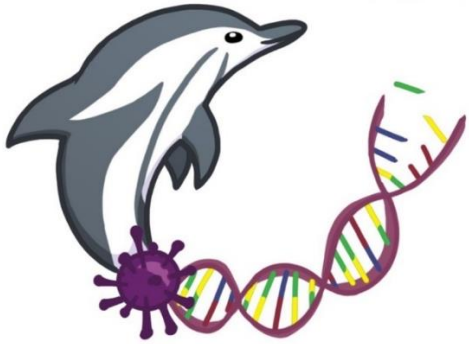
Interaction between genetic variation and susceptibility to environmental challenge in Mediterranean dolphins

ESTEVES-DA-SILVA, CATIA,SOFIA

Interaction between genetic variation and susceptibility to environmental challenge in Mediterranean dolphins



Cátia Sofia Esteves da Silva

Submitted in accordance with the requirements for the degree of

Doctor of Philosophy

Durham University

Faculty of Science

School of Biological and Biomedical Sciences

August, 2023

Declaration

The candidate confirms that this thesis has not previously been submitted for a degree at Durham University or any other University. The work submitted is her own and that appropriate credit has been given where reference has been made to the work of others.

Dedication

To my younger self.

"The most important things you learn on this road,

are the ones you can't put on your CV."

– Anja Le Grange, a colleague that crossed my path at the right time.

Acknowledgements

This thesis is the result of all the support that I received, in so many different ways, during these past years. Firstly, I would like to thank my supervision team, Prof. A. R. Hoelzel, Dr. Andreanna Welch, Dr. André Moura, Dr. Sara Novais, for helping me grow, in all the different ways that I needed. A special thanks to the main supervisor, Prof. Rus, whom I admire greatly, for never given up on me, and for all the teaching and guidance, even at a distance, with so much kindness and lightness. I will always be grateful for all your scientific knowledge, but it was your experience on sharing it, while kindly motivating me, that resulted in the best lesson and mentoring model to me.

Although not part of the supervision team, I would also like to include here my deep appreciation to Dr. George Gkafas and Dr. Menno de Jong, for the excellent guidance, help, patience, and empathy. They were my great mentors, to whom I could never thank enough.

I would like to thank the institutions involved: Fundação para a Ciência e Tecnologia, Lincoln University, Durham University, and Cetemares – Polytechnic of Leiria, for all the contribution to complete this project. This work had the support of Fundação para a Ciência e Tecnologia, through the grant SFRH/BD/118195/2016, to which I am thankful. I want to thank specially to my first year examiners of Lincoln University: Prof. Matthew Goddard, and Dr. Matthew Simmonds, for the great teaching and feedback provided. Also, my deep appreciation to the wellbeing departments of Durham and Lincoln Universities, in particular to Dr. David Lowther from the University College – Durham, and to the PGR tutors of Lincoln University,

Dr. Stefan Millson and Dr. Sandra Varga, for really seeing and supporting me, even at a distance.

Nothing would have been possible without samples, so I would like to show my gratitude to Dr. Stefania Gaspari who kindly collected and provided the sample archive used in this study. Collaborations are key for science improvement and I would also like to extend my appreciation to Tilen Genov and Morigenos team, for the way I was received in the Dolphin Research Course, 2018, and the prompt availability showed for future collaborations.

To all the Molecular Ecology group, for being like supervisors and also amazing friends/colleagues/housemates, with whom I had some of the best experiences of my life. Thank you for always receiving me so well and making me feel I was part of something. A big thank you to Erandi and Monica, Ale and Rodrigo, Andrea and Nilo, and Vania – my amazing Mexican/Spanish/Brazilian hosts/families. Also a very special thanks to my "bioinformatics team", mostly to the never-ending patience of Menno, George Gkafas, André Gomes, Biagio, Vania, and Erandi. I could never thank you enough for being there for me, even with all the difficulties of being on the other side of the screen, like most of the times.

To Lemos Lab team, which was my first scientific home for more than 10 years, a deep appreciation, specially to Marco and Sara, for all the opportunities granted that contributed to my growth at all levels. A special thanks also to Carina, Rafael and Luís...for your humanity without filters. And João Reboleira, the so talented human being that promptly picked up my epiphany drawings, transforming them in the beautiful logo of this thesis.

To all the friends and family that, regardless of the distance or my life decisions, such the ones related with this project, tried to really understand, help me, and keep in touch. You will never know how many times your simple gestures of caring made a huge difference. On this, a particular thank you to Sofia Libório, Francelly Sosa, Joana Falcão, Cláudio Brandão, Simone Teixeira, Alfredo Rodrigues, Sofia Aranha, Pedro Ramalhete, Ana Faria, Gonçalo Silva, Joana Manecas, Erandi, Tiago Chança, Bruno Martins, Rafaela Pardal, Steeve Carvalho, and my therapists. Thank you for being there with all the patience, helping me dive into myself and resurface with more knowledge, tools and confidence.

I would also like to thank my family, for supporting me no matter my decisions, as long as they were in tuned with myself. All of my achievements are also yours and I could not be prouder of my origins. My cats, the best company and greatest zen masters I could have, are also included in this category as the most present beings and tireless "therapists", during these past years. My deep gratitude to all the people that cared for them while I was at Durham or away doing supplementary courses.

Finally, to me. I want to thank myself for, in the right time, had allowed me the chance to look and dive in the most needed way possible: inside.

Abstract

Top predators such as marine mammals are particularly sensitive to perturbations in the ecosystem due to their longevity and food web position. The adaptation of wild animals to environmental change is thus crucial and facilitated by genetic diversity. Pathogens represent a significant mortality source in wild animals, with host resistance and survival being mediated by molecular mechanisms and genetic components. Since pathogens represent a strong selective pressure for their hosts and new pathogens are constantly introduced into host populations, such that pathogen dynamics are particularly good models to study adaptation.

The present study investigated the genetic diversity in the Mediterranean Sea, as well as the genomic basis of striped dolphin (*Stenella coeruleoalba*) vulnerability to morbillivirus disease – a highly contagious and deadly virus that infects mammals, and has been considered a serious ecological threat.

The striped dolphin has a world-wide distribution, and is the most abundant cetacean species in the Mediterranean Sea – the stage of several morbillivirus epizootic outbreaks over the past 33 years. This Delphinidae species has been reported as one of the most susceptible cetaceans to morbillivirus disease, and even possibly a reservoir for this virus, which could have serious implications for endangered species conservation.

To assess aspects that could be influencing the vulnerability of striped dolphin to morbillivirus infection, different approaches were taken

in this study, including assessing connectivity and the association of death during epizootics with inbreeding/genotype analysis at susceptibility-related candidate genes. Patterns of genetic diversity and population structure can contribute to influence disease spread and species fitness. Thus, population structure was evaluated within the East and West regions of the Western Mediterranean basin, as well as between Mediterranean and Atlantic individuals. Using random genome-wide scans, measures of heterozygosity-fitness correlations, inbreeding, and potential candidate genes associated with morbillivirus susceptibility, were investigated for this species. Additionally, signs of selection associated with morbillivirus disease were also assessed using a hybridization-based capture procedure with specific genomic regions, including immune system related genes, as targets.

