THESIS / THÈSE

MASTER IN BIOLOGY OF ORGANISMS AND ECOLOGY

Population genetics of the invasive snail Myosotella

Cherpion, Laura

Award date:
2024

Awarding institution:
University of Namur
Université Catholique de Louvain

Link to publication

Université catholique de Louvain
École de biologie

Université de Namur
Département de biologie

Population genetics of the invasive snail *Myosotella*

Cherpion Laura

Mémoire présenté en vue de l'obtention du diplôme
de Master en Biologie des Organismes et Ecologie

Promoteur : Alice Dennis (UNamur)

Année académique 2023-2024

Abstract :

An invasive species can have a negative impact on the biodiversity, the ecosystem and the economy. Snails in the genus *Myosotella* are native to coastal Europe, but they are also invasive. That means that we find *Myosotella* all over the world. One theory is that they were spread by the help of human activities in different countries. We know well that local adaptations, gene flow and selection play an important role in the colonization of invasive species. Study the biogeography and the populations genetic is therefore necessary to better understand this invasive genus. The objective of this master thesis is to study the population genetics of *Myosotella* all over the world. We focus on the invasive aspect of the snails *Myosotella* and try to better understand their origin. We studied individuals coming from 16 different location by extracting DNA, Sanger sequencing and analyzing sequence variation. Visualization with a haplotype network and a phylogenetic tree have helped determine if there is a particular genotype which is worldwide spread or not. The results show three different groups which are divided into three putative species. Other groups of Azorean samples are also present in the middle of the haplotype network which could correspond to a fourth species. We compared our results with the distribution of species, and we suggest that one of the species is *M. myosotis*. These snails can have been transferred from one place to another by human activity which can explain why the sequences are similar. Further analyses should be done concerning the morphology as it is different within and between species, but also concerning the local adaptations of those snails.

Keywords: Populations genetic, *Myosotella*, invasive, distribution , genetic structure, morphology

Introduction

Populations genetics and biogeography of invasive species

An invasive species is defined as a species that is introduced, established and spread in environments outside their native range or in environments where the species do not evolve in the first place (Prentis *et al.*, 2008). It is interesting to study invasive species because they can have a negative impact on the biodiversity (Allendorf and Lundquist, 2003; Lee, 2002), the ecosystem (Lee, 2002) and the economy (Allendorf and Lundquist, 2003; Sakai *et al.*, 2001). Because of this, invasive species have been studied in many disciplines in order to manage their impacts (Sakai *et al.*, 2001). Moreover, processes that allow species to become a successful invader are not well understood (Tsutsui *et al.*, 2000) and studying invasive species can help us to better understand evolutionary processes in general (van Boheemen *et al.*, 2019; Verhoeven *et al.*, 2011). In fact, traits enabling invasiveness are important, because invasive species can have a great potential for adaptations, and those adaptations can in turn help understand why some species are invasive (Prentis *et al.*, 2008). An invasion of a species is generally described to occur in two steps: First, the introduction, survival and establishment of the species. Second, the invasive species must spread and replace the

native species. The genetic analysis of invasive species can allow us to predict the success of those two steps (Allendorf and Lundquist, 2003).

In the study of biogeography, the distribution and genetic structure of populations is compared across large distances in the world (Golikov *et al.*, 1990). This allows the limitations of dispersion of species to be inferred. This means how species spread in new environment and established in those environments. Across the globe, there is a limit to the dispersion of species and we can use models to predict how this progresses (Wiens, 2011). In reality, this is a principle that draws from many different approaches (Craw, 1984), including geography, biology (Muller-Hohenstein, 2001), and ecology (Riddle *et al.*, 2008; Wiens, 2011), but also to the phylogenetic relationships among species (Wiens and Donoghue, 2004). These phylogenies can be used to determine the different connections existing between areas (Wiens and Donoghue, 2004). In fact, in the biogeography studies researchers have to take historical and evolutionary but also the ecological aspects in consideration (Muller-Hohenstein, 2001). Thanks to this, we can calculate the incidence and the total number of invasion at different places in the world (Binggeli, 1996), better understand evolutionary processes (Wiens and Donoghue, 2004), and analyze the dispersion of invasive species (Dark, 2004).

A local adaptation is defined as an evolution or an adaptation due to a geographically varying selection (Sanford and Kelly, 2011; Tiffin and Ross-Ibarra, 2014). It arises by local selection for genotype with greater fitness in a particular environment (Si *et al.*, 2014). There is an interplay between gene flow and selection that ultimately determines the extent of the local adaptation (Savolainen *et al.*, 2007). Actually, the gene flow is a mechanism that has a restrictive effect on local adaptations because this mechanism is opposed to local selection (Lenormand, 2002). Selection is an evolutionary process that can act on countless loci in the genome, but it can also be disrupted by the homogenizing forces of migration (Bürger and Akerman, 2011). Without migration, local selective processes allow species to adapt to their environment (Lenormand, 2002; McDermott and McDonald, 1993). In addition, we can see a divergence in the selection over space if there is a gradient in the environment. Therefore, the populations can evolve differently (Sanford and Kelly, 2011). Afterwards, gene flow results from migration of alleles from one genetic pool to another (Tigano and Friesen, 2016). Sometimes it can lead to the establishment of another population in another region (McDermott and McDonald, 1993). Gene flow can be beneficial because it can restore the genetic variation in a local population (Lenormand, 2002). As a matter of fact, the gene flow can contribute to the rise of the genetic diversity and it can promote a rapid evolution (López-Goldar and Agrawal, 2021). Otherwise, gene flow can disrupt evolution in individuals by inhibition adaptations to different conditions in the environment or it can promote evolution by extending new genes and new mix of genes (Slatkin, 1987). Additionally, as explained by Sanford and Kelly, a limited gene flow can increase the chances of local adaptions (Hedgecock, 1986; Sanford and Kelly, 2011). Nevertheless, low gene flow can retard the local adaptations of species if the selection is not strong enough (Caprio and Tabashnik, 1992). In addition, gene flow can promote local adaptation if the different populations meet the same changes in the environment (López-Goldar and Agrawal, 2021). Even if diverse populations are linked with a high level of gene flow, it is possible to maintain differentiation thanks to a strong selection

after the dispersion (Sanford and Kelly, 2011). On the contrary, gene flow can be negative for local adaptations. There is a loss of adaptive alleles if selection doesn't act strongly enough and it can have a homogenizing effect and lead to limited local adaptations (Tigano and Friesen, 2016). If the level of gene flow is too high it can possibly neutralize the effect of selection on individuals because it will decrease the genetic differentiation among populations (Hamrick, 1982). Finally, we know that the environment plays a significant role in local adaptations. In reality, the environment in which the species develop themselves are heterogenous (Sanford and Kelly, 2011). The interaction between biotic and abiotic factors effect in the environment of the species, leading to a complex spatial pattern in adaptation and/or selective pressures. Then, the change in the biotic or abiotic parameters of the environment will have an impact on local adaptations (López-Goldar and Agrawal, 2021). It is also important to mention that two evolutionary pressures, local adaptations and dispersal evolution, play an important role in the spread of invasive species (Andrade-Restrepo *et al.*, 2019). For invasive species, differences in ecology between the native and the introduced places can lead to the development of local adaptations. We know as well that species which spread and invade new places over different part of the world will face new environmental conditions. (Colautti and Lau, 2016). Therefore, local adaptations are considered an important element in the invasion of species and depending on the situations, the level of the impact of local adaptations on the success of the invasion will be different. Local adaptations can also enhance the expansion of individuals (Li *et al.*, 2015). Moreover, when a species colonizes a new environment, if local adaptations evolve faster, then this can lead to a better survival rate and also a better reproductive capacity.

Those two elements are required to enhance invasiveness (facilitate the colonization of environment) (Colautti and Lau, 2016).

The basic principle of population genetics is that we can compare the genetic structure of populations, and observe changes over time and space (Ewens, 2013). We can also study local adaptations of individuals by identifying traits that are adaptive or associated with different ecological variables (Tiffin and Ross-Ibarra, 2014). Furthermore, we can understand the evolution of adaptive mechanisms and use molecular technics to better understand how these genes contribute to the invasion of species. Studying invasive species by using molecular techniques can help us measure the dynamics of populations and directly inform some management, prevention and restorations plans (Le Roux and Wieczorek, 2009).

The genus Myosotella

The model used here is a genus of snail named *Myosotella (*Phylum *mollusca*) (Rudy Jr *et al.*, 2013). The genus belongs to the order Basommatophora (González *et al.*, 2016; Mitov, 2016), family Ellobiidae, and subfamily Pythiinae (Antonio M. de Frias, 1996; González *et al.*, 2016; Mitov, 2016). This genus is composed of several species (de Frias Martins and Mendes, n.d.) (as *M. myosotis* and *M. denticulata* (Morton, 1955)). However, sometimes we can misidentify different snails because individuals may look alike and have some differences (Rudy Jr *et al.*, 2013). Different names have also be given to the same species (synonyms: *Ovatella myosotis* (Paulson, 1957)) and the same name has be given to diverse species (de Frias Martins and Mendes, n.d.). Within this diversity of reported species, there is a great deal of diversity concerning morphology. There is a high plasticity in the proportions of the shells (relative length and width), the colour, and

also the apertural dentition of *Myosotella myosotis*. Morphological comparisons of shell shape have suggested that the shells can have different morphology. In fact, we find a high diversity for the species within the genus *Myosotella* relating to the shells of the snails (Brown and Lydeard, 2010; de Frias Martins and Mendes, n.d.; Herbert, 2012; Mitov, 2016). There even is a lot of diversity within the same species (Herbert, 2012) or between the different species of the genus (Morton, 1955). A previous work found out that there is a local and regional variety for the shells of the snails depending on the salinity of the environment. The snails living in an environment with a high salinity have a shell that is thicker, bigger, and longer, while the snails living in a low-salinity environment have a thinner and shorter shell. Furthermore, the snails found at different places possess different shells (Mitov, 2016). If we have a closer look at *Myosotella myosotis*, one of the species that is well described in the genus (look at the picture in (Rudy Jr *et al.*, 2013)), its height can be 12 mm long and 6 mm wide (Mitov, 2016). According to Rudy Jr *et al.*, the snails can reach a size of 8 mm. The shell can be of different colours (black or brown) and can have different shapes (2013). On the other hand, *M. denticulata* is considered shorter than *M. myosotis* and the two species don't have the same number of labral denticles (Herbert, 2012).

Myosotella possess lungs, meaning they are pulmonated gastropods (An, 2006; Carlton, 1992; Rudy Jr *et al.*, 2013). The genus is typically a halophilous species which is tolerant of a range of salinities (Pérez-Quintero, 2007) and it is considered to be euryhaline (Pérez-Quintero, 2013). On the subject of the reproductive system, they are hermaphroditic (Rudy Jr *et al.*, 2013). Additionally, the reproductive strategy of species will have an effect on the dispersal of species but also on the phylogeographic patterns. One of those reproductive strategies is the direct development larvae (crawling, non-pelagic larva which to not feed) (Blakeslee *et al.*, 2021). It is predicted that there is a lack of genetic structure in the marine species developing with a planktonic larva, compared to species that develop directly. In support of this, the levels of dispersal of the larva of sedentary marine species can be related with the population genetic structure (Kyle and Boulding, 2000). The larvae of the species in genus *Myosotella* are not planktonic (Al-Khafaji, 2021). So, the larvae of individuals develop directly (Blakeslee *et al.*, 2021).

The species in the genus *Myosotella* are native to coastal Europe (Berman, 1989; Carlton, 1992; Cohen and Zabin, 2009; Ruthensteiner and Stocker, 2009). For example, we find *M. myosotis* in Ireland, Scotland, the United-Kingdom, Denmark and in the countries along the Mediterranean sea, Black sea, Aegean sea, the Marmara (Mitov, 2016), and also in Belgium (**Figure 1 and 2**) (Bruyndoncx *et al.*, 2002)**.** On the other hand, *Myosotella* is not found near the tropics (Antonio M. de Frias, 1996). About the habitat of *Myosotella*, this is a genus that relies less on water proximity and is considered as supratidal or estuary species (Antonio M. de Frias, 1996) but it can also be considered as an intertidal species (Al-Khafaji, 2021). Moreover, concerning the environment of *Myosotella*, we know that *Myosotella* lives in salt marshes. Salt marshes are places showing various conditions because it regroups terrestrial and ocean characteristics (Vernberg, 1993). In addition to that, these snails can live in the intertidal spaces (Al-Khafaji, 2021) and according to Romero *et al.*, (2016), the different species living in intertidal space can better adapt themselves because they have various characteristics in order to tolerate changes in the environment. For *M. myosotis*, these snails live in wetland and marine coastal and

supratidal zones (Slapnik, R. 2018), in salt marshes or estuarine mudflats (Berman, 1989), or in the supralittoral fringe of lagoons and estuaries (Marcus and Marcus, 1965). The snails are also found among the roots of plants and under stones or debris at high spring tide level. It seems to be the same habitat order for the various places colonized by *M. myosotis* (Herbert, 2012). For *M. denticulata,* they live in the same places as *M. myosotis*, but we also find them in Marine Neritic zones (**Figure 1**) (Seddon, M.B., Killeen, I. & Slapnik, R.,2018, and in places with a lot of rocks and crevices (Sánchez *et al.*, 2021).
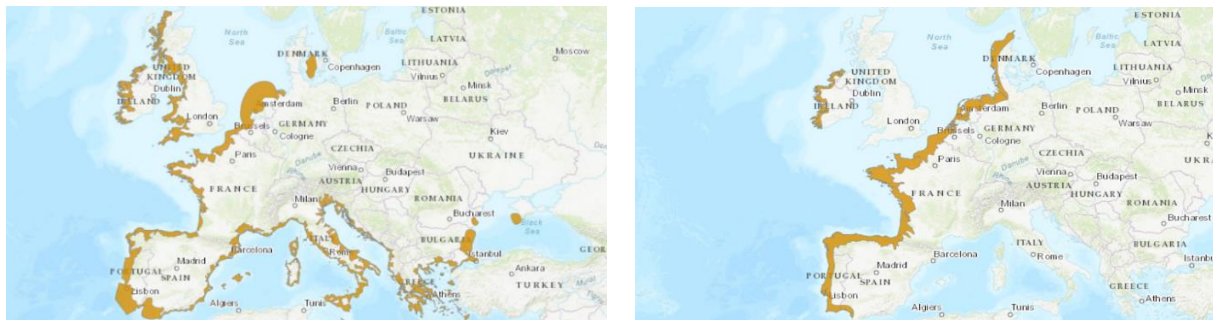
Figure 1: The first map represents the repartition of *Myosotella myosotis* (Slapnik, R., 2018) and the second shows the repartition *Myosotella denticulata* (B) (Seddon, M.B., Killeen, I. & Slapnik, R., 2018). *M. myosotis* is found in Mediterranean Europe and in the coast of Iberian peninsula (Sánchez *et al.*, 2021) but also in the Atlantic coast (Slapnik, R., 2018). For *M. denticulata*, we find those snails on the Atlantic coast and on the Iberian coasts (Sánchez *et al.*, 2021).

Figure 2: Map representing the distribution of *M. myosotis* in 2013 in the world (Mendes, 2013).

Despite its wide native range, this genus is an invasive species which has spread all over the world (Al-Khafaji, 2021; Herbert, 2012). Researchers found *M. myosotis* in the North and south East Pacific, also in the North west Atlantic. These snails are also present in New Zealand, Australia and in Iceland (Herbert, 2012). We can find them in California (Ray, 2005a), they have been introduced in Africa (Herbert, 2012; Peer *et al.*, 2018) but we cannot find them in Alaska (Ray, 2005b). We also find them in Iraq and Turkey (Al-Khafaji, 2021), and in Canada (Gillespie *et al.*, 2007).

The reasons for *Myosotella*'s occurrences all over the world is believed to be human activity (Herbert, 2012). In fact, as said before, *Myosotella* doesn't have a planktonic larva, meaning that individuals are spread by the movement of either adults or eggs (Al-Khafaji, 2021). Typically, estuary and marine invasive species are introduced by shipping into other places. In reality, they may be transported by attaching themselves to the surface of boats or by ballast water (Ansrp, 2005). During the 18th century, *Myosotella myosotis* was believed to have been introduced in the Atlantic with ballast stones (Carlton, 1992). *M. Myosotis* would have been introduced in New England with the maritime industry. Moreover, *Ovatella* has also been introduced in America with the movement of commercial Atlantic oysters by ballast water (Berman, 1989).

Interestingly, studies conducted on *M. myosotis* on several sites in the North-East Atlantic where the species has been introduced have shown that it had no negative impact on the environment. In fact, *M. myosotis* did not exploit the resources of the snail species already present and used different resources. However, at this time, they did not know if the species would have an impact on the ecosystems in southern Africa. They were particularly worried for estuarine Tomichia species found only in southern Africa but more studies are needed (Herbert, 2012).

The objective of this master thesis is to study the population genetics of *Myosotella* all over the world. We will focus on the invasive aspect of the snails *Myosotella* and try to better understand it. In order to do that, we will try to find out if there is a particular genotype which is worldwide spread or not.

Material and methods

Collections

We analyzed samples from different parts of the world. In total, we attempted to sequence 119 samples of *Myosotella* which have been collected in nine countries and 29 locations, representing two continents. We used samples from USA (Connecticut) Canada (Pender Island), Turkey (Kusadasi), Spain (Gijon), UK (E of Harwich and Bermuda), Italy (Pialassa dei Piomboni), Malta, France (L'horn, L'île Grand and Kerbrat and Cap d'ail) and diverse samples come from Portugal (Porto Novo Vimeiro and multiple places in the Azores) (**Figure 3**). The snails were provided by our supervisor Alice Dennis and collaborator Antonio Frias Martin. They were all collected in the field and preserved in ethanol. In our database, the samples are encoded by numbers from 6000 to 6309 (**Annex 1**).
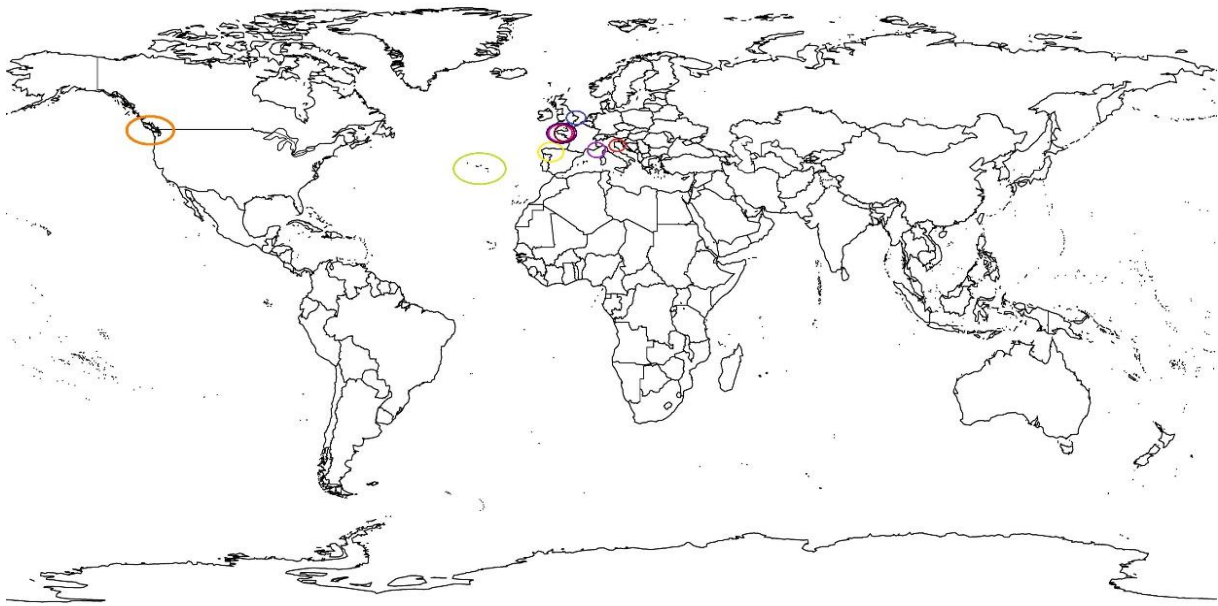
Figure 3: Map representing all the places where successfully amplified samples were collected (**Annex 1**). In orange: Canada, in green: the Azores, in red: Italy, in light blue: UK, in yellow: Spain, in light purple: Cap d'ail France, in pink: L'horn France, in dark purple: Kerbrat France, in brown: L'île Grand France.

