Chapter 1

General Introduction

Yacine Ben Chehida
Glossary

**Adaptive introgression**: Gene flow of a part of the genome from one species into another that confer an adaptive function in the recipient species.

**Antitropical distribution**: Geographical distribution represents a distribution in which closely related taxa occur on either side of the equator but are absent from the tropics.

**Barrier loci**: see **Speciation gene**.

**Biogeography**: The study of the distribution of organisms in space and time.

**Coalescent theory**: Mathematical theory describing alleles genealogy from a sample taken from a population going backward in time until their most recent common ancestor.

**Demes**: A local random mating unit.

**Disruptive selection**: Selection in favor of extreme phenotypes and against those intermediate between them.

**Dobzhansky Muller incompatibility (DMI)**: Reduction in the fitness of a hybrid because of interaction between alleles at two or more loci in one parent population with specific alleles at other loci in the other parent population.

**Effective population size** (**Ne**): Measure of the power of genetic drift. Often defined as the number of individuals in an ideal population producing the same rate of genetic drift as the real population.

**Extinction vortex**: As populations decline, a reinforcement between genetic and demographic processes such as environmental stochasticity, demographic stochasticity, inbreeding depression or genetic drift can occur which precipitates the population size downward and thus drives the species towards extinction.

**Founder effect**: Foundation of a new population by a restricted number of individuals (the founders) carrying only a fraction of the total genetic variation from the source population.

**Genetic drift**: Random changes in allele frequencies within a population.

**Hybridization**: Production of offspring by interbreeding between members of genetically distinct populations or species.

**Incomplete lineage sorting**: Persistence of a genetic polymorphism through a speciation event, so that fixation occurs only in the descendant species, or in their descendants after subsequent speciation.

**Introgression**: Gene flow of a part of the genome from one donor species to recipient species.

**Landscape genetics**: Union of the field of population genetics with landscape ecology aiming to study the interaction between landscape features and microevolutionary processes.
**Linkage disequilibrium**: A non-random association between alleles at different loci in a population.

**Monophyly**: Taxon containing all the lineages that descended from a common ancestor.

**Paraphyly**: Taxon containing the most recent common ancestor, but not all of its descendants.

**Polyphyly**: Taxon containing lineages that do not share a unique common ancestor.

**Phylogeography**: The study of the evolutionary processes responsible for the geographic distributions of genealogical lineages at the intra- and interspecific levels.

**Recombination**: Biological process occurring during the meiosis resulting in the reshuffling of the allelic combination. The produced offspring has a different combination of alleles that differs from those found in either parent.

**Sexual selection**: Difference in reproductive success between members of the same sex.

**Speciation genes**: Genes or loci differentiated between populations and that contribute significantly to the evolution of reproductive isolation.
Population differentiation across contiguous geographical areas due to barriers to gene flow (i.e., allopatric divergence) is expected to be less frequent in the marine realm, especially for highly mobile marine pelagic species such as cetaceans, sharks, and turtles (Palumbi, 1994). The high dispersal capability of these species and the scarcity of obvious physical barriers to dispersal should favor in theory population homogeneity over large geographic scales and limit population differentiation in their natural range (Bowen, et al., 2016; Bailleul, et al., 2018). Nevertheless, intraspecific population differentiation in marine pelagic species has been documented (Fontaine, et al., 2007; Louis, et al., 2014a) and can be strong, even at small geographical scales (Bowen, et al., 2016). Resource specializations and local adaptation to particular environments are among the most often reported biological processes shaping genetic differentiation and dispersal patterns for such species (Louis, et al., 2014a; Louis, et al., 2014b). Geographic and temporal variation in abiotic factors have been correlated with population divergence (Fontaine, et al., 2014), so are biotic factors, such as food availability, feeding preference, predators, competitors, and behavioral traits (e.g., social structure, foraging strategies, migration pattern, and philopatry; O’Corry Crowe, 2007).

Dispersal and population structure of mobile pelagic species can lead to population divergence, specialization, and potentially also to reproductive isolation or speciation (i.e., the process by which biological entities evolve to become distinct species). These processes are influenced by (1) local variations in environmental conditions (biotic and abiotic), (2) past environmental changes that shaped historical population/species distributions, and (3) by the ongoing environmental changes that lead to a redistribution of species with the current global warming. All these considerations raise the following questions:

*How do populations become genetically distinct? What are the underlying mechanisms driving population/species dispersal, local adaptation, and reproductive isolation? Ultimately, how can speciation occur in highly mobile marine species? How important are past events, such as past climate changes, in shaping present population structure and diversification patterns for such a species? To what extent changes in the population demography and connectivity can influence it? What is the contribution of local adaptation? What is the real impact of the current global changes on these species? How can such knowledge inform conservation decisions?*

This thesis aims to discuss some of these questions on cetaceans. Cetaceans represent an infraorder within mammals totally adapted to life in aquatic environments (Berta, et al., 2015a). They constitute the largest group of marine mammals with officially 90 species according to
the Committee on Taxonomy of the Society for Marine Mammalogy\(^1\) and include whales, dolphins, and porpoises. Cetaceans are good examples of species to address the previously mentioned questions due to their broad range and their high dispersal capabilities. Focusing on a group of small cetaceans, the porpoise family (Phocoenidae), this thesis will attend to provide a better understanding of these key questions in evolutionary biology. In this introduction, I will first introduce the process of speciation illustrating theoretical consideration with examples in cetacean. In the second part, I will explain how past environmental changes have driven the biogeography and the phylogeography of cetacean. I will also discuss some fundamental phylogeography, population genetics, and conservation genetics principles necessary for the comprehension of this thesis. In the third part, I will present the biological model I am focusing on in the framework of my PhD. In the last part, I will present an outline of the chapters of this thesis.

1. Speciation

1.1 Definition and general concept

Speciation represents the ensemble of mechanisms that lead to the formation of new species. In the context of this thesis, I will mainly use the biological concept of species, which defines species in terms of interbreeding (de Queiroz, 2005). A species represents a group of interbreeding individuals that are reproductively isolated from other such individuals and whose offspring are viable and fertile in natural conditions. Therefore, the mechanism underlying all speciation events is reproductive isolation (Bush, 1975). Reproductive isolation represents the biological mechanisms leading to the creation of barriers to gene exchanges between groups (individuals/populations). If these isolation barriers hold long enough, genetic differences between groups accumulate via mutations, selection, and drift and may lead to biological incompatibilities. At a certain point, as incompatibilities cumulate, the differences are large enough to prevent groups from breeding or having viable offspring. At that point, the groups belong to distinctive species.

Isolation barriers are generally separated into pre-zygotic and post-zygotic barriers. Pre-zygotic barriers encompass all the mechanisms that prevent the formation of a zygote, i.e. the mating between individuals or the fusion of gametes. The nature of these barriers can be geographical, ecological, behavioral, physiological, temporal, mechanical, and/or gametic (Sætre and Ravinet, 2019). Post-zygotic barriers represent the ensemble of mechanisms that occur after the zygote has formed. These mechanisms either reduce the viability or the reproductive abilities of the hybrids. These barriers can depend on the environment (ecological inviability or behavioral sterility) or not (hybrid inviability/sterility due to genetic incompatibilities or inability to produce viable gametes). In most organisms, pre-zygotic barriers evolve first and faster (Coyne and Orr, 1997; Mendelson, 2003). The expected order of apparition of such pre- and/or post-zygotic barriers under different speciation scenarios has been reviewed by Seehausen et al. (2014).

Species hybridization is common in nature (Taylor and Larson, 2019). Therefore, interbreeding is not an absolute criterion to define species and in certain cases, the biological species concept may be seen as too rigid. That is why modern evolutionary biologists prefer to consider speciation as a continuous process (Mallet, et al., 2007; Nosil, 2012b). There is accumulating evidence indicating that biological divergence during speciation varies continuously between pairs of conspecific groups (reviewed in Nosil, 2012b; Stankowski and Ravinet, 2021). Such evidence varies, for example, in the strength of reproductive isolation,
hybridization frequency, the fitness of the hybrids, level of gene flow, the geographic difference in allelic frequencies, degree of individual clustering in molecular markers, phenotypic variations, or extent of genetic lineage sorting. Therefore, speciation can be thought of as a gradual process, commonly called the “speciation continuum” (Nosil, 2012b; Stankowski and Ravinet, 2021), varying from totally undifferentiated entities to totally reproductively isolated ones (Figure 1.1). Each conspecific group represents different points along this speciation continuum, where the degree of observed reproductive isolation is informative about how far speciation has proceeded.

![Figure 1.1. Representation of three points across "the speciation continuum". The biological divergence between separating taxa can be measured using different types of metrics (here represented as the level of genomic divergence, shared polymorphism, reproductive isolation, genotypic clustering lineage sorting, or allele frequency distribution across a geographic cline), that are all complementary proxy of how far speciation has proceeded. Modified from Nosil et al. (2012c).]
Despite several attempts to fix a cut-off point of genetic divergence to define a species, there is currently no consensus (Nosil, et al., 2003). Indeed, the extent of genetic exchange observed between subspecies and species decreases gradually with genetic distance, without any obvious break allowing to precisely define a sharp species boundary (Roux, et al., 2016). However, there is an intermediate interval of genomic divergence, named the "grey zone" of speciation (Figure 1.1 and 1.2). In this range, the notion of species is extremely debatable as for a similar amount of genetic divergence, the reproductive isolation between congeneric groups can be absent, partial or total, and where different species concepts do not provide the same species delineation (De Queiroz, 2007; Roux, et al., 2016). A comparison of 61 pairs of animals suggested that this “grey zone” ranges from 0.5% to 2% of net synonymous divergence (Roux, et al., 2016). Below and above this range of genomics divergence, pairwise comparisons are strongly suggested to be intraspecific (lack of reproductive barriers) and interspecific (strong barriers to gene flow), respectively (Figure 1.2b).

![Figure 1.2. The grey zone of speciation. a) Sketch illustrating the concept of grey zone of speciation. Modified from Bailleul et al. (2018). b) Probability of gene flow as a function of the genetic divergence along the speciation continuum. The area in grey represents the grey zone of speciation. Each dot is for one observed pair of populations/species. Modified from Roux et al. (2016).](image)

1.2 Geographical modes of speciation

Multiple mechanisms can prevent or decrease individuals' mating or survival, and initiate barriers by restricting gene flow between populations, which can trigger the speciation processes. Generally, gene flow restriction among populations results from their geographic separation, but, as I will discuss, other mechanisms can also have this effect. Therefore, traditionally speciation was primarily presented and discussed in terms of geography. Four main
geographical models of speciation are commonly recognized in the literature (Figure 1.3): allopatric, parapatric, sympatric, and peripatric (Coyne and Orr, 2004).

**Figure 1.3.** Geographic modes of speciation. From²

In allopatric speciation, a geographic barrier prevents breeding, and thus gene flow, between groups of individuals of the same species. The different groups diverge independently due to their geographical separation and become more and more differentiated from each other with time. When the differences get to a certain point, pre- and post-zygotic reproductive barriers appear, which may prevent mating between groups when they come back into contact, i.e. after secondary contact. The cause of this geographical isolation is generally the subdivision of a uniform geographical area by a geographic barrier, for example, due to geological or climatic origin. In the marine environment, a notorious example is the case of allopatric speciation driven by the formation of the Isthmus of Panama ~3.2 million years ago (Mya; Odea, et al., 2016). The rise of the Isthmus of Panama led to the closure of the Central American seaway that allowed faunal interchange between the Pacific and Atlantic Oceans when it was open. This event completely isolated the marine taxa that once inhabited the Central American seaway. It had a tremendous effect on those taxa by triggering their diversification by allopatric

divergence (Bowen, et al., 2016; Odea, et al., 2016). Today, many marine sister species inhabit the water mass separated by the Isthmus of Panama including marine mammals (Uhen, et al., 2010).