This study suggests stronger structure between Atlantic and Mediterranean individuals than between sampled regions within the Mediterranean region, suggesting high levels of connectivity within the western basin. Levels of inbreeding for the Mediterranean striped dolphin were low, and genome-wide heterozygosity were found to be within reported values for mammals. Genome-wide heterozygosity-fitness analyses did not reveal significant correlations between diversity levels and fitness associated with surviving morbillivirus infection. However, possible signs of morbillivirus susceptibility were found with the identification of particular loci with functions that were consistent with morbillivirus infection/disease. The targeted capture approach

corroborated the absence of a clear general effect, even among multiple loci with shared immune system functions with only one locus (HAVCR2) showing a strong correlation with morbillivirus exposure.

These time-series data contribute to the scientific knowledge on striped dolphin susceptibility to morbillivirus disease, as well as its geographical diversity, conferring relevant information that could be useful to future studies and conservation measures for this species.

Table of Contents

List of Figures

List of Tables

Chapter 1

General Introduction

1.1 Evolution of host-pathogen interactions

Genetic diversity is a crucial factor in wild animals' adaptation to environmental change. Host-pathogen dynamics are considered excellent models to study patterns and processes of adaptation in the wild, since pathogens represent clear and quantifiable selective pressures for their hosts (Dronamraju, 2004). Several classic studies that have enabled considerable advances in our understanding of adaptive mechanisms have been based on host-pathogen interactions (Flor, 1955; Hughes and Nei, 1988, 1992; Woolhouse et al., 2002; Ebert, 2008; Penczykowski et al., 2016; Zueva et al., 2018). High degree of specificity between parasites' and hosts' genotypes often leads to parallel evolution between host and pathogen complementary genetic systems. For example, a recent study focusing on parasite-driven selection in Atlantic salmon and its flatworm *Gyrodactylus salaris*, identified 57 candidate genes related to immune function (Zueva et al., 2018). Therefore, although host-pathogen interactions frequently involve a complex genetic basis, it is often possible to identify sets of genes that are good candidates to further understand the regulation of immunity and pathogen tolerance/susceptibility in wild animals.

Pathogen resistance involves a complex combination of physical, cellular and molecular mechanisms (Janeway et al., 2001), and determining the precise mechanisms involved in individual host-pathogen systems can be very challenging (Woolhouse et al., 2002). However, several cases are known to involve interactions between self and non-self molecules, with pathogen specificity involving a well-known genetic basis

(Janeway et al., 2001). For example, vertebrate antibodies will bind to antigens through their variable domains, which are produced by 3 different gene families through a process known as V(D)J recombination, a discovery that was rewarded with the 1987 Nobel Prize for Physiology or Medicine attributed to Susumu Tonegawa. Similarly, various antigens are presented to T-cells through binding to transmembrane proteins coded by genes of the Major Histocompatibility Complex (MHC). MHC gene variation has been studied extensively in wild animals (Piertney and Oliver, 2006). Considered to have the most polymorphic loci in the vertebrate genome (Yuhki and O'Brien, 1990; Gaudieri et al., 2000), maintained by balancing selection (Hughes and Yeager, 1998) and mediated by host-pathogen interactions and/or mate choice (Piertney and Oliver, 2006; Spurgin and Richardson, 2010), low levels of MHC genetic diversity may increase the susceptibility of a population to infectious diseases, leading to catastrophic impacts (Klein and Sato, 1998; O'Brien and Yuhki, 1999; Bowen et al., 2002). Thus, it's considered an ideal model for the study of adaptation in the wild (Hughes and Nei, 1988, 1992).

1.2 Host-pathogen gene co-evolution models

Two broad classes of host-pathogen models generate specific predictions regarding genetic variations in response to pathogenic pressure: the gene-for-gene and the matching-allele (Woolhouse et al., 2002). The gene-for-gene hypothesis was introduced by Harold Henry Flor (Flor, 1942; Flor, 1955) to explain the inheritance of disease in studies of plants, suggesting that "for each gene that conditions reaction in the host there is a corresponding gene in the parasite that conditions pathogenicity". Studying the interaction between flax and the fungal pathogen flax rust, Flor (1942, 1955) detected that when flax genes for resistance were dominant, rust genes for virulence were recessive. Thus, molecules encoded by dominant avirulence (Avr) genes in pathogens, trigger defence responses in the hosts carrying the corresponding resistance (R) genes. In broader terms, the resistance in the host and the ability to cause disease in the parasite is controlled by matching genes. Each gene must be identified only by its counterpart of the host-parasite system. Later, a more detailed definition was introduced by Person and collaborators (1962): "A gene-for-gene relationship exists when the presence of a gene in one population is contingent on the continued presence of a gene in another population, and where the interaction between the two genes leads to a single phenotypic

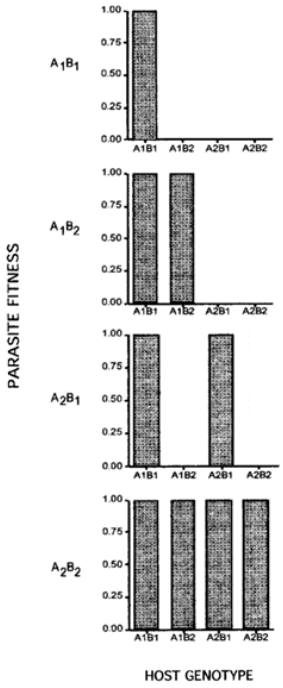


Figure 1 - Fitness of four pathogens on each host genotype under gene-for-gene model. (Agrawal and Lively, 2002).

expression by which the presence or absence of the relevant gene in either organism may be recognized." (Flor, 1971; Agrawal and Lively, 2002; Kaur, et al., 2021). This gene-for-gene interaction between resistance and avirulence genes, matches antigen–antibody interaction processes (Vander Biezen and Jones 1998; Kaur, et al., 2021). This means that, as an evolutionary adaptation, host and parasite developed complementary genetic systems based on susceptibility and resistance match of genes (Figure 1). As a result, the fluctuation of host allele frequencies, depending on the relative frequency of resistant and susceptible genotypes is the expected prediction by this model (Agrawal and Lively, 2002).

In contrast, the matching-allele model requires a self/non-self-recognition system (Agrawal and Lively, 2002), first demonstrated in a study with *Daphnia magna* - *Pasteuria ramosa* interaction (Luijckx et al., 2013). The host's immune system is eluded when a match between a specific allele of a pathogen virulence gene and a compatible allele in the host's resistance gene is observed, making certain pathogen alleles 'specialists' on specific host alleles. In other words, there must be a specific virulence/susceptibility combination between the host/pathogen alleles for infection to be successful (Figure 2). At a population level, this means that the pathogen allele that successfully infects the most frequent susceptibility allele in the host population, will

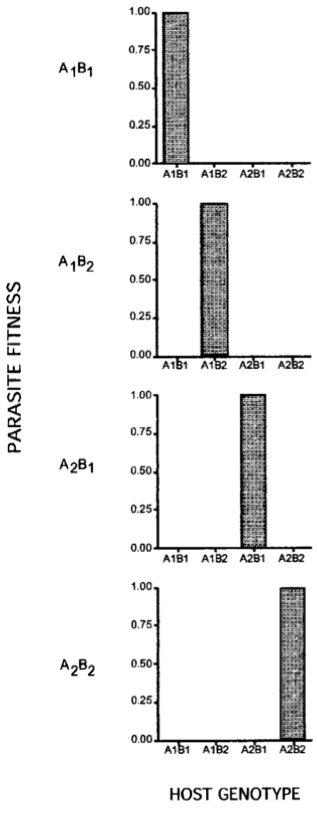


Figure 2 - Fitness of four pathogens on each host genotype under matching alleles model (Agrawal and Lively, 2002).

be favoured and increase in frequency through natural selection. The key feature of this model is that the virulence/susceptibility dynamic can only be changed by the appearance of novel genotypes in the host population. Thus, the spread of the virulent allele through a population of hosts will only be reduced if novel genetic varieties appear in the host population.