DNA extraction and evaluation

The first step of the experiment was to extract DNA from each individual. We used extraction kits purchased from a company. We tried two different kits: the Nucleospin kit (Macherey-Nagel, Düren) and DNEasy blood and tissue kit (Qiagen, Hilden). After trial extractions, we chose the kit Dneasy blood and tissue kit because it allowed us to extract more DNA than the other. We followed the standard protocol of the kit. The only difference is that we incubated the tubes overnight at 56° C for the tissue digestion step.

After the extractions, we used a Nanodrop (Thermo Fisher Scientific, Waltham) to verify the quality and quantity of DNA. The Nanodrop is a Spectrophotometer and allows us to analyze DNA, RNA or proteins (Desjardins *et al.*, 2009). By using the Nanodrop, we can get an approximation of the DNA quantity in our samples and also the presence contaminants inside our samples. This is based on the molecular absorbance in function of the wavelength. In addition, by analyzing this curve, we can see if we have a good quality of DNA. Specifically, we used the concentration of our samples and the wavelength ratio 260/280 (this has to turn around 2 to inform a good quality of DNA).

PCR and Sanger Sequencing

Successful DNA extractions were next amplified with PCR. PCR allow us to amplify a stretch of DNA in our samples (Chase *et al.*, 1998). First of all, it is important to mention that we study the gene CO1. We are using a mitochondrial gene. We use this due to the property of this molecule that is

advantageous for evolutionary studies. The mitochondria are organelles located in the cell. They have an oxidative phosphorylation function and can also make ATP, and possess their own genome that is inherited from the mother only (Castro *et al.*, 1998). This mitochondrial genome is double stranded and circular (Castro *et al.*, 1998; Gaitán-Espitia *et al.*, 2013; Grande *et al.*, 2008) and can evolve 5 to 10 time more rapidly than the DNA of the nucleus, in part because the errors are not as well repaired during the replication (Castro *et al.*, 1998). Mitochondrial genes are widely used to build phylogenetic trees (Grande *et al.*, 2008). We are going to study the mitochondrial gene named CO1, which is also used as a "barcode" gene to identify species (Castle *et al.*, 2021; Dawnay *et al.*, 2007). By sequencing this same gene, we are able to compare among other species available in public databases.

To achieve it, we had to select the right primers. In this experiment, the primers that we used first were the "universal metazoan" primers HCO1 and LCO1 (Folmer *et al.*, 1994). We tried one other primer which is MCO1 (Dennis and Hellberg, 2010), it worked but it was not as successful as HCO1 and LCO1. However, we also tried other primers because our sequences from Sanger sequencing from HCO1 and LCO1 were not good. First, we tried primers for other genes: the 16S RNA (16LRN13398 and 16SRHTB) and 18S RNA (18SR174 and 18SF466). Then, we designed new primers for CO1 using the Primers design program implemented in Geneious Prime 2023.1.2 (https://www.geneious.com/). For this, we used the BLAST® program (NCBI, Bethesda, US) and we search for the entire mito-genomes of *Myosotella* in GenBank (https://blast.ncbi.nlm.nih.gov/Blast.cgi). We used the only two available: the mito-genome of a Spanish (NC_012434.1) and a

Azorean (JN606067.1) individual. Using these aligned sequences, we designed four sets of primers: MyA405F and MyA118R, MyA899F and MyA1515R, MyB429F and MyB1123R, MyB933F and MyB1519R. At the end we only kept the primers MyA1515R and MyA429F. At the end, we only kept the primers MyB429F and MyB1123R because this pair was the most successful (**Annex 2**). In addition, we saw double bands when we used these combinations of primers: MyB933F, MyB1519R and MyA405F, MyA118R.

All PCRs were run in 50 µl reactions using Promega reagents as follows:10 µl of 5X buffer, 1 µl of DNTP, 0.24 µl of Gotaq, 4 µl of each primer, 1 µl of MgCl2, 2 µl of DNA, and 27,76 of H20. We also worked to find the best conditions for PCR to maximize the success of the experiment. For example, we changed the temperature and the amount of DNA in the solutions. We tested several extension temperatures varying from 48°C to 56°C. We also added MgCl2 (1 and 2 µl) to increase the number of cycles at 40 cycles instead of 30 cycles in the PCR. After trials, we chose to use 1 µl MgCl2 and for the primers MyB429F and MyB1123R, we settled the different parameters for the PCR: 2.30 min denaturation step at 94°C;40 cycles of [94°C for 45 s, 55°C for 1 min, and 72°C for 1 min]; elongation step at 72°C for 10 min;18°C to finish.

PCR products were visualized on a 1.5% agarose gel. We prepared the agarose gel and loaded the PCR solutions into individual the wells of the gel. The gel allows us to see if the PCR worked before sending the sample into sequencing. If the PCR worked, we should observe a single band on the gel at 700 bp (**Annex 3**). If there are bands, we could send the DNA to Sanger sequencing. We did many agarose gels to check the various PCR tests that we have done. We also tried gel excisions for the

couple of primers MyA405F and MyA118R, but it didn't work.

In successful PCRs we proceeded to cleaning excess dNTPs from the samples before sending them for Sanger sequencing. We followed the exoSAP-IT PCR clean up Reagent indications (Thermo Fisher Scientific, Waltham). Samples were sent for sequencing in both primer directions to (Genewiz (https://www.genewiz.com/en-GB/) or Eurofins (https://www.eurofins.be/fr/)) Sanger sequencing. This is a method that allow us to determine the sequences of DNA. In total, we sequenced 189 PCR products.

Manual curation of Sanger Sequences: editing, alignment and evaluations

After all data collection, we analyzed the different results. First, we used Geneious to edit the raw Sanger sequences. Thus, we manually corrected the sequences (**Annex 4**). Then, we created an alignment of sequences from the same individual, compared them and made a consensus of those sequences for the samples that we have in both directions. We also compared our sequences with *Myosotella* consensus sequences of the gene CO1 that we found in GenBank. In addition, we utilized the blastn (blast for the nucleotides) with the highly similar sequences (megablast) in order to confirm if the sequences were *Myosotella* or not.

Analysis

Alignment

In order to construct the haplotype network, we aligned all the sequences of the different individual that worked to Sanger sequencing and the two sequences from GenBank. Then, we trimmed all the sequences to the same length, and deleted the samples that were too short (**Annex 6 and Annex 7).**

Haplotype Network

Next, we visualized the data using a haplotype network. This was implemented in R (R Core Team (2022)) with R studio (RStudio Team (2020)) with the packages pegas (Toparslan *et al.*, 2020), plyr (Wickham and Wickham, 2020), and reshape (Wickham, 2007) (**Annex 5**). This allows us to visualize the relationships among haplotypes. We can obtain an idea of the population structure with haplotype networks. We can see the different geographical locations containing each the haplotype and the number of mutations that separate them.

Phylogenetic tree

We created a Bayesian phylogenetic tree for all the samples that came from the sequencing and had good sequences to analyze. The phylogenetic tree can help us to see the relations between the different individuals from the different locations. We also used the blast program to obtain outgroup sequences from GenBank. We used a sequence of an individual whose sequencing worked. Then, we utilized the blastn (blast for the nucleotides) with the highly similar sequences (megablast). After that, we obtained the sequences that produced significant alignments (that resembles the sequence encoded) and we chose different species for comparing with *Myosotella*. We used *Polygyra cereolus* (NC_032036.1) and *Brachypodella costata* (KT602792.1) for that.

Trees were constructed using the program MrBayes (Ronquist *et al.*, 2012) in Geneious with the parameters: 1 million chain length, the invgamma variation rate and the GTR substitution model. In order to decide on these parameters, we used the program MEGA (Tamura *et al.*, 2021). Thanks to that, we were able to say that the Bayesian inference was the more adapted to

our data. We ran the tree four times in order to see if there were any variations between those trees.

Results

Sampling/PCRs done/ successful number/ locations/ primers

As said in the material and methods part, we analyzed 119 samples. We also used different primer pairs. In total, we performed 48 PCRs. We performed 20 PCRs with the primers HCO1 and LCO1 and 1 PCR with the primer HCO1 and MCO1. Then, 1 PCR with the primers 16S (16LRN13398 and 16SRHTB) and 1 with the primers 18S (18SR174 and 18SF466). After this, we designed new primers and we did 12 PCR with the primers set three (MyB429F and MyB1123R), 1 PCR primer set four (MyB933F and MyB1519R), 5 PCR with the set of primers two (MyA899F and MyA1515R) and 7 PCR with the first set of primers (MyA405F and MyA118R). We sequenced 189 PCR products in total. Of that number, only 53 sequences worked. At the end, 50 samples were good enough to be taken in our analyses.

Thus, we used 50 samples from six countries, 16 locations, representing two continents. We have samples from Canada (Pender Island), Spain (Gijon), UK (E of Harwich), Italy (Pialassa dei Piomboni), France (L'horn, L'île Grand and Kerbrat and Cap d'ail) and diverse samples come from Portugal (Porto Novo Vimeiro and multiple places in the Azores). We also used the two sequences from GenBank which come from Spain (Pontevedra, NW Spain (Grande *et al.*, 2004)) and the Azores. It is important to mention that many samples did not work. In addition, with the primer LCO1 and HCO1, the sample MbB6000 from Connecticut in USA showed other molluscs sequences when we basted it (blastn) and with MY6019 from L'île Grand, Brittany in France which were sequenced twice showed bacteria sequences when we blasted (Blastn) it. We had to remove the Connecticut samples for which we did PCR with the primers MyB429F and MyB1123R, from our analyses. We blasted the sequences of those samples with Blast, and we obtained sequences of a bacteria named Gammaproteobacteria bacterium and sequences other molluscs.

Haplotype network

The haplotype network (**Figure 4**) can be used to see the haplotypes used to construct the phylogeny. The little bars crossing the connections between circles correspond to inferred mutational steps between the haplotypes. The samples' locations are represented by distinct colors. The different numbers on the circle correspond to the different sites where the individuals have difference in the genetic structure (**Annex 8**). We decided to divide the haplotype into separate groups due to the difference in the number of mutations between the groups. We noticed that there are five different groups (A, B1, B2A, B2B and C). As we can see on the figure, we have 19 haplotype circles in total. There are 7 circles in the group C, 1 circle in the group B1 and B2A, 2 circles in the group B2B, and 8 circles in the group A. The size of the circle indicates the number of individuals possessing that haplotype. The study of the haplotype network will allow us to see the differences between individuals inside the groups. We didn't keep the exact same individuals as in the tree because some sequences were too short. Thus, samples My6285 (UK) and My6290 (Portugal) were deleted. The group C contains the samples from the Azores. We have samples from sites 1, site 3 and 4. Most of the samples from the site 2 and one

sample from France are in this category as well. The group B1 is composed of the samples from Canada, one sample from Horta, Faial (Azores) and the sequence from the Azores that we took from GenBank. In the middle, we see two separate groups: B2A which contains one sample from Horta, Faial (Azores) and B2B which contains two samples belonging to the site 2 (Azores) and one sample from W of Ponta (Azores). The last group A regroups the samples from France, Italy, Spain, UK, one sample from W of Ponta from the Azores and one sample from Areia Larga Madalena Pico from the Azores. The Spanish sequences from GenBank is also present in this section.

Phylogeny

The bayesian phylogeny (**Figure 5**) was constructed using the 50 different samples that were successfully sequenced. Thanks to this analyze, we can estimate the relationship between the groups. We see the different measures showing the support for that particular branch, ranging from 0.5379 to 1. Concerning the content of the tree, we used two groups as outgroups. Those two outgroups are different species than *Myosotella*. The two species used as outgroup are *Polygyra cereolus* and *Brachypodella costata*. There are gastropods in the group of heterobranchia. On the phylogenetic tree, we notice that there are three different groups (labeled A,B1 and C). The group C contains the different samples of the Azores. In this group, there are all the samples of the sites 1, 3 and 4. We also find

most of the site 2 samples, one sample coming from France and the sample from Porto Novo/ Vimeiro from Portugal. In group B1, the sister group of group C, we have all the samples coming from Canada and the Azorean sequence coming from GenBank. Two samples from Horta, Faial (Azores) are also found with one sample from Canada. Then, some samples from site 2 and W of Ponta (Azores) are represented in this section. In addition, we can see the groups B2A and B2B, corresponding to the haplotype network. Group B2A found in the haplotype network corresponds to the sample My6300 from Horta, Faial (Azores) which is considered similar to My6301 here in the tree. Moreover, the group B2B regroups the samples My6299, My6075 and My6096. The group A regroups mostly all the samples coming from France. We have the samples coming from Italy, Spain and UK. We also have one sample from W of Ponta in the Azores, one sample from Areia larga in the Azores and the Spanish sequences from GenBank. Moreover, we can look at the percent of identity between samples coming from the three groups (**Annex 9**). For example, between the sample My6003 from France (Kerbrat Brittany) (A) and the sample My6202 from Canada (B1), there is 83.2% of resemblance. After, between the sample My6106 from the Azores (Site 4) (C) and the samples My6202 from Canada (B1), there is 83,150% of resemblance. In last, between the sample My6003 from France (Kerbrat Brittany) (A) and the sample My6106 from the Azores (Site 4) (C), there is 84,190% of resemblance.
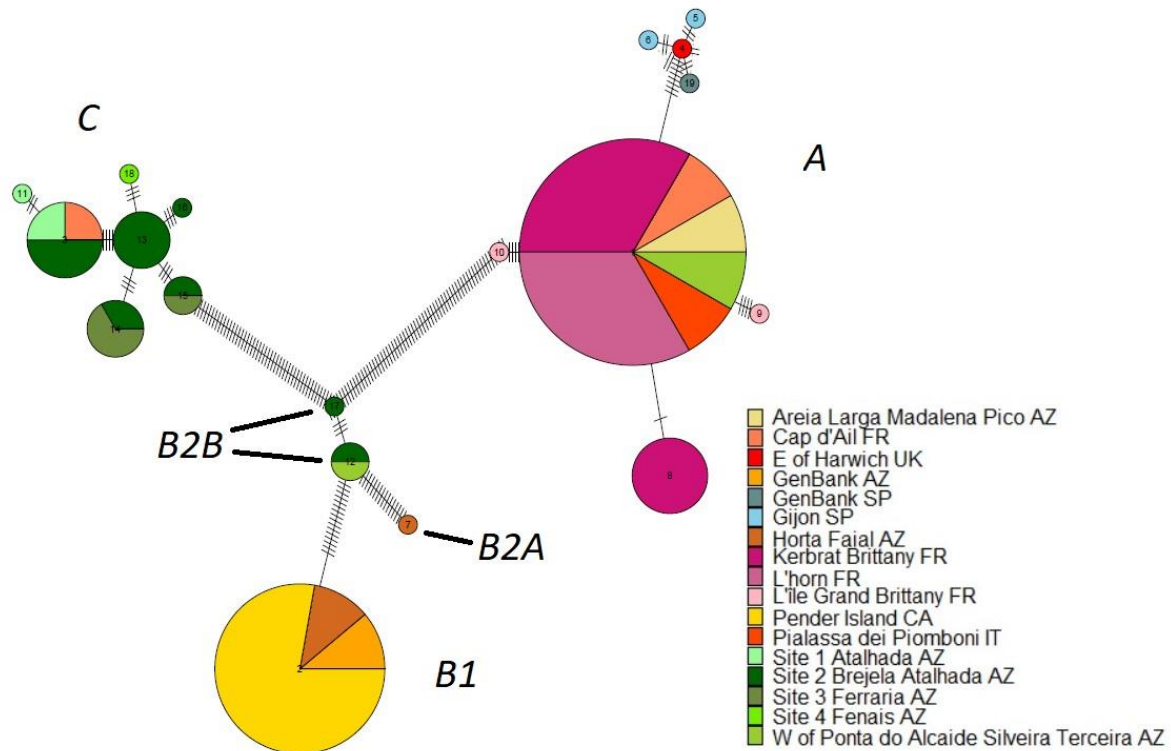
Figure 4: Haplotype network showing the relations between the individuals coming from various locations. The samples coming from different places are represented by different color. The different countries are written with abbreviations as AZ for Azores, CA for Canada, SP for Spain, IT for Italy, FR for France and UK for United Kingdom. The diverse groups containing the samples are written as well (A, B1, B2A, B2B and C).

Figure 5: Phylogenetic tree showing the link between the different individuals coming from different locations. We can see the number of the samples as their locations. The different groups are also represented (A, B1, B2A, B2B and C).