Peripatric speciation is a special case of allopatric speciation where a small population splits from the main range and experiences a **founder effect**. Peripatric speciation can be distinguished from allopatric speciation by the size of the isolated population. This small population is the founder population that will evolve separately to become a new species. Small populations can diverge rapidly as they are strongly subjected to **genetic drift**, which leads to a rapid fixation of alleles and the accumulation of deleterious mutations (Barton and Charlesworth, 1984). Furthermore, if the isolated population colonizes a novel environment it can be subjected to different selection pressures as well. Due to their high dispersal abilities, cetacean species may be prone to evolve by peripatric speciation, particularly species living in large groups. Those groups are particularly susceptible to colonize new environments and to fragmentation, leading few individuals to get isolated from the main group and which theoretically leads to peripatric speciation (Möller, et al., 2008). Evidence of this mode of speciation is limited and remains so far controversial (Barton and Charlesworth, 1984; Slatkin, 1996).

Sympatric speciation occurs when reproductive isolation appears in the absence of geographical barriers, so that groups of individuals diverge in the same geographic range but become reproductively isolated until they are not able to breed anymore. Due to the geographical proximity of the individuals, sympatric speciation takes place in the presence of gene flow which has a homogenizing effect. Genetic divergence occurs because gene flow is overridden by strong divergent selection. In general, the reproductive isolation leading to a sympatric speciation event is related to competition, mate choice, temporal factors, or when new niches become available within the same area (Bolnick and Fitzpatrick, 2007). Though an increasing number of studies support its plausibility (Turner and Burrows, 1995; Barluengua, et al., 2006; Bolnick and Fitzpatrick, 2007), this mode of speciation remains extremely controversial (Jiggins, 2006). First, the right conditions for sympatric speciation to happen in nature are rare. For example, it requires a specific balance between selection and **recombination**. The problem is that at any point in time, recombination can break up **linkage disequilibrium** between beneficial alleles associated with environmental adaptation and with reproductive isolation, hence preventing the completion of the speciation process (Bolnick and Fitzpatrick, 2007; Foote, 2018). Second, a major difficulty when studying sympatric speciation is to unequivocally distinguish divergence with gene flow from secondary contact (Foote,
In this regard, many sister species that are sympatric today have speciated in allopatry and then expanded their ranges. This means that current sympatry, in itself, is not evidence that sympatric speciation took place. In the future, the incorporation of paleogenomics samples could help to settle the debate (Foote, 2018).

Parapatric speciation is another mode of speciation where reproductive isolation appears while gene flow is still occurring. It is an intermediate between allopatric and sympatric speciation. In parapatric speciation, individuals are distributed along partially overlapping continuous ranges or in two adjacent populations. Individuals do not mate randomly across their whole distribution but only with their geographical neighbors. Therefore, in this scenario, divergence may happen because of reduced gene flow within or between the population(s). In general, it begins with variable environmental conditions across the populations’ range (e.g., environmental cline), leading to disruptive selective pressures across their distribution resulting in the local adaptation to different environments. Divergent selection increases local adaptation that in turn causes a change in genotypic and phenotypic frequencies and increases selection against migrants in the different environmental conditions. With time, differentiation between populations increases and becomes sufficient to lead eventually to the formation of two distinct species. Parapatric speciation often results in the formation of a contact zone with hybrids (Hewitt, 2001). This mode of speciation is expected to be more common than sympatric speciation because it involves less gene flow between the diverging populations. An example of incipient parapatric speciation has been documented in three species of lizards spread across different soil types (white sands vs dark soil) in New Mexico (Rosenblum and Harmon, 2010). Depending on the type of soil these lizards inhabit, they display morphological and genetic differences reflecting adaptation to their respective ecological niches (different soil types; Rosenblum, 2010). It has been further shown for the three species that the white sands and dark soil populations are partially reproductive isolated (Hardwick, et al., 2013).

1.3 The mechanistic cause of speciation: a modern view

The geographical modes of speciation are useful when developing theoretical models that explore potential mechanisms driving speciation. However, they are a proxy for how diverging lineages are able to exchange genes. In this regard, they constitute a simplistic approximation of how speciation proceeds as in practice populations may experience alternating phases of allopatry, parapatry, and sympatry during speciation, with different barriers arising at different times. Also, these modes of speciation do not offer good enough explanations of how gene exchanges are prevented which as I discussed previously is the key
when studying speciation. Therefore, nowadays the critical issue in understanding speciation is to understand how gene flow is reduced at the earliest stages when two populations start to diverge. This is paramount because when gene flow is high enough it can override the effect of divergent selection or genetic drift, and thus stops reproductive isolation and speciation to occur. As there is growing evidence of the persistent gene flow between sister species, it implies that divergence often arises while genetic material is still exchanged (Pinho and Hey, 2010). Over the last years, the study of sympatric and parapatric speciation has received a lot of attention because, on the contrary to allopatric speciation, it involves speciation with gene flow (Foote, 2018). These modes of speciation provide important insights into the role of selection in driving speciation. Thus, while traditionally the study of speciation was primarily interested in understanding how different geographic scenarios could lead to speciation; the field of speciation has shifted progressively from this geographic view toward a more mechanistic view of the speciation by trying to determine the intrinsic et extrinsic processes that reduced gene exchange between diverging populations (Pinho and Hey, 2010; Sousa and Hey, 2013).

**Figure 1.4.** Alternative models of speciation under the "modern view". All models assume that an ancestral population (grey) splits into two sister populations (orange and blue) at time \( T_2 \). Arrows indicate the presence of gene flow. a) Divergence model assuming no gene flow. b) Divergence model assuming constant gene flow through time. c) Divergence model assuming populations started diverging in the presence of gene flow which ceased at \( T_1 \). d) Divergence model assuming started diverging in the absence of gene flow, followed by a secondary contact i.e. assume a return of gene flow at \( T_1 \). Figure inspired by Sousa and Hey (2013).

Under this new framework, independently of the geographical context, the focus is to understand whether genetic divergence occurs in the absence of gene flow (i.e. strict isolation model; Figure 1.4a) or in the presence of gene flow (isolation-with-migration models or IM, Figure 1.4b-d). The second category of models (IM’s) is subdivided into divergence with constant gene flow (Figure 1.4b), divergence with ancestral gene flow (Figure 1.4c), and secondary contact (Figure 1.4d; Sousa and Hey, 2013).
This modern approach to speciation took largely advantage of the genomics revolution because it has been observed that the different models of isolation can leave distinct signatures in the genetic polymorphism along the genome and that gene flow and divergence are heterogeneous across the whole genome (Seehausen, et al., 2014; Roux, et al., 2016; Miles, et al., 2017). This stimulated the new field of speciation genomics that tries to identify and quantify the genetic signatures left by different modes of speciation, and also the genomic locations of loci that may have contributed to the evolution of the divergence and reproductive isolation between populations and species. This offers a direct insight into how the evolutionary forces shape the genomics polymorphism, and the architecture of the genetic divergence and gene flow between groups, thereby facilitating our comprehension of the conditions that can drive speciation. A large emphasis has been given to the involvement of selective processes in disrupting gene flow at specific loci in the genome (the so-called barrier loci). This new field provided rich insights into the mechanisms causing speciation and has led to a reassessment of some of the most compelling empirical case studies of speciation (Martin, et al., 2013; Fontaine, et al., 2015; Malinsky, et al., 2015).

1.4 The cause of speciation

There are many mechanisms that can cause speciation (Coyne and Orr, 2004). Below I will quickly review some of them that can be relevant for cetacean species.

Several authors have proposed that speciation can be driven by chromosomal rearrangement with some suggesting even that chromosomal rearrangements are playing a major role in most of the speciation events (Rieseberg, 2001; Wellenreuther and Bernatchez, 2018). Many models of chromosomal speciation assume that the chromosomal changes accumulating among diverging populations impair heterozygous hybrids' fertility/viability (Lande, 1979). An alternative hypothesis is that chromosomal inversions can reduce gene flow by suppressing recombination and extend the effects of linked isolation genes (Rieseberg, 2001). Inversions, for example, have a pervasive role in evolution because this suppression of recombination can protect combinations of favorable alleles facilitating adaptation to local conditions (Wellenreuther and Bernatchez, 2018). When such inversion confers a strong advantage to a species in a specific ecological niche, it can be transmitted to closely related species via a process called adaptive introgression. Then such a species can expand its niche and hence diversify. Adaptive introgression is a major driver of insect diversity (Feder, et al., 2003; Fontaine, et al., 2015; Miles, et al., 2017; Fuller, et al., 2019). Some inversions are also directly implicated in speciation because they can comprise genes involved in reproductive
isolation. For example, the 3Ra inversion of the mosquito *Anopheles funestus* is implicated in a strong assortative mating between populations (Ayala, et al., 2013). This inversion offers a direct substrate for a new speciation event because it contributes to up to 92% of the reproductive isolation between *Anopheles funestus* populations (Ayala, et al., 2013).

Speciation by genetic drift is, as its name implies, a mode of speciation where biological divergence is primarily driven by genetic drift. When populations get separated they accumulate genetic differences via genetic drift leading progressively to an increase of their reproductive isolation. Speciation by genetic drift is particularly likely when separated populations become small in size in such a way that genetic drift is stronger than selection. In such a scenario, drift can rapidly fix alleles leading to the accumulation of reproductive incompatibilities between the diverging populations. In some cases, it results in the so-called **Dobzhansky-Muller incompatibilities** (Presgraves, 2010). The fixation of chromosomal rearrangements by drift in a population can lead to the rapid apparition of postzygotic isolation in hybrids because of the low fertility of chromosomes heterozygotes (Lande, 1979; Hoffmann and Rieseberg, 2008).

**Sexual selection** is known to be a strong driver of speciation (Turner and Burrows, 1995; van Doorn, et al., 2009). In many taxa, one sex chooses its mates based on variation in secondary sexual traits. Strong sexual selection can drive the rapid evolution of secondary sexual traits and cause rapid evolution of prezygotic isolation between populations (Ryan and Rand, 1993). Certain groups of animals, such as cichlid fishes (Gante and Salzburger, 2012; Wagner, et al., 2012) and birds of paradise (Irestedt, et al., 2009), have radiated in part due to strong sexual selection. The strong sexual dimorphism and sperm competition reported in many cetacean species (Würsig, et al., 2016; Mesnick and Ralls, 2018) suggest that sexual selection can play an important role in the diversification of this group (Fordyce, 2018).

