic Information System. Open Source Geospatial Foundation. URL http://qgis.osgeo.org.

Queiroga H, Costa R, Leonardo N, Soares D, Cleary DF (2011) Morphometric variation in two intertidal littorinid gastropods. *Contributions to Zoology* **80**, 201-211.

Quesada H, Gallagher C, Skibinski DAG, Skibinski DOF (1998) Patterns of polymorphism and gene flow of gender-associated mitochondrial DNA lineages in European mussel populations. *Molecular Ecology* **7**, 1041-1051.

R Development Core Team (2011) R: A language and environment for statistical computing. R Foundation for Statistical Computing. URL http://www.R-project.org/, Vienna, Austria.

Rambaut A, Suchard MA, Xie D, Drummond AJ (2014) Tracer v1.6, Available from http://beast.bio.ed.ac.uk/Tracer.

Ramos-Esplá AA, Bitar G, Khalaf G*, et al.* (2014) Ecological characterization of sites of interest for conservation in Lebanon: Enfeh Peninsula, Ras Chekaa cliff, Raoucheh, Saida, Tyre and Nakoura. In: *MedMPAnet Project*. RAC/SPA - UNEP/MAP, Tunis, 146 p.

Rannala B, Hartigan J (1996) Estimating gene flow in island populations. *Genetical Research* **67**, 147-158.

Ratnasingham S, Hebert PDN (2007) BOLD: The Barcode of Life Data System (http://www.barcodinglife.org). *Molecular Ecology Notes* **7**, 355-364.

Ravinet M, Westram A, Johannesson K*, et al.* (2016) Shared and nonshared genomic divergence in parallel ecotypes of *Littorina saxatilis* at a local scale. *Molecular Ecology* **25**, 287-305.

Rawlings TA, MacInnis MJ, Bieler R, Boore JL, Collins TM (2010) Sessile snails, dynamic genomes: gene rearrangements within the mitochondrial genome of a family of caenogastropod molluscs. *BMC Genomics* **11**, 440.

Raybould AF, Clarke RT, Bond JM, Welters RE, Gliddon CJ (2002) Inferring patterns of dispersal from allele frequency data, 89-110.

Reddy SB, Rosenberg NA (2012) Refining the relationship between homozygosity and the frequency of the most frequent allele. *Journal of Mathematical Biology* **64**, 87-108.

Reid DG (1989) The comparative morphology, phylogeny and evolution of the gastropod family Littorinidae. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences* **324**, 1-110.

Reid DG, Dyal P, Williams ST (2010) Global diversification of mangrove fauna: a molecular phylogeny of *Littoraria* (Gastropoda: Littorinidae). *Molecular Phylogenetics and Evolution* **55**, 185-201.

Reid DG, Dyal P, Williams ST (2012) A global molecular phylogeny of 147 periwinkle species (Gastropoda, Littorininae). *Zoologica Scripta* **41**, 125-136.

Reid DG, Lal K, Mackenzie-Dodds J*, et al.* (2006) Comparative phylogeography and species boundaries in *Echinolittorina* snails in the central Indo-West Pacific. *Journal of Biogeography* **33**, 990-1006.

Reid DG, Rumbak E, Thomas RH (1996) DNA, morphology and fossils: phylogeny and evolutionary rates of the gastropod genus *Littorina*. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **351**, 877-895.

Reyes A, Gissi C, Pesole G, Saccone C (1998) Asymmetrical directional mutation pressure in the mitochondrial genome of mammals. *Molecular Biology and Evolution* **15**, 957-966.

Ribeiro ÂM, Foote AD, Kupczok A*, et al.* (2017) Marine genomics: News and views. *Marine Genomics* **31**, 1-8.

Rice WR (1989) Analyzing tables of statistical tests. *Evolution* **43**, 223-225.

Robinson JD, Diaz-Ferguson E, Poelchau MF*, et al.* (2010) Multiscale diversity in the marshes of the Georgia coastal ecosystems LTER. *Estuaries and Coasts* **33**, 865-877.

Rogers AR (1995) Genetic evidence for a Pleistocene population explosion. *Evolution* **49**, 608-615.

Rolán-Alvarez E (2007) Sympatric speciation as a by-product of ecological adaptation in the Galician Littorina saxatilis hybrid zone. *Journal of Molluscan Studies* **73**, 1-10.