Discussion

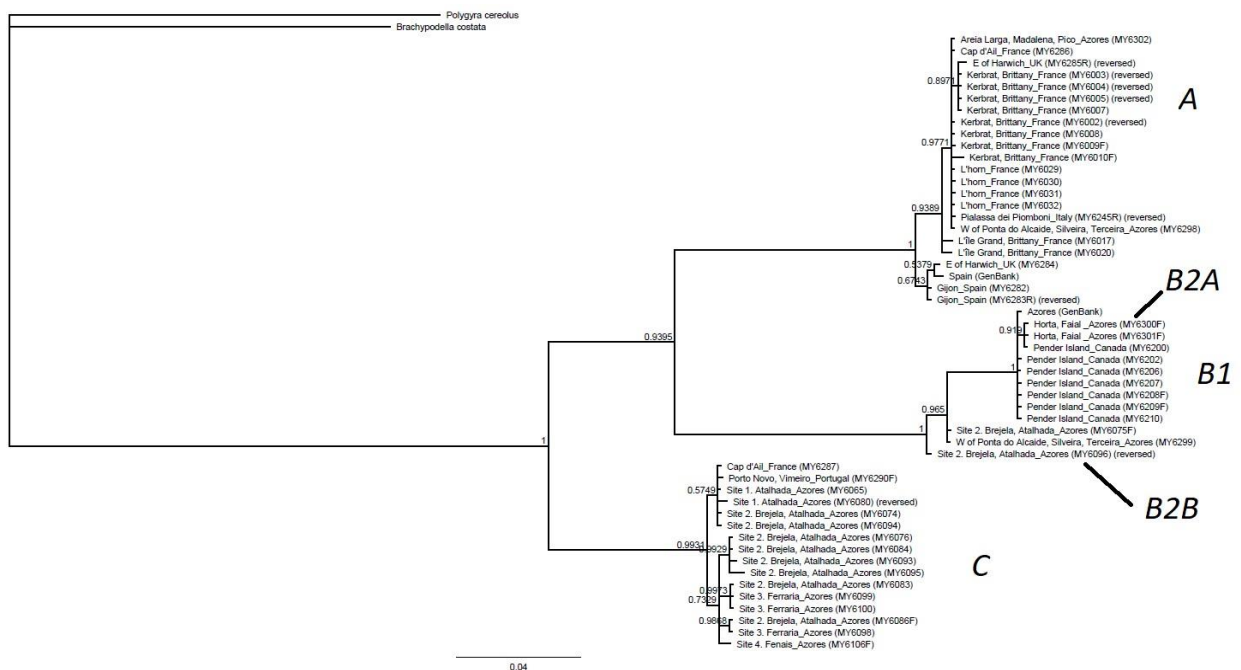We analyzed 50 samples from six countries and 16 different locations. If we compare the phylogenetic tree and the haplotype network, we find out that the results are mostly the same. We observe different groups (A, B1, B2A, B2B, and C) concerning the genetic structure of our different samples coming from multiple locations. The group C is composed of samples from the Azores from sites 1,3 and 4. We have most of the samples from the site 2 and one sample coming from France. The group B1 contains all the samples coming from Canada, samples from the Azores and the Azorean sequences coming from GenBank. The group A is mostly composed by the samples coming from France. We have the samples coming from Italy, Spain and UK. One sample from Areia larga and one from W of Ponta in the Azores and the Spanish sequences from GenBank are also represented in this association. If we look at the groups A, B1 and C, we can conclude that these might be three different species.

There is one significant difference between the phylogenetic tree and the haplotype network. We see two groups in the middle of the haplotype network. The group B2B which contains samples belonging to the site 2 of the Azores with one sample from W of Ponta in the Azores and the group B2A which contains one sample from Horta, Faial in the Azores. In the tree, the group B2A is considered similar to the other sample from Horta, Faial and to one sample from Canada (B1) and the group B2B is separated find in the group B1 but is separated from the central branch. Inspection of the DNA alignment shows that the sequence in B2A is very similar, but shorter. Based on this, we can conclude that the haplotype network is not accurate and that the individual found in this group should be with the

other samples of the group B1 as see in the phylogenetic tree. This difference could be explained by the fact that we used all the sequences length to do the tree but shorter sequences for the haplotype as all the sequences have to be cut at the same length. Despite this, we can still either assume that the group B2B correspond to a fourth species or maybe that it could be the same species, but isolated population of the group B1. This is possible that the species face different environmental conditions and that we could see a divergence between individuals. We can clearly see the divergence between these groups when we look at the phylogenetic tree and the haplotype network.

Moreover, if we have a closer look to the haplotype network, we perceive in the group C that we have 7 short haplotype circles compared to the two others where we have less haplotype circles (B1, B2A and B2B) or one bigger haplotype circle (A). That means that we have more genetic diversity between the individuals coming from the Azores and for the groups B1,B2A and B2B, we have less genetic diversity in the genetic structure. This supports the idea of having four possible species.

For the names of these possible species, we are not really able to give names for all the groups at this time. As said in the Introductions, a misidentification of the different snails is possible because individuals may look alike (Rudy Jr *et al.*, 2013). If we look at the **Annex 1,** the names of species have been linked with some locations, but it seems that the four species found in our trees don't correspond to the species that have been identified. In fact, the sample My6302 is according to A.F Martins, *M. bicolor* and according to A. Dennis, the samples from France (Kerbrat, L'horn and Île Grande) would be *M. denticulata*. Those two supposed species are found in the same group in the phylogenetic tree (group A). However, in this group, we see that all the

samples coming from continental Europe are in the same groups excepted one sample. Because this is a monophyletic group, we might be able to conclude that these are *M. denticulata* or *M. myosotis*. In fact, the distribution of the different species inside the genus *Myosotella* may help us to determine which species it is. We know that *M. denticulata* is found on Atlantic coast and on the Iberian coasts (Sánchez *et al.*, 2021). On the other hand, we know that *M. myosotis* is found in Mediterranean Europe, in the coast of Iberian peninsula (Sánchez *et al.*, 2021) and in the Atlantic coast (Slapnik, R., 2018). We can suggest that the group A correspond to this species because we find samples from the Mediterranean Europe and *M. denticulata* is not found there. However, according to Herbert (2012), there is a difference in the height and in the habitat of *M. denticulata* in comparison of *M. myosotis*. In addition to that, according to our collaborator Antonio Frias Martin, the morphology of the penile complex help to identify species (de Frias Martins and Mendes, n.d.). For the other groups, we are not able to confirm any species, but we can suppose that it doesn't correspond to *M. myosotis* or *M. denticulata*. At the end, we suggest that further analyses are needed. We need to analyze the internal and external morphologies of the diverse species but also their habitats. Those elements may be able to help us identify the species.

Additionally, concerning the relations between the separate groups, we can look at the phylogenetic tree. We notice that the group B1 is as closer to the group A than the group C. That means that the genetic structure of the population coming from Canada is closely related to the populations coming from the Azores and the Europe. So, the individuals from Canada have maybe been imported from the Azores or continental Europe. As we know, the larvae of *Myosotella* are not pelagic. In other words, the larvae is not capable to move from one place to another. So, the transfer of individual happens by the transportation of some organisms by boat due to human activity. We can also suppose that this introduction could be recent because the low diversity that we find in this group, *Myosotella* might be in the first step of invasion steps (Allendorf and Lundquist, 2003). For group A, we see that all of the samples coming from continental Europe are in the same group excepted one sample. On the other hand, the samples from the Azores are found mostly in the group C but they are also found in the two other groups. In addition to finding samples from the Azores, in the group A, we have all the samples coming from continental Europe excepted one samples from Cap d'ail in France which is found in the group C. It also could be explained by the transport of some species from one place to another by human activity.

Concerning the population genetic structure, we can do hypothesis with gene flow or migrations. In the literature they say that gene flow is opposed to the selection (Lenormand, 2002). It can have a homogenizing effect, limit local adaptations (Tigano and Friesen, 2016) and decrease the genetic differentiation among populations if the level of gene flow is too important (Hamrick, 1982). It could be interesting to have some nuclear markers, and sequences from more individuals, in order to confirm if there is gene flow or not (Neigel, 1997). In addition, multiple introduction can have an impact on the genetic structure throughout the range of species by homogenizing it, and this has the ability to rise genetic variation (Zalewski *et al.*, 2010). Multiple migrations can lead to the similar genetic structure (Abdelkrim *et al.*, 2005). In the case of *Myosotella*, if the exchanges of individuals occur all the time between the different

places, it might explain why individuals coming from places far apart are similar thanks to multiple migrations events.

As explained in the introduction, the reproductive strategy of species influence the dispersal of species and the phylogeographic patterns. We said that there are two strategies. One with a planktotrophic development and one with a direct development. In fact, Blakeslee *et al* found that in Littorina snails, the species with a direct development show a stronger genetic structure and large geographic extent. They concluded that the genetic structure of populations in species with a planktonic development had a greater connectivity and a smaller quantity of genetic differentiation (Blakeslee *et al.*, 2021). As *Myosotella* have a direct development, this could explain the large repartition of the genus all over the world. In addition, a strong genetic structure is also able to explain that the dispersion of the genetic variation between the different populations is broad. We can suppose that there is a high distribution which cause the three populations to be isolated. Moreover, if there is more genetic differentiation between the populations in direct development species, that could mean that there is less gene flow in those populations. This could also explain the isolation between the three groups. Additionally, selection may also keep a level of differentiation between species due to the difference of environment. As we know that invasive species can face to different environment conditions as it colonize different places (Colautti and Lau, 2016) and that the different species within the genus *Myosotella* are found in different habitats (Herbert, 2012). Thus, it could explain a variation in the genetic structure of the different populations.

On the other hand, in the literature, it has been proposed that genetic variation is important when talking about invasive species because it can be used to study the adaptions of species (Prentis *et al.*, 2008). Our analyzes are not enough to see the potential local adaptations or gene flow. In fact, studies of only one gene can't give us enough information concerning the local adaptations or the gene flow in the different populations. In order to study this, we must study more than one gene. We know that local adaptations are considered an important element in the invasion of species and depending on the situations, the level of the impact of local adaptations on the success of the invasion will be different. Actually, local adaptations are also important for invasive species because they can amplify the expansion of individuals (Li *et al.*, 2015). Moreover, when a species migrates to a new environment, high genetic diversity can aid in faster local adaptation, and maybe to a better survival rate and also a better reproductive capacity. Two elements that are required to improve the invasive aspect (Colautti and Lau, 2016). As a matter of fact, it could be interesting to study the genetic variation of the native populations and the introduced populations in order to better understand the invasive species (Zhao *et al.*, 2015). So, for further investigations, studying those two points could be relevant to better understand the genetic structure of an invasive species.

Interpretation of these results should also take their limitations into account, especially since this was a species that posed technical problems for the Sanger sequencing. During our experiments, we had some problems with the primers. In fact, it first seemed that the sequencing was not working with our samples because the primers were not enough suitable for our samples. Then, after the design of the new primers, we saw that all the samples didn't work with

the primers that we originally chose. More-over, the fact that we sequenced bacteria instead of molluscs shows that the "universal" primers were not suitable for our snail. It could be interesting to the designed new primers in order to have more samples. It could also be interesting to change the parameters of the PCR to maximize the results. Furthermore, many Sanger sequences were not successful. This should be explored to determine whether new primers can be designed, or if the bacterial contamination that was amplified is important to these snails.

It is also important to view these results in light of potential problems with using just one gene. As we know, we can use mitochondrial information to study the phylogeny. One explanation can be that the mitochondrial genome of molluscs do not always follow the expected rules. This has been documented in many taxa. As said in Ghiselli *et al.*, 2021, molluscs genomes show a fluctuation of size, there are high levels of gene boundaries that overlap, there is a high disparity about the composition in nucleotide which can lead to asymmetry. There are changes concerning the transcriptional orientation of genes like inversion which can lead to place those genes on strands differing by nucleotides compositions. There are also many rearrangements of genomes, duplications or reduction of some genes and insertion of new genes. Additionally, some molluscs as bivalve can show a doubly uniparental inheritance system. Typically, the mitochondrial genome is inherited only from the mother. In the case of doubly uniparental inheritance, both parents contribute to the transfer of the mitochondrial genome instead of only inherit the genome from the female (Ghiselli *et al.*, 2021). We can suppose that species inside the genus *Myosotella* can use both strategies. In fact, if a part of individuals inside

the genus *Myosotella* inherit the mitochondrial genome from both parents and other individuals inherit the mitochondrial genome from only the mother, it could explain why we can have distinct groups in our analyses. The sequences could show differences if this were occurring. The fact that the primers were not suitable for all the samples because the CO1 sequences contains a lot of variations. We could suggest that the distinct primers that we took and the primers that we design were too specific for all the samples we have. This hypothesis can explain why some of the samples worked well with the primers we used. Also, we could have other species inside the genus *Myosotella*, and the primers are suitable only for a couple of species.

Lastly, one other explanation for these results could be that some sequences are the NUMTs, also known as pseudogenes. These are cases where fragments of mitochondrial DNA which are inserted in the nuclear genome. This is a process that occurs when double stranded breaks are fixed in the genome. This could be a problem for the PCR analyses because the primers can be targeted to mitochondrial part but also to the NUMTs (Schultz and Hebert, 2022). Sometimes, the NUMTs can drive to some errors in phylogenetics trees if it is undiagnosed NUMTs (Ghiselli *et al.*, 2021; Schultz and Hebert, 2022). It can mostly occur when studying only one gene (Ghiselli *et al.*, 2021). NUMTS have the potential to exaggerate the visible species and the variation at CO1 among individuals of the same species. On the other hand, they say that it is rare that targeting the NUMTs instead of the mitochondrial sequences because the number of copy of mitochondrial gene CO1 is higher than for the NUMTs (Schultz and Hebert, 2022). So, an exaggeration of the richness of the haplotype or in

the number of species may be possible. It could also possibly be an explanation concerning the technical problem with the primers.

Conclusion

In conclusion, The objective of this master thesis was to study the population genetics of *Myosotella* all over the world. We tried to find out if there is a particular genotype which is spread worldwide, or not. As known, *Myosotella* is an invasive genus found all over the world. Thanks to our analyzes, we can conclude that there are three possible species in we look at the haplotype network and the phylogenetic tree. Moreover, there are a few samples in the middle of the haplotype network that correspond might correspond to a fourth species. We can suggest the name of *Myosotella myosotis* for the group containing most of the samples from Europe. For the other groups, we don't know what species it is. We can propose that there could be recent migration of individuals from one place to another. This might explain why we find the Azorean samples in every group, why the Canadian samples are closely related to the other groups, and why we can find one samples coming from France in the Azorean group. About the genetic structure of populations, we can suggest that there is some gene flow

within the diverse groups and that multiple introductions of species have an impact on the genetic structure of population. Moreover, selection can also keep a contrast between species due to the difference of environment. As we only just study the mitochondrial gene CO1, we had primers problems during our experiments. It might be explained by the different property of the mitochondrial genome or by the doubly uniparental inheritance instead. Finaly, the NUMTs could explain the disfunction of our primer and could also exaggerate the number of haplotypes. So, maybe there is an exaggeration about the number of species. For further experiments, studying the external and internal morphology of the individuals but also their habitat in order to identify the species could be relevant. It could also be interesting to study the local adaptations and the gene flow in the different population of *Myosotella,* and to examine this using a larger number of both mitochondrial and nuclear markers.

Annexes

Annex 1:

Table regrouping all the information concerning the different samples that were successfully sequenced in the experiment.

Country	Location	Dates	GPS	GPS	N	codes	Presumed species	Extraction	PCR	Collected by	Notes
Italy	Pialassa dei Piomboni	28 July 2023			1	My 6245		1	4		
Canada	Port Browning, Pender Island	May 2022	48.77	-123.26	7	My 6200-6210		1	2/6200+6202-1/6210-10/6209-11/6206+6207-12/6208		Dried out on desk. Added ethanol and put in freeyer
France	Cap d'Ail	02 July 2014			2	My 6286-6287		1	2 for 6286 and 1 for 6287	A.F. Martins	1.Specimens from Antonio Frias Martins (initial samples in Feb 2023)
Portugal	Porto Novo, Vimeiro, Portugal	23 December 1994			1	My 6290		1	2		3.Specimens from Antonio Frias Martins (initial samples in Feb 2023)
Spain	Gijon, N Spain	26 October 2016			2	My 6282-6283		1	2/samples	A.F. Martins	4. Specimens from Antonio Frias Martins (initial samples in Feb 2023)
UK	E of Harwich	09 September 2014			2	My 6284-6285		1	3/6284-4/6285	A.F. Martins	5. Specimens from Antonio Frias Martins (initial samples in Feb 2023)
Azores, Portugal	W of Ponta do Alcaide, Silveira, Terceira				2	My 6298-6299		1	2/6298-4/6299	A.F. Martins	10. Specimens from Antonio Frias Martins (initial samples in Feb 2023)
Azores, Portugal	Horta, Faial				2	My 6300-6301		1	2/6300+6301	A.F. Martins	11. Specimens from Antonio Frias Martins (initial samples in Feb 2023)
France	Kerbrat, Brittany				4	My 6029-6032	M. denticulata	1	11/6029-16/6030-13/6031-6/6032	A. Dennis	

Country	Location	Date	Latitude	Longitude		My numbers	Species			Collector	Notes
France	L'Horn, Kerbrat, near Plougoulm, Brittany	19 March 2023	48.685.548	-405.336	8	My 6002-6010	M. denticulata	1	14/6002+6004+6005-25/6003 1/6007+6008+6009+6010	A. Dennis	Salinity approx 3ppT. Snails under a piece of rusty metal, and on nearby muddy surface. Further upstream, marsh opens up and gets flat. Here many more, under Halimione portulacoides (aka sea purslane, and evergreen plant)
France	Île Grande, Brittany. South side of bridge	22 March 2023	487.962	-35.736	2	My 6017-6020	M. denticulata	1	13/6017-12/6020	A. Dennis	Under grass (lawn?) at path
Azores, Portugal	Atalhada, w of Lagoa, Sao Miguel. In front of Observatorio Vlcanologico e Geotermico. Site 1	8 May 2023	37.742	-25.589	2	My 6065/6080		1	7/6065-1/6080	A. Dennis and A. F. Martins	Also collected Ovatella, Pseudomelampus, Pedipes, maybe Auriculnella
Azores, Portugal	Brejela, Atalhada, W of Lagao, Sao Miguel site 2	8 May 2023			3	My 6083-6086		1	3/6083-5/6084-13/6086	A. Dennis and A. F. Martins	Group 1: under grass
Azores, Portugal	Brejela, Atalhada, W of Lagao, Sao Miguel. site 2	8 May 2023			7	My 6074-6076/6093-6096		1	2/6074-10/6075-12/6076-18/6093-4/6094-13/6095+6096	A. Dennis and A. F. Martins	Group 2: under rocks
Azores, Portugal	Ponta da Ferraria (by hot baths), Sao Miguel. Site 3	9 May 2023	37.858	-25.852	3	My 6098-6100		1	2/6098+6100-1/6099	A. Dennis and A. F. Martins	Is there a tendancy of lighter shells in lower intertidal individuals?
Azores, Portugal	Fenais da Luz, Sao Miguel. Site 4	9 May 2023	37.82	-25.64	1	My 6106		1	2	A. Dennis and A. F. Martins	Also collected Pedipes and Ovatella (?)
Azores, Portugal	Areia Larga, Madalena, Pico (tube 1)	3 October 2012			1	My 6302	M. bicolor	1	2	Andreia Brazao Cunha	Specimens from Antonio Frias Martins (second set of samples during my visit in May 2023)

Annex 2:

Table of all the primers that have been used.

| Primers	DNA sequences	References
LCO1	GGTCAACAAATCATAAAGATATTGG	(Folmer *et al.*, 1994)
HCO1	TAAACTTCAGGGTGACCAAAAAATCA	(Folmer *et al.*, 1994)
MCO1	GGTCAACAAATCATAAGGATATTGG	(Dennis and Hellberg, 2010)
16LRN 13398	CGC CTG TTT AAC AAA AAC AT	(Koufopanou *et al.*, 1999)
16SRHTB	ACG CCG GTT TGA ACT CAG ATC	(Koufopanou *et al.*, 1999)
18SR174	GTA TTA GCT CTA GAA TTG CC	(Mikkelsen *et al.*, 2019)
18SF466	AAA CGG CTA CCA CAT CCA AG	(Mikkelsen *et al.*, 2019)
MyA405F	TGGTGCGTCGGTGGATTTAG	Present study
MyA118R	TGAAAGTGTGCCACCACGTA	Present study
MyA899F	TCACCGCTGCCACCATAATTAT	Present study
MyA1515R	CACGGTAGTCTCCAGGTTGC	Present study
MyB429F	CTTTTCGCTACACCTTGCTGG	Present study
MyB1123R	CGTAATGGAAGTGGGCTACCA	Present study
MyB933F	TGGAATTAAGGTGTTCAGGTGACT	Present study
MyB1519R	CTGTAACCGTCGTCTCCAGG	Present study

Annex 3:

Picture showing an electrophoresis gel containing a ladder (for scale) and 7 successful PCR samples. The black bands show that there is DNA in our samples. So, we can send theses samples to Sanger sequencing.