Hybrid speciation can be broadly defined as "the hybridization between two or more distinct lineages that contribute to the origin of a new species" (Counterman, 2016). It results in a hybrid lineage reproductively isolated from the parental species. Hybrid speciation can occur through the doubling of the number of chromosomes (allopolyploidy) in hybrids or not (homoploid). In allopolyploid hybrid speciation, because hybrids have a different ploidy, they can directly isolate from parent species via mechanisms such as improper chromosome pairing during meiosis or increased genetic incompatibilities in backcrosses (Otto and Whitton, 2000). Speciation by allopolyploidy is extremely common in plants but has also been reported in animals (Otto and Whitton, 2000; Abbott, et al., 2013), especially in fishes (Leggatt and Iwama, 2003). Homoploid hybrid speciation is less common because, on the contrary to allopolyploidy, the isolation from parental species is not instantaneous, so it requires reproductive isolation in
the presence of gene flow (Barton, 2001). One hypothesis is that recombination restores with
time the viability and/or fertility of F1 hybrids which can mate with each other but are isolated
from the parental species due to genetic incompatibilities (Counterman, 2016). Homoploid
hybrid speciation has been documented in many groups such as plants (Gross and Rieseberg,
2004), insects (Jiggins, et al., 2008), fishes (Selz, et al., 2013), or mammals (Chafin, et al.,
2020).

Natural selection can directly favor traits that cause reproductive isolation and
speciation via a process called reinforcement (Hopkins, 2016). Because interspecific hybrids
often have a reduced viability/fertility, individuals that do not hybridize will have a higher
fitness. Therefore, in such a context, natural selection should favor traits that increase
assortative mating (so against hybrids) and hence facilitating reproductive isolation. Under
these circumstances, any trait directly strengthening prezygotic isolation between lineages will
provide a net advantage and will be selected by reinforcement. However, for reinforcement to
occur, a reduced hybrid fitness must have already evolved between lineages. Reinforcement is
expected to be more common in sympatry where taxa are more likely to hybridize than in
allopatry to avoid the cost associated with hybridization (Yukilevich, 2012).

Ecological speciation is another mode of speciation driven by natural selection. I will
describe it in more detail in the next section.

1.5 Focus on ecological speciation

Ecological speciation is the process “by which reproductive isolation evolves between
populations as a result of ecologically-based divergent selection between environments” (Nosil,
2012c). In this case, speciation occurs because traits under divergent selection are correlated or
coupled with those affecting reproductive isolation. “Ecological” is defined here as an extrinsic
factor (abiotic and/or biotic) promoting divergent selection by acting in contrasting directions
in different lineages. Thus, ecological speciation is opposed to all other modes of speciation
where (1) the divergence occurs due to random genetic changes (speciation by genetic drift), or
(2) any kind of selection where populations experience similar selective pressures but fix
different advantageous mutations or traits that are incompatible with one another.

It is important to point out that ecological speciation is a mechanistic way of defining
speciation. Thus, it does not preclude other modes of geographical speciation. Indeed,
ecological speciation can occur in any geographical context as long as ecological divergent
selection is the mechanism driving divergence (Schluter, 2001). However, in an allopatric
context, it is difficult to tease apart the effect of genetic drift from the one of natural selection in driving the divergence between populations (Bierne, et al., 2011; Foote, 2018). In contrast, natural selection is necessarily the evolutionary force driving divergence in sympatry or parapatry (i.e. all the modes of divergence that imply gene flow). Thus, it represents an ideal setting to study ecological speciation and understand how it can gradually lead to the apparition of reproductive isolation despite the presence of genetic exchanges.

A simplistic scenario of ecological speciation would consider two populations living in different habitats. Due to ecological differences between habitats, each population would undergo divergent selective forces. Thus, the respective adaptation of each population to its habitat might cause the accumulation of genetics, morphologic, and/or behavioral differences between populations. If ecological adaptations correlate or couple with reproductive traits (e.g., behavioral, sexual selection, color patterns), then hybridization and thus genetic exchanges become counter selected. With time and reinforcement, these differences may prevent individuals from the two populations to mate and decrease hybrids’ fitness. Therefore, gene flow between such populations might cease, which might push the speciation forward. Evidence of ecological speciation has been demonstrated for model and no-model organisms via several theoretical models and laboratory experiments (Rundle and Nosil, 2005; Nosil, 2012a; Nosil, et al., 2012). The most famous examples of ecological speciation include cichlids (Winkelmann, et al., 2014), sticklebacks (Gasterosteus aculeatus; Hendry, et al., 2009), apple maggot (Rhagoletis pomonella; Jiggins and Bridle, 2004), and stick insects (Timema cristinae; Nosil, et al., 2012).

Ecological speciation generally occurs via two mechanisms. The first involves pleiotropy and assumes that the loci under divergent ecological selection represent the same genes causing reproductive isolation (also called barrier or speciation genes). For example, genes affecting body size can also be affecting mate choice and thus be involved in sexual isolation (Servedio, et al., 2011). Under the second mechanism, genes under divergence ecological selection are close (but physically at a different position) from the genes causing reproductive isolation (Nosil, 2012a). Due to their physical proximity, linkage disequilibrium is expected between those genes. Therefore, in such a context, we expect an increase of the divergence in speciation genes by genetic hitchhiking when the divergence of genes under contrasting ecological selection increases.

In this thesis, an ecotype is defined as a genetically distinct group of individuals presenting specialized ecological adaptations to their specific environment. When studying ecological speciation, the definition of ecotypes can help to understand the process governing
the divergence of individuals in different environments (Thorpe, et al., 2005). For killer whales (Orcinus orca), for example, many ecotypes have been described (de Bruyn, et al., 2013). Three distinct ecotypes are generally recognized in harbor porpoise (Phocoena phocoena) (Fontaine, 2016; IMR-NAMMCO, 2019). One of them, Phocoena phocoena meridionalis, displays distinct morphological (Smeenk, et al., 1992) and ecological features suggesting strong dietary adaptation (Pinela, et al., 2010). In a similar way, the narrow-ridged finless porpoises (Neophocaena asiaeorientalis) represent probably one of the best-documented cases of incipient speciation related to divergent habitat specialization. They are currently divided into two distinct ecotypes: the Yangtze (Neophocaena asiaeorientalis asiaeorientalis) and the East Asian (Neophocaena asiaeorientalis sunameri) finless porpoises (Jefferson and Wang, 2011). Both ecotypes inhabit different environments as the Yangtze finless porpoise is adapted to freshwater waters whereas the East Asian finless porpoise prefers marine habitats. Numerous morphological (Anli and Kaiya, 1995; Gao and Zhou, 1995) and genetic (Lin, et al., 2014; Zhou, et al., 2018) differences between the two ecotypes have been reported reflecting their ecologically differentiated habitat specializations. Comparative population genomics analyses of their whole genomes indicated that genes related to freshwater adaptation are under strong selection in Yangtze finless porpoise (Zhou, et al., 2018). Local adaptation of each lineage to these different conditions likely drives their genetic distinctiveness. More generally, all the examples introduced in this paragraph represent plausibly the initial stages of new species driven by ecological speciation.

1.6 Case of the speciation in the marine environment

The physical and chemical properties of the oceans (including light, primary production, depth, temperature, and salinity) influence the process of speciation in the marine realm (Palumbi, 1992). In marine systems, the lack of obvious barriers to gene flow makes widely separated regions potentially connected. These assumptions challenge the idea of allopatric speciation, especially in marine taxa with large dispersal abilities. In this regard, many species show little genetic population structure and appear to act as large random mating units (i.e., panmictic) in the marine environment (Palumbi, 1992; Bailleul, et al., 2018). However, many vertebrate and invertebrate taxa have an elevated number of species. These observations suggest that speciation is effectively occurring despite the large potential for long distance gene flow (Palumbi, 1994; Puebla, 2009). Furthermore, ecological speciation seems to play a critical role in the marine environment, as many sister taxa inhabit environments with different ecological
characteristics (Palumbi, 1994; Puebla, 2009). Another likely explanation is that allopatric speciation is occurring but at a larger geographic scale than in terrestrial environments. Indeed, when a species has strong dispersal abilities, reproductive isolation can occur at a very large geographical scale via allopatric or parapatric speciation (Hellberg, 2009). The greater the dispersal ability of the species, the larger the scale will be. However, all marine species have limits on their dispersal abilities which are influenced by the local oceanographic conditions that can generate opportunities for genetic divergence, reproductive isolation, and speciation (Palumbi, 1994).

Though not always evident, major geographic barriers exist also in the marine realm. For example, lands prevent the dispersal of all marine organisms. Furthermore, the current dynamic strongly determines the biogeographic limits of many marine taxa (Palumbi, 1994). Strong currents such as the Antarctic Circumpolar Current represent a barrier to gene flow for many small organisms such as small vertebrates, invertebrates, and planktonic organisms (Dueñas, et al., 2016; Ben Chehida, et al., 2019). While currents might not represent a direct geographical barrier for cetaceans, they influence indirectly their distributions and promote their divergence by impacting dispersal, abundance, and distribution of the primary production and feeding areas.

1.7 A few examples in cetaceans

The study of speciation in cetacean is still in its infancy (see also Westbury et al., 2020). The killer whale is so far among the best studied group. The ecological success and rapid radiation of killer whales over the last 350 kilo years (kya; Morin, et al., 2015) is plausibly driven by ecological and allopatric divergence processes. Killer whales occur in the Antarctic, Pacific, and Atlantic Oceans (de Bruyn, et al., 2013). They are divided into different ecotypes (often referred to as types). Each one displays distinctive social, morphological, and genetic features. These differences are associated with specific ecological and behavioral adaptations such as variable hunting strategies, communications dialects, and/or diet specialization (de Bruyn, et al., 2013) transmitted through culture (Foote, et al., 2016).

In the North Pacific, for example, the transient type feeds on marine mammals whereas the resident is a fish specialist (de Bruyn, et al., 2013). Foote et al. (2016) showed that genes associated with these various diets display a clear signature of selection. Therefore, these types might be rapidly diverging by exploiting and specializing to different ecological niches. Such local adaptation is probably also accompanied by reproductive isolation. These authors proposed that the divergent selection associated with the exploitation of new niches initiates
reproductive isolation. Then, once reproductive isolation reaches a certain level, additional differences can rapidly accumulate in other genes involved in reproductive isolation, which strengthens incompatibilities between hybrids by reinforcement.

Furthermore, the fact that killer whales are found in sympatry in many locations leads further support that sympatric speciation (driven by ecological divergent selection) might be occurring (Moura, et al., 2015). However, as noted by Foote & Morin (2015) it is so far impossible to exclude the alternative hypothesis that the diversification of the killer whales results primarily from divergence during allopatric phases (via selection or drift) alternating with phases of secondary contact. Under this scenario, a likely hypothesis is that the phases of allopatry initiated the ecological divergence, and then this initial divergence is strengthened in phases of sympathy due to competition for resources (Foote, et al., 2011) or reinforcement (Foote, et al., 2016). To this day, whether the initial divergence occurred in allopatry or sympathy remains unknown but a mixture of both processes is also plausible. This hypothesis is further supported by a recent genomic study analyzing a global data set of killer whales revealing that lineage divergence occurred with gene flow (Foote, et al., 2019). These results likely reflect a pattern of recurrent vicariance and admixture related to habitat contraction and expansion stimulated by the Quaternary glacial cycles.