Rolán E, Groh K (2005) *Malacological fauna from the Cape Verde archipelago. Part 1. Polyplacophora and Gastropoda*, 1st edn. ConchBooks, Hackenheim, Germany.

Ronquist F, Teslenko M, van der Mark P*, et al.* (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* **61**, 539-542.

Rosenberg NA, Jakobsson M (2008) The relationship between homozygosity and the frequency of the most frequent allele. *Genetics* **179**, 2027-2036.

Rosewater J (1981) The family Littorinidae in tropical West Africa. *Atlantide Report* **13**, 7-48.

Ross HA, Murugan S, Sibon Li WL (2008) Testing the reliability of genetic methods of species identification via simulation. *Systematic Biology* **57**.

Rousset F (2008) genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources* **8**, 103-106.

Rousset F (2013) Exegeses on maximum genetic differentiation. *Genetics* **194**, 557-559.

Rowe E, Mariano AJ, Ryan EH (2013) *"The North Atlantic Current." Ocean Surface Currents.* http://oceancurrents.rsmas.miami.edu/atlantic/north-atlantic.html

Rubinoff D (2006a) DNA barcoding evolves into the familiar. *Conservation Biology* **20**, 1548-1549.

Rubinoff D (2006b) Utility of mitochondrial DNA barcodes in species conservation. *Conservation Biology* **20**, 1026-1033.

Rueffler C, Van Dooren TJM, Leimar O, Abrams PA (2006) Disruptive selection and then what? *Trends in Ecology & Evolution* **21**, 238-245.

Ryman N, Leimar O (2009) GST is still a useful measure of genetic differentiation — a comment on Jost's D. *Molecular Ecology* **18**, 2084-2087.

Saccone C, Lanave C, De Grassi A (2006) Metazoan OXPHOS gene families: Evolutionary forces at the level of mitochondrial and nuclear genomes. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1757**, 1171-1178.

Santos S, Cruzeiro C, Olsen JL, van der Veer HW, Luttikhuizen PC (2012) Isolation by distance and low connectivity in the peppery furrow shell *Scrobicularia plana* (Bivalvia). *Marine Ecology Progress Series* **462**, 111-124.

Sato M, Sato K (2013) Maternal inheritance of mitochondrial DNA by diverse mechanisms to eliminate paternal mitochondrial DNA. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1833**, 1979-1984.

Schneider S, Excoffier L (1999) Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to Human mitochondrial DNA. *Genetics* **152**, 1079-1089.

Schwarz H (2017) On the usefulness of useless knowledge. *Nature Reviews Chemistry* **1**, 0001.

Searle R (1980) Tectonic pattern of the Azores spreading centre and triple junction. *Earth and Planetary Science Letters* **51**, 415-434.

Selkoe KA, Toonen RJ (2011) Marine connectivity: a new look at pelagic larval duration and genetic metrics of dispersal. *Marine Ecology Progress Series* **436**, 291-305.

Seo T-K, Thorne JL, Hasegawa M, Kishino H (2002) Estimation of effective population size of HIV-1 within a host: a pseudomaximum-likelihood approach. *Genetics* **160**, 1283-1293.

Shanks AL (2009) Pelagic larval duration and dispersal distance revisited. *The Biological Bulletin* **216**, 373-385.

Silva ACF, Mendonça V, Paquete R, Barreiras N, Vinagre C (2015) Habitat provision of barnacle tests for overcrowded periwinkles. *Marine Ecology* **36**, 530-540.

Silva SE, Silva IC, Madeira C*, et al.* (2013) Genetic and morphological variation in two littorinid gastropods: evidence for recent population expansions along the East African coast. *Biological Journal of the Linnean Society* **108**, 494-508.

Simon C, Frati F, Beckenbach A*, et al.* (1994) Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Annals of the Entomological Society of America* **87**, 651-701.

Slatkin M (1985a) Gene flow in natural populations. *Annual Review of Ecology and Systematics* **16**.

Slatkin M (1985b) Rare alleles as indicators of gene flow. *Evolution* **39**, 53-65.

Slatkin M, Maddison WP (1989) A cladistic measure of gene flow inferred from the phylogenies of alleles. *Genetics* **123**, 603-613.

Small KS, Brudno M, Hill MM, Sidow A (2007) Extreme genomic variation in a natural population. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 5698-5703.

Smith BT, Klicka J (2013) Examining the role of effective population size on mitochondrial and multilocus divergence time discordance in a songbird. *PLoS ONE* **8**, e55161.