Annex 4:

A)
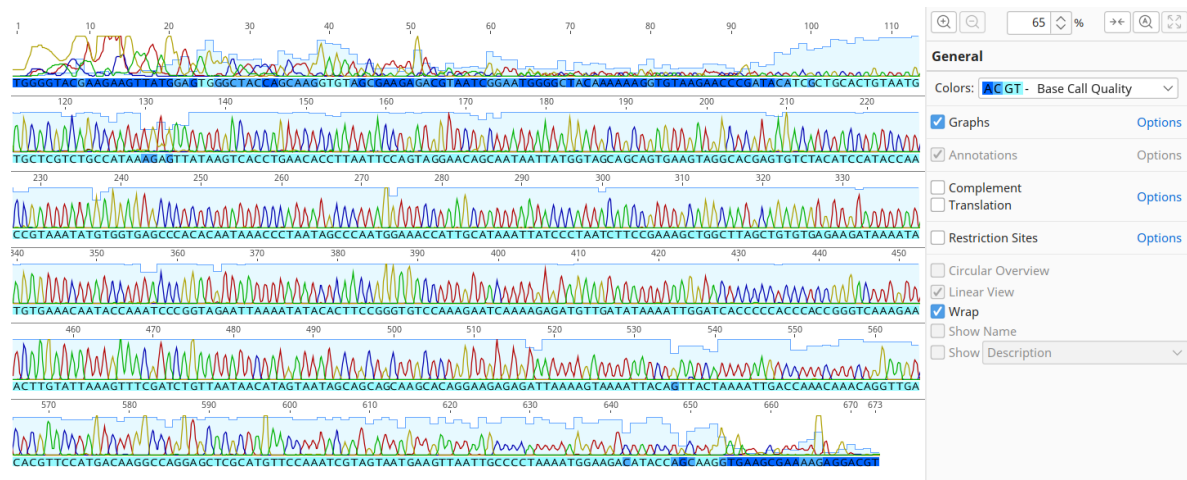
B)
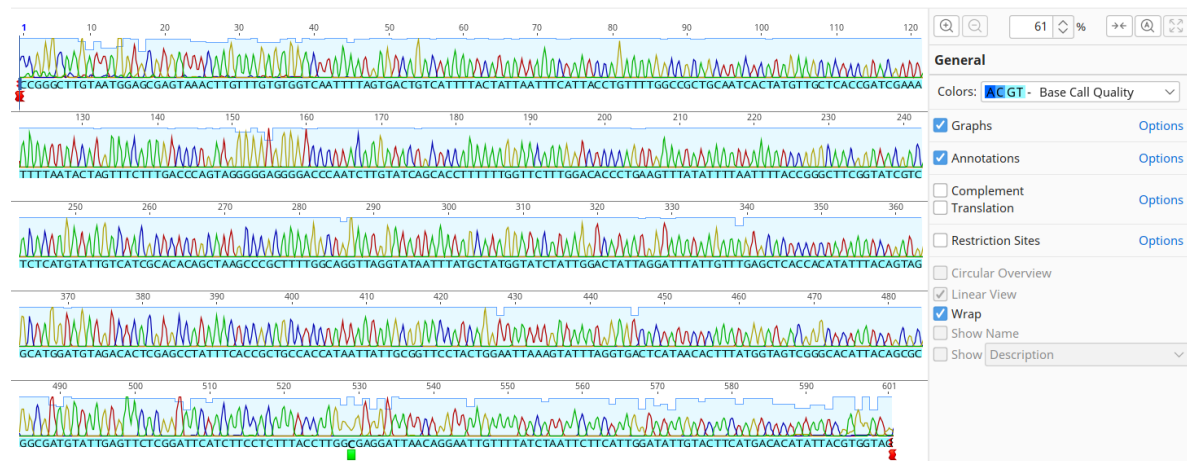
Pictures showing Sanger sequences before manual curation in A and after correction in B.

Annex 5:

R code used in our analyses to make the haplotype network.

```
#make sure your environment is clear
rm(list=ls(all=TRUE))

# Input data:
# a) a fasta alignment of a single gene
# b) a csv file with the sample name and location
   # Column names in the csv file should be "sample" and "site".
   # THE ORDER OF SAMPLES MUST BE THE SAME IN BOTH THE ALIGNMENT (FASTA) AND
SITES (CSV) FILES!
```

```
# Load the packages pegas, plyr, and reshape in R.
install.packages("pegas") #remember, only need to install packages once
install.packages("plyr")
install.packages("reshape")
library(pegas) #call these libraries every time you need to use them
library(plyr)
library(reshape)

# Determine your workspace directory (i.e the location where your FASTA and CSV File reside):
# can be looked up -> right-click on folder with files -> trades -> copy location-path
setwd("C:/Users/HP/Downloads/MT_R")


# Load the sample alignment in R:
alignment<-read.dna("final seq Copy.fasta", format="fasta")

al <- as.alignment(alignment)# Load the sample-sites file:
sites<-read.csv("final.csv",header = TRUE,dec = ".")
sites

# Calculate haplotypes and frequencies:
h <- haplotype(alignment)
h

# Create label vector for the haplotypes (replaces roman default labels):
labels <- as.character(c(1:nrow(h)))
labels
rownames(h) <- labels

# Build the haplotype network based on the haplotype frequencies.
net<-haploNet(h)
net

# Are the labels the same on the alignment and the sites file?
# (necessary to match the sites in the rigth order to the samples)
identical(as.character(sites$Samples),labels(alignment))

# Add a column for haplotypes (prep for assigning the haplotypes to the sites):
sites<-cbind(sites,haplotype=rep(NA, nrow(sites)))
sites

# Assign haplotypes to the samples automatically using a loop:
for (i in 1:length(labels(h))){
        sites$haplotype[attr(h, "index")[[i]]]<-i
}
```

```
# Build a matrix of frequencies that will tell R how to fill pies on haplotype network.
# Requires the packages 'plyr' and 'reshape'
# pegas-package can only colorate in "rainbow"
dfcount<-ddply(sites,.(Sites,haplotype),summarise,freq=length(Sites))
dfcount

dc <- cast(haplotype ~ Sites, data = dfcount, value = "freq", fill = 0)
dc
dc$haplotype<-NULL
dc<-as.matrix(as.data.frame(dc))
dc

# Plot the haplotype network
# Here you can fit following parameters:
#    attr(net, "freq"): size of the pie charts, a numer before attr()= multiply the pies with the number
#    scale.ratio= distance between the haplotypes
#    cex = scale of the labels
#    labels= FALSE -> no labels; TRUE -> labels (1:n)
#    font= scale of the font
#    legend= FALSE -> no automatically legend; TRUE -> automatically legend


#For the colours in the haplotype network
test=data.frame(location=unique(sites$Sites))
country<-c(rep("Azores",2), "France", "UK", "Spain", "Azores", rep("France",3), "Canada", "Italy",
rep("Azores",4), "Spain","Azores")
test$country <- country
test
test=test[order(test$country),]
test

palette1=colorRampPalette(c("lightgoldenrod","orange",
"chocolate","palegreen","darkgreen","darkolivegreen4","chartreuse2","yellowgreen"))(8)
palette2=c("gold")
palette3=colorRampPalette(c("coral","deeppink3","hotpink3","lightpink"))(4)
palette4=c("orangered")
palette5=c("skyblue","paleturquoise4")
palette6=c("red")
palette=c(palette1,palette2,palette3,palette4,palette5,palette6)

test$col=palette
test=test[order(test$location),]
palette=test$col
```

#Do a visual haplotype
plot(net, size=3.5*attr(net, "freq"), scale.ratio = 0.5, cex = 0.9 ,
    labels=TRUE, pie = dc, font=20, asp = 1.5, fast=F,show.mutation=1.5, legend=T, bg=palette)

Annex 6:

Sequences used to make the haplotype network.

>Areia_Larga,_Madalena,_Pico_Azores_(MY6302) Alignment of 2 sequences: MY6302F-MyB429F_2.ab1,
MY6302R-MyB1123R_2.ab1 (reversed)
AATTTTAGTGAC-
TGTCGTTTTACTATTAATTTCATTACCTGTTTTGGCCGCTGCAATCACTATGTTGCTCACCGATCGAA
ATTTTAATACTAGTTTCTTTGACCCAGTAGGGGGAGGGGACCCAATCTTGTATCAGCACCTTTTTTG
GTTCTTTGGACACCCTGAAGTTTATATTTTAATTTTACCGGGCTTCGGTATCGTCTCTCATGTATTGT
CATCGCACACAGCTAAGCCCGCTTTTGGCAGGTTAGGTATAATTTATGCTATGGTATCTATTGGACT
ATTAGGATTTATTGTTTGAGCTCACCACATATTTACAGTAGGCATGGATGTAGACACTCGAGCCTAT
TTCACCGCTGCCACCATAATTATTGCGGTTCCTACTGGAATTAAAGTATTTAGGTGACTCATAACAC
TTTATGGTAGTCG-GGCACATTACA-
GCGCGGCGATGTATTGAGTTCTCGGATTCATCTTCCTCTTTACCTTGGGAGGATTA-ACAGGAAT
>Azores_(GenBank) Myosotella myosotis mitochondrion, complete genome
AATTTTAGTAAC-
TGTAATTTTACTTTTAATCTCTCTTCCTGTGCTTGCTGCTGCTATTACTATGTTATTAACAGATCGAA
ACTTTAATACAAGTTTCTTTGACCCGGTGGGTGGGGGTGATCCAATTTTATATCAACATCTCTTTTG
ATTCTTTGGACACCCGGAAGTGTATATTTTAATTCTACCGGGATTTGGTATTGTTTCACATATTTTAT
CTTCTCACACAGCTAAGCCAGCTTTCGGAAGATTAGGGATAATTTATGCAATGGTTTCCATTGGGCT
ATTAGGGTTTATTGTGTGGGCTCACCACATATTTACGGTTGGTATGGATGTAGACACTCGTGCCTAC
TTCACTGCTGCTACCATAATTATTGCTGTTCCTACTGGAATTAAGGTGTTCAGGTGACTTATAACTC
TTTATGGCAGACG-AGCACATTACA-
GTGCAGCGATGTATTGGGTTCTTGGGTTCATTTTTCTTTTTACATTAGGAGGATTG-ACAGGAAT
>Cap_d'Ail_France_(MY6286) Alignment of 2 sequences: MY6286F-MyB429F_3.ab1, MY6286R-
MyB1123R_3.ab1 (reversed)
AATTTTAGTGAC-
TGTCGTTTTACTATTAATTTCATTACCTGTTTTGGCCGCTGCAATCACTATGTTGCTCACCGATCGAA
ATTTTAATACTAGTTTCTTTGACCCAGTAGGGGGAGGGGACCCAATCTTGTATCAGCACCTTTTTTG
GTTCTTTGGACACCCTGAAGTTTATATTTTAATTTTACCGGGCTTCGGTATCGTCTCTCATGTATTGT
CATCGCACACAGCTAAGCCCGCTTTTGGCAGGTTAGGTATAATTTATGCTATGGTATCTATTGGACT
ATTAGGATTTATTGTTTGAGCTCACCACATATTTACAGTAGGCATGGATGTAGACACTCGAGCCTAT
TTCACCGCTGCCACCATAATTATTGCGGTTCCTACTGGAATTAAAGTATTTAGGTGACTCATAACAC
TTTATGGTAGTCG-GGCACATTACA-
GCGCGGCGATGTATTGAGTTCTCGGATTCATCTTCCTCTTTACCTTGGGAGGATTA-ACAGGAAT
>Cap_d'Ail_France_(MY6287) Alignment of 2 sequences: MY6287F-MyB429F_1.ab1, MY6287R-
MyB1123R_1.ab1 (reversed)
GATTTTAGTTAC-
AGTTGTGTTACTTTTAATCTCTCTTCCTGTTCTGGCTGCTGCCATTACGATGCTGCTAACCGATCGAA
ATTTTAATACTAGTTTCTTTGATCCGGTGGGCGGAGGAGACCCGATCTTGTATCAACACCTTTTTTG
GTTTTTTGGGCACCCGGAGGTTTATATTTTAATCCTCCCGGGCTTCGGAATTGTTTCTCATGTGCTGT

CTTCACACACAGCTAAACCAGCTTTTGGGACCTTGGGGATAATTTATGCTATAGTATCTATTGGACT
ATTAGGATTTATTGTATGGGCCCACCACATATTTACGGTTGGGATAGACGTAGATACTCGGGCATA
CTTCACTGCTGCCACTATAATTATTGCGGTACCCACTGGAATTAAAGTATTTAGGTGGCTTATAACA
CTATACGGTAGACG-AGCTCACTATA-
GTGCAGCGATGTATTGAGTTCTTGGGTTCATCTTTCTATTTACTTTGGGGGGGCTT-ACAGGAAT
>E_of_Harwich_UK_(MY6284) Alignment of 2 sequences: MY6284R-MyB429F_6_R.ab1, MY6284R-
MyB1123R_6.ab1 (reversed)
AATTTTAGTCAC-
TGTCGTTTTACTATTAATTTCCTTACCTGTTTTGGCTGCTGCAATTACTATGTTGCTCACCGATCGAA
ATTTTAATACTAGTTTCTTTGACCCAGTGGGGGGAGGGGACCCAATCTTGTATCAGCACCTTTTTTG
GTTCTTTGGACACCCTGAAGTTTATATTTTAATTTTACCGGGCTTCGGTATCGTCTCTCATGTATTGT
CATCGCATACAGCTAAGCCCGCTTTTGGTAGGTTAGGTATAATTTATGCTATGGTATCTATTGGACT
ATTAGGATTTATTGTTTGAGCTCACCACATATTTACAGTAGGCATGGATGTAGACACTCGAGCCTAT
TTCACCGCTGCCACCATAATTATTGCGGTTCCTACTGGAATTAAAGTATTTAGGTGACTCATAACAC
TTTATGGTAGTCG-GGCACACTACA-
GCGCGGCGATGTATTGAGTTCTCGGATTCATCTTCCTCTTTACCTTGGGAGGATTA-ACAGGAAT
>Gijon_Spain_(MY6282) Alignment of 2 sequences: MY6282F-MyB429F_3.ab1, MY6282R-
MyB1123R_3.ab1 (reversed)
AATTTTAGTCAC-
TGTCGTTTTACTATTAATTTCCTTACCTGTTTTGGCTGCTGCAATTACTATGTTGCTCACCGATCGAA
ATTTTAATACTAGTTTCTTTGACCCAGTGGGGGGAGGGGACCCAATCTTGTATCAGCACCTTTTTTG
GTTCTTTGGACACCCTGAAGTTTATATTTTAATTTTACCGGGCTTCGGTATCGTCTCTCATGTATTGT
CATCGCACACAGCTAAGCCCGCTTTTGGTAGGTTAGGTATAATTTATGCTATGGTATCTATTGGACT
ATTAGGATTTATTGTTTGAGCTCACCACATATTTACAGTAGGCATGGATGTAGACACTCGAGCCTAT
TTCACCGCTGCCACCATAATTATTGCGGTTCCTACTGGAATTAAAGTATTTAGGTGACTCATAACAC
TTTATGGTAGTCG-GGCACACTACA-
GCGCGGCGATGTATTGAGTTCTCGGATTCATTTTCCTCTTTACCTTGGGAGGATTA-ACAGGAAT
>Gijon_Spain_(MY6283R)_(reversed)
AATTTTAGTCAC-
TGTCGTTTTACTATTAATTTCCTTACCTGTTTTGGCTGCTGCAATTACTATGTTGCTCACCGATCGAA
ATTTTAATACTAGTTTCTTTGACCCAGTGGGGGGAGGGGACCCAATCTTGTATCAGCACCTTTTTTG
GTTCTTTGGACACCCTGAAGTTTATATTTTAATTTTACCGGGCTTCGGTATCGTCTCTCATGTATTGT
CATCGCACACAGCTAAGCCCGCTTTTGGTAGGTTAGGTATAATTTATGCTATGGTATCTATTGGACT
ATTAGGATTTATTGTTTGAGCTCACCACATATTTACAGTAGGCATGGATGTAGACACTCGAGCCTAT
TTCACCGCTGCCACCATAATTATTGCGGTTCCTACTGGAATTAAAGTATTTAGGTGACTCATAACAC
TTTATGGTAGTCG-GGCACACTACA-
GCGCGGCGATGTATTGAGTTCTCGGATTCATTTTCCTCTTTACCTTGGGAGGATTACACAGGA--
>Horta,_Faial__Azores_(MY6300F)
-ATTTTAGTAAC-
TGTAATTTTACTTTTAATCTCTCTTCCTGTGCTTGCTGCTGCTATTACTATGTTATTAACAGATCGAA
ACTTTAATACAAGTTTCTTTGACCCGGTGGGTGGGGGTGATCCAATTTTATATCAACATCTCTTTTG
ATTCTTTGGACACCCGGAAGTGTATATTTTAATTCTACCGGGATTTGGTATTGTTTCACATATTTTAT
CTTCTCACACAGCTAAGCCAGCTTTCGGAAGATTAGGGATAATTTATGCAATGGTTTCCATTGGGCT
ATTAGGGTTTATTGTGTGGGCTCACCACATATTTACGGTTGGTATGGATGTAGACACTCGTGCCTAC
TTCACTGCTGCTACCATAATTATTGCTGTTCCTACTGGAATTAAGGTGTTCAGGTGACTTATAACTC

TTTATGGCAGACG-AGCACATTACA-
GTGCAGCGATGTATTGGGTTCTTGGGTTCATTTTTCTTTTTACATTAGGAGGATTG-ACAGGAAT

>Horta,_Faial__Azores_(MY6301F)
AATTTTAGTAAC-
TGTAATTTTACTTTTAATCTCTCTTCCTGTGCTTGCTGCTGCTATTACTATGTTATTAACAGATCGAA
ACTTTAATACAAGTTTCTTTGACCCGGTGGGTGGGGGTGATCCAATTTTATATCAACATCTCTTTTG
ATTCTTTGGACACCCGGAAGTGTATATTTTAATTCTACCGGGATTTGGTATTGTTTCACATATTTTAT
CTTCTCACACAGCTAAGCCAGCTTTCGGAAGATTAGGGATAATTTATGCAATGGTTTCCATTGGGCT
ATTAGGGTTTATTGTGTGGGCTCACCACATATTTACGGTTGGTATGGATGTAGACACTCGTGCCTAC
TTCACTGCTGCTACCATAATTATTGCTGTTCCTACTGGAATTAAGGTGTTCAGGTGACTTATAACTC
TTTATGGCAGACG-AGCACATTACA-
GTGCAGCGATGTATTGGGTTCTTGGGTTCATTTTTCTTTTTACATTAGGAGGATTG-ACAGGAAT