The mechanisms presented above are probably relevant to many cetacean species. Local adaptation to different habitats can occur, reflecting dietary, foraging strategy, cultural, mating, environmental, physiological, and/or morphological specialization (Fontaine, et al., 2007; Louis, et al., 2014a; Zhou, et al., 2018). Through time, the evolution of contrasting adaptation increases the genetic divergence between populations and can induce reproductive isolation. This mechanism has been also documented in bottlenose dolphins (*Tursiops truncatus*). The combination of genetic modeling, stable isotopes, and stomach content showed that the genetic differentiation between the coastal and pelagic populations of the North-East Atlantic is driven by ecological niche specializations (Louis, et al., 2014a). It was also shown that coastal ecotypes of common bottlenose dolphins have emerged independently in different ocean basins from pelagic populations, such parallel evolution was repetitively driven by the local specialization to the coastal environment (Louis, et al., 2020a).

2. **Geography and genetics**

2.1 *The ecological determinant of cetacean distribution*

Like any organisms, cetacean distributions depend on their ecological tolerance and
biogeographic history (Brown and Lomolino, 1998). Cetaceans had to adapt to several properties of the marine environment not experienced by terrestrial mammals. These include increased buoyancy due to the high density of the water, resistance to swimming induced by viscous forces between water molecules, osmotic challenges, and the potential for substantial heat loss due to the conductive property of the water compared with the air. Despite these constraints, current biogeographic distributions of cetaceans differ markedly from one species to another (Kaschner, et al., 2011; Pompa, et al., 2011). They occupy a wide diversity of habitats, from marine to freshwater habitats, from tropical to polar regions, and from shallow coastal areas to deep oceanic waters. Some species have a very wide distribution and others are very restricted. The killer whale and the Risso’s dolphin (Grampus griseus), for example, have a cosmopolitan distribution. In contrast, some species have an extremely narrow range and are only present in a particular kind of habitat. For instance, the vaquita (Phocoena sinus) is only present in an area of fewer than 2,000 km² in the subtropical region of the Gulf of California. Between these extremes, all kinds of distribution ranges are found in other cetacean species (Ridgway and Harrison, 1998).

Sea surface temperature, presence of ice, and primary production are the main ecological features varying over oceanic distances that strongly influence the patterns of cetacean distribution (Berta, et al., 2015c). In addition, anthropogenic activities (pollution, bycatch, habitat modification, and whaling), as well as human-mediated global change are directly or indirectly affecting the distribution of many cetacean species (MacLeod, 2009; Kaschner, et al., 2011; Weir and Pierce, 2012).

Temperature is a major limiting abiotic factor because water temperature in the sea is much colder than the body temperature of the cetaceans. Given water conducts heat more than twenty times than air, heat loss is a preponderant challenge for cetacean species. To minimize heat loss, most cetaceans have large body sizes, rounded shapes (minimizing surface to volume ratios), and good insulation via a thick blubber layer. The capacity of each cetacean species to tolerate low water temperature is therefore paramount to explain their ecological preferences and distribution.

Geographically, surface ocean temperatures tend to be highest near the equator and decrease with increasing latitude. Consequently, extensive sea ice forms only in polar and subpolar zones. Seasonal alternation of freezing and melting of sea ice restricts access to the highest latitudes for most cetaceans to the summer. This phenomenon forces polar and subpolar cetacean species to move during the coldest months of the year to ice free regions (Berta, et al., 2015c).
Food availability is also critical for cetaceans. It is mainly determined by the primary production, which in turn is strongly determined by the availability of sunlight, nutrients, and currents. Therefore, rates of marine primary production vary greatly temporally and geographically (Figure 1.5; Kaschner, et al., 2006; Kaschner, et al., 2011; Berta, et al., 2015c).

Figure 1.5. General pattern of seasonal variation in marine productivity for four different marine production systems. From Berta et al. (2015c).

Currents and winds can create local hotspots of primary production by bringing cold waters rich in nutrients to the surface. Upwellings, for example, occur when strong winds push surface water offshore bringing cold nutrient-rich waters from the bottom to the photic zone. This process fertilizes the surface waters and stimulates primary production supporting higher trophic levels (e.g., consumers and top predators like cetaceans). Upwellings are among the richest ecosystems. In the Atlantic Ocean, they mostly occurring along the west coasts of Africa. Upwellings attract cetacean species (Croll, et al., 2005), such as harbor porpoises, common dolphin (*Delphinus delphis*), or bottlenose dolphin (Méndez-Fernandez, et al., 2013), and whale species.

Climate change is also impacting the distribution of many cetaceans (MacLeod, 2009; Sydeman, et al., 2015). In the Arctic, the effects of climatic change are known to be particularly strong (Shukla, et al., 2019). This has two major effects on the distribution of cetaceans. First, the overall rise of temperature reduces the extent of area covered by sea ice which frees new habitats for cetaceans. In this regard, some cetacean species are shifting their range poleward.
Second, climate change modifies the primary production and changes the prey availability for top predators like cetaceans. Consequently, many species are forced to change their distribution to adjust to the distribution of their prey or even to change their diet accordingly (Heide-Jørgensen, et al., 2011; Ramp, et al., 2015).

2.2 Biogeography

From a historical perspective, cetacean biogeography is the legacy of past environmental changes (Steeman, et al., 2009; Marx and Uhen, 2010) and variation in food availability (Marx and Uhen, 2010). At the macro-evolutionary scale, major environmental changes and tectonic events have modified the presence of corridors in the oceans, sea level, global direction of currents, primary productivity, and food distribution. All these events have directly influenced the dispersion, habitat colonization, radiation, and extinction of cetaceans. These factors determined the species diversity, abundance, and distribution (Steeman, et al., 2009; Marx and Uhen, 2010). For example, Pastene et al. (2007) suggested that the separation between the common minke whale (*Balaenoptera bonaerensis*) and the Antarctic minke whale (*Balaenoptera acutorostrata*) occurred ~5 Million years ago (Mya) and has been enhanced by a modification of ocean temperature. Climate changes, and especially the changing temperature and wind regimes altered the upwellings, leading to a redistribution of feeding areas, which in turn promoted the allopatric divergence of minke whales. Nowadays, the distribution of these minke whales coincides with regions of intense upwelling. More generally, Pastene et al. (2007) proposed the hypothesis that global warming fragments upwelling areas, which facilitate the separation between conspecific cetacean populations. Fragmented upwelling areas should support, in theory, smaller ecosystems. This effect translates into smaller isolated populations that “drift” away faster (via genetic drift), leading to an accelerated genetic divergence among conspecific populations (Pastene, et al., 2007). Change in upwelling distribution has been also suggested to explain the radiation of both mysticetes (Pastene, et al., 2007; Steeman, et al., 2009) and odontocetes (Banguera-Hinestroza, et al., 2014).

At a micro-evolutionary scale, biogeographic and phylogeographic patterns observed in cetacean are mostly the legacy of Quaternary environmental variations that caused continuous alteration of marine ecosystems (see below). Thanks to their extraordinary ability to colonize new ecological niches, cetacean rapidly diversified by adjusting to these changing conditions (Fontaine, et al., 2014; Lin, et al., 2014; Foote, et al., 2016).
2.3 The importance of food

Feeding specialization is one of the main drivers underlying current cetacean diversity in terms of species richness, size, behavior, morphology (e.g. body sizes, shapes, and color), ecology, and distribution (Slater, et al., 2010). In fact, their high disparity of body sizes, shapes, and color suggests that their radiation has an ecological basis. The radiation of modern cetaceans is characterized by the colonization of new vacant ecological niches in response to past climate change and ocean restructuration resulting in the acceleration of their diversification (Pastene, et al., 2007; Steeman, et al., 2009). The fossil record of cetaceans shows major changes in species richness related to past variation in primary production. This supports the hypothesis that food resources have been instrumental in driving cetacean diversity (Fordyce, 2018).

Mysticetes and odontocetes show specific feeding adaptations to effectively capture prey and exploit the energy. While mysticetes are mass filter feeders, odontocetes use echolocation to efficiently search for prey (Berta, et al., 2015b). These feeding differences strongly influence their life-history traits. Indeed, mysticetes can fast for several months and migrate thousands of kilometers between their tropical breeding grounds and polar feeding grounds. Therefore, the capacity to accumulate important stocks of energy is crucial to them, which is made easier when the body size is massive (Slater, et al., 2017). In contrast, odontocetes are active predators feeding on larger prey than mysticetes. This is made possible by the presence of teeth that they use to catch prey. It has been suggested that the rapid diversification of the Delphinidae during the Pliocene has been driven, at least in part, by their ability to exploit new ecological niches associated with the acquisition of echolocation as a key novelty (McGowen, et al., 2009). An example of such an ecological diversification process is still ongoing as observed in the specialization to different types of food resources in the bottlenose dolphins (Louis, et al., 2014a) and killer whales (Foote, et al., 2016).

2.4 Phylogeography

i. Quaternary glaciations shaped current phylogeographic pattern

Quaternary environmental cycles are one of the main drivers of the phylogeographic patterns currently observed within cetacean species (O'Corry Crowe, 2007). During the Quaternary, the Earth’s climate experienced deep environmental changes marked by the climatic oscillations of glacial and interglacial periods (Ruddiman, 2001). The onset of these Quaternary glacial-interglacial climate oscillations resulted in a new climatic state dominated
by colder, more variable temperatures, and large sea-level fluctuations. The fluctuation of the climate has been mediated by the Milankovitch cycles and changes in global circulation. Each cycle has greatly varied in amplitude, with a shift towards longer and deeper glaciations beginning approximately 0.9 Mya (Ruddiman, 2001). The expansion of the ice sheet in the poles during glacial periods was accompanied by a reduction in the temperature (up to -10°C) and a drop of the sea level (up to ~100m; Figure 1.6, Lisiecki and Raymo, 2007). These environmental changes impacted the connection between continents, induced a redistribution of ocean current circulation, modified marine productivity, expansion of the polar front toward low latitudes, and reduced the neritic zones leading to a significant loss of coastal marine habitats (Ruddiman, 2001).

The dynamics of Quaternary glaciations deeply modified the distribution of marine species suitable habitats which in turn affected speciation, extinction, migration, and ultimately the current biodiversity (Svenning, et al., 2015). Marine vertebrates were particularly impacted (Pimiento, et al., 2017). Species that survive to the Quaternary glaciations showed three types of responses: (1) adaptation to the new environment, (2) range shift by dispersion to colonize new habitats with more suitable conditions, and (3) distribution contraction and survival in refugia (Hewitt, 2000; Lister, 2004). Refugia represent areas relatively buffered from the environmental changes occurring during the glaciation that enable the persistence of species. These different kinds of changes left an imprint at all levels of biodiversity and are major historical determinants of a species' geographical range (Hewitt, 2000; Svenning, et al., 2015). The present communities, species, and spatial genetic structure are, in a large part, the legacy of the environmental changes occurring during the Quaternary ice ages.

ii. The last glacial maximum (LGM)

The last glaciation was the most intense and reached its maximum extent during the Last Glacial Maximum between 26 -19 thousand years ago (Kya; Clark, et al., 2009; Gowan, et al., 2021). In the Northern Hemisphere, the polar front oscillated between latitudes of 50° and 70° North between the LGM and the retreat of the ice sheet in the Holocene (~12 Kya; Figure 1.6). The sea level was about 120 m lower during the LGM compared with its present level (Schäfer-Neth and Paul, 2004). In Europe, for example, the polar front reached the north of the Bay of Biscay. These deep modifications have profoundly modified the distributions and the genetic variability of the marine fauna and flora (Hewitt, 2000). Therefore, the large extension of the ice cover forced many species to shift southwards their distribution ranges towards warmer latitudes or contract their range in specific glacial refugia to track suitable habitats (Hewitt, 2000).
iii. **Genetic consequences of the Quaternary glaciations**

The profound changes induced during the LGM are expected to leave footprints in the genomes, as the observed patterns of genetic variation could reflect the evolutionary history of species (Hewitt, 2000). For example, populations in refugia likely remained relatively small and isolated for a significant amount of time, deeply impacting patterns of genetic differentiation between populations due to genetic drift and mutations. Likewise, changes in connectivity (or gene flow) among populations after the retreat of the ice sheet and the recolonization of numerous habitats impacted the pattern of genetic variation and

---

3 https://www.futura-sciences.com/planete/definitions/climatologie-periode-glaciaire-1068/ (Date of access: 02-11-2021).
differentiation. Hence, the phylogeographic analysis of such genetic signatures allows reconstructing the evolutionary history of a taxon of interest (Hewitt, 2000). In particular, spatial analyses of genetic variation and structure can help identify the location of refugia or colonization routes taken by populations following the ice retreat. Other past demographic changes influencing genetic variation include founding events, range shift, population fragmentation, population contraction, and expansion. Genetic patterns expected under each of these scenarios have been widely reviewed (Hewitt, 2000; Arenas, et al., 2012).