The matching-allele model is thus conceptually related to the Red Queen hypothesis, devised to explain the emergence of sexual reproduction (Agrawal and Lively, 2002; Benton, 2009; Lively, 2010; Brockhurst et al., 2014). The name derived from a quote in Lewis Carroll's "Through the Looking-Glass", where the Red Queen says to Alice: "it takes all the running you can do, to keep in the same place" (Carroll, 1871), as the hypothesis states that to maintain resistance against pathogens over time, the appearance of novel genetic variation must occur regularly.

1.3 Morbillivirus

Morbilliviruses are single-stranded RNA viruses, highly infectious, pleomorphic, lymphotropic and epitheliotropic, which means that the initial replication occurs in lymphoid tissues and the dissemination is facilitated by lymphocytes through the lymphatic system, spreading to epithelial cells, causing profound immune system impairment (Van Bressem et al., 2014). Broncho-pneumonia is associated with an acute state, followed by non-suppurative demyelinating meningoencephalitis and colonization of the brain by opportunistic pathogens (sub-acute state), and secondary infections and brain lesions (chronic state). Several host

specific strains of morbillivirus are described, being classified into a total of 6 classes depending on their preferred host species: the measles virus (typically affecting humans), rinderpest virus (reindeers), peste-des-petits-ruminants virus (ruminants), canine distemper virus (dogs), phocine distemper virus (seals and otariids), and cetacean morbillivirus (whales and dolphins) (Amarasinghe et al., 2019). However, research has shown that cetacean morbillivirus (CeMV) infections are neither host-restricted nor location-restricted (Jo et al., 2018a, 2018b), which implies a progressive widening of its range, distribution and trans-oceanic spread (Jo et al., 2018a, 2018b; Mira et al., 2019). Cross-species transmission is thought to be determined by the presence of similar host proteins and cell receptors in the different groups (Shimizu et al., 2013). This is further supported by the close genetic relationship between cetacean and ruminant morbilliviruses, suggesting the existence of a common ancestor between the viruses of these groups (Liu et al., 2016), mirroring the close phylogenetic relationship between cetaceans and artiodactyls (Montgelard et al., 1997).

1.4 Morbillivirus infection in cetaceans

Infection with morbillivirus is a primary mass mortality agent in marine mammals worldwide (Van Bressem et al., 2014), and is considered one of the most pathogenic viruses in these organisms. Factors related to the particular ecological niche of marine mammals, such as wide distribution, migration, reproduction, density and social interactions, are

key when it comes to favouring the transmission and circulation of the virus among the individuals and species.

Dolphin morbillivirus (DMV), first observed in Atlantic bottlenose dolphins (*Tursiops truncatus*) in 1982 (Hersh et al., 1990; Duignan et al., 1996), are known to have caused epizootics of high mortality in various places around the world (Figure 3) and is increasingly the focus of considerable attention in terms of their effects on cetacean fitness and conservation status (Aguilar and Gaspari 2012; Van Bressem et al., 2014; Casalone et al., 2014; Di Guardo and Mazzariol, 2016). Extensive research effort has focused on understanding the mechanisms underlying increased risk of epizootics in dolphins worldwide (Van Bressem et al., 2014; Jo et al., 2018a, 2018b). Although the morbillivirus strains are well characterized (e.g. Bellière et al., 2011; Zinzula et al., 2022), several features of its pathogenesis, such as the genetic basis of morbillivirus susceptibility in dolphins, is still poorly understood (Di Guardo and Mazzariol, 2016; Zinzula et al., 2022).

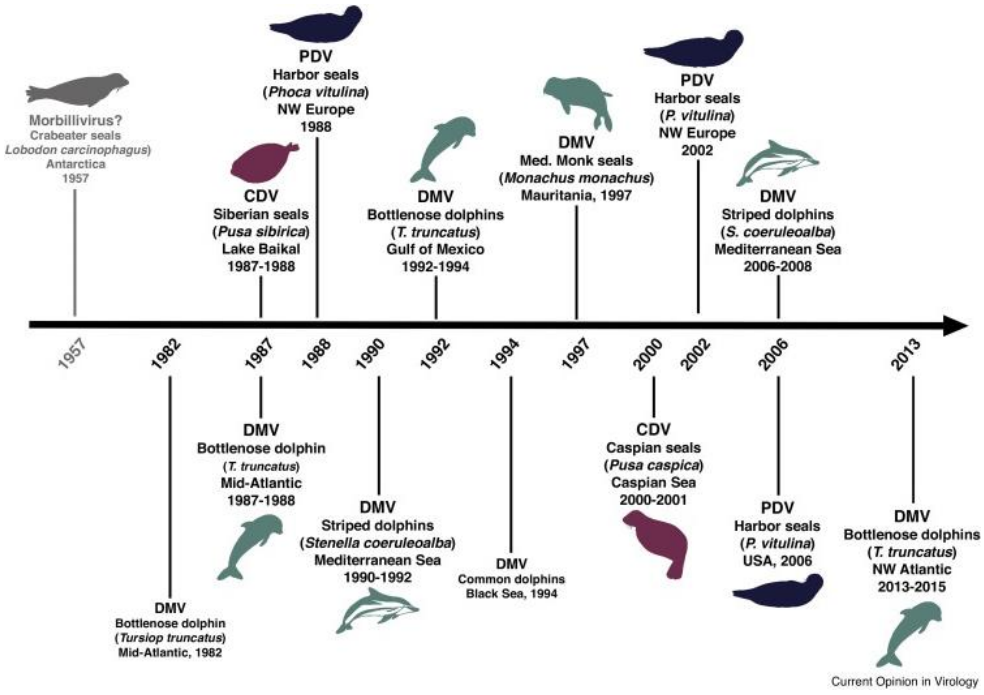


Figure 3 – Timeline of major morbilliviruses outbreaks in marine mammals. Dolphin morbillivirus (DMV), canine distemper (CDV), and phocine distemper (PDV) (Jo et al., 2018a).

Several epizootics have been described in Mediterranean striped dolphins (*Stenella coerueoalba*) since 1990 (Domingo et al., 1990; Di Guardo and Mazzariol, 2013), being associated with considerable increases in mortality and stranding frequency across the region (e.g. Valsecchi et al., 2004; Raga et al., 2008). The cycle of occurrence of epizootic episodes was suggested to be from 3-5 years (Di Guardo and Mazzariol, 2013). Although other Mediterranean cetaceans have also been affected, none appear to have experienced the levels of mortality documented in striped dolphin (Di Guardo and Mazzariol, 2013). Striped dolphin susceptibility to morbillivirus, has been attributed to high inbreeding – stranded animals tended to show high levels of inbreeding (Valsecchi et al., 2004), combined

with high pollutant load (Aguilar and Borrell, 1994, 2005) and vertical transmission, from mother to embryo/fetus/calf has also been suggested (Di Guardo and Mazzariol, 2016). Other potential factors that have not been tested include high population densities or lack of genetic variation at immune system genes. However, to our knowledge, no research to date has carried out a comprehensive whole genome assessment of changes in genetic composition as this species experiences morbillivirus epizootics (Valsecchi et al., 2004; Gaspari et al., 2007).