>Kerbrat,_Brittany_France_(MY6002)_(reversed) Alignment of 2 sequences: MY6002F_EF72218663.ab1,
MY6002R_EF72218664.ab1 (reversed)
AATTTTAGTGAC-
TGTCGTTTTACTATTAATTTCATTACCTGTTTTGGCCGCTGCAATCACTATGTTGCTCACCGATCGAA
ATTTTAATACTAGTTTCTTTGACCCAGTAGGGGGAGGGGACCCAATCTTGTATCAGCACCTTTTTTG
GTTCTTTGGACACCCTGAAGTTTATATTTTAATTTTACCGGGCTTCGGTATCGTCTCTCATGTATTGT
CATCGCACACAGCTAAGCCCGCTTTTGGCAGGTTAGGTATAATTTATGCTATGGTATCTATTGGACT
ATTAGGATTTATTGTTTGAGCTCACCACATATTTACAGTAGGCATGGATGTAGACACTCGAGCCTAT
TTCACCGCTGCCACCATAATTATTGCGGTTCCTACTGGAATTAAAGTATTTAGGTGACTCATAACAC
TTTATGGTAGTCG-GGCACATTACA-
GCGCGGCGATGTATTGAGTTCTCGGATTCATCTTCCTCTTTACCTTGGGAGGATTA-ACAGGAAT

>Kerbrat,_Brittany_France_(MY6003)_(reversed) Alignment of 2 sequences: MY6003F_EF72218665.ab1,
MY6003R_EF72218666.ab1 (reversed)
AATTTTAGTGAC-
TGTCATTTTACTATTAATTTCATTACCTGTTTTGGCCGCTGCAATCACTATGTTGCTCACCGATCGAA
ATTTTAATACTAGTTTCTTTGACCCAGTAGGGGGAGGGGACCCAATCTTGTATCAGCACCTTTTTTG
GTTCTTTGGACACCCTGAAGTTTATATTTTAATTTTACCGGGCTTCGGTATCGTCTCTCATGTATTGT
CATCGCACACAGCTAAGCCCGCTTTTGGCAGGTTAGGTATAATTTATGCTATGGTATCTATTGGACT
ATTAGGATTTATTGTTTGAGCTCACCACATATTTACAGTAGGCATGGATGTAGACACTCGAGCCTAT
TTCACCGCTGCCACCATAATTATTGCGGTTCCTACTGGAATTAAAGTATTTAGGTGACTCATAACAC
TTTATGGTAGTCG-GGCACATTACA-
GCGCGGCGATGTATTGAGTTCTCGGATTCATCTTCCTCTTTACCTTGGGAGGATTA-ACAGGAAT

>Kerbrat,_Brittany_France_(MY6004)_(reversed) Alignment of 2 sequences: MY6004F_EF72218667.ab1,
MY6004R_EF72218668.ab1 (reversed)
AATTTTAGTGAC-
TGTCATTTTACTATTAATTTCATTACCTGTTTTGGCCGCTGCAATCACTATGTTGCTCACCGATCGAA
ATTTTAATACTAGTTTCTTTGACCCAGTAGGGGGAGGGGACCCAATCTTGTATCAGCACCTTTTTTG
GTTCTTTGGACACCCTGAAGTTTATATTTTAATTTTACCGGGCTTCGGTATCGTCTCTCATGTATTGT
CATCGCACACAGCTAAGCCCGCTTTTGGCAGGTTAGGTATAATTTATGCTATGGTATCTATTGGACT
ATTAGGATTTATTGTTTGAGCTCACCACATATTTACAGTAGGCATGGATGTAGACACTCGAGCCTAT
TTCACCGCTGCCACCATAATTATTGCGGTTCCTACTGGAATTAAAGTATTTAGGTGACTCATAACAC
TTTATGGTAGTCG-GGCACATTACA-
GCGCGGCGATGTATTGAGTTCTCGGATTCATCTTCCTCTTTACCTTGGGAGGATTA-ACAGGAAT

>Kerbrat,_Brittany_France_(MY6005)_(reversed) Alignment of 2 sequences: MY6005F_EF72218645.ab1, MY6005R_EF72218646.ab1 (reversed)

AATTTTAGTGAC-
TGTCATTTTACTATTAATTTCATTACCTGTTTTGGCCGCTGCAATCACTATGTTGCTCACCGATCGAA
ATTTTAATACTAGTTTCTTTGACCCAGTAGGGGGAGGGGACCCAATCTTGTATCAGCACCTTTTTTG
GTTCTTTGGACACCCTGAAGTTTATATTTTAATTTTACCGGGCTTCGGTATCGTCTCTCATGTATTGT
CATCGCACACAGCTAAGCCCGCTTTTGGCAGGTTAGGTATAATTTATGCTATGGTATCTATTGGACT
ATTAGGATTTATTGTTTGAGCTCACCACATATTTACAGTAGGCATGGATGTAGACACTCGAGCCTAT
TTCACCGCTGCCACCATAATTATTGCGGTTCCTACTGGAATTAAAGTATTTAGGTGACTCATAACAC
TTTATGGTAGTCG-GGCACATTACA-
GCGCGGCGATGTATTGAGTTCTCGGATTCATCTTCCTCTTTACCTTGGGAGGATTA-ACAGGAAT
>Kerbrat,_Brittany_France_(MY6007) Alignment of 2 sequences: MY6007F-MyB429F_1.ab1, MY6007R-MyB1123R_1.ab1 (reversed)

AATTTTAGTGAC-
TGTCATTTTACTATTAATTTCATTACCTGTTTTGGCCGCTGCAATCACTATGTTGCTCACCGATCGAA
ATTTTAATACTAGTTTCTTTGACCCAGTAGGGGGAGGGGACCCAATCTTGTATCAGCACCTTTTTTG
GTTCTTTGGACACCCTGAAGTTTATATTTTAATTTTACCGGGCTTCGGTATCGTCTCTCATGTATTGT
CATCGCACACAGCTAAGCCCGCTTTTGGCAGGTTAGGTATAATTTATGCTATGGTATCTATTGGACT
ATTAGGATTTATTGTTTGAGCTCACCACATATTTACAGTAGGCATGGATGTAGACACTCGAGCCTAT
TTCACCGCTGCCACCATAATTATTGCGGTTCCTACTGGAATTAAAGTATTTAGGTGACTCATAACAC
TTTATGGTAGTCG-GGCACATTACA-
GCGCGGCGATGTATTGAGTTCTCGGATTCATCTTCCTCTTTACCTTGGGAGGATTA-ACAGGAAT
>Kerbrat,_Brittany_France_(MY6008) Alignment of 2 sequences: MY6008F-MyB429F_1.ab1, MY6008R-MyB1123R_1.ab1 (reversed)

AATTTTAGTGAC-
TGTCGTTTTACTATTAATTTCATTACCTGTTTTGGCCGCTGCAATCACTATGTTGCTCACCGATCGAA
ATTTTAATACTAGTTTCTTTGACCCAGTAGGGGGAGGGGACCCAATCTTGTATCAGCACCTTTTTTG
GTTCTTTGGACACCCTGAAGTTTATATTTTAATTTTACCGGGCTTCGGTATCGTCTCTCATGTATTGT
CATCGCACACAGCTAAGCCCGCTTTTGGCAGGTTAGGTATAATTTATGCTATGGTATCTATTGGACT
ATTAGGATTTATTGTTTGAGCTCACCACATATTTACAGTAGGCATGGATGTAGACACTCGAGCCTAT
TTCACCGCTGCCACCATAATTATTGCGGTTCCTACTGGAATTAAAGTATTTAGGTGACTCATAACAC
TTTATGGTAGTCG-GGCACATTACA-
GCGCGGCGATGTATTGAGTTCTCGGATTCATCTTCCTCTTTACCTTGGGAGGATTA-ACAGGAAT
>Kerbrat,_Brittany_France_(MY6009F)

AATTTTAGTGAC-
TGTCGTTTTACTATTAATTTCATTACCTGTTTTGGCCGCTGCAATCACTATGTTGCTCACCGATCGAA
ATTTTAATACTAGTTTCTTTGACCCAGTAGGGGGAGGGGACCCAATCTTGTATCAGCACCTTTTTTG
GTTCTTTGGACACCCTGAAGTTTATATTTTAATTTTACCGGGCTTCGGTATCGTCTCTCATGTATTGT
CATCGCACACAGCTAAGCCCGCTTTTGGCAGGTTAGGTATAATTTATGCTATGGTATCTATTGGACT
ATTAGGATTTATTGTTTGAGCTCACCACATATTTACAGTAGGCATGGATGTAGACACTCGAGCCTAT
TTCACCGCTGCCACCATAATTATTGCGGTTCCTACTGGAATTAAAGTATTTAGGTGACTCATAACAC
TTTATGGTAGTCG-GGCACATTACA-
GCGCGGCGATGTATTGAGTTCTCGGATTCATCTTCCTCTTTACCTTGGGAGGATTA-ACAGGAAT
>Kerbrat,_Brittany_France_(MY6010F)

AATTTTAGTGAC-
TGTCGTTTTACTATTAATTTCATTACCTGTTTTGGCCGCTGCAATCACTATGTTGCTCACCGATCGAA

ATTTTAATACTAGTTTCTTTGACCCAGTAGGGGGAGGGGACCCAATCTTGTATCAGCACCTTTTTTG
GTTCTTTGGACACCCTGAAGTTTATATTTTAATTTTACCGGGCTTCGGTATCGTCTCTCATGTATTGT
CATCGCACACAGCTAAGCCCGCTTTTGGCAGGTTAGGTATAATTTATGCTATGGTATCTATTGGACT
ATTAGGATTTATTGTTTGAGCTCACCACATATTTACAGTAGGCATGGATGTAGACACTCGAGCCTAT
TTCACCGCTGCCACCATAATTATTGCGGTTCCTACTGGAATTAAAGTATTTAGGTGACTCATAACAC
TTTATGGTAGTCG-GGCACATTACA-
GCGCGGCGATGTATTGAGTTCTCGGATTCATCTTCCTCTTTACCTTGGGAGGATTA-ACAGGAAT
>L'horn_France_(MY6029) Alignment of 2 sequences: MY6029F-MyB429F_5.ab1, MY6029R-
MyB1123R_5.ab1 (reversed)
AATTTTAGTGAC-
TGTCGTTTTACTATTAATTTCATTACCTGTTTTGGCCGCTGCAATCACTATGTTGCTCACCGATCGAA
ATTTTAATACTAGTTTCTTTGACCCAGTAGGGGGAGGGGACCCAATCTTGTATCAGCACCTTTTTTG
GTTCTTTGGACACCCTGAAGTTTATATTTTAATTTTACCGGGCTTCGGTATCGTCTCTCATGTATTGT
CATCGCACACAGCTAAGCCCGCTTTTGGCAGGTTAGGTATAATTTATGCTATGGTATCTATTGGACT
ATTAGGATTTATTGTTTGAGCTCACCACATATTTACAGTAGGCATGGATGTAGACACTCGAGCCTAT
TTCACCGCTGCCACCATAATTATTGCGGTTCCTACTGGAATTAAAGTATTTAGGTGACTCATAACAC
TTTATGGTAGTCG-GGCACATTACA-
GCGCGGCGATGTATTGAGTTCTCGGATTCATCTTCCTCTTTACCTTGGGAGGATTA-ACAGGAAT
>L'horn_France_(MY6030) Alignment of 2 sequences: My6030_3F-MyB429F.ab1, My6030_3R-
MyB1123R.ab1 (reversed)
AATTTTAGTGAC-
TGTCGTTTTACTATTAATTTCATTACCTGTTTTGGCCGCTGCAATCACTATGTTGCTCACCGATCGAA
ATTTTAATACTAGTTTCTTTGACCCAGTAGGGGGAGGGGACCCAATCTTGTATCAGCACCTTTTTTG
GTTCTTTGGACACCCTGAAGTTTATATTTTAATTTTACCGGGCTTCGGTATCGTCTCTCATGTATTGT
CATCGCACACAGCTAAGCCCGCTTTTGGCAGGTTAGGTATAATTTATGCTATGGTATCTATTGGACT
ATTAGGATTTATTGTTTGAGCTCACCACATATTTACAGTAGGCATGGATGTAGACACTCGAGCCTAT
TTCACCGCTGCCACCATAATTATTGCGGTTCCTACTGGAATTAAAGTATTTAGGTGACTCATAACAC
TTTATGGTAGTCG-GGCACATTACA-
GCGCGGCGATGTATTGAGTTCTCGGATTCATCTTCCTCTTTACCTTGGGAGGATTA-ACAGGAAT
>L'horn_France_(MY6031) Alignment of 2 sequences: MY6031F-MYB429F1.ab1, MY6031R-
MYB1123R1.ab1 (reversed)
AATTTTAGTGAC-
TGTCGTTTTACTATTAATTTCATTACCTGTTTTGGCCGCTGCAATCACTATGTTGCTCACCGATCGAA
ATTTTAATACTAGTTTCTTTGACCCAGTAGGGGGAGGGGACCCAATCTTGTATCAGCACCTTTTTTG
GTTCTTTGGACACCCTGAAGTTTATATTTTAATTTTACCGGGCTTCGGTATCGTCTCTCATGTATTGT
CATCGCACACAGCTAAGCCCGCTTTTGGCAGGTTAGGTATAATTTATGCTATGGTATCTATTGGACT
ATTAGGATTTATTGTTTGAGCTCACCACATATTTACAGTAGGCATGGATGTAGACACTCGAGCCTAT
TTCACCGCTGCCACCATAATTATTGCGGTTCCTACTGGAATTAAAGTATTTAGGTGACTCATAACAC
TTTATGGTAGTCG-GGCACATTACA-
GCGCGGCGATGTATTGAGTTCTCGGATTCATCTTCCTCTTTACCTTGGGAGGATTA-ACAGGAAT
>L'horn_France_(MY6032) Alignment of 2 sequences: MY6032F_EF72218706.ab1,
MY6032R_EF72218705.ab1 (reversed)
AATTTTAGTGAC-
TGTCGTTTTACTATTAATTTCATTACCTGTTTTGGCCGCTGCAATCACTATGTTGCTCACCGATCGAA
ATTTTAATACTAGTTTCTTTGACCCAGTAGGGGGAGGGGACCCAATCTTGTATCAGCACCTTTTTTG
GTTCTTTGGACACCCTGAAGTTTATATTTTAATTTTACCGGGCTTCGGTATCGTCTCTCATGTATTGT

CATCGCACACAGCTAAGCCCGCTTTTGGCAGGTTAGGTATAATTTATGCTATGGTATCTATTGGACT
ATTAGGATTTATTGTTTGAGCTCACCACATATTTACAGTAGGCATGGATGTAGACACTCGAGCCTAT
TTCACCGCTGCCACCATAATTATTGCGGTTCCTACTGGAATTAAAGTATTTAGGTGACTCATAACAC
TTTATGGTAGTCG-GGCACATTACA-
GCGCGGCGATGTATTGAGTTCTCGGATTCATCTTCCTCTTTACCTTGGGAGGATTA-ACAGGAAT
>L'île_Grand,_Brittany_France_(MY6017) Alignment of 2 sequences: MY6017F-MyB429F_5.ab1, MY6017R-MyB1123R_5.ab1 (reversed)
AATTTTAGTGAC-
TGTCGTTTTACTATTAATTTCATTACCTGTTTTGGCTGCTGCAATCACTATGTTGCTCACCGATCGAA
ATTTTAATACTAGTTTCTTTGACCCAGTAGGGGGAGGGGACCCAATCTTGTATCAGCACCTTTTTTG
GTTCTTTGGACACCCTGAAGTTTATATTTTAATTTTACCGGGCTTCGGTATCGTCTCTCATGTATTGT
CATCGCACACAGCTAAGCCCGCTTTTGGCAGGTTAGGTATAATTTATGCTATGGTATCTATTGGACT
ATTAGGATTTATTGTTTGAGCTCACCACATATTTACAGTAGGCATGGATGTAGACACTCGAGCATAT
TTCACCGCTGCCACCATAATTATTGCGGTTCCTACAGGAATTAAAGTATTTAGGTGACTCATAACAC
TTTATGGTAGTCG-GGCACACTACA-
GCGCGGCGATGTATTGAGTTCTCGGATTCATCTTCCTCTTTACCTTGGGAGGATTA-ACAGGAAT
>L'île_Grand,_Brittany_France_(MY6020) Alignment of 2 sequences: MY6020F-MyB429F_5.ab1, MY6020R-MyB1123R_5.ab1 (reversed)
AATTTTAGTGAC-
TGTCGTTTTACTATTAATTTCATTACCTGTTTTGGCTGCTGCAATCACTATGTTGCTCACCGATCGAA
ATTTTAATACTAGTTTCTTTGACCCAGTAGGGGGAGGGGACCCAATCTTGTATCAGCACCTTTTTTG
GTTCTTTGGACACCCTGAAGTTTATATTTTAATTTTACCGGGCTTCGGTATCGTCTCTCATGTATTGT
CATCGCACACAGCTAAGCCCGCTTTTGGCAGGTTAGGTATAATTTATGCTATGGTATCTATTGGACT
ATTAGGATTTATTGTTTGAGCTCACCACATATTTACAGTAGGCATGGATGTAGACACTCGAGCCTAT
TTCACCGCTGCCACCATAATTATTGCGGTTCCTACTGGAATTAAAGTATTTAGGTGACTCATAACAC
TTTATGGTAGTCG-GGCACACTACA-
GTGCGGCGATGTATTGAGTTCTTGGATTCATCTTCCTCTTTACCTTGGGAGGATTA-ACAGGAAT
>Pender_Island_Canada_(MY6200) Alignment of 2 sequences: MY6200F-MyB429F_6.ab1, MY6200R-MyB1123R_6.ab1 (reversed)
AATTTTAGTAAC-
TGTAATTTTACTTTTAATCTCTCTTCCTGTGCTTGCTGCTGCTATTACTATGTTATTAACAGATCGAA
ACTTTAATACAAGTTTCTTTGACCCGGTGGGTGGGGGTGATCCAATTTTATATCAACATCTCTTTTG
ATTCTTTGGACACCCGGAAGTGTATATTTTAATTCTACCGGGATTTGGTATTGTTTCACATATTTTAT
CTTCTCACACAGCTAAGCCAGCTTTCGGAAGATTAGGGATAATTTATGCAATGGTTTCCATTGGGCT
ATTAGGGTTTATTGTGTGGGCTCACCACATATTTACGGTTGGTATGGATGTAGACACTCGTGCCTAC
TTCACTGCTGCTACCATAATTATTGCTGTTCCTACTGGAATTAAGGTGTTCAGGTGACTTATAACTC
TTTATGGCAGACG-AGCACATTACA-
GTGCAGCGATGTATTGGGTTCTTGGGTTCATTTTTCTTTTTACATTAGGAGGATTG-ACAGGAAT
>Pender_Island_Canada_(MY6202) Alignment of 2 sequences: MY6202F-MyB429F_6.ab1, MY6202R-MyB1123R_6.ab1 (reversed)
AATTTTAGTAAC-
TGTAATTTTACTTTTAATCTCTCTTCCTGTGCTTGCTGCTGCTATTACTATGTTATTAACAGATCGAA
ACTTTAATACAAGTTTCTTTGACCCGGTGGGTGGGGGTGATCCAATTTTATATCAACATCTCTTTTG
ATTCTTTGGACACCCGGAAGTGTATATTTTAATTCTACCGGGATTTGGTATTGTTTCACATATTTTAT
CTTCTCACACAGCTAAGCCAGCTTTCGGAAGATTAGGGATAATTTATGCAATGGTTTCCATTGGGCT
ATTAGGGTTTATTGTGTGGGCTCACCACATATTTACGGTTGGTATGGATGTAGACACTCGTGCCTAC