As recently reported in killer whales (Foote, et al., 2019), a typical pattern is the observation of a leading-edge effect following the colonization of a new habitat freed from ice (Excoffier and Ray, 2008). A leading-edge effect is a progressive decrease of genetic diversity from a source refugia towards populations at the colonization front. This phenomenon is the consequence of the gradual loss of diversity due to genetic drift associated with successive founder effects (Excoffier and Ray, 2008). Another very common example of genetic signature frequently documented in cetaceans is the presence of star-like topology of mitochondrial DNA haplotype networks that is also interpreted as evidence of lineage expansions associated with a post-glacial recolonization of a habitat (Rosel, et al., 1999b; O'Corry Crowe, et al., 2003; Louis, et al., 2020b).

In its early days, research in phylogeography was primarily descriptive. One of the biggest issues was that it provided no effective way to assess how likely an evolutionary history might explain the observed patterns of genetic diversity (Knowles, 2009). As an answer, the rapid emergence of simulation of population genetic data using the coalescent theory opened up new avenues for more rigorous statistical ways to test explicitly different evolutionary scenarios. In turn, this accelerated the development of a plethora of new statistical methods applied to phylogeography (Knowles, 2009; Nielsen and Beaumont, 2009; Beaumont, et al., 2010). One of these methods, the Approximate Bayesian Computation (ABC), is discussed in further detail in Box 1. ABC methods were applied in many studies related to cetaceans, for example, to distinguish between different demographic scenarios (Fontaine, et al., 2012), to establish relationship among populations (Louis, et al., 2014a), infer effective population size (Fontaine, et al., 2012), detect admixture (Louis, et al., 2014a) or estimate divergence time among populations (Fontaine, et al., 2014; Rivera-León, et al., 2019).
Box 1: Some useful tools in phylogeography.

Tremendous progress was made in phylogeography with the increasing application of the **coalescent theory** allowing an efficient simulation-based inference approach leading to the emergence of the field of statistical phylogeography (Knowles, 2009). This new field enabled phylogeography to move from a descriptive approach simply based on descriptive inferences made from phylogenetic trees and descriptive statistics to a more statistically robust simulation-based framework relying on hypothesis testing. This field largely depends on simulations of genetic data under various models to build statistical inferences and evaluate the support for alternative evolutionary hypotheses. Different kinds of statistical toolboxes, mostly likelihood- or Bayesian-based, can then be applied to compare the probability of different competing models of interests and estimate their parameter values (Beaumont and Rannala, 2004). The first common type of method assumes simple models and estimates 1) the likelihood of the models and 2) the parameter values that fit best the data (Drummond, et al., 2005; Beerli and Palczewski, 2010; Hey, 2010) using Markov chain Monte Carlo techniques (Beaumont and Rannala, 2004). Making use of all the data available, this kind of approach is particularly performant to study simple models. However, the calculation of the likelihood can become mathematically intractable for complex evolutionary models (Marjoram, et al., 2003). To bypass the exact likelihood calculations, a second type of approach called approximate Bayesian computation (ABC) is now commonly used (Lopes and Beaumont, 2010). ABC methodology relies on the usage of summary statistics and simulations. In a nutshell, summary statistics are calculated from simulated data sets under alternative models which are compared with the summary statistics calculated from the observed data. Using a unified statistical framework, ABC related methods can compare several evolutionary models and provide an estimation of the model's parameters. A major advantage of ABC methodology over the full-likelihood methods is that it can accommodate a large spectrum of complex evolutionary models (Lopes and Beaumont, 2010). However, ABC comes with some limitations too (Sunnåker, et al., 2013). For example, because it uses summary statistics, a part of the information embedded in the genetic sequence can be lost (Robert, et al., 2011). Moreover, ABC is also computationally intensive (Sunnåker, et al., 2013), but the recent application of machine learning techniques in ABC can alleviate some of these issues (Pudlo, et al., 2016; Raynal, et al., 2017).

Additional insights into phylogeographic studies have also come from ecological niche modeling (ENM) techniques (Alvarado-Serrano and Knowles, 2014). ENMs are a class of correlative models that combine occurrence and environmental data to make predictions on the ecological conditions suitable for a species and thus infer the relative species suitability to different habitats. ENMs can inform about the past, present, and future distribution of a species based on climatic and environmental data coming from present records as well as from paleomodels or forecasting models. These ENMs can improve the comprehension of the process shaping genetic variation across landscapes. They can be used to identify the potential location of past populations, localize potential glacial refugia, guide evolutionary model’s formulation, test or validate biogeographic hypothesis inferred from genetic data, or predicting how distributional shifts might leave a particular genetic signature (Knowles, et al., 2007). The inclusion of ENM in phylogeographic studies was applied to several cetacean species (Foote, et al., 2013; Nykänen, et al., 2019; Louis, et al., 2020b). For example, combining phylogeographic approaches with ENMs, Foote et al. (2013) showed that following the ice retreat during the Pleistocene–Holocene transition, the suitable habitat of the bowhead whales (*Balaena mysticetus*) increased threefold. As an answer to these changes, bowhead whales shifted their distribution northward. Their model also predicted a rapid diminution of the suitable habitat for bowhead whales by 2100.
2.5 *Spatial genetic structure in cetacean populations*

In the literature, there are many biological definitions of a population (Waples and Gaggiotti, 2006), however in the context of this thesis, I will refer to the evolutionary definition that corresponds to a group of individuals of the same species living close enough together and that can potentially mate with any member of the other sex (Waples and Gaggiotti, 2006). As populations can be geographically structured into demes more or less isolated, and connected by the individual dispersal, the number of individuals connecting discrete populations defines the concept of population connectivity. In an evolutionary paradigm, the exchange of individuals between populations is associated with gene flow, i.e. the successful transfer of genetic material from one population to the other. Likewise, migration is often defined as the movement of individuals (or gametes) from one population to another, but migration may or may not result in gene flow. Unless otherwise stated, I will adopt in this thesis the population genetic definition of migration in which migration implies gene flow. Finally, another important concept is the effective number of migrants per generation which represents the number of migrants that disperse every generation and contribute to the genetic makeup of the receiving population (Hudson, et al., 1992).

The population geographic distribution and habitat structure impact both the connectivity among populations and thus their level of genetic differentiation (Manel and Holderegger, 2013). In landscape genetics, a common practice is to consider genetic structure together with the structure of suitable habitats to reveal the impact of environmental variation in shaping the genetic connectivity among populations or demes. Habitat structure can be fragmented or continuous and homogeneous over vast areas. This structure determines the size and the connectivity of populations and thus determines the respective influence of genetic drift and migration. Large favorable habitats with highly connected populations can support abundant populations. In such a context, the predominance of gene flow will mitigate the effects of genetic drift, thus homogenize allelic frequencies and limit the differentiation between populations. In contrast, fragmented habitats will only allow small and isolated populations to survive. This will promote local allele fixation by genetic drift reducing genetic diversity within populations and increased genetic differentiation between populations (Bohonak, 1999).

When populations are separated by a barrier to gene flow, populations evolve independently and allele frequencies will diverge over time due to genetic drift (and sometimes also due to selection) via a process called lineage sorting. This process will increase genetic differentiation between populations. Over time, the phylogenetic relationships between lineages present in the populations will first be polyphyletic, then paraphyletic, and finally
monophyletic (Neigel and Avise, 1986). In the last step, after a sufficient amount of time, when no alleles are shared between populations, all lineages within the same population become more closely related to each other than they are to individuals of separate populations. The speed at which lineage sorting occurs is highly dependent on the effective population size \((N_e)\) of the diverging populations. Interestingly, Hudson and Coyne (2002) calculated that autosomal loci will stop sharing alleles between two lineages after 9 to 12 \(N_e\) generations on average under a simple isolation model.

The dispersal capacity is another fundamental characteristic that directly affects the rate of gene flow among spatially distributed populations. Dispersal ability is expected to enhance gene flow and thus, to impede genetic differentiation between populations (Bohonak, 1999). In this regard, a negative relationship between dispersal capacity and genetic differentiation has been documented for several organisms such as reef fishes (Riginos, et al., 2014) or birds (Belliure, et al., 2000). As all species have limited dispersal capabilities, genetic differentiation between individuals or populations is expected to increase with the geographical distance separating them generating a pattern of isolation by distance (or IBD; Rousset, 1997). Under IBD, a positive correlation is expected between genetic and geographic distance among demes. Conversely to IBD, discontinuous relationships between genetic and geographic distance reveal jumps in genetic distances which are possibly indicative of barriers to gene flow or a secondary contact between previously isolated populations.

The behavior of a species can also promote or reduce the spatial genetic structure. For example, dispersal patterns in nature are often sex-dependent (Perrin and Mazalov, 1999). Sex-biased dispersal represents the situation where individuals of one sex disperse more than individuals from the other sex (Greenwood, 1980). This can strongly influence the genetic structure of populations (Prugnolle and de Meeus, 2002). It is especially the case when one sex is philopatric, i.e. stay or return to its natal site to breed while individuals from the other sex disperse. Philopatry is expected to amplify genetic differentiation among populations (Prugnolle and de Meeus, 2002). For mammals, in general, females are more philopatric and thus gene flow is often male mediated (Greenwood, 1980). Philopatry and sex-biased dispersion have been documented for several cetacean species (Escorza-Treviño and Dizon, 2000; Möller and Beheregaray, 2004). In cetacean, female philopatry is presumably mediated by the cultural transmission from mothers to offspring, which is facilitated by the relatively long period of maternal care in these species (O’Corry Crowe, 2007). Female philopatry can be revealed by comparing bi-parentally (e.g. nuclear markers like microsatellites or SNPs) and maternally (e.g., mtDNA) inherited genetic markers (Box 2) by exploiting the fact that under female
Box 2: Applications of molecular markers in cetacean phylogeography.

A molecular marker is a specific polymorphic locus from the whole genome used to detect and compare the genetic variations among individuals, populations, or species (Allendorf, 2017). The first studies using molecular markers on marine mammals were published in the late 70's with the development of allozymes. These studies assessed the level of genetic diversity and differentiation between populations (Bonnell and Selander, 1974).