By studying the genetic basis of ecologically relevant functional traits, through the identification of genetic footprints of selection, one can understand the patterns of local adaptation, and in the case of host-pathogens relationships, to predict the long-term viability of wild species. However, it has been recognized that studies on immune system genes in wild cetaceans has been excessively restricted to a small set of well described genes (Acevedo-Whitehouse and Cunningham, 2006). Therefore, a comprehensive analysis of immune system genetic variation in this species will be important not only for the understanding of striped dolphin susceptibility to morbillivirus, but also to our understanding of the genetic basis of immune responses in cetaceans.

1.5 Known genetics of morbillivirus resistance

Two molecules are well known to be involved in the host-virus interactions between morbillivirus and their mammal hosts and have thus been prime candidates for studies in the genetic basis of susceptibility to

the virus (Delpeut et al., 2014; Stejskalova et al., 2017). These are the signalling lymphocyte activation molecule (SLAMF1) – a morbillivirus receptor present on lymphoid cells; and the poliovirus-related receptor-4 (NECTIN4) – a morbillivirus receptor on epithelial cells and a tumour-associated marker for carcinomas (Delpeut et al., 2014; Noyce et al., 2011). SLAMF1 contains a binding site domain for morbillivirus and therefore facilitates virus entry into the cell (early stage), while NECTIN4 acts as an epithelial receptor spreading the virus from airway epithelial cells (later stage), facilitating the virus exit and transmission to other individual hosts (Delpeut et al., 2014; Jo et al., 2018a, 2018b). A scheme of this virus life cycle can be seen in Figure 4, as well as an example of its endocytic entry in human cells in Figure 5.

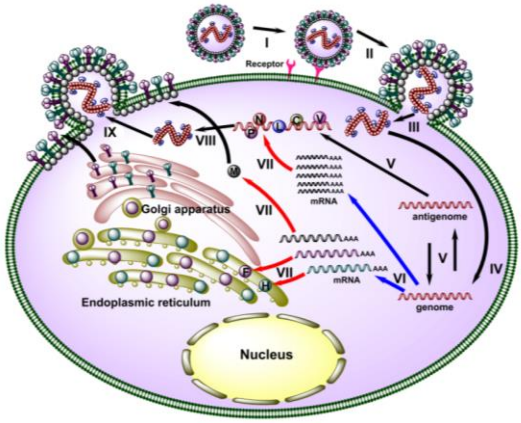


Figure 4 - Morbillivirus life cycle representation in a cell, including the main 9 steps: (I) viral attachment to receptor, (II) membrane fusion, (III) release of nucleocapsid, (IV) nucleocapsid uncoating, (V) reversible synthesis of genome and antigenome, (VI) viral mRNA transcription, (VII) viral protein expression, (VIII) genome encapsidation and (IX) progeny budding from cellular membrane. Nucleocapsid (N), phosphoprotein (P), matrix (M), fusion (F), hemagglutinin (H), large (L) represent the structural morbilliviral proteins encoded by its genome. V, C are non-structural proteins encoded by P (Liu et al., 2016).

The first step of a viral replication cycle is characterized by attachment of the H protein to the SLAMF1 receptor, followed by membrane fusion mediated by the F protein. This causes the release of the nucleocapsid into the cell, which activate cascading biochemical reactions from nucleocapsid uncoating to progeny budding that will be able to infect neighbouring cells (Liu et al., 2016).

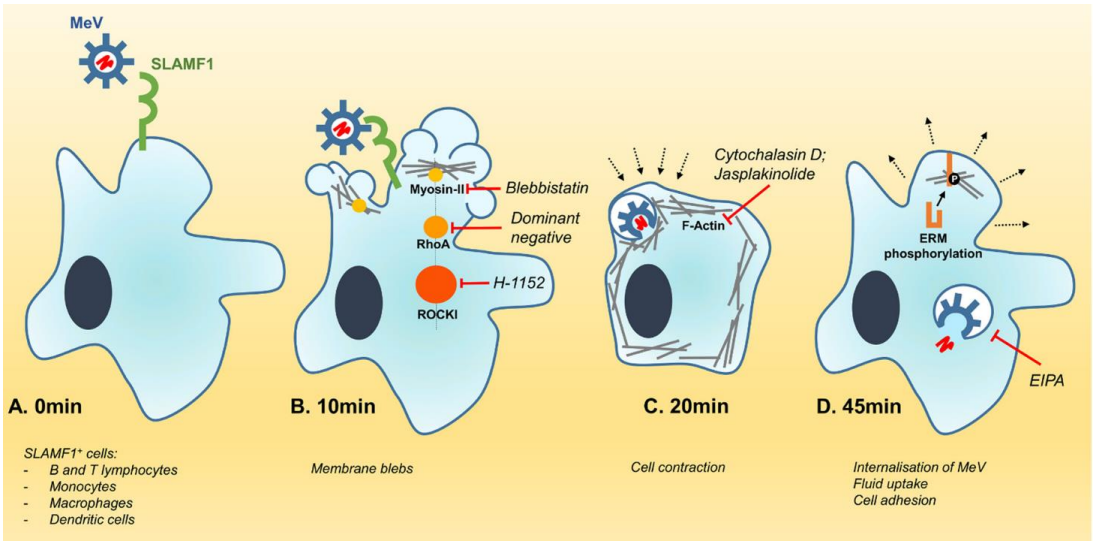


Figure 5 – Endocytic entry in human cells by MeV. (A) MeV binds to SLAMF1 (CD150) receptor through a specific interaction between the viral hemagglutinin and SLAMF1. (B) The last interaction induces the formation of membrane blebs, associated with the RhoA-ROCK-myosin II axis. If inhibited, it reduces MeV entry. (C) An acute retraction of the cell starts to occur by the cortical cytoskeleton. The perturbation of this process also reduces MeV entry. (D) After 45 minutes, fluid uptake and the internalisation of MeV occurs, a process sensitive to inhibition with EIPA drug (a traditional inhibitor of macropinocytosis), reducing MeV entry and infectious. Cells re-establish their characteristic through ERM phosphorylation (proteins that regulate membrane blebbing) (Gonçalves-Carneiro, 2017).

Although SLAMF1 and NECTIN4 are considered relatively conserved within species of the same order (Jo et al., 2018a), NECTIN4 appears to be more so than SLAMF1, suggesting it could be important in determining species specific binding ability to morbillivirus (Shimizu et al., 2013; Delpeut et al., 2014; Feng et al., 2016). A study assessing the potential vulnerability of Delphinidae family to morbillivirus infections focusing in the cellular receptor structure for this virus, suggested that a higher binding affinity between SLAMF1 and this virus (Van Bressem et al., 2009) may result in intensified viral infectivity and tissue propagation (Shimizu et al., 2013). When comparing the amino acid composition of SLAMF1 between 26 cetacean species, variation was found in six residues. In bottlenose and striped dolphins, which have had various described cases of morbillivirus mass mortality events throughout the world (Figure 3), substitutions at five amino acid residues were found (E68G, I74V, R90H, V126I, and Q130H) relative to other species, with three of them (E68G, R90H, and Q130H) causing changes in protein electric charges, which could possibly influence the virus affinity (Van Bressem et al., 2009). However, there is still lack of evidence regarding the relative susceptibilities between individual polymorphisms (Stejskalova et al., 2017).