TTCACTGCTGCTACCATAATTATTGCTGTTCCTACTGGAATTAAGGTGTTCAGGTGACTTATAACTC
TTTATGGCAGACG-AGCACATTACA-
GTGCAGCGATGTATTGGGTTCTTGGGTTCATTTTTCTTTTTACATTAGGAGGATTG-ACAGGAAT
>Pender_Island_Canada_(MY6206) Alignment of 2 sequences: MY6206F-MYB429F1.ab1, MY6206R-
MYB1123R1.ab1 (reversed)
AATTTTAGTAAC-
TGTAATTTTACTTTTAATCTCTCTTCCTGTGCTTGCTGCTGCTATTACTATGTTATTAACAGATCGAA
ACTTTAATACAAGTTTCTTTGACCCGGTGGGTGGGGGTGATCCAATTTTATATCAACATCTCTTTTG
ATTCTTTGGACACCCGGAAGTGTATATTTTAATTCTACCGGGATTTGGTATTGTTTCACATATTTTAT
CTTCTCACACAGCTAAGCCAGCTTTCGGAAGATTAGGGATAATTTATGCAATGGTTTCCATTGGGCT
ATTAGGGTTTATTGTGTGGGCTCACCACATATTTACGGTTGGTATGGATGTAGACACTCGTGCCTAC
TTCACTGCTGCTACCATAATTATTGCTGTTCCTACTGGAATTAAGGTGTTCAGGTGACTTATAACTC
TTTATGGCAGACG-AGCACATTACA-
GTGCAGCGATGTATTGGGTTCTTGGGTTCATTTTTCTTTTTACATTAGGAGGATTG-ACAGGAAT
>Pender_Island_Canada_(MY6207) Alignment of 2 sequences: MY6207F-MyB429F_5.ab1, MY6207R-
MyB1123R_5.ab1 (reversed)
AATTTTAGTAAC-
TGTAATTTTACTTTTAATCTCTCTTCCTGTGCTTGCTGCTGCTATTACTATGTTATTAACAGATCGAA
ACTTTAATACAAGTTTCTTTGACCCGGTGGGTGGGGGTGATCCAATTTTATATCAACATCTCTTTTG
ATTCTTTGGACACCCGGAAGTGTATATTTTAATTCTACCGGGATTTGGTATTGTTTCACATATTTTAT
CTTCTCACACAGCTAAGCCAGCTTTCGGAAGATTAGGGATAATTTATGCAATGGTTTCCATTGGGCT
ATTAGGGTTTATTGTGTGGGCTCACCACATATTTACGGTTGGTATGGATGTAGACACTCGTGCCTAC
TTCACTGCTGCTACCATAATTATTGCTGTTCCTACTGGAATTAAGGTGTTCAGGTGACTTATAACTC
TTTATGGCAGACG-AGCACATTACA-
GTGCAGCGATGTATTGGGTTCTTGGGTTCATTTTTCTTTTTACATTAGGAGGATTG-ACAGGAAT
>Pender_Island_Canada_(MY6208F)
AATTTTAGTAAC-
TGTAATTTTACTTTTAATCTCTCTTCCTGTGCTTGCTGCTGCTATTACTATGTTATTAACAGATCGAA
ACTTTAATACAAGTTTCTTTGACCCGGTGGGTGGGGGTGATCCAATTTTATATCAACATCTCTTTTG
ATTCTTTGGACACCCGGAAGTGTATATTTTAATTCTACCGGGATTTGGTATTGTTTCACATATTTTAT
CTTCTCACACAGCTAAGCCAGCTTTCGGAAGATTAGGGATAATTTATGCAATGGTTTCCATTGGGCT
ATTAGGGTTTATTGTGTGGGCTCACCACATATTTACGGTTGGTATGGATGTAGACACTCGTGCCTAC
TTCACTGCTGCTACCATAATTATTGCTGTTCCTACTGGAATTAAGGTGTTCAGGTGACTTATAACTC
TTTATGGCAGACG-AGCACATTACA-
GTGCAGCGATGTATTGGGTTCTTGGGTTCATTTTTCTTTTTACATTAGGAGGATTG-ACAGGAAT
>Pender_Island_Canada_(MY6209F)
AATTTTAGTAAC-
TGTAATTTTACTTTTAATCTCTCTTCCTGTGCTTGCTGCTGCTATTACTATGTTATTAACAGATCGAA
ACTTTAATACAAGTTTCTTTGACCCGGTGGGTGGGGGTGATCCAATTTTATATCAACATCTCTTTTG
ATTCTTTGGACACCCGGAAGTGTATATTTTAATTCTACCGGGATTTGGTATTGTTTCACATATTTTAT
CTTCTCACACAGCTAAGCCAGCTTTCGGAAGATTAGGGATAATTTATGCAATGGTTTCCATTGGGCT
ATTAGGGTTTATTGTGTGGGCTCACCACATATTTACGGTTGGTATGGATGTAGACACTCGTGCCTAC
TTCACTGCTGCTACCATAATTATTGCTGTTCCTACTGGAATTAAGGTGTTCAGGTGACTTATAACTC
TTTATGGCAGACG-AGCACATTACA-
GTGCAGCGATGTATTGGGTTCTTGGGTTCATTTTTCTTTTTACATTAGGAGGATTG-ACAGGAAT

>Pender_Island_Canada_(MY6210) Alignment of 2 sequences: MY6210F-MYB429F2.ab1, MY6210R-MYB1123R2.ab1 (reversed)

AATTTTAGTAAC-
TGTAATTTTACTTTTAATCTCTCTTCCTGTGCTTGCTGCTGCTATTACTATGTTATTAACAGATCGAA
ACTTTAATACAAGTTTCTTTGACCCGGTGGGTGGGGGTGATCCAATTTTATATCAACATCTCTTTTG
ATTCTTTGGACACCCGGAAGTGTATATTTTAATTCTACCGGGATTTGGTATTGTTTCACATATTTTAT
CTTCTCACACAGCTAAGCCAGCTTTCGGAAGATTAGGGATAATTTATGCAATGGTTTCCATTGGGCT
ATTAGGGTTTATTGTGTGGGCTCACCACATATTTACGGTTGGTATGGATGTAGACACTCGTGCCTAC
TTCACTGCTGCTACCATAATTATTGCTGTTCCTACTGGAATTAAGGTGTTCAGGTGACTTATAACTC
TTTATGGCAGACG-AGCACATTACA-
GTGCAGCGATGTATTGGGTTCTTGGGTTCATTTTTCTTTTTACATTAGGAGGATTG-ACAGGAAT
>Pialassa_dei_Piomboni_Italy_(MY6245R)_(reversed)

AATTTTAGTGAC-
TGTCGTTTTACTATTAATTTCATTACCTGTTTTGGCCGCTGCAATCACTATGTTGCTCACCGATCGAA
ATTTTAATACTAGTTTCTTTGACCCAGTAGGGGGAGGGGACCCAATCTTGTATCAGCACCTTTTTTG
GTTCTTTGGACACCCTGAAGTTTATATTTTAATTTTACCGGGCTTCGGTATCGTCTCTCATGTATTGT
CATCGCACACAGCTAAGCCCGCTTTTGGCAGGTTAGGTATAATTTATGCTATGGTATCTATTGGACT
ATTAGGATTTATTGTTTGAGCTCACCACATATTTACAGTAGGCATGGATGTAGACACTCGAGCCTAT
TTCACCGCTGCCACCATAATTATTGCGGTTCCTACTGGAATTAAAGTATTTAGGTGACTCATAACAC
TTTATGGTAGTCG-GGCACATTACA-
GCGCGGCGATGTATTGAGTTCTCGGATTCATCTTCCTCTTTACCTTGGGAGGATTA-ACAGGAAT
>Site_1._Atalhada_Azores_(MY6065) Alignment of 2 sequences: MY6065F-MyB429F_2.ab1, MY6065R-MyB1123R_2.ab1 (reversed)

GATTTTAGTTAC-
AGTTGTGTTACTTTTAATCTCTCTTCCTGTTCTGGCTGCTGCCATTACGATGCTGCTAACCGATCGAA
ATTTTAATACTAGTTTCTTTGATCCGGTGGGCGGAGGAGACCCGATCTTGTATCAACACCTTTTTTG
GTTTTTTGGGCACCCGGAGGTTTATATTTTAATCCTCCCGGGCTTCGGAATTGTTTCTCATGTGCTGT
CTTCACACACAGCTAAACCAGCTTTTGGGACCTTGGGGATAATTTATGCTATAGTATCTATTGGACT
ATTAGGATTTATTGTATGGGCCCACCACATATTTACGGTTGGGATAGACGTAGATACTCGGGCATA
CTTCACTGCTGCCACTATAATTATTGCGGTACCCACTGGAATTAAAGTATTTAGGTGGCTTATAACA
CTATACGGTAGACG-AGCTCACTATA-
GTGCAGCGATGTATTGAGTTCTTGGGTTCATCTTTCTATTTACTTTGGGGGGGGCTT-ACAGGAAT
>Site_1._Atalhada_Azores_(MY6080)_(reversed) Alignment of 2 sequences: MY6080F_EF72218627.ab1, MY6080R_EF72218628.ab1 (reversed)

GATTTTAGTTAC-
AGTTGTATTACTTTTAATCTCTCTTCCTGTTCTGGCTGCTGCCATTACGATGCTGCTAACCGATCGAA
ATTTTAATACTAGTTTCTTTGATCCGGTGGGCGGAGGAGACCCGATCTTGTATCAACACCTTTTTTG
GTTTTTTGGGCACCCGGAGGTTTATATTTTAATCCTCCCAGGCTTCGGAATTGTTTCTCATGTGCTGT
CTTCACACACAGCTAAACCAGCTTTTGGGACCTTGGGGATAATTTATGCTATAGTATCTATTGGACT
ATTAGGATTTATTGTATGGGCCCACCACATATTTACGGTTGGGATAGACGTAGATACTCGGGCATA
CTTCACTGCTGCCACTATAATTATTGCGGTACCCACTGGAATTAAAGTATTTAGGTGGCTTATAACA
CTATACGGTAGACG-AGCTCACTATA-
GTGCAGCGATGTATTGAGTTCTTGGGTTCATCTTTCTATTTACTTTGGGGGGGGCTT-ACAGGAAT
>Site_2._Brejela,_Atalhada_Azores_(MY6074) Alignment of 2 sequences: MY6074F_EF72218205.ab1, MY6074R_EF72218206.ab1 (reversed)

GATTTTAGTTAC-
AGTTGTGTTACTTTTAATCTCTCTTCCTGTTCTGGCTGCTGCCATTACGATGCTGCTAACCGATCGAA
ATTTTAATACTAGTTTCTTTGATCCGGTGGGCGGAGGAGACCCGATCTTGTATCAACACCTTTTTTG
GTTTTTTGGGCACCCGGAGGTTTATATTTTAATCCTCCCGGGCTTCGGAATTGTTTCTCATGTGCTGT
CTTCACACACAGCTAAACCAGCTTTTGGGACCTTGGGGATAATTTATGCTATAGTATCTATTGGACT
ATTAGGATTTATTGTATGGGCCCACCACATATTTACGGTTGGGATAGACGTAGATACTCGGGCATA
CTTCACTGCTGCCACTATAATTATTGCGGTACCCACTGGAATTAAAGTATTTAGGTGGCTTATAACA
CTATACGGTAGACG-AGCTCACTATA-
GTGCAGCGATGTATTGAGTTCTTGGGTTCATCTTTCTATTTACTTTGGGGGGGCTT-ACAGGAAT
>Site_2._Brejela,_Atalhada_Azores_(MY6075F)
AATTTTAGTGAC-
TGTAATCTTACTTTTAATCTCTCTTCCTGTGCTTGCTGCTGCTATTACTATGTTATTAACAGATCGAA
ACTTTAATACAAGTTTCTTTGACCCGGTGGGGGGGAGGTGATCCAATCTTATATCAACATCTCTTTTG
ATTCTTTGGACACCCGGAAGTTTATATTTTAATTTTACCGGGCTTTGGTATTGTTTCACATATTTTAT
CTTCTCACACAGCTAAGCCAGCTTTCGGAAGATTAGGGATAATTTATGCAATGGTTTCCATTGGGCT
ATTAGGGTTTATTGTGTGGGCTCACCACATATTTACGGTTGGTATGGATGTAGACACTCGTGCCTAC
TTCACTGCTGCTACCATAATTATTGCTGTTCCTACTGGAATTAAAGTGTTTAGGTGACTTATAACCC
TCTATGGTAGACG-AGCACATTACA-
GTGCAGCGATGTATTGGGTTCTTGGGTTCATTTTCCTCTTTACATTAGGAGGGTTG-ACAGGAAT
>Site_2._Brejela,_Atalhada_Azores_(MY6076) Alignment of 2 sequences: MY6076F-MyB429F_4.ab1,
MY6076R-MyB1123R_4.ab1 (reversed)
GATTTTAGTTAC-
AGTTGTTTTACTTTTAATCTCTCTTCCTGTTCTTGCTGCCGCCATTACGATGCTGCTAACCGATCGAA
ATTTTAATACTAGTTTCTTTGATCCGGTGGGCGGAGGAGACCCGATCTTGTATCAACACCTTTTTTG
GTTTTTTGGGCACCCGGAGGTTTATATTTTAATCCTCCCGGGCTTCGGAATTGTTTCTCATGTGCTGT
CTTCACACACAGCTAAACCAGCTTTTGGGACCTTGGGGATAATTTATGCTATAGTATCTATTGGACT
ATTAGGATTTATTGTATGAGCCCACCACATATTTACGGTTGGGATAGACGTAGATACTCGGGCATA
CTTCACTGCTGCCACTATAATTATTGCGGTACCCACTGGAATTAAAGTATTTAGGTGGCTTATAACA
CTATACGGTAGACG-AGCTCACTATA-
GTGCAGCGATGTATTGAGTTCTTGGGTTCATCTTTCTATTTACTTTGGGGGGGCTT-ACAGGAAT
>Site_2._Brejela,_Atalhada_Azores_(MY6083)_ Alignment of 2 sequences: MY6083F-MYB429F4.ab1,
MY6083R_EF72218210.ab1 (reversed)
GATTTTAGTTAC-
AGTTGTTTTACTTTTAATCTCTCTTCCTGTTCTTGCTGCTGCCATTACGATGCTGCTAACCGATCGAA
ATTTTAATACTAGTTTCTTTGATCCGGTGGGCGGAGGAGACCCGATCTTGTATCAACACCTTTTTTG
GTTTTTTGGGCACCCGGAGGTTTATATTTTGATCCTCCCGGGCTTCGGAATTGTTTCTCATGTGCTGT
CTTCACACACAGCTAAACCAGCTTTTGGGACCTTGGGGATAATTTATGCTATAGTATCTATTGGACT
ATTAGGATTTATTGTATGAGCCCACCACATATTTACGGTTGGGATAGACGTAGATACTCGGGCATA
CTTCACTGCTGCTACTATAATTATTGCGGTACCCACTGGAATTAAAGTATTTAGGTGGCTTATAACA
CTATACGGTAGACG-AGCTCACTATA-
GTGCAGCGATGTATTGAGTTCTTGGGTTCATCTTTCTATTTACTTTGGGGGGGCTT-ACAGGAAT
>Site_2._Brejela,_Atalhada_Azores_(MY6084) Alignment of 2 sequences: MY6084F-MyB429F_5.ab1,
MY6084R-MYB1123R4.ab1 (reversed)
GATTTTAGTTAC-
AGTTGTTTTACTTTTAATCTCTCTTCCTGTTCTTGCTGCCGCCATTACGATGCTGCTAACCGATCGAA
ATTTTAATACTAGTTTCTTTGATCCGGTGGGCGGAGGAGACCCGATCTTGTATCAACACCTTTTTTG