Introduced in the 80's, mtDNA has been among the most popular genetic markers. Its unique properties (i.e. multicopy, circular, cytoplasmic DNA, high mutation rate, maternal inheritance in mammals) made it a powerful marker to investigate ecological and evolutionary questions. Particularly, the cytochrome-b and the D-loop were the most commonly used markers and they provided valuable information about the evolutionary history of several cetacean species (Rosel, et al., 1995; Rosel, et al., 1999b; Hoelzel, et al., 2002). Forty years after its introduction, mtDNA is still one of the most frequently employed molecular markers in the study of cetacean (Rosel, et al., 2017a; Rosel, et al., 2017b). Since the ~2010, the sequencing of complete mitochondrial genomes offered a new alternative to single mtDNA genes, at a reasonable and relatively low cost. Complete mitogenomes represent a richer source of information and allow more robust phylogenetic inference (Morin, et al., 2010; Vilstrup, et al., 2011; Louis, et al., 2020b).

However, mtDNA has also important limitations (Galtier, et al., 2009). For example, as a non-recombining molecule, it provides a single sampling point to describe the evolutionary and genealogical history of populations or species. Furthermore, mtDNA is maternally inherited so it provides a partial picture of the species evolutionary history. It is therefore customary to complement the genetic inference made from mtDNA with additional markers such as microsatellites to provide a more accurate picture of the evolutionary history of a species (Garrick, et al., 2015). Microsatellites are bi-parentally inherited short stretches of repeated DNA and show very variability (mutation rate ~10^{-3} to 10^{-4} mutation per generation). This high level of polymorphism enabled to easily identify individuals and provided the opportunity to address new questions related to the conservation genetics and the biology of cetaceans. The last decades saw a burgeoning number of phylogeographic studies combining both nuclear and mitochondrial DNA markers (Rosel, et al., 1999a; Fontaine, et al., 2014; Morin, et al., 2015).

The recent advent of next-generation sequencing approaches has enabled the identification of a large number of independent single nucleotide polymorphisms (SNPs) and has greatly improved the resolution of phylogeographic studies. New methods for assessing large numbers of polymorphic sites, such as Restriction site Associated DNA Sequencing (Davey and Blaxter, 2011) or even whole genome shotgun sequencing, have largely gained in popularity providing good statistical power for genetic assessment of populations while providing a better representation of the genomic variation.
philopatry gene flow is expected to be principally male mediated (Prugnolle and de Meeus, 2002). Therefore, genetic differentiation recorded at mtDNA should be much higher than at nuclear loci. Contrasted levels of genetic differentiation between mtDNA and microsatellites have been reported in numerous cetacean species and taken as evidence of female philopatry (Rosel, et al., 1999a; Escorza-Treviño and Dizon, 2000; Escorza-Treviño, et al., 2004).

Cetaceans have a very high capacity for long-range movements, so one would expect that populations of cetaceans would display shallow spatial genetic structure. This is the case for some species like the Gray's beaked whale (*Mesoplodon grayi*; Thompson, et al., 2015). However, cetaceans can also be strongly genetically structured even at relatively small geographical scales (Chivers, et al., 2002; O'Corry Crowe, 2007). Various types of factors are known to influence their spatial genetic structure. Indeed, biotic factors (habitat specialization and resource availability), behavioral factors (sex-biased dispersal, philopatry, kin association, and culture), and anthropic factors (human-mediated habitat disturbance) can generate strong population subdivision at a short geographical scale. Although our understanding is still incomplete, abiotic factors (e.g., temperature, bathymetry, or salinity) can directly reduce gene flow between cetacean populations. These factors are also known to influence indirectly the genetic structure of cetaceans by conditioning the availability of their food (Fontaine, et al., 2007; Amaral, et al., 2012).

### 2.6 Conservation genetics

We are currently witnessing the sixth mass extinction characterized by the rapid loss of many species as a result of human activities and climate change (Ozanne, et al., 1997). In this respect, marine mammals, and especially cetacean species have been long considered as charismatic flagship species highlighting this crisis with direct hunting, by-catch, and habitat destruction (Albert, et al., 2018). The goal of conservation biology is to protect and maintain biodiversity at all levels of organization (from genes to ecosystems). Different evolutionary
processes determine or influence the diversity at all these hierarchical levels and the field of conservation aims not just to preserve them, but also the evolutionary processes that created and maintain them (Hendry, et al., 2011). Within conservation biology, the sub-field of conservation genetics focuses on the application of genetic tools and methods coming from evolutionary disciplines such as molecular ecology, phylogeny, phylogeography, or population genetics to address conservation problems (Hays and Fagan, 2016). Conservation genetics studies can help designing strategies to protect cetacean species impacted by human activities (Morin and Dizon, 2018). Indeed, genetic data can be used to address fundamental biological questions related to population structure, connectivity, degree of isolation/fragmentation, population size, historical variations, population trends, assess the level of inbreeding, reconstruct pedigrees, or study adaptation which in turn provides invaluable information to guide effective conservation plans (Frankham, et al., 2019). All this information is necessary to determine population status and assess whether demographic trends and the threats are sustainable or not for species survival. The type of genetic data and the study design required to obtain such information depends on the particular conservation goals sought (Moritz, 1994a; Moritz, 2002). However, gathering biological data on cetaceans can be logistically difficult and expensive. The advantage provided by indirect genetic inferences over other more direct approaches (e.g., field surveys, tagging) is that DNA can be easily collected from non-invasive sampling strategy or even post-mortem. Especially studying demography or individual movements are key information for conservation (Van Dyke, 2008). Classical direct approaches such as mark–release–recapture or (radio-/GPS-) telemetry are generally logistically difficult to implement and expensive. They also suffer from spatial observation bias due to insufficiently large study areas (Vandewoestijne, et al., 2004) leading to an incorrect estimation of the dispersal pattern (Baguette, 2003). This can compromise the elaboration of effective conservation measures. Although they do not replace direct approaches, indirect genetic methods provide tools to estimate several demographic parameters important to conservation and bypass certain limitations of the direct approaches (Frankham, et al., 2019).

One of the primary goals of conservation genetics is to define conservation units on which efforts are focused. In the case of threatened cetacean species, inadequate conservation units can nullify or impede the efforts made to protect a species or a threatened group. It is therefore paramount to ensure that the group defined are relevant for each specific conservation goal (Taylor and Dizon, 1999). From an economical perspective, defining appropriate units is also crucial because quotas on cetacean harvest or incidental catches are based on these definitions (Taylor and Dizon, 1999; Morin and Dizon, 2018). Generally, management efforts
target the population level to define conservation units, nevertheless, the precise definition of these populations is still controversial (Taylor and Dizon, 1999; Moritz, 2002; Thomassen, et al., 2011). Nowadays, defining populations based on direct approaches is difficult, especially for cetacean species. Therefore, genetic tools and techniques remain essential to define cetacean populations and understand their spatial structure (Martien, et al., 2019). In exploited species, an accurate understanding of such population structure is essential for estimating abundance and determining whether catches numbers are sustainable or not. For threatened species, such knowledge is critical for assessing population trends and establishing conservation policies (Moritz, 1994a; Moritz, 2002; Morin and Dizon, 2018).

There are two principal types of conservation units recognized by management agencies: Evolutionary Significant Units (ESUs) and Managements units (MUs; also called demographically independent population). An ESU forms a set of demographically distinct populations that share a common long-term history but are distinct from other ESUs (Moritz, 1994b). There are debates about what counts exactly as "evolutionary significant" and several definitions of what constitutes an ESU exists (reviewed in Funk, et al., 2012). A fundamental aspect of ESUs is that it implies some substantial level of genetic differentiation, low gene flow, and some degree of local adaptation (i.e. a certain level of ecological and genetic distinctiveness). In addition, some definitions assume (at least in theory) a substantial level of reproductive isolation (Waples, 1991; Funk, et al., 2012). The criterion proposed by Moritz to concretely diagnose ESUs is the reciprocal monophyly of mitochondrial haplotypes and the presence of significant differences in allelic frequencies at nuclear loci (Moritz, 1994b). The main advantage of this criterion is that it offers a simple way to define conservation units. But, this criterion is often considered highly debatable (Moritz, 2002). MtDNA is maternally inherited so it provides a partial picture of the species evolutionary history. In addition, mitochondrial reciprocal monophyly is not necessarily synonymous with reproductive isolation. The reverse is also true as incomplete lineage sorting, introgression, selection, and other biological processes can induce other topologies than monophyly while lineages are substantially reproductively isolated (Galtier, et al., 2009). Finally, ESUs are supposed to be locally adapted, but the molecular markers (Box 2) used to identify such populations are often assumed to evolve nearly neutrally and thus do not necessarily reflect adaptive divergence between populations (McKay and Latta, 2002). In cetaceans, conservation units have been defined almost exclusively based on genetic evidence (Dalebout, et al., 2007; Morin, et al., 2016).

Contrary to ESUs, MUs do not require any degree of local adaptation/reproductive
isolation or mitochondrial monophyly. They just require demographic independence. This implies, that if gene flow is sufficiently reduced between populations to generate significant differences in frequencies using mtDNA or nuclear markers, they can be considered as an MU because they are functionally independent (Moritz, 1994a). So, MUs refer to populations that are demographically distinct from one another. While MUs are often seen as a short-term management objective, ESUs correspond to conservation goals in the longer term allowing the preservation of evolutionary potential (Moritz, 2002). Often the phylogenetic method employed to identify ESUs will fail to detect MUs. The application of other clustering approaches (Pritchard, et al., 2000; Jombart, et al., 2010) is consequently necessary to detect significant differences in allelic frequencies and identify MUs. However, setting a genetic limit at which one or more populations are considered as demographically independent is necessarily arbitrary (and thus contentious). Also, when the split between populations is very recent, molecular markers might not have the statistical power to detect demographic independence. Power analyses (Ryman and Palm, 2006) and genetic simulations (Peng and Amos, 2008) can be used to test whether the molecular markers used have sufficient power to discriminate the populations or not (see chapter 5).