Although considered two of the main markers of morbillivirus infection, a wider range of genes must be analysed, given that virus susceptibility and transmission is likely influenced by multiple and yet unidentified cellular receptors (Delpeut et al., 2014). Recent studies have

identified that in addition to SLAMF1 and NECTIN4, several other genes (such as BSG, CD147, TLR3, TLR7, TLR8, SLC11A1, NCR1, MAPK8, FBXW11, INADL, ANK3, ACOX3 and MHC-DQB) may also play a role in determining variation in host susceptibility to morbillivirus infection in cetaceans (Stejskalova et al., 2017; Batley et al., 2019; Batley et al., 2021). Stejskalova and collaborators (2017) identified polymorphisms in eleven immunity-related genes of striped dolphin and the harbour porpoise (*Phocoena phocoena*) from the Mediterranean Sea, using next generation sequencing, and suggested they might have a role in morbillivirus infection. All genes, previously selected based on their potential role in morbillivirus infection/immunity, contained polymorphisms that showed signals of selection based on dN/dS tests, with the exception of MHC-DQB. Comparing non-survivors against putative survivors of a CeMV epizootic in Indo-Pacific bottlenose dolphins (*Tursiops aduncus*) from South Australia, Batley and colleagues (2019) carried out gene association studies through reduced representation sequencing (RRS), using ddRAD sequencing, and later a whole-genome approach (Batley et al. 2021). The first study identified five candidate genes whose functions are associated with stress, immune responses and pain (Batley et al., 2019). However, these studies are limited by low sample size, analysing only one CeMV epizootic, and low representation of immune system genes. The latter aspect improved in the most recent study (Batley et al., 2021). Their ddRAD assay covered only 1% of the dolphin genome, of which only a very small fraction would have included sequences from immune system genes.

More recently, and clearly emphasizing the importance of whole genome assessments, the authors were able to identify 50 genes related to innate and adaptive immune responses to morbillivirus (Batley et al., 2021).

Long-term studies are rare due to the logistical constrains involved, but represent a valuable contribution to a better understanding of how genetic variation may change following environmental changes. These allow specific correlations to be made between genetic variation and environmental pressures as they occur, and thus enable a better understanding of the factors involved in driving genetic variation and structure in the wild. Additionally, genome wide assessments of functional genes are also rare, and most studies regarding dolphin-morbillivirus interaction have been focusing on a small set of well described immune system genes. Thus, by covering an 18-year period including both epizootic and non-epizootic periods, and analysing a comprehensive set of immune system genes of striped dolphin during that period, the present study overcomes the limitations of previous studies, and fills existing knowledge gaps regarding the genetic basis of morbillivirus resistance in this species.

Furthermore, because morbillivirus is lymphotropic, infection can cause a reduction in cellular immunity response (Liu et al., 2016), and mortality can thus result from complex interactions between the virus and other opportunistic infections. In fact, striped dolphins stranded in the Mediterranean during morbillivirus epizootics, are commonly found with a range of other opportunistic infections (e.g. Van Bressem et al., 2009, 2014; Profeta et al., 2015). Therefore, mortality due to morbillivirus in

dolphins must be analysed in the context of susceptibility to other opportunistic infections. In vertebrates, the MHC genes are known to play an important role in the immune responses to infectious diseases (Edwards and Potts, 1996). Distributed between two classes (I – intracellular antigens and II – extracellular antigens), MHC genes encode cell-surface glycoproteins which recognize and bind to pathogen derived antigens, leading to activation of the appropriate immune system responses by T-lymphocytes (Dengjel et al., 2005; Piertney and Oliver, 2006). It has been suggested that in mammals, genetic variation in the MHC, maintained by heterozygote advantage or frequency dependent selection, is linked to adaptation to a diversity of pathogens (Hughes and Nei, 1988; Hill et al., 1991; Hughes and Nei, 1992; Edwards and Potts, 1996; Piertney and Oliver, 2006). Although MHC has been the subject of numerous studies regarding pathogen susceptibility and adaptation in wild animals, it is currently accepted that it is insufficient to fully explain host-pathogen dynamics in wildlife (Acevedo-Whitehouse and Cunningham, 2006). Other genes, such as Toll-Like Receptors (TLR), Interleukins, and Gamma-Fibrinogen (FGG) are involved in important immune functions (Turner et al., 2011; Webster et al., 2011; Tschirren et al., 2013). FGG in particular, is involved in the blood clotting process and has been shown to have experienced strong levels of selection in cetaceans during their transition from land to sea, which was hypothesised to be due to the different pathogenic environment between these two very different ecosystems (Moura et al., 2013). Physiologically, cetaceans form scar tissue

considerably quicker than other mammals, which could be a way to minimize the chances that water borne diseases enter the blood stream through open wounds. Host-pathogen dynamics are thus a known selective pressure in cetaceans, with well described effects in the observed patterns of genetic diversity in this group (Dronamraju, 2004).

A recent work, analysing the genetic structure of Mediterranean striped dolphins across time and different geographic regions, revealed cyclical fluctuations over 21 years, that could be associated with the occurrence of morbillivirus epizootics (Figure 6; Gaspari et al., 2019).