Smith MA, Woodley NE, Janzen DH, Hallwachs W, Hebert PDN (2006) DNA barcodes reveal cryptic host-specificity within the presumed polyphagous members of a genus of parasitoid flies (Diptera: Tachinidae). *Proceedings of the National Academy of Sciences of the United States of America* **103**, 3657-3662.

So JJ, Uthicke S, Hamel J-F, Mercier A (2011) Genetic population structure in a commercial marine invertebrate with long-lived lecithotrophic larvae: *Cucumaria frondosa* (Echinodermata: Holothuroidea). *Marine Biology* **158**, 859-870.

Soares P, Ermini L, Thomson N*, et al.* (2009) Correcting for purifying selection: an improved human mitochondrial molecular clock. *The American Journal of Human Genetics* **84**, 740-759.

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312-1313.

Stamatakis A (2016) The RAxML v8.2.X Manual. Heidelberg Institute for Theoretical Studies. Available from http://www.exelixis-lab.org/.

Stöger I, Schrödl M (2013) Mitogenomics does not resolve deep molluscan relationships (yet?). *Molecular Phylogenetics and Evolution* **69**, 376-392.

Strasburg JL, Rieseberg LH (2009) How robust are "Isolation with Migration" analyses to violations of the IM Model? A simulation study. *Molecular Biology and Evolution* **27**, 297-310.

Subramanian S, Lambert DM (2011) Time dependency of molecular evolutionary rates? Yes and no. *Genome Biology and Evolution* **3**, 1324-1328.

Sun Se, Li Q, Kong L*, et al.* (2016) DNA barcoding reveal patterns of species diversity among northwestern Pacific molluscs. *Scientific Reports* **6**, 33367.

Sun Y-B, Shen Y-Y, Irwin DM, Zhang Y-P (2011) Evaluating the roles of energetic functional constraints on teleost mitochondrial-encoded protein evolution. *Molecular Biology and Evolution* **28**, 39-44.

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585-595.

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.

Templeton AR, Hemmer H, Mace G*, et al.* (1986) Local adaptation, coadaptation, and population boundaries. *Zoo Biology* **5**, 115-125.

Terranova MS, Lo Brutto S, Arculeo M, Mitton JB (2007) A mitochondrial phylogeography of *Brachidontes variabilis* (Bivalvia: Mytilidae) reveals three cryptic species. *Journal of Zoological Systematics and Evolutionary Research* **45**, 289-298.

Thoday JM (1959) Effects of disruptive selection. *Heredity* **13**, 187-203.

Thomas JA, Welch JJ, Lanfear R, Bromham L (2010) A generation time effect on the rate of molecular evolution in invertebrates. *Molecular Biology and Evolution* **27**, 1173-1180.

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**, 4673-4680.

Thompson RC, Crowe TP, Hawkins SJ (2002) Rocky intertidal communities: past environmental changes, present status and predictions for the next 25 years. *Environmental Conservation* **29**, 168-191.

Tourasse NJ, Li W-H (2000) Selective constraints, amino acid composition, and the rate of protein evolution. *Molecular Biology and Evolution* **17**, 656-664.

Treml EA, Roberts JJ, Chao Y*, et al.* (2012) Reproductive output and duration of the pelagic larval stage determine seascape-wide connectivity of marine populations. *Integrative and Comparative Biology* **52**, 525-537.

Tufto J, Engen S, Hindar K (1996) Inferring patterns of migration from gene frequencies under equilibrium conditions. *Genetics* **144**, 1911-1921.

Turner TF, McPhee MV, Campbell P, Winemiller KO (2004) Phylogeography and intraspecific genetic variation of prochilodontid fishes endemic to rivers of northern South America. *Journal of Fish Biology* **64**, 186-201.

Uiblein F (2016) Oceans of hypotheses – comprehensive alpha-taxonomic accounts and integrative taxonomic revisions in Marine Biology Research. *Marine Biology Research* **12**, 783-784.

United Nations (2017) The First Global Integrated Marine Assessment. World Ocean Assessment I (ed. Division for Ocean Affairs and the Law of the Sea), p. 1752. Cambridge University Press, Cambridge.

Uthicke S, Benzie JAH (2003) Gene flow and population history in high dispersal marine invertebrates: mitochondrial DNA analysis of *Holothuria nobilis* (Echinodermata: Holothuroidea) populations from the Indo-Pacific. *Molecular Ecology* **12**, 2635-2648.