GTTTTTTGGGCACCCGGAGGTTTATATTTTAATCCTCCCGGGCTTCGGAATTGTTTCTCATGTGCTGT
CTTCACACACAGCTAAACCAGCTTTTGGGACCTTGGGGATAATTTATGCTATAGTATCTATTGGACT
ATTAGGATTTATTGTATGAGCCCACCACATATTTACGGTTGGGATAGACGTAGATACTCGGGCATA
CTTCACTGCTGCCACTATAATTATTGCGGTACCCACTGGAATTAAAGTATTTAGGTGGCTTATAACA
CTATACGGTAGACG-AGCTCACTATA-
GTGCAGCGATGTATTGAGTTCTTGGGTTCATCTTTCTATTTACTTTGGGGGGGCTT-ACAGGAAT
>Site_2._Brejela,_Atalhada_Azores_(MY6086F)
GATTTTAGTTAC-
AGTTGTTTTACTTTTAATCTCTCTTCCTGTTCTTGCTGCTGCCATTACGATGCTGCTAACCGATCGAA
ATTTTAATACTAGTTTCTTTGATCCGGTGGGCGGAGGAGATCCGATCTTGTATCAACACCTTTTTTG
GTTTTTTGGGCACCCGGAGGTTTATATTTTAATCCTCCCGGGCTTCGGAATTGTTTCTCATGTGCTGT
CTTCACACACAGCTAAACCAGCTTTTGGGACCTTGGGGATAATTTATGCTATAGTATCTATTGGACT
ATTAGGATTTATTGTATGAGCCCACCACATATTTACGGTTGGGATAGACGTAGATACTCGGGCATA
CTTCACTGCTGCCACTATAATTATTGCGGTACCCACCGGAATTAAAGTATTTAGGTGGCTTATAACA
CTATACGGTAGACG-AGCTCACTATA-
GTGCAGCGATGTATTGAGTTCTTGGGTTCATCTTTCTATTTACTTTGGGGGGGCTT-ACAGGAAT
>Site_2._Brejela,_Atalhada_Azores_(MY6093) Alignment of 2 sequences: My6093_3F-MyB429F.ab1,
My6093_3R-MyB1123R.ab1 (reversed)
GATTTTAGTTAC-
AGTTGTTTTACTTTTAATCTCTCTTCCTGTTCTTGCTGCCGCCATTACGATGCTGCTAACCGATCGAA
ATTTTAATACTAGTTTCTTTGATCCGGTGGGCGGAGGAGACCCGATCTTGTATCAACACCTTTTTTG
GTTTTTTGGGCACCCGGAGGTTTATATTTTAATCCTCCCGGGCTTCGGAATTGTTTCTCATGTGCTGT
CTTCACACACAGCTAAACCAGCTTTTGGGACCTTGGGGATAATTTATGCTATAGTATCTATTGGACT
ATTAGGATTTATTGTATGAGCCCACCACATATTTACGGTTGGGATAGACGTAGATACTCGGGCATA
CTTCACTGCTGCCACTATAATTATTGCGGTACCCACTGGAATTAAAGTATTTAGGTGGCTTATAACA
CTATACGGTAGACG-AGCTCACTATA-
GTGCAGCGATGTATTGAGTTCTTGGGTTCATCTTTCTATTTACTTTGGGGGGGCTT-ACAGGAAT
>Site_2._Brejela,_Atalhada_Azores_(MY6094) Alignment of 2 sequences: MY6094F-MyB429F_5.ab1,
MY6094R-MyB1123R_5.ab1 (reversed)
GATTTTAGTTAC-
AGTTGTGTGTTACTTTTAATCTCTCTTCCTGTTCTGGCTGCTGCCATTACGATGCTGCTAACCGATCGAA
ATTTTAATACTAGTTTCTTTGATCCGGTGGGCGGAGGAGACCCGATCTTGTATCAACACCTTTTTTG
GTTTTTTGGGCACCCGGAGGTTTATATTTTAATCCTCCCGGGCTTCGGAATTGTTTCTCATGTGCTGT
CTTCACACACAGCTAAACCAGCTTTTGGGACCTTGGGGATAATTTATGCTATAGTATCTATTGGACT
ATTAGGATTTATTGTATGGGCCCACCACATATTTACGGTTGGGATAGACGTAGATACTCGGGCATA
CTTCACTGCTGCCACTATAATTATTGCGGTACCCACTGGAATTAAAGTATTTAGGTGGCTTATAACA
CTATACGGTAGACG-AGCTCACTATA-
GTGCAGCGATGTATTGAGTTCTTGGGTTCATCTTTCTATTTACTTTGGGGGGGCTT-ACAGGAAT
>Site_2._Brejela,_Atalhada_Azores_(MY6095) Alignment of 2 sequences: My6095_3F-MyB429F.ab1,
My6095_3R-MyB1123R_R.ab1 (reversed)
GATTTTAGTTACATGTTGTTTTACTTTTAATCTCTCTTCCTGTTCTTGCTGCCGCCATTACGATGCTG
CTAACCGATCGAAATTTTAATACTAGTTTCTTTGATCCGGTGGGCGGAGGAGACCCGATCTTGTATC
AACACCTTTTTTGGTTTTTTGGGCACCCGGAGGTTTATATTTTAATCCTCCCGGGCTTCGGAATTGTT
TCTCATGTGCTGTCTTCACACACAGCTAAACCAGCTTTTGGGACCTTGGGGATAATTTATGCTATAG
TATCTATTGGATTATTAGGATTTATTGTATGAGCCCATCACATATTTACGGTTGGGATAGACGTAGA
TACTCGGGCATACTTCACTGCTGCCACTATAATTATTGCGGTACCCACTGGAATTAAAGTATTTAGG

TGGCTTATAACACTATACGGTAGACG-AGCTCACTATA-
GTGCAGCGATGTATTGAGTTCTTGGGTTCATCTTTCTATTTACTTTGGGGGGGCTT-ACAGGAAT
>Site_2._Brejela,_Atalhada_Azores_(MY6096)_(reversed) Alignment of 2 sequences:
MY6096F_EF72218118.ab1, MY6096R_EF72218117.ab1 (reversed)
AATTTTAGTGAC-
AGTAATCTTACTTTTAATCTCTCTTCCTGTGCTTGCTGCTGCTATTACTATGYTATTAACAGATCGAA
ACTTTAATACAAGTTTCTTTGWCCCGGTGGGGGGGAGGDGATCCAATCTTATATCAACATCTCTTTTG
ATTYTTTGGVCACCCGGARGTTTATATTTTAATYTTMCCGGGCTTBGGTATTGTTTCACATATTTTAT
CTTCTCACACAGCTAAGCCAGCTTTCGGAAGATTAGGGATAATTTATGCAATGGTTTCCATTGGGCT
ATTAGGGTTTATTGTGTGGGCTCACCACATATTTACGGTTGGTATGGATGTAGACACTCGTGCMTA
CTTCACTGCTGCCACCATAATTATTGCKGTTCCYACYGGAATTAAAGTRTTTAGGTGRCTTATAACC
CTMTAYGGTAGACG-AGCWCAYTAYA-
GTGCAGCGATGTATTGRGTTCTTGGGTTCATCTTCCTMTTTACWTTRGGRGGGYTK-ACAGGAAT
>Site_3._Ferraria_Azores_(MY6098) Alignment of 2 sequences: MY6098F-MyB429F_4.ab1, MY6098R-
MYB1123R3.ab1 (reversed)
GATTTTAGTTAC-
AGTTGTTTTACTTTTAATCTCTCTTCCTGTTCTTGCTGCTGCCATTACGATGCTGCTAACCGATCGAA
ATTTTAATACTAGTTTCTTTGATCCGGTGGGCGGAGGAGATCCGATCTTGTATCAACACCTTTTTTG
GTTTTTTGGGCACCCGGAGGTTTATATTTTAATCCTCCCGGGCTTCGGAATTGTTTCTCATGTGCTGT
CTTCACACACAGCTAAACCAGCTTTTGGGACCTTGGGGATAATTTATGCTATAGTATCTATTGGACT
ATTAGGATTTATTGTATGAGCCCACCACATATTTACGGTTGGGATAGACGTAGATACTCGGGCATA
CTTCACTGCTGCCACTATAATTATTGCGGTACCCACCGGAATTAAAGTATTTAGGTGGCTTATAACA
CTATACGGTAGACG-AGCTCACTATA-
GTGCAGCGATGTATTGAGTTCTTGGGTTCATCTTTCTATTTACTTTGGGGGGGCTT-ACAGGAAT
>Site_3._Ferraria_Azores_(MY6099) Alignment of 3 sequences: MY6099F_EF72218203.ab1, MY6099F-
MYB429F3.ab1, MY6099R_EF72218204.ab1 (reversed)
GATTTTAGTTAC-
AGTTGTTTTACTTTTAATCTCTCTTCCTGTTCTTGCTGCTGCCATTACGATGCTGCTAACCGATCGAA
ATTTTAATACTAGTTTCTTTGATCCGGTGGGCGGAGGAGACCCGATCTTGTATCAACACCTTTTTTG
GTTTTTTGGGCACCCGGAGGTTTATATTTTGATCCTCCCGGGCTTCGGAATTGTTTCTCATGTGCTGT
CTTCACACACAGCTAAACCAGCTTTTGGGACCTTGGGGATAATTTATGCTATAGTATCTATTGGACT
ATTAGGATTTATTGTATGAGCCCACCACATATTTACGGTTGGGATAGACGTAGATACTCGGGCATA
CTTCACTGCTGCTACTATAATTATTGCGGTACCCACTGGAATTAAAGTATTTAGGTGGCTTATAACA
CTATACGGTAGACG-AGCTCACTATA-
GTGCAGCGATGTATTGAGTTCTTGGGTTCATCTTTCTATTTACTTTGGGGGGGCTT-ACAGGAAT
>Site_3._Ferraria_Azores_(MY6100)_ Alignment of 2 sequences: MY6100F-MyB429F_1.ab1, MY6100R-
MyB1123R_1_R.ab1 (reversed)
GATTTTAGTTAC-
AGTTGTTTTACTTTTAATCTCTCTTCCTGTTCTTGCTGCTGCCATTACGATGCTGCTAACCGATCGAA
ATTTTAATACTAGTTTCTTTGATCCGGTGGGCGGAGGAGACCCGATCTTGTATCAACACCTTTTTTG
GTTTTTTGGGCACCCGGAGGTTTATATTTTGATCCTCCCGGGCTTCGGAATTGTTTCTCATGTGCTGT
CTTCACACACAGCTAAACCAGCTTTTGGGACCTTGGGGATAATTTATGCTATAGTATCTATTGGACT
ATTAGGATTTATTGTATGAGCCCACCACATATTTACGGTTGGGATAGACGTAGATACTCGGGCATA
CTTCACTGCTGCTACTATAATTATTGCGGTACCCACTGGAATTAAAGTATTTAGGTGGCTTATAACA
CTATACGGTAGACG-AGCTCACTATA-
GTGCAGCGATGTATTGAGTTCTTGGGTTCATCTTTCTATTTACTTTGGGGGGGCTT-ACAGGAAT

>Site_4._Fenais_Azores_(MY6106F)

GATTTTAGTTAC-
AGTTGTTTTACTTTTAATCTCTCTTCCTGTTCTTGCTGCTGCCATTACGATGCTGCTAACCGATCGAA
ATTTTAATACTAGTTTCTTTGATCCGGTGGGCGGAGGAGACCCGATCTTGTATCAACACCTTTTTTG
GTTTTTTGGGCACCCGGAGGTTTATATTTTAATCCTCCCAGGCTTCGGAATTGTTTCTCATGTGCTGT
CTTCACACACAGCTAAACCAGCTTTTGGGACCTTGGGGATAATTTATGCTATAGTATCTATTGGACT
ATTAGGATTTATTGTATGAGCCCACCACATATTTACGGTTGGGATAGACGTAGATACTCGGGCATA
CTTCACTGCTGCCACCATAATTATTGCGGTACCCACTGGAATTAAAGTATTTAGGTGGCTTATAACA
CTATACGGTAGACG-AGCTCACTATA-
GTGCAGCGATGTATTGAGTTCTTGGGTTCATCTTTCTATTTACTTTGGGGGGGCTT-ACAGGAAT

>Spain_(GenBank) Myosotella myosotis mitochondrion, complete genome

AATTTTAGTCAC-
TGTCGTTTTACTATTAATTTCCTTACCTGTTTTGGCTGCTGCAATTACTATGTTGCTCACCGATCGAA
ATTTTAATACTAGTTTTTTTGACCCAGTGGGGGGGAGGGGACCCAATCTTGTATCAGCACCTTTTTTG
GTTCTTTGGACACCCTGAAGTTTATATTTTAATTTTACCGGGCTTCGGTATCGTCTCTCATGTATTGT
CATCGCATACAGCTAAGCCCGCTTTTGGTAGGTTAGGTATAATTTATGCTATGGTATCTATTGGACT
ATTAGGATTTATTGTTTGAGCTCACCACATATTTACAGTAGGCATGGATGTAGACACTCGAGCCTAT
TTCACCGCTGCCACCATAATTATTGCGGTTCCTACTGGAATTAAAGTATTTAGGTGACTCATAACAC
TTTATGGTAGTCG-GGCACACTACA-
GCGCGGCGATGTATTGAGTTCTCGGATTCATTTTCCTCTTTACCTTGGGAGGATTA-ACAGGAAT

>W_of_Ponta_do_Alcaide,_Silveira,_Terceira_Azores_(MY6298) Alignment of 2 sequences: MY6298F-MyB429F_2.ab1, MY6298R-MyB1123R_2.ab1 (reversed)

AATTTTAGTGAC-
TGTCGTTTTACTATTAATTTCATTACCTGTTTTGGCCGCTGCAATCACTATGTTGCTCACCGATCGAA
ATTTTAATACTAGTTTCTTTGACCCAGTAGGGGGGAGGGGACCCAATCTTGTATCAGCACCTTTTTTG
GTTCTTTGGACACCCTGAAGTTTATATTTTAATTTTACCGGGCTTCGGTATCGTCTCTCATGTATTGT
CATCGCACACAGCTAAGCCCGCTTTTGGCAGGTTAGGTATAATTTATGCTATGGTATCTATTGGACT
ATTAGGATTTATTGTTTGAGCTCACCACATATTTACAGTAGGCATGGATGTAGACACTCGAGCCTAT
TTCACCGCTGCCACCATAATTATTGCGGTTCCTACTGGAATTAAAGTATTTAGGTGACTCATAACAC
TTTATGGTAGTCG-GGCACATTACA-
GCGCGGCGATGTATTGAGTTCTCGGATTCATCTTCCTCTTTACCTTGGGAGGATTA-ACAGGAAT

>W_of_Ponta_do_Alcaide,_Silveira,_Terceira_Azores_(MY6299) Alignment of 2 sequences: MY6299F-MyB429F_4_R.ab1, MY6299R-MyB1123R_4.ab1 (reversed)

AATTTTAGTGAC-
TGTAATCTTACTTTTAATCTCTCTTCCTGTGCTTGCTGCTGCTATTACTATGTTATTAACAGATCGAA
ACTTTAATACAAGTTTCTTTGACCCGGTGGGGGGGAGGTGATCCAATCTTATATCAACATCTCTTTTG
ATTCTTTGGACACCCGGAAGTTTATATTTTAATTTTACCGGGCTTTGGTATTGTTTCACATATTTTAT
CTTCTCACACAGCTAAGCCAGCTTTCGGAAGATTAGGGATAATTTATGCAATGGTTTCCATTGGGCT
ATTAGGGTTTATTGTGTGGGCTCACCACATATTTACGGTTGGTATGGATGTAGACACTCGTGCCTAC
TTCACTGCTGCTACCATAATTATTGCTGTTCCTACTGGAATTAAAGTGTTTAGGTGACTTATAACCC
TCTATGGTAGACG-AGCACATTACA-
GTGCAGCGATGTATTGGGTTCTTGGGTTCATTTTCCTCTTTACATTAGGAGGGTTG-ACAGGAAT

Annex 7:

Table listing the sites, the countries and the code of our samples that we used for our R code. The different countries are written with abbreviations: AZ for Azores, CA for Canada, SP for Spain, IT for Italy, FR for France and UK for United Kingdom.

Sites	Countries	Samples
Areia Larga Madalena Pico	AZ	MY6302
GenBank	AZ	GenBank
Cap d'Ail	FR	MY6286
Cap d'Ail	FR	MY6287
E of Harwich	UK	MY6284
Gijon	SP	MY6282
Gijon	SP	MY6283
Horta Faial	AZ	MY6300
Horta Faial	AZ	MY6301
Kerbrat Brittany	FR	MY6002
Kerbrat Brittany	FR	MY6003
Kerbrat Brittany	FR	MY6004
Kerbrat Brittany	FR	MY6005
Kerbrat Brittany	FR	MY6007
Kerbrat Brittany	FR	MY6008
Kerbrat Brittany	FR	MY6009
Kerbrat Brittany	FR	MY6010
L'horn	FR	MY6029
L'horn	FR	MY6030
L'horn	FR	MY6031
L'horn	FR	MY6032
L'île Grand Brittany	FR	MY6017
L'île Grand Brittany	FR	MY6020
Pender Island	CA	MY6200
Pender Island	CA	MY6202
Pender Island	CA	MY6206
Pender Island	CA	MY6207
Pender Island	CA	MY6208
Pender Island	CA	MY6209
Pender Island	CA	MY6210
Pialassa dei Piomboni	IT	MY6245
Site 1 Atalhada	AZ	MY6065
Site 1 Atalhada	AZ	MY6080
Site 2 Brejela Atalhada	AZ	MY6074
Site 2 Brejela Atalhada	AZ	MY6075
Site 2 Brejela Atalhada	AZ	MY6076
Site 2 Brejela Atalhada	AZ	MY6083
Site 2 Brejela Atalhada	AZ	MY6084

Site 2 Brejela Atalhada	AZ	MY6086
Site 2 Brejela Atalhada	AZ	MY6093
Site 2 Brejela Atalhada	AZ	MY6094
Site 2 Brejela Atalhada	AZ	MY6095
Site 2 Brejela Atalhada	AZ	MY6096
Site 3 Ferraria	AZ	MY6098
Site 3 Ferraria	AZ	MY6099
Site 3 Ferraria	AZ	MY6100
Site 4 Fenais	AZ	MY6106
GenBank	SP	GenBank
W of Ponta do Alcaide Silveira Terceira	AZ	MY6298
W of Ponta do Alcaide Silveira Terceira	AZ	MY6299

Annex 8:

Matrix containing the number of the haplotypes for the different samples.

	Sites	Samples	haplotype
1	Areia Larga Madalena Pico AZ	MY6302	1
2	GenBank AZ	GenBank	2
3	Cap d'Ail FR	MY6286	1
4	Cap d'Ail FR	MY6287	3
5	E of Harwich UK	MY6284	4
6	Gijon SP	MY6282	5
7	Gijon SP	MY6283	6
8	Horta Faial AZ	MY6300	7
9	Horta Faial AZ	MY6301	2
10	Kerbrat Brittany FR	MY6002	1
11	Kerbrat Brittany FR	MY6003	8
12	Kerbrat Brittany FR	MY6004	8
13	Kerbrat Brittany FR	MY6005	8
14	Kerbrat Brittany FR	MY6007	8
15	Kerbrat Brittany FR	MY6008	1
16	Kerbrat Brittany FR	MY6009	1
17	Kerbrat Brittany FR	MY6010	1
18	L'horn FR	MY6029	1
19	L'horn FR	MY6030	1
20	L'horn FR	MY6031	1
21	L'horn FR	MY6032	1
22	L'île Grand Brittany FR	MY6017	9
23	L'île Grand Brittany FR	MY6020	10
24	Pender Island CA	MY6200	2
25	Pender Island CA	MY6202	2

26	Pender Island CA	MY6206	2
27	Pender Island CA	MY6207	2
28	Pender Island CA	MY6208	2
29	Pender Island CA	MY6209	2
30	Pender Island CA	MY6210	2
31	Pialassa dei Piomboni IT	MY6245	1
32	Site 1 Atalhada AZ	MY6065	3
33	Site 1 Atalhada AZ	MY6080	11
34	Site 2 Brejela Atalhada AZ	MY6074	3
35	Site 2 Brejela Atalhada AZ	MY6075	12
36	Site 2 Brejela Atalhada AZ	MY6076	13
37	Site 2 Brejela Atalhada AZ	MY6083	14
38	Site 2 Brejela Atalhada AZ	MY6084	13
39	Site 2 Brejela Atalhada AZ	MY6086	15
40	Site 2 Brejela Atalhada AZ	MY6093	13
41	Site 2 Brejela Atalhada AZ	MY6094	3
42	Site 2 Brejela Atalhada AZ	MY6095	16
43	Site 2 Brejela Atalhada AZ	MY6096	17
44	Site 3 Ferraria AZ	MY6098	15
45	Site 3 Ferraria AZ	MY6099	14
46	Site 3 Ferraria AZ	MY6100	14
47	Site 4 Fenais AZ	MY6106	18
48	GenBank SP	GenBank	19
49	W of Ponta do Alcaïde Silveira Terceira AZ	MY6298	1
50	W of Ponta do Alcaïde Silveira Terceira AZ	MY6299	12

Annex 9:

Matrix showing the percent of resemblance between our different samples.