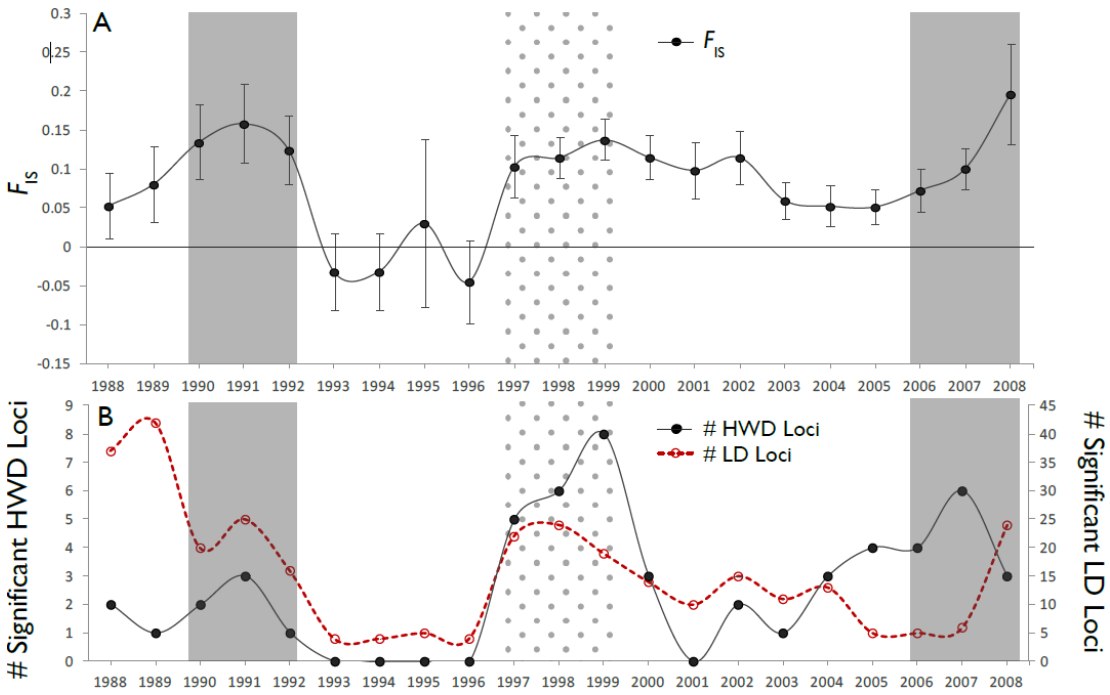


Figure 6 – Genetic variation of Mediterranean striped dolphins across time. Grey bars represent the years of morbillivirus epizootics that are well described, while the grey dots area represent an epizootic less described in the literature. A) F_{IS} – inbreeding coefficient; B) # HWD loci – number of loci with significant deviations from Hardy-Weinberg equilibrium; # LD loci – number of loci pairs with significant tests for linkage disequilibrium. Adapted from Gaspari et al., 2019.

The inbreeding coefficient (F_{IS}) fluctuations correlate with the two known epizootic periods – 1990-1992 and 2006-2008, (Van Bressem et al., 2014) and a possible third one 1997-1999 (Van Bressem et al., 2001), as do increases in the number of microsatellite loci with significant linkage disequilibrium and deviations from Hardy-Weinberg, all pointing to disruption in the equilibrium of genetic variation. These cyclical changes in genetic composition associated with morbillivirus episodes were suggested to reflect a genetic component to morbillivirus survival and could therefore represent selective sweeps – increase of an allele until fixation, accompanied by a large reduction of genetic variation in nearby chromosome regions. However, it was not possible to confirm this hypothesis in that study, because selective sweeps are difficult to detect based on microsatellite genotypes alone. Genome wide analysis of relevant functional genes, is thus important to overcome these strong limitations and fill the existing knowledge gaps on morbillivirus resistance in this species.

In addition, previous work on genetic variation in active sites for both SLAMF1 and DQB1 genes, for striped dolphin across the same timeframe, showed that changes could also be observed in functional variation coincident with the various epizootics (Figure 7).

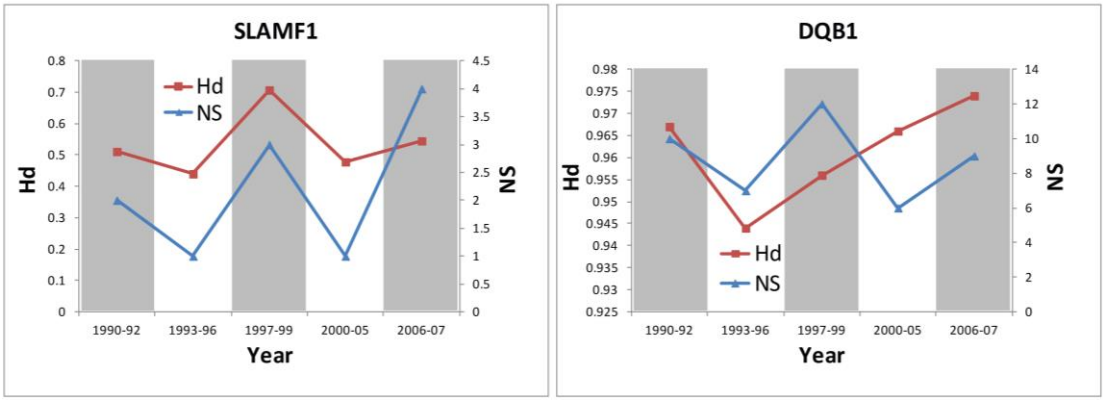


Figure 7 – Genetic diversity of SLAMF1 and DQB1 in striped (*Stenella coeruleoalba*), across epizootic (grey bars) and non-epizootic events (white bars). Haplotype diversity (Hd), and number of non-synonymous (NS) changes. Grey vertical areas represent the three better documented morbillivirus epizootics (Gaspari et al, 2019). Adapted from Charlotte Dooley, Mbio report, University of Lincoln.

SLAMF1 exon 2 presented a pattern of increased diversity during epizootic events, and an abrupt decrease in the non-epizootic period. This frequency dependent fluctuation of allele frequencies matches the predictions made by gene-for-gene model, of a cyclical variation in genetic composition over time. Interestingly, the known mechanism of interaction between the SLAMF1 receptor and the virus, also matched the modelled relationships in the gene-for-gene model, with cyclical fluctuations on genetic composition over time, depending on the relative frequency of resistant and susceptible genotypes. Being an entry gate for morbillivirus into the cell, SLAMF1 diversity has been found to be associated with the binding ability to this virus (Feng et al., 2016). The higher the diversity, the more intensified and widespread the viral activity will be (Van Bressem et al., 2009; Shimizu et al., 2013). This is likely because higher

diversity might increase the likelihood of a genotype match between host and virus, based on complementary genetic systems of susceptibility-resistance matches of genes between the host and parasite, as postulated by the gene-for-gene model. On the other hand, DQB1 patterns show a constant increase of diversity over the multiple epizootics, more consistent with the predictions of the matching-allele model. The DQB1 gene is part of the MHC gene complex, and plays a central role in the immune system. Because MHC molecules function by binding specific antigen molecules, higher genetic variation of MHC genes in a population enables the effective recognition of a broader range of pathogen antigens, with the respective activation of the appropriate immune response. Therefore, pathogens are only able to expand through a population if novel genetic variation is introduced, while host resistance will only increase if novel MHC variants are introduced repeatedly, in accordance with the matching-allele model. These models, whose key feature is a specific match between the host and parasite, provide clear predictions of how genetic variation is expected to change in response to the various epizootics in the present study system, with the respective pattern being strictly dependent upon each gene function.

1.6 Genetic diversity in the Mediterranean striped dolphin

The striped dolphin is a cosmopolitan pelagic odontocete, highly social, and the most common cetacean in the Mediterranean Sea, feeding on pelagic and bathypelagic species such as fish and cephalopod species (Aguilar, 2000). Its sexual maturity is reached at 11-12 years and its lifespan exceeds the age of 45 years (Aguilar, 2000).

Ecological processes such as the local environment and species-specific resource requirements can induce small-scale population differentiation in cetaceans (Hoelzel, 2009). For example, orcas (*Orcinus orca*) which have different 'residents' and 'transients' ecotypes based on prey source specialization, mating systems, vocal behaviour and social organisation, show genetic differentiation among these populations (Deecke et al. 2005; Hoelzel et al. 2007; Morin et al. 2010; de Bruyn et al. 2013). This, together with natural selection, differentiation by drift and life history of orcas, may be leading to speciation (Moura et al., 2014). This means that, although still considered a single species, the differences and interbreed avoidance between ecotypes, is leading to a sympatric speciation – a genetic divergence between 'residents' and 'transients'.

Previous studies on spatial genetic structure and diversity of striped dolphins, based on mtDNA and microsatellites, suggested weak differentiation for this species within the Mediterranean, but stronger evidence for separation between the Mediterranean and the Atlantic

(Figure 8; García-Martínez et al., 1999; Gkafas et al., 2017; Gaspari et al., 2019).