Van den Broeck H, Breugelmans K, De Wolf H, Backeljau T (2008) Completely disjunct mitochondrial DNA haplotype distribution without a phylogeographic break in a planktonic developing gastropod. *Marine Biology* **153**, 421-429.

van Tienderen KM, van der Meij SET (2017) Extreme mitochondrial variation in the Atlantic gall crab *Opecarcinus hypostegus* (Decapoda: Cryptochiridae) reveals adaptive genetic divergence over Agaricia coral hosts. *Scientific Reports* **7**, 39461.

Vargas J, Pérez M, Toro J, Astorga MP (2015) Presence of two mitochondrial genomes in the mytilid *Perumytilus purpuratus*: Phylogenetic evidence for doubly uniparental inheritance. *Genetics and Molecular Biology* **38**, 173-181.

Vergara-Chen C, González-Wangüemert M, Marcos C, Pérez-Ruzafa Á (2010) Genetic diversity and connectivity remain high in *Holothuria polii* (Delle Chiaje 1823) across a coastal lagoon-open sea environmental gradient. *Genetica* **138**, 895-906.

Virgilio M, Backeljau T, Nevado B, De Meyer M (2010) Comparative performances of DNA barcoding across insect orders. *BMC Bioinformatics* **11**, 206.

Vollmer SV, Palumbi SR (2007) Restricted gene flow in the Caribbean staghorn coral *Acropora cervicornis*: implications for the recovery of endangered reefs. *Journal of Heredity* **98**, 40-50.

Wade CM, Mordan PB, Naggs F (2006) Evolutionary relationships among the Pulmonate land snails and slugs (Pulmonata, Stylommatophora). *Biological Journal of the Linnean Society* **87**, 593-610.

Wakeley J (2001) The coalescent in an island model of population subdivision with variation among demes. *Theoretical Population Biology* **59**, 133-144.

Wan Q-H, Wu H, Fujihara T, Fang S-G (2004) Which genetic marker for which conservation genetics issue? *Electrophoresis* **25**, 2165-2176.

Wang J (2012) On the measurements of genetic differentiation among populations. *Genetics Research* **94**, 275-289.

Wares JP, Barber PH, Ross-Ibarra J, Sotka EE, Toonen RJ (2006) Mitochondrial DNA and population size. *Science* **314**, 1388-1390.

Wares JP, Cunningham CW (2001) Phylogeography and historical ecology of the North Atlantic intertidal. *Evolution* **55**, 2455-2469.

Wares JP, Goldwater DS, Kong BY, Cunningham CW (2002) Refuting a controversial case of a human-mediated marine species introduction. *Ecology Letters* **5**, 577-584.

Waters JM, McCulloch GA, Eason JA (2007) Marine biogeographical structure in two highly dispersive gastropods: implications for trans-Tasman dispersal. *Journal of Biogeography* **34**, 678-687.

Watterson G (1975) On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology* **7**, 256-276.

Wecker P, Fournier A, Bosserelle P*, et al.* (2015) Dinoflagellate diversity among nudibranchs and sponges from French Polynesia: Insights into associations and transfer. *Comptes Rendus Biologies* **338**, 278-283.

Weersing K, Toonen RJ (2009) Population genetics, larval dispersal, and connectivity in marine systems. *Marine Ecology Progress Series* **393**, 1-12.

Weir BS, Cockerham C (1984) Estimating F-Statistics for the Analysis of Population Structure. *Evolution* **38**, 1358-1370.

Weiss S, Ferrand N (2007) *Phylogeography of southern European refugia* Springer, Netherlands.

Wessel P, Smith WHF (1996) A global, self-consistent, hierarchical, high-resolution shoreline database. *Journal of Geophysical Research: Solid Earth* **101**, 8741-8743.

Whitlock MC (2011) G'_{ST} and D do not replace F_{ST}. *Molecular Ecology* **20**, 1083-1091.

Whitlock MC, McCauley DE (1999) Indirect measures of gene flow and migration: $F_{ST} \neq$ 1/(4Nm+1). *Heredity* **82**, 117-125.

Wiemers M, Fiedler K (2007) Does the DNA barcoding gap exist? – a case study in blue butterflies (Lepidoptera: Lycaenidae). *Frontiers in Zoology* **4**, 8.

Wilke T, Pfenninger M (2002) Separating historic events from recurrent processes in cryptic species: phylogeography of mud snails (*Hydrobia* spp.). *Molecular Ecology* **11**, 1439-1451.