Bibliography

Abdelkrim, J., Pascal, M., Calmet, C., Samadi, S., 2005. Importance of assessing population genetic structure before eradication of invasive species: examples from insular Norway rat populations. Conservation Biology 19, 1509–1518.

Al-Khafaji, K., 2021. First record of the invasive snail Myosotella myosotis (Draparnaud, 1801) in coast Shatt al-Arab estuary/Al-Faw District, Basrah-Iraq. Egyptian Journal of Aquatic Biology and Fisheries 25, 321–328.

Allendorf, F.W., Lundquist, L.L., 2003. Introduction: population biology, evolution, and control of invasive species. Conservation Biology 24–30.

An, A.I.S., 2006. State Management Plan for Aquatic Invasive Species in South Carolina.

Andrade-Restrepo, M., Champagnat, N., Ferrière, R., 2019. Local adaptation, dispersal evolution, and the spatial eco-evolutionary dynamics of invasion. Ecol Lett 22, 767–777. https://doi.org/10.1111/ele.13234

Ansrp, T., 2005. Approved for Public Release Distribution Unlimited.

Antonio M. de Frias, M., 1996. Anatomy and systematics of the western Atlantic Ellobiidae (Gastropoda: Pulmonata).

Berman, J., 1989. The ecology of biological invasions: interactions between native and introduced salt marsh gastropods (PhD Thesis). University of Oregon.

Binggeli, P., 1996. A taxonomic, biogeographical and ecological overview of invasive woody plants. Journal of Vegetation Science 7, 121–124. https://doi.org/10.2307/3236424

Blakeslee, A.M.H., Miller, A.W., Ruiz, G.M., Johannesson, K., André, C., Panova, M., 2021. Population structure and phylogeography of two North Atlantic Littorina species with contrasting larval development. Mar Biol 168, 117. https://doi.org/10.1007/s00227-021-03918-8

Brown, K.M., Lydeard, C., 2010. Mollusca Gastropoda. Ecology and Classification of North American Freshwater Invertebrates.

Bruyndoncx, L., Jordaens, K., Ysebaert, T., Meire, P., Backeljau, T., 2002. Molluscan diversity in tidal marshes along the Scheldt estuary (The Netherlands, Belgium). Hydrobiologia 474, 189–196.

Bürger, R., Akerman, A., 2011. The effects of linkage and gene flow on local adaptation: A two-locus continent–island model. Theoretical Population Biology 80, 272–288. https://doi.org/10.1016/j.tpb.2011.07.002

Caprio, M.A., Tabashnik, B.E., 1992. Gene flow accelerates local adaptation among finite populations: simulating the evolution of insecticide resistance. Journal of Economic Entomology 85, 611–620.

Carlton, J.T., 1992. Introduced marine and estuarine mollusks of North America: an end-of-the-20th-century perspective. Journal of shellfish research 11, 489–505.

Castle, D., Hebert, P.D.N., Clare, E.L., Hogg, I.D., Tremblay, C., 2021. Capturing the value of biosurveillance "big data" through natural capital accounting. Big Earth Data 5, 352–367. https://doi.org/10.1080/20964471.2021.1946290

Castro, J.A., Picornell, A., Ramon, M., 1998. Mitochondrial DNA: a tool for populational genetics studies. Int Microbiol 1, 327–332.

Chase, M.R., Etter, R.J., Rex, M.A., Quattro, J.M., 1998. Extraction and Amplification of Mitochondrial DNA from Formalin-Fixed Deep-Sea Mollusks. BioTechniques 24, 243–247. https://doi.org/10.2144/98242bm16

Cohen, A.N., Zabin, C.J., 2009. Oyster shells as vectors for exotic organisms. Journal of Shellfish Research 28, 163–167.

Colautti, R.I., Lau, J.A., 2016. CONTEMPORARY EVOLUTION DURING INVASION: EVIDENCE FOR DIFFERENTIATION, NATURAL SELECTION, AND LOCAL ADAPTATION, in: Barrett, S.C.H., Colautti, R.I., Dlugosch, K.M., Rieseberg, L.H. (Eds.), Invasion Genetics. John Wiley & Sons, Ltd, Chichester, UK, pp. 101–121. https://doi.org/10.1002/9781119072799.ch6

Craw, R.C., 1984. Biogeography and biogeographical principles. New Zealand Entomologist 8, 49–52.

Dark, S.J., 2004. The biogeography of invasive alien plants in California: an application of GIS and spatial regression analysis. Diversity and Distributions 10, 1–9. https://doi.org/10.1111/j.1472-4642.2004.00054.x

Dawnay, N., Ogden, R., McEwing, R., Carvalho, G.R., Thorpe, R.S., 2007. Validation of the barcoding gene COI for use in forensic

genetic species identification. Forensic Science International 173, 1–6. https://doi.org/10.1016/j.forsciint.2006.09.013

de Frias Martins, A.M., Mendes, A.R.M., n.d. Do cosmopolitans Do cosmopolitans speciate? Anatomical diversity of? Anatomical diversity of Myosotella in Azores.

Dennis, A.B., Hellberg, M.E., 2010. Ecological partitioning among parapatric cryptic species: NICHE PARTITIONING AMONG CRYPTIC SPECIES. Molecular Ecology 19, 3206–3225. https://doi.org/10.1111/j.1365-294X.2010.04689.x

Desjardins, P., Hansen, J.B., Allen, M., 2009. Microvolume Protein Concentration Determination Using the NanoDrop 2000c Spectrophotometer. JoVE 1610. https://doi.org/10.3791/1610

Ewens, W.J., 2013. Population Genetics. Springer Science & Business Media.

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. Mol Mar Biol Biotechnol 3, 294–299.

Gaitán-Espitia, J.D., Nespolo, R.F., Opazo, J.C., 2013. The Complete Mitochondrial Genome of the Land Snail Cornu aspersum (Helicidae: Mollusca): Intra.

Ghiselli, F., Gomes-dos-Santos, A., Adema, C.M., Lopes-Lima, M., Sharbrough, J., Boore, J.L., 2021. Molluscan mitochondrial genomes break the rules. Phil. Trans. R. Soc. B 376, 20200159. https://doi.org/10.1098/rstb.2020.0159

Gillespie, G.E., Phillips, A.C., Paltzat, D.L., Therriault, T.W., 2007. Distribution of non-indigenous intertidal species on the Pacific coast of Canada. Nippon Suisan Gakkaishi (Japanese Edition) 73, 1133.

Golikov, A.N., Dolgolenko, M.A., Maximovich, N.V., Scarlato, O.A., 1990. Theoretical approaches to marine biogeography. Marine ecology progress series. Oldendorf 63, 289–301.

González, V.L., Kayal, E., Halloran, M., Shrestha, Y., Harasewych, M.G., 2016. The complete mitochondrial genome of the land snail Cerion incanum (Gastropoda: Stylommatophora) and the phylogenetic relationships of Cerionidae within Panpulmonata. Journal of Molluscan Studies 82, 525–533.

Grande, C., Templado, J., Cervera, J.L., Zardoya, R., 2004. Molecular phylogeny of euthyneura (Mollusca: Gastropoda). Molecular biology and evolution 21, 303–313.

Grande, C., Templado, J., Zardoya, R., 2008. Evolution of gastropod mitochondrial genome arrangements. BMC Evolutionary Biology 8, 1–15.

Hamrick, J.L., 1982. Plant population genetics and evolution. American Journal of Botany 69, 1685–1693.

Hedgecock, D., 1986. Is gene flow from pelagic larval dispersal important in the adaptation and evolution of marine invertebrates? Bulletin of marine science 39, 550–564.

Herbert, D., 2012. *Myosotella myosotis* (Mollusca: Ellobiidae) — an overlooked, but well-established introduced species in South Africa. African Journal of Marine Science 34, 459–464. https://doi.org/10.2989/1814232X.2012.716374

Koufopanou, V., Reid, D.G., Ridgway, S.A., Thomas, R.H., 1999. A Molecular Phylogeny of the Patellid Limpets (Gastropoda: Patellidae) and Its Implications for the Origins of Their Antitropical Distribution. Molecular Phylogenetics and Evolution 11, 138–156. https://doi.org/10.1006/mpev.1998.0557

Kyle, C.J., Boulding, E.G., 2000. Comparative population genetic structure of marine gastropods (Littorina spp.) with and without pelagic larval dispersal. Marine Biology 137, 835–845. https://doi.org/10.1007/s002270000412

Le Roux, J., Wieczorek, A.M., 2009. Molecular systematics and population genetics of biological invasions: towards a better understanding of invasive species management. Annals of Applied Biology 154, 1–17.

Lee, C.E., 2002. Evolutionary genetics of invasive species. Trends in Ecology & Evolution 17, 386–391. https://doi.org/10.1016/S0169-5347(02)02554-5

Lenormand, T., 2002. Gene flow and the limits to natural selection. Trends in ecology & evolution 17, 183–189.

Li, X.-M., She, D.-Y., Zhang, D.-Y., Liao, W.-J., 2015. Life history trait differentiation and local adaptation in invasive populations of Ambrosia artemisiifolia in China. Oecologia 177, 669–677. https://doi.org/10.1007/s00442-014-3127-z

López-Goldar, X., Agrawal, A.A., 2021. Ecological Interactions, Environmental Gradients, and Gene Flow in Local Adaptation. Trends in

Plant Science 26, 796–809. https://doi.org/10.1016/j.tplants.2021.03.006

Marcus, E. du B.-R., Marcus, E., 1965. On Brazilian supratidal and estuarine snails. Boletim da Faculdade de Filosofia, Ciências e Letras, Universidade de São Paulo. Zoologia 25, 19–103.

McDermott, J.M., McDonald, B.A., 1993. Gene flow in plant pathosystems. Annual review of phytopathology 31, 353–373.

Mendes, A.R.M., 2013. O complexo Myosotella'myosotis/denticulata': uma perspetiva morfológica, molecular e ecológica (PhD Thesis).

Mikkelsen, N.T., Todt, C., Kocot, K.M., Halanych, K.M., Willassen, E., 2019. Molecular phylogeny of Caudofoveata (Mollusca) challenges traditional views. Molecular Phylogenetics and Evolution 132, 138–150. https://doi.org/10.1016/j.ympev.2018.10.037

Mitov, P.G., 2016. New data on Myosotella myosotis (Draparnaud, 1801)(Pulmonata: Basommatophora: Ellobiidae) from Bulgaria. Acta zool. bulg 68, 2016.

Morton, J.E., 1955. The functional morphology of the British Ellobiidae (Gastropoda, Pulmonata) with special reference to the digestive and reproductive systems. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences 239, 89–160.

Muller-Hohenstein, K., 2001. Biogeography, in: Smelser, N.J., Baltes, P.B. (Eds.), International Encyclopedia of the Social & Behavioral Sciences. Pergamon, Oxford, pp. 1201–1206. https://doi.org/10.1016/B0-08-043076-7/02599-7

Neigel, J.E., 1997. A Comparison of Alternative Strategies for Estimating Gene Flow from Genetic Markers. Annu. Rev. Ecol. Syst. 28, 105–128. https://doi.org/10.1146/annurev.ecolsys.28.1.105

Paulson, E.G., 1957. TA^ üOJteMY OF SALT MARSH SNAIL, OVA^TELLA yfgi ftiede MYOSOTIS, IN CENTRAL CALIFORNIA. NAUTILUS 71, 1.

Peer, N., Rajkaran, A., Miranda, N., Taylor, R., Newman, B., Porri, F., Raw, J., Mbense, S., Adams, J., Perissinotto, R., 2018. Latitudinal gradients and poleward expansion of mangrove ecosystems in South Africa: 50 years after Macnae's first assessment. African Journal of Marine Science 40, 101–120. https://doi.org/10.2989/1814232X.2018.1466728

Pérez-Quintero, J.C., 2013. Mollusc communities along upstream–downstream gradients in small coastal basins of the south-western Iberian Peninsula. Hydrobiologia 703, 165–175.

Pérez-Quintero, J.C., 2007. Diversity, habitat use and conservation of freshwater molluscs in the lower Guadiana River basin (SW Iberian Peninsula). Aquatic Conservation: Marine and Freshwater Ecosystems 17, 485–501.

Prentis, P.J., Wilson, J.R.U., Dormontt, E.E., Richardson, D.M., Lowe, A.J., 2008. Adaptive evolution in invasive species. Trends in Plant Science 13, 288–294. https://doi.org/10.1016/j.tplants.2008.03.004

Ray, G.L., 2005a. Invasive marine and estuarine animals of California.

Ray, G.L., 2005b. Invasive Marine and Estuarine Animals of the Pacific Northwest and Alaska: Defense Technical Information Center, Fort Belvoir, VA. https://doi.org/10.21236/ADA440459

Riddle, B.R., Dawson, M.N., Hadly, E.A., Hafner, D.J., Hickerson, M.J., Mantooth, S.J., Yoder, A.D., 2008. The role of molecular genetics in sculpting the future of integrative biogeography. Progress in Physical Geography 32, 173–202.

Romero, P.E., Pfenninger, M., Kano, Y., Klussmann-Kolb, A., 2016. Molecular phylogeny of the Ellobiidae (Gastropoda: Panpulmonata) supports independent terrestrial invasions. Molecular phylogenetics and evolution 97, 43–54.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. Syst Biol 61, 539–542. https://doi.org/10.1093/sysbio/sys029

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/.

Rudy Jr, P., Rudy, L.H., Shanks, A., Butler, B., 2013. Myosotella myosotis. Oregon Estuarine Invertebrates, Second Edition.

Ruthensteiner, B., Stocker, B., 2009. Genital system anatomy and development of Ovatella myosotis by three-dimensional computer visualization. Acta Zoologica 90, 166–178.

Sakai, A.K., Allendorf, F.W., Holt, J.S., Lodge, D.M., Molofsky, J., With, K.A., Baughman, S., Cabin, R.J., Cohen, J.E., Ellstrand, N.C., 2001. The population biology of invasive species. Annual review of ecology and systematics 32, 305–332.

Sánchez, O., Robla, J., Arias, A., 2021. Annotated and Updated Checklist of Land and Freshwater Molluscs from Asturias (Northern Spain) with Emphasis on Parasite Transmitters and Exotic Species. Diversity 13, 415.

Sanford, E., Kelly, M.W., 2011. Local Adaptation in Marine Invertebrates. Annu. Rev. Mar. Sci. 3, 509–535. https://doi.org/10.1146/annurev-marine-120709-142756

Savolainen, O., Pyhäjärvi, T., Knürr, T., 2007. Gene flow and local adaptation in trees. Annu. Rev. Ecol. Evol. Syst. 38, 595–619.

Schultz, J.A., Hebert, P.D.N., 2022. Do pseudogenes pose a problem for metabarcoding marine animal communities? Molecular Ecology Resources 22, 2897–2914. https://doi.org/10.1111/1755-0998.13667

Seddon, M.B., Killeen, I. & Slapnik, R. 2018. *Myosotella denticulata. The IUCN Red List of Threatened Species* 2018: e.T170601A86023858. 30 March 2023. https://dx.doi.org/10.2305/IUCN.UK.2018-1.RLTS.T170601A86023858.en.

Si, C.-C., Dai, Z.-C., Lin, Y., Qi, S.-S., Huang, P., Miao, S.-L., Du, D.-L., 2014. Local adaptation and phenotypic plasticity both occurred in Wedelia trilobata invasion across a tropical island. Biol Invasions 16, 2323–2337. https://doi.org/10.1007/s10530-014-0667-4

Slapnik, R. 2018. *Myosotella myosotis. The IUCN Red List of Threatened Species* 2018: e.T170600A85576079. 30 March 2023. https://dx.doi.org/10.2305/IUCN.UK.2018-1.RLTS.T170600A85576079.en.

Slatkin, M., 1987. Gene flow and the geographic structure of natural populations. Science 236, 787–792.

Tamura, K., Stecher, G., Kumar, S., 2021. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. Mol Biol Evol 38, 3022–3027. https://doi.org/10.1093/molbev/msab120

Tiffin, P., Ross-Ibarra, J., 2014. Advances and limits of using population genetics to understand local adaptation. Trends in Ecology & Evolution 29, 673–680. https://doi.org/10.1016/j.tree.2014.10.004

Tigano, A., Friesen, V.L., 2016. Genomics of local adaptation with gene flow. Molecular Ecology 25, 2144–2164. https://doi.org/10.1111/mec.13606

Toparslan, E., Karabag, K., Bilge, U., 2020. A workflow with R: Phylogenetic analyses and visualizations using mitochondrial cytochrome b gene sequences. PLoS ONE 15, e0243927. https://doi.org/10.1371/journal.pone.0243927

Tsutsui, N.D., Suarez, A.V., Holway, D.A., Case, T.J., 2000. Reduced genetic variation and the success of an invasive species. Proceedings of the National Academy of Sciences 97, 5948–5953. https://doi.org/10.1073/pnas.100110397

van Boheemen, L.A., Atwater, D.Z., Hodgins, K.A., 2019. Rapid and repeated local adaptation to climate in an invasive plant. New Phytologist 222, 614–627. https://doi.org/10.1111/nph.15564

Verhoeven, K.J.F., Macel, M., Wolfe, L.M., Biere, A., 2011. Population admixture, biological invasions and the balance between local adaptation and inbreeding depression. Proc. R. Soc. B. 278, 2–8. https://doi.org/10.1098/rspb.2010.1272

Vernberg, F.J., 1993. Salt-marsh processes: a review. Environmental Toxicology and Chemistry: An International Journal 12, 2167–2195.

Wickham, H., 2007. Reshaping data with the reshape package. Journal of statistical software 21, 1–20.

Wickham, H., Wickham, M.H., 2020. Package 'plyr.' Obtenido Httpscran Rproject Orgwebpackagesdplyrdplyr Pdf.

Wiens, J.J., 2011. The niche, biogeography and species interactions. Phil. Trans. R. Soc. B 366, 2336–2350. https://doi.org/10.1098/rstb.2011.0059

Wiens, J.J., Donoghue, M.J., 2004. Historical biogeography, ecology and species richness. Trends in Ecology & Evolution 19, 639–644. https://doi.org/10.1016/j.tree.2004.09.011

Zalewski, A., Michalska-Parda, A., Bartoszewicz, M., Kozakiewicz, M., Brzeziński, M., 2010. Multiple introductions determine the genetic structure of an invasive species

population: American mink Neovison vison in Poland. Biological Conservation 143, 1355–1363. https://doi.org/10.1016/j.biocon.2010.03.0 09

Zhao, S.Y., Sun, S.G., Dai, C., Gituru, R.W., Chen, J.M., Wang, Q.F., 2015. Genetic variation and structure in native and invasive *S olidago canadensis* populations. Weed Research 55, 163–172. https://doi.org/10.1111/wre.12130