Besides identifying conservation units and resolving taxonomics uncertainties, the field of conservation genetics aims at contributing to our understanding of the challenges facing small and fragmented cetacean populations (Gilpin and Soulé, 1986). For example, the loss of genetic diversity associated with small or declining populations results in a cascade of negative genetic and demographic effects that can lead to the so-called extinction vortex that can precipitate populations towards extinction (Blomqvist, et al., 2010). These effects are exacerbated in fragmented populations (Frankham, et al., 2019). In the case of cetacean, Vachon et al. (2018) showed that nuclear genetic diversity correlated positively with population abundance while mitochondrial genetic diversity correlated positively with species range. These results suggest that an accurate estimation of the genetic diversity can be a useful tool to shed some light on issues related to habitat fragmentation or population decline. Therefore, cetacean conservation policies should aim to preserve genetic diversity while minimizing population fragmentation (Frankham, et al., 2019). The vaquita or the Maui’s dolphin (Cephalorhynchus hectori maui) have extremely low population sizes (<60 individuals) and are directly concerned by the issues related to small population sizes. Likewise, cetacean species inhabiting the freshwater environment such as the Yangtze finless porpoises or the Indus River Dolphin (Platanista gangetica minor) are also seeing their habitat increasingly destroyed or fragmented due to human activities and are currently highly threatened (Reeves, 2003).
A fundamental principle in conservation genetics relates to the fact higher genetic diversity can lead to an increase in fitness and ultimately can help the survival of a species (Gilpin and Soulé, 1986). Therefore, level genetic diversity is a proxy of the extinction risk of a population/species (Laikre, 2010). There is however evidence that such a link is not necessarily straightforward (Robinson, et al., 2019; Teixeira and Huber, 2021). Furthermore, genetic diversity reflects the evolutionary potential of a population, i.e. determines the capacity of a population to adapt to changing conditions over the long term (Steeves, et al., 2017). In the context of the current climatic change, populations with a low genetic diversity might fail to adapt to current and future changes. As previously discussed, cetaceans are particularly vulnerable to the current global change. Accurately quantifying the genetic diversity of cetacean species can be useful to help mitigate the threats they might face in the future. In particular, the combination of recent genomics tools together with other techniques such as ENMs (see Box 1) and time series can be instrumental to understand how cetacean are adapting to the modification of their environment, dissect the molecular basis of evolution under global change, and forecast how species will respond to future changes (Razgour, et al., 2019; Waldvogel, et al., 2020).

3. **The porpoise family**

3.1 **Taxonomy and morphology**

This thesis focuses on the porpoise family (Phocoenidae; Figure 1.7 and 1.8). Porpoises belong to the infraorder of Cetacea, which to one of the ten families of the parvorder of Odontoceti (toothed whales). Porpoises, dolphins, and monodontids form the super-family of Delphinoidea that likely descended from the Kentriodontidae, an extinct family of odontocetes (Racicot, 2018). Phocoenidae are most closely related to the Monodontidae (Figure 1.7), and these two families constitute a separate clade that split from the Delphinidae (Figure 1.7) within the last 18 million years and diverged from one another within the last 15 million years (McGowen, et al., 2009; Steeman, et al., 2009; McGowen, et al., 2019).

Morphologically, porpoises are distinguished from other cetaceans by diagnostic features such as their small size (≤ 2.5 meters), their robust body shape, their lack of external rostrum, the protuberance of their premaxillae, and the spatulated shape of their teeth (Read, 2009). Furthermore, the skulls and postcranial skeletons of porpoises exhibit a pronounced paedomorphosis, which means that the adult porpoises display similar features to those of juvenile individuals of close cetacean species (Read, 2009). These features separate them from
other Delphinoidea that are generally larger and tend to have long rostrum and cone-shaped teeth.

Figure 1.7. Cetacean phylogeny according to Mcgowen et al. (2011). The Phocoenidae are highlighted by a red square. Figure modified from Mcgowen et al. (2011).

The earliest known porpoise is Salumiphocaena stocktoni which was discovered in the Pacific Ocean and dated to approximately 11 million years ago during the Late Miocene (Barnes, 1985). However, so far, it remains controversial whether their origin is tropical (Gaskin, 1982) or temperate (Barnes, 1985). In total, at least 18 species of porpoises have been described, including 12 that are currently extinct (Murakami, et al., 2012b; Marx, et al., 2016).

According to several genetic studies, all the extant porpoises have originated within the past few million years (Rosel, et al., 1995; McGowen, et al., 2009; Steeman, et al., 2009; McGowen, et al., 2019). They officially include seven species (Figure 1.8): the harbor porpoise, the spectacled porpoise (Phocoena. dioptrica), the Burmeister’s porpoise (Phocoena
spinipinnis), the vaquita, the Dall’s porpoise (*Phocoenoides dalli*), and two species of finless porpoises (*Neophocaena* sp.).

**Figure 1.8.** Species of the porpoise family. **a**) Indo-Pacific finless porpoise (*Neophocoena phocaenoides*); **b**) Narrow-ridged finless porpoise (*Neophocoena asiaeorientalis*); **c**) Vaquita (*Phocoena sinus*); **d**) Spectacled porpoise (*Phocoena dioptrica*); **e**) Burmeister's porpoise (*Phocoena spinipinnis*); **f**) Dall's porpoise (*Phocoenoides dalli*); **g**) Harbor porpoise (*Phocoena phocoena*). Illustrations by Uko Gorter⁴.

Based on morphological and genetic evidence (Wang, et al., 2008; Lin, et al., 2014), finless porpoises are indeed recognized as two distinct biological species: the Indo-Pacific finless porpoise (*Neophocoena phocaenoides*), and the narrow-ridged finless porpoise (*Neophocoena asiaeorientalis*). It is worth noting that a recent genomic study recommends separating the narrow-ridged finless porpoise into two species (Zhou, et al., 2018).

Interestingly, many porpoise species are also divided into distinct subspecies, ecotypes, and/or morphotypes that are genetically and/or morphologically different (Gaskin, 1984; Escorza-Treviño, et al., 2004; Rosa, et al., 2005). The taxonomy of the porpoise family has a complex, changing, and confusing history. While morphological and genetic evidence often place the finless porpoise as the most basal member of the family, the remaining taxonomic relationships between porpoises are still subject to debate. Phylogenies based on morphological data generally support the grouping of the Dall’s and the spectacled porpoise (Barnes, 1985;

Conversely, most of the phylogenies based on molecular data suggest grouping the Dall’s and harbor porpoises as a monophyletic group, and the vaquita, Burmeister’s, and spectacled porpoises as another monophyletic group (Figure 1.9; see Rosel et al., 1995; McGowen et al., 2009). However, the observation of hybrids between Dall’s and harbor porpoises has reinforced the hypothesis supporting the monophyletic grouping of these two species (Willis et al., 2004).

**Figure 1.9.** Species Phylogenetic relationships among species of porpoises according to Berta et al. (2015c) and the associated current IUCN (International Union for Conservation of Nature) status at the subspecies level. LC: Least Concern; NT: Near Threatened; VU: Vulnerable; EN: Endangered; CR: Critically Endangered. Figure modified from Berta et al. (2015c).

### 3.2 Geographical distribution

Porpoises inhabit a broad range of climatic zones displaying a high disparity in terms of the types of habitat occupied, and in terms of the spread of their geographical distribution (Teilmann and Sveegaard, 2019). Indeed, porpoises successfully adapted to cold (harbor and spectacled porpoises), temperate (harbor porpoise, Dall’s porpoise, vaquita, and Burmeister’s), and tropical (finless porpoises) waters (Figure 1.10). The two finless porpoise species are coastal species. While the Indo-Pacific finless porpoise only occurs continually from the Indian Ocean to the warm waters of the Persian Gulf, the narrow-ridged finless porpoise is found in the Sea of Japan, Yellow Sea, and even the freshwater of the Yangtze River. The remaining porpoise species display a typical **antitropical distribution** (Figure 1.10). The harbor porpoise
is the most widely distributed porpoise found mainly in the coastal regions of the Black Sea, the Baltic Sea, the North Atlantic, and the North Pacific. The Dall’s porpoise has also a large distribution and inhabits coastal and pelagic waters of the North Pacific. On the contrary, the vaquita has the most restricted geographical distribution of all the cetaceans, as it is restricted to the upper part of the Gulf of California (Teilmann and Sveegaard, 2019). The Burmeister’s porpoise occupies the coastal water of South America, ranging from the Peruvian waters in the Pacific to the south of Brazil in the Atlantic. Finally, the spectacled porpoise occurs in the extreme south waters of the American continent and all-around Antarctica in the Southern Ocean. However, due to the scarcity of knowledge about this species, the precise range of the spectacled porpoise remains unknown (Natalie and Goodall, 2009; Pimper, et al., 2012). Due to its wide distribution, it is highly plausible that separate populations exist in the Southern Ocean.

Interestingly, the distribution of some porpoise species overlaps, such as in the coastal area of the North Pacific for the Dall's and harbor porpoises, and theoretically in the southern region of South America for the Burmeister's and spectacled porpoises. These overlaps could be at the origin of hybridization between the overlapping species, as reported for example between Dall's and harbor porpoises (Willis, et al., 2004).

3.3 Conservation

The coastal habitats occupied by many porpoises and their feeding preferences bring them in close proximity to areas of intense human activities such as oil and mineral exploitation,
dredging, shipping, wind farming, pollution, and fishery (Reeves, 2003). Populations from all porpoise species are severely impacted by entanglement in gillnet fisheries that kills several thousands of individuals each year (Jefferson and Curry, 1994). Besides, habitat degradations/destructions, global changes, and direct hunting also impact many porpoise species (Read, 2009). However, the nature of human activity exerting conservation threats varies from species to species, as does the conservation status of each species (Figure 1.9).

Classified as Critically Endangered by the IUCN red list of threatened species since 1996 (Rojas-Brach and Taylor, 2017), the vaquita represents the world’s most endangered marine mammals with less than 30 individuals remaining in 2017 (Morell, 2017). Bycatch represents the most serious threat for this species due to the illegal fishing of the Totoaba fish (Totoaba macdonaldi) and to a lesser extent to legal fisheries for shrimps and finfishes (Rojas-Brach, et al., 2018). The gillnets used to catch the Totoaba are the most lethal to the vaquita because their mesh size is about the size of the head of the vaquita (which facilitates entanglement; Rojas-Brach, et al., 2018). Pollution and habitat destruction also contribute to the species collapse. Discovered in 1958 (Norris and McFarland, 1958), the vaquita lost more than 95% of its population size since 2011 (37% decline just from 2011 to 2013; Rojas-Brach, et al., 2018). Since 1996, management and conservation actions have increased but these have not been effectively implemented. In 2015, following the recommendation of the International Committee for the Recovery of the Vaquita (CIRVA), the Mexican Government implemented a 2-year fishing ban in the zone covering the full range of the vaquita. But as stressed by the CIRVA, unless radical measures to protect the vaquita are taken, the species will soon become extinct (Morell, 2017). In this context, an emergency action plan to protect the remaining vaquita in a semi-natural environment, called VaquitaCPR, was implemented in October 2017 but failed (Pennisi, 2017). Currently, a total cessation of all gillnet fishing in the entire habitat of the vaquita represents the only remaining option to save this species (Rojas-Brach, et al., 2019). Despite the extremely low population census size of the vaquita (Morell, 2017), such drastic measures are expected to increase its chance of recovery, as it shows no apparent sign of inbreeding depression, and the calves seem robust and healthy (Rojas-Brach, et al., 2019).

A similar concern has been raised for all known finless porpoise lineages. Classified as Critically Endangered by the IUCN red list of threatened species, the Yangtze finless porpoise (Neophocaena asiaeorientalis asiaeorientalis) is threatened by dredging, pollution, habitat fragmentation, and the intense vessel traffic occurring in the Yangtze River (Wang, 2009; Morin, et al., 2020b). Demographic modeling showed that, if the reported average decline rate of 13.73% per year from 2006 to 2012 is maintained, the Yangtze finless porpoise will
disappear in the next 50 years (Huang, et al., 2017). However, thanks to the implementation of strong ex-situ and in-situ conservation measures and fishing bans (Wang, 2009; Huang, et al., 2020), the species is showing the first promising signs of recovery (Huang, et al., 2020). Likewise, the Indo-Pacific and the East Asian finless porpoises are both at risk due to recurrent entanglement in a variety of types of fishing gear, serious problems of habitat loss and degradation due to mangrove deforestation and shrimp farming, vessel strikes, and environmental contamination and are classified by the IUCN red list of threatened species as Endangered and Vulnerable, respectively (Wang and Reeves, 2017a; Wang and Reeves, 2017b). For example, the bycatch rate of East Asian finless porpoises in the Japanese water might represent 8% of the total population (Jefferson and Hung, 2004).