Wilkins JS (2010) What is a species? Essences and generation. *Theory in Biosciences* **129**, 141-148.

Wilkins JS (2011) Philosophically speaking, how many species concepts are there? *Zootaxa* **2765**, 58-60.

Williams ST (2016) Molluscan shell colour. *Biological Reviews*, 1-20.

Williams ST, Foster PG, Littlewood DTJ (2014) The complete mitochondrial genome of a turbinid vetigastropod from MiSeq Illumina sequencing of genomic DNA and steps towards a resolved gastropod phylogeny. *Gene* **533**, 38-47.

Williams ST, Reid DG (2004) Speciation and diversity on tropical rocky shores: a global phylogeny of snails of the genus *Echinolittorina*. *Evolution* **58**, 2227-2251.

Williams ST, Reid DG, Littlewood DTJ (2003) A molecular phylogeny of the Littorininae (Gastropoda: Littorinidae): unequal evolutionary rates, morphological parallelism, and biogeography of the Southern Ocean. *Molecular Phylogenetics and Evolution* **28**, 60-86.

Wright S (1931) Evolution in mendelian populations. *Genetics* **16**, 97-159.

Wright S (1949) The genetical structure of populations. *Annals of Eugenics* **15**, 323-354.

Wright S (1978) *Evolution and genetics of populations, Vol. 4. Variability within and among natural populations* University of Chicago Press, Chicago.

Xu B, Yang Z (2013) pamlX: a graphical user interface for PAML. *Molecular Biology and Evolution* **30**, 2723-2724.

Yang Z (1998) Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Molecular Biology and Evolution* **15**, 568-573.

Yang Z, Nielsen R (1998) Synonymous and nonsynonymous rate variation in nuclear genes of mammals. *Journal of Molecular Evolution* **46**, 409-418.

Yeates DK, Seago A, Nelson L*, et al.* (2011) Integrative taxonomy, or iterative taxonomy? *Systematic Entomology* **36**, 209-217.

Zeng K, Fu Y-X, Shi S, Wu C-I (2006) Statistical tests for detecting positive selection by utilizing high-frequency variants. *Genetics* **174**, 1431-1439.

Zhang J, Kapli P, Pavlidis P, Stamatakis A (2013) A general species delimitation method with applications to phylogenetic placements. *Bioinformatics* **29**, 2869-2876.

Curriculum vitae

Séverine FOURDRILIS

French, Belgium
full professional proficiency in English
☎ +32(0)2.627.42.85   📱 +32(0)478.49.70.36
severine.fourdrilis@gmail.com
www.linkedin.com/in/fourdrilis

MARINE BIOLOGIST

molecular biology • coral reefs

RESEARCH STATEMENT

→ Following the completion of my PhD, I am very keen to carry out post-doctoral research to investigate how coral reefs diversity is maintained through **climate change and anthropogenic disturbances**

→ I am a passionate advocate of oceans and want to draw upon my molecular biology skills for **sustainable development and management of coral reef ecosystems**

→ **Integrative approaches** such as coupling genoproteomics and transcriptomics, **from genes to ecosystems**, is a key approach to fully understand individual interactions in relation to their ecosystem

SKILLS

Proteomics   2D-DIGE • HPLC • MASS SPECTROMETRY • WESTERN BLOTTING • ELISA • SCANNER TYPHOON9400 • BIOINFORMATICS DecyderD, MassLynx, Flex Analysis

Genetics   PCR • ELECTROPHORESIS • DNA SEQUENCING • POPULATION GENETICS • PHYLOGENY • BIOINFORMATICS Arlequin, Beast, DnaSP, jModeltest, Mega, Migrate-n, MrBayes, Network, Paup, R, Seqscape

Ecology   GLOBAL FUNCTIONING OF FRESHWATER, MARINE AND TERRESTRIAL ECOSYSTEMS • WATER ANALYSIS • LANDSCAPE RESTORATION • GIS • MALACOLOGY • ICHTHYOPLANKTON STAINING • SCLEROCHRONOLOGY • PIT-TAGGING • SCANNING ELECTRON MICROSCOPY • BIOSTATISTICS

PROFESSIONAL EXPERIENCE

2011-present   **PhD** in evolutionary biology at the Royal Belgian Institute of Natural Sciences, a world-class research institute (supervisors: Dr. Backeljau from the RBINS, Dr. Mardulyn from the University of Brussels)