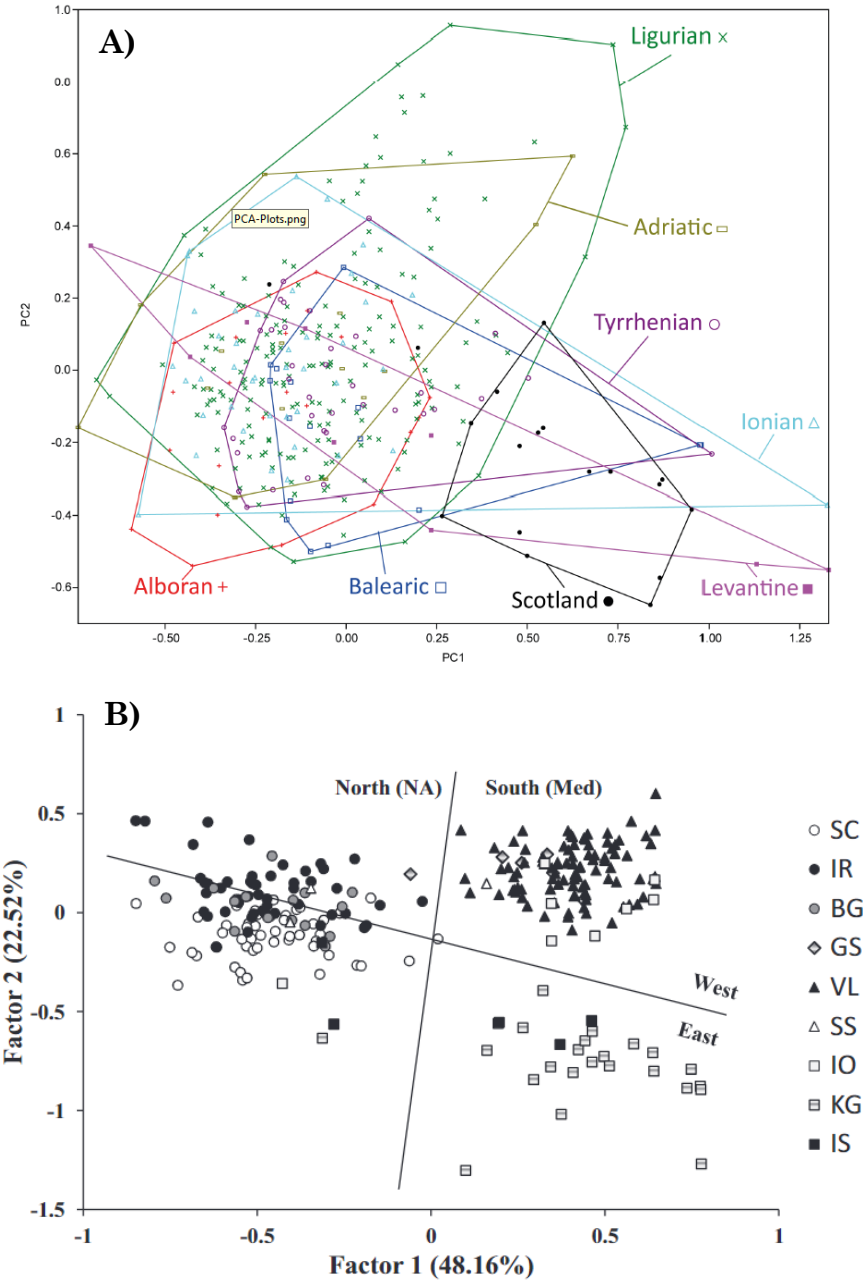


Figure 8 – Differentiation analysis within the Mediterranean region and between the later and Atlantic by A) principal component analysis on individual microsatellite genotypes (Gaspari et al., 2019) and B) factorial correspondence analysis (Gkafas et al., 2017) of striped dolphin population (NA: North Atlantic, SC: Scotland, IR: Ireland, BG:

Biscay Gulf, GS: Gibraltar Strait, VL: Valencia, SS: Strait of Sicily, IO: Ionian Sea, KG: Korinthiakos Gulf, IS: Israel).

Genetic diversity influences the fitness and adaptive potential of the individuals and can be caused by different factors such as mutation, natural selection, gene flow and genetic drift (Amos and Harwood, 1998). Environmental factors may be more challenging, particularly in a changing environment where new selective pressures can emerge, if the genetic diversity of a population is low. Relationships between levels of genetic diversity and fitness, with special focus in cases of high levels of heterozygosity and pathogen susceptibility, have been suggested and correlated with genetic drift and inbreeding (David, 1998; Tsitrone et al., 2001; Chapman et al., 2009; Gkafas et al., 2020). Over the past 3 decades, the striped dolphin has been facing increased environmental pressure in the Mediterranean Sea, from morbillivirus epizootics. As mentioned above, pathogenic challenges can have a strong effect on individual survival as well as impacts on genetic diversity. A recent study, focusing on patterns of genetic diversity of this Mediterranean species, suggested the association between significant temporal cyclical fluctuations in genetic composition with the occurrence of morbillivirus epizootics (Figure 6, Gaspari et al., 2019). Changes in group composition (adults and juveniles) were also suggested to be related to specific morbillivirus resistance dynamics. Compared to mysticetes, the higher susceptibility to infectious diseases in odontocetes has also been suggested to be partially attributed

to their evolutionary history (Braun et al., 2015; Di Guardo and Mazzariol, 2019). The transition from terrestrial to an aquatic environment also yielded the loss of protein functions, which in some cases can result in deleterious consequences. In that scenario, cetacean ancestors suffered genomic changes that ultimately led to the loss of a higher number of functional genes, many related to the immune system response against pathogens (Huelsmann et al., 2019). Later, by the time odontocetes diverged from mysticetes, the functionality of Mx1 and Mx2 – genes related to antiviral immunity, were reported to be lost in the toothed whales (Braun et al., 2015; Di Guardo and Mazzariol, 2019). That, together with other relevant ecological and behavioural factors that distinguish these groups, has been associated with the different levels of susceptibility and resistance to this virus (Braun et al., 2015; Jo et al., 2018a, 2018b; Di Guardo and Mazzariol, 2019). Genetic features associated with the susceptibility or resistance to this virus, have been the basis of recent studies of other odontocete species, mostly suggesting promising candidate genes involved with innate and adaptive immune response (Batley et al., 2019; Batley et al., 2021). The presence and dominance, at the population level, of specific phenotypes, such as the suggested case of the lymphoid tissue characteristics comparing striped dolphin and bottlenose dolphin (Centelleghe et al., 2019), can also play an important role in shaping the response to this infection.

Nonetheless, either at individual or population levels it is also important to consider that the immune status can be influenced by other

factors, like the case of chemical pollution, recognized as one of the main threats to cetaceans (Temple and Terry, 2007; Fossi et al., 2013; Desforges et al., 2016). The Mediterranean Sea is well-known for its intense anthropogenic pressure, as well as its characteristic configurations, such as being a semi-enclosed area surrounded by highly industrialized manufacturers, which undoubtably predispose it to the presence of high levels of chemical pollution (Fowler, 1987; Lejeusne et al., 2010), which in turn contributes to the presence of the virus and the severity of its impact on the hosts (Kannan et al., 1993).