Despite their wide distribution and abundance, harbor porpoise has been severely affected by anthropogenic activities. First, it has been largely hunted until the early twentieth century for its meat and blubber. For example, the sub-species isolated in the Black Sea has been brought close to the extinctions until a total ban on cetacean hunting has been implemented in 1990 (Birkun Jr and Frantzis, 2008). For that reason, the Black Sea sub-species P. p. relicta is still classified as Endangered in the IUCN red list (Birkun Jr and Frantzis, 2008). Secondly, as harbor porpoise feeds mainly on commercial fishes such as herring and cod, their diet preferences and coastal distribution make them particularly inclined to interact with humans. Harbor porpoises have been bycaught in a variety of fishing gears but gillnets with large mesh size set for species such as cod, hake, turbot, and monkfish are responsible for the majority of harbor porpoise bycatch (IMR-NAMMCO, 2019). Consequently, harbor porpoise represents the most frequently caught cetacean in gillnets (Hammond, et al., 2002; Read, et al., 2006) with bycatch rates locally exceeding 6% of the total population abundance in the North Atlantic (Tregenza, 1997; IMR-NAMMCO, 2019). Many other threats impair the survival of harbor porpoises, such as degradation of their habitats, the accumulation of heavy metals (Das, et al., 2004), the intense fishing of porpoises’ main prey, and the current global change (IMR-NAMMCO, 2019). As a consequence, due to the growing concerns about its status, the harbor porpoise is subject to several international agreements for its conservation (e.g. ASCOBANS, IUCN, and Habitat – Natura 2000). However, as the harbor porpoise is still abundant in most of its range, immediate concerns are therefore not about the short-term viability of the species (IMR-NAMMCO, 2019), but more about its viability in some areas where they are now highly threatened (e.g. in the waters around Iberia (Pierce, et al., 2020), the Baltic Sea (Hammond, et al., 2008), and the Black Sea (Birkun Jr and Frantzis, 2008).

Dall’s porpoises are also abundant in the North Pacific with a total population size
estimated at over 1.2 million individuals in 1993 (Buckland, 1993). Although a reassessment of IUCN status for this species in 2017 has set it to Least Concern, several sources of concerns exist (Jefferson and Braulik, 2018). For example, as they are intensively hand-harpoon hunted in Japan, over 17,000 animals are still killed each year in Japan and sold in the Japanese market (Jefferson and Braulik, 2018). The long-term effects of this intense exploitation are poorly understood. The other threats to Dall's porpoises include high incidental catches in salmon and squid fisheries that take several thousand individuals every year (Jefferson and Braulik, 2018) and environmental contaminants by organochlorines (Subramanian, et al., 1988; Buckland, 1993). As for direct catches, the impact of the incidental catches on the Dall's porpoises has not been assessed. Overall, the IUCN considers the population trends of the Dall's porpoises as unknown (Jefferson and Braulik, 2018). In fact, in general, the knowledge about this species is still limited.

Due to the difficulty to detect the Burmeister's porpoises in the wild (Ridgway and Harrison, 1998), no comprehensive information is available on the abundance of this species. Localized surveys in Southern Chile and Argentina suggest that Burmeister's porpoises occur in small groups and at relatively low densities (< 2.4 animals per 100 km; Félix, et al., 2018). The Burmeister's porpoises are hunted in the Peruvian and Chilean waters for their meat, but the magnitude and effects of these catches on natural populations remain totally unknown (Félix, et al., 2018). Burmeister's porpoises are also regularly incidentally caught in gillnet fisheries in Peru and Chile, killing thousands of individuals each year (Mangel, et al., 2010). Due to the growing impact of fisheries and the direct harvest for human consumption, the Burmeister's porpoises have been recently classified as Nearly Threatened by the IUCN red list of threatened species even if the population trend status is set to unknown (Félix, et al., 2018).

The spectacled porpoise is by far the most mysterious porpoise species. Currently, the general knowledge about this species is almost nonexistent. In the past, this species was probably harpooned by Tierra del Fuego natives and by whalers. But there are no estimates of abundance available for this species. Due to its Antarctic circumpolar distribution (Natalie and Goodall, 2009), the spectacled porpoise is mostly known from exceptional cases of strandings and bycatches in commercial fisheries (Dellabianca, et al., 2018). The species is believed to be one of the least impacted porpoise species by human activities due to its remote distribution from the mainland. Spectacled porpoises are thus currently considered as of Least Concern from a conservation point of view by the IUCN. Considering the growing exploitation of the Southern Ocean (Kock, 2007), commercial fisheries can lead to an increased number of harmful interactions in the future. However, the current total lack of data and knowledge regarding
population structure and demographic trends about the spectacled porpoise is an obvious obstacle to define any future conservation plan.

One of the goals of this thesis is to provide a better understanding of the biology of some of these porpoises using indirect genetic approaches. Genetic approaches can provide complementary information to classical direct approaches to study demography (see the section on conservation genetics). Porpoises are protected and elusive species, thus performing experiments and monitoring in the field or semi-controlled conditions individual porpoises can be extremely delicate. There are also many difficulties related to the sampling of this type of species because spatial scales are often important and the logistics behind the marine fieldwork is often considerable (Berta, et al., 2015d). Therefore, studies of the structure and demography of these small cetaceans are largely dependent on uncontrolled sampling, coming principally from strandings and incidental catches (Berta, et al., 2015d). Although not ideal, genetic approaches can still make use of this type of sampling, using for example carcass drift prediction models (Peltier, et al., 2012; Louis, et al., 2014a) and make reliable inferences of different parameters useful to devise conservation actions (Berta, et al., 2015d). That is why, for the best-known species, i.e. the finless and the harbor porpoises, genetic approaches are instrumental to fill the gap of direct approaches by giving concrete information on the population structure, migration pattern, and demographic trends (IMR-NAMMCO, 2019; Huang, et al., 2020). For example, during the Joint IMR/NAMMCO International Workshop on the Status of Harbour Porpoises in the North Atlantic (IMR-NAMMCO, 2019) it was discussed that genetic analyses should be applied to fill the knowledge gap on the population structure of harbor porpoise from East Greenland, Faroe Islands and Northwest Africa. Likewise, for Dall's porpoises, 11 MUs were defined in part based on genetic evidence (IWC, 2002). Also, the extraordinary genomics resources recently acquired for the vaquita are intended, first and foremost, to assist their conservation (Morin, et al., 2020a). Finally, as the knowledge about the southern species is really lagging behind, future genetic studies can help to fill this gap and can provide valuable information to guide future conservation actions.

4. Outline of the thesis

One of the aims of this thesis is to investigate the evolution of the porpoise family across different evolutionary time scales: (1) the macro-evolution (i.e. the phylogeny and phylogeography) of the porpoise family; and (2) the microevolutionary processes that take place within some of these species, including the patterns of genetic variations within species,
population structure, and the demographic evolution at the population and subspecies levels. This knowledge will shed some light on the ecological and evolutionary processes taking place in these porpoises, and the contrasted patterns related to their distinct ecology and geographic distributions. In this regard, the thesis is divided into a macro-evolutionary chapter (Chapter 2) and four micro-evolutionary chapters (Chapter 3-6). The other major aim of this thesis is to gather knowledge intended to guide the conservation of certain members of the porpoise family. That is why the theme of conservation biology is a recurrent one discussed in all the chapters.

In Chapter 2, I discuss the phylogenetic relationships at the species and subspecies level among porpoises to 1) reconstruct the history of their divergence, and 2) improve our understanding of their biogeography and phylogeography. Using complete mitogenomes and samples from all the known porpoise lineages, I explore the processes underlying the diversity of this iconic group of small cetaceans. The level of mitochondrial diversity observed will be interpreted in light of the conservation issues highlighted in many porpoise taxa. As this chapter outlines the overall mechanisms influencing the diversification dynamics for all species of the porpoise family, it lays the groundwork for the next chapters (Chapter 3-6) that dissect in more detail the phylogeographic patterns for different lineages highlighted in this first chapter.

Chapter 3 focuses on the influence of the past climate change on the evolutionary history of the North Atlantic harbor porpoises. Phylogeography theory indicates that post-LGM recolonization of the North Atlantic waters is expected to be accompanied by a typical pattern of leading-edge effect. This hypothesis is investigated by combining population genetics, phylogenetics, and habitat modeling (ENM) approaches. The spatial population structure highlighted by analyzing mtDNA and microsatellites revealed in this chapter also provides valuable insight into the conservation of the harbor porpoises.

Chapter 4 assesses how the ongoing climate change is affecting an isolated population of harbor porpoises in Iberian waters using genetic time series. Climate change impact is expected to be especially visible at the distribution edge of a species where suboptimal conditions exist. To test this prediction, the Iberian harbor porpoise is used as a model because it occurs at the margin between two biogeographical zones where the ecological conditions are changing. I used genetic time series to assess the change in genetic composition over the last 30 years at the mitochondrial level. The phylogenetic relationships, the genetic diversity and the demographic trends are analyzed to depict a general overview of the population's genetic dynamics. The results provided in this chapter might play a crucial role in devising effective conservation plans essential to mitigate the strong impact of human activities on the Iberian harbor porpoises.
Chapter 5 goals were two-fold. The first objective is to investigate the genetic structure of the endangered isolated sub-species of harbor porpoise across the entire subspecies range in the Black Sea and adjacent waters. As morphological differences between the Black Sea and Azov Sea individuals were previously reported, I thereby assessed whether these differences were supported by genetic differences. Since no genetic differentiation was observed despite the morphological differences, as a second objective I devised a simulation framework to study the theoretical conditions under which genetic data would be unable to capture possible demographic differentiation, an issue often referred to as the "grey zone" of population differentiation. In some cases, this phenomenon leads to an absence of apparent population genetic structure between demographically independent populations.

Chapter 6 investigates the processes underlying population structure, connectivity, and demography in the critically endangered Yangtze finless porpoise. Relying on microsatellites and mtDNA, I employed different genetic approaches to identify the genetic structure and to estimate the genetic diversity, and effective population size for each population identified. Finally, I reconstructed the most likely demographic history leading to the current structure of the Yangtze finless porpoise using an ABC simulation-based statistical framework. From a practical perspective, the results of this chapter are used to guide conservation actions in the field.

Chapter 7 synthesizes and discusses the implications of all previous chapters.
5. References


Pimper, L.E., Goodall, R.N.P. and Remis, M.I. First mitochondrial DNA analysis of the spectacled porpoise (Phocoena dioptrica) from Tierra del Fuego, Argentina. *Mammalian Biology - Zeitschrift für Säugetierkunde*


Rosel, P.E., *et al.* A review of molecular genetic markers and analytical approaches that have been used for delimiting marine mammal subspecies and species. *Marine Mammal Science* 2017b;33(Special Issue):56-75.


Thorpe, R.S., Reardon, J.T. and Malhotra, A. Common Garden and Natural Selection Experiments Support Ecotypic Differentiation in the Dominican Anole (Anolis oculatus). The American Naturalist 2005;165(4):495-