POPULATION GENETICS • PHYLOGENY • MARINE CONNECTIVITY • INTERNATIONAL CONFERENCES • PEER-REVIEWED PUBLICATION

Project & Student management

2009	**R&D SCIENTIST** in charge of the implementation of proteomic facilities at the Laboratory of Functional and Evolutionary Entomology

PROTEOMICS • TERMITES FARMING • EXTRACTION OF TERMITE GUT ENZYMES INVOLVED IN LIGNOCELLULOSE DEGRADATION

2008	**R&D SCIENTIST** developing protocols for allergen quantification in food, at the Mass Spectrometry Laboratory (supervisor: Dr. De Pauw) within a multi-partners project

PROTEOMICS • LC-MS/MS • REPORT WRITING *http://hdl.handle.net/2268/64291* • POSTER COMMUNICATION *http://hdl.handle.net/2268/59124* • METHOD DEVELOPMENT

2007	**R&D SCIENTIST** at the Laboratory of Bioenergetics and Cellular Physiology (supervisor: Dr. Sluse)

PROTEOMICS • FISH DISSECTION • ECOTOXICOLOGY OF METHYLMERCURY ON LIVER ENERGY METABOLISM *http://hdl.handle.net/2268/66420*

2006	**TOUR GUIDE** at the Scientific Culture Centre, University of Brussels

COMMUNICATING SCIENCE TO NON-SCIENTIFIC PUBLIC

2004	**TRAINEESHIP** at the French National Institute for Agricultural Research

PIT TAGGING DATA ANALYSIS

2003	**TRAINEESHIP** at the University of Liege

SCANNING ELECTRON MICROSCOPY ON CORAL MUCUS

2001	**TRAINEESHIP** (1 month) Institute of Research for Development

SCLEROCHRONOLOGY PRACTICE ON FISH OTOLITHS

Genetics

Proteomics

Versatility in biology

Teamwork

Fieldwork & Laboratory

Writing skills French/English

ACADEMIC ACHIEVEMENTS

2015	**COURSE** (9 weeks) edX online platform (*verified certificate*) **Tropical coastal ecosystems (coral reefs, mangroves) management**

2013	**COURSE** (35h) Ecole Pratique des Hautes Etudes (EPHE) **Coral reef ecosystems**

2008	**COURSE** (70h) Belgian Development Agency **Addressing international cooperation**

2004	**COURSE** (30h) Digital campus « Environment and

Sustainable development

Marine ecosystems

Development » (ENVAM), University of Rennes I
Littoral morphology

2003   **COURSE** (10 days) Marine Research Station STARESO,
            Corsica
            Mediterranean ecosystems

2001-2005   **BACHELOR AND MASTER OF BIOLOGY**, Universities
            of Rennes I (France) and Liege (Belgium, Erasmus
            agreement). Skills gained during the Master's thesis:

            AUTONOMY • TROPICAL ZOOLOGY • ICHTHYOPLANKTON
            TISSUE STAINING AND MORPHOMETRY

PUBLICATIONS

1. Kirsch S., **Fourdrilis S. (co-author)**, Dobson R., Scippo M.L., Maghuin-Rogister G., De Pauw
    E. (2009) Quantitative methods for food allergens: a review. *Analytical and Bioanalytical
    Chemistry* 395 (1): 57-67. http://rd.springer.com/article/10.1007%2Fs00216-009-2869-7

2. **Fourdrilis S.**, Mardulyn P., Hardy O.J., Jordaens K., De Frias Martins A.M., Backeljau T.
    (2016) Mitochondrial DNA hyperdiversity and its potential causes in the marine periwinkle
    Melarhaphe   neritoides   (Mollusca: Gastropoda).   *PeerJ*   4(e2549).
    https://peerj.com/articles/2549/?td=fb

3. **Fourdrilis S.**, Mardulyn P., Backeljau T. (*in preparation*) Implications of mtDNA hyperdiversity
    for assessing population genetic differentiation and connectivity: deceiving haplotypic
    differentiation in a panmictic periwinkle in the North East Atlantic.

4. **Fourdrilis S.**, Backeljau T. (*in preparation*) Comparative mitogenomics of Littorinidae
    (Mollusca: Gastropoda) and molecular phylogeny of Littorinimorpha.

INTERESTS

Aquariology, Scuba diving (beginner), Naturalism, DIY, Conversations about
environmental causes