Therefore, given that the Mediterranean population is relatively isolated and with apparent no substructure, and that changes in genetic composition and functional variation (Figure 7) are suggested to be associated with morbillivirus epizootics, this species in this region represent a unique model to study genetic adaptation to morbillivirus resistance. Furthermore, because this species may represent a potential morbillivirus reservoir in the Mediterranean Sea, this study also has important conservation value.

1.7 Aim of the study

The present project investigates the genetic diversity of the Mediterranean striped dolphin, as well as patterns of genetic adaptation in genes of the immune system, relative to well-described morbillivirus epizootic events. Understanding these processes is difficult due to the lack of wild models where changes in genetic composition can be tracked as the

organisms experience pathogenic challenge (Penczykowski et al., 2016). This study addressed those challenges by carrying out "real-time" genetic characterization of the micro-evolutionary patterns involved in the adaptation to morbillivirus infection.

For this purpose, using an exome-capture approach, changes in genetic variation at selected protein coding regions (exons) of immune system genes were assessed for specimens of an existing archive of striped dolphin samples, spanning a period of 18 years, during which time the populations were naturally exposed to 2-3 morbillivirus epizootics. Samples from the same periods were also genotyped, through a double digest Restriction Site Associated DNA Sequencing (RADseq) approach, so non-functional regions of the genome could also be included. This way, both functional and neutral genetic variation were tracked in "real-time", as the population experienced well described pathogenic challenges, also allowing a more detailed look into the genomic diversity among *S. coeruleoalba* in the Mediterranean Sea.

The main questions addressed were as follows:

1. Is there genomic differentiation and geographic structure among *S. coeruleoalba* populations in the Mediterranean Sea?
2. Is genomic diversity (e.g. high inbreeding or selection at certain physiological traits) associated with epizootic susceptibility in *S. coeruleoalba*?

3. Are morbillivirus epizootic events leaving specific genomic signatures in *S. coeruleoalba* that are unlikely to be explained by other evolutionary forces?

The expected results for each task were as follow:

1. Patterns of genetic diversity and spatial structure within the Mediterranean Sea, if present, were expected to be revealed by the presence of clear separated clusters between regions. This would be revealed by population genetic analyses including assessing inbreeding levels, linkage disequilibrium, allelic diversity, heterozygosity levels, principal component analysis (PCA), and demographic history among others. Structuring patterns were compared with previous studies.
2. If physiological traits other than immune system function are involved in epizootic susceptibility, then these were expected to be disproportionally represented in genome-wide scans for selection between epizootics. Furthermore, if effects such as high inbreeding levels were involved in determining epizootic susceptibility, then estimates of (e.g.) inbreeding from non-functional regions of the genome should correlate with periods of known epizootics relative to non-epizootics.
3. Patterns of genetic variation at genes with potential functions in resistance to morbillivirus epizootics are expected to follow the predictions of well-known host-pathogen co-evolution models. The

exact model is expected to depend on each mechanism of immunity involved in the product of each individual gene (although this might not be known for certain genes analysed). Some genes may show a pattern of cyclical variation in allele frequencies between different epizootics, as predicted by the gene-for-gene model. Other genes may show a pattern of repeated introduction of new alleles as expected under a matching-alleles model. Furthermore, if such changes reflect an adaptive process, then genetic variation at such genes is expected to deviate from neutral expectations.

Chapter 2

Genomic diversity among *Stenella coeruleoalba* in the Mediterranean Sea

2.1 Introduction

Survival, adaptability and long-term persistence of populations is inherently associated with genetic diversity. As the essence of molecular ecology studies, genetic analysis has been widely applied to identify signs of diversity, structure, local adaptation and selection, which is of utmost importance, particularly when it comes to the study of wild and highly mobile species. It is known that, despite the apparent absence of physical barriers to dispersal in the marine environment, population structure can be present even at small scales (Norris, 2000; Bierne et al., 2003). Several factors have been suggested to generate population structure, namely environmental pressures – exerting differential adaptation, habitat requirements, abrupt regional demographic variations or isolation-by-distance (Bahri-Sfar et al., 2000; Rolland et al., 2007; Galarza et al., 2009; Mendez et al. 2011; Amaral et al. 2012; Casado-Amezúa et al., 2012). Population structure can be determined when it decreases the level of gene flow, and thus the genetic diversity of a population and its chances to succeed (McKnight et al., 2017).

Given the geographically complex characteristics of Mediterranean Sea and the high levels of biodiversity present, it is not surprising that this is a well-studied area, where population structure patterns have been identified for various marine species (Natoli et al. 2005, 2006; Carreras-Carbonell et al. 2006; Carreras et al., 2006; Domingues et al. 2007; Gaspari et al. 2007; Rolland et al., 2007; Fontaine et al., 2014; El Ayari et al., 2019; Violi et al., 2023).

The striped dolphin (*Stenella coeruleoalba*, Meyen 1833), the most common and abundant cetacean in the Mediterranean Sea (Archer and Perrin, 1999; Gaspari et al., 2007), has been the study model of increasing genetic research in this region, particularly during the last three decades, severely affected by a series of morbillivirus epizootics (Di Guardo and Mazzariol, 2013). Genetic differences between Atlantic and Mediterranean striped dolphin species is well supported (Bourret et al., 2007; Garcia-Martinez et al., 1999; Gaspari et al., 2007; Gkafas et al., 2017). For comparisons within the Mediterranean Sea, the presence of population structure and geographic subdivision has been reported, especially between the eastern and western basins, but the level of differentiation is less consistently reported (Valsecchi et al., 2004; Gaspari et al., 2007, 2019; Gkafas et al., 2017). Levels of population structure are much higher in some other delphinid species across the same regions (Natoli et al., 2005, 2006; Gaspari et al., 2013, 2015). Lower genetic diversity (suggesting inbreeding), has been associated with an increased susceptibility to morbillivirus (Valsecchi et al., 2004).

Although most studies on population genetics in the past have been based on mitochondrial DNA (mtDNA) or microsatellites, there is more recently a focus on genome-wide analyses. Double digest restriction-site associated DNA sequencing (ddRADseq) is a fractional genome sequencing strategy that allows the identification, verification and scoring of a large number of single nucleotide polymorphisms (SNPs) (Peterson et al., 2012). This method does not require a reference genome (though having one helps

identify more SNPs), is less expensive than whole-genome sequencing and allows the adjustment of the genome coverage degree by selecting various restriction enzymes and fragment sizes.

In this study, population structure of the Mediterranean striped dolphin was analysed, using the ddRADseq technology. Samples were included from an 18-year period, covering multiple epizootic events and including primarily samples from the Spanish coast near Valencia, but also smaller numbers from other Mediterranean regions and the North Atlantic Ocean. This study assesses population structure, primarily in the western basin of the Mediterranean, diversity and demographic history.

2.2 Methods

2.2.1 Sample Selection and Preparation

Tissue samples were collected from 154 dead, stranded or fisheries bycatch related adult striped dolphins, between 1990 to 2008, along the Mediterranean coastal region (Figure 9). Information on each individual was limited and important factors such as age, cause of death and contamination load were not available. From a total of 154 individuals, 70 samples from Valencia were already analysed and published in previous studies (Gkafas et al., 2020), using the same restriction enzymes and protocols, and so could be directly compared. Samples were collected and preserved in 20% DMSO NaCl 5M buffer and stored at -20ºC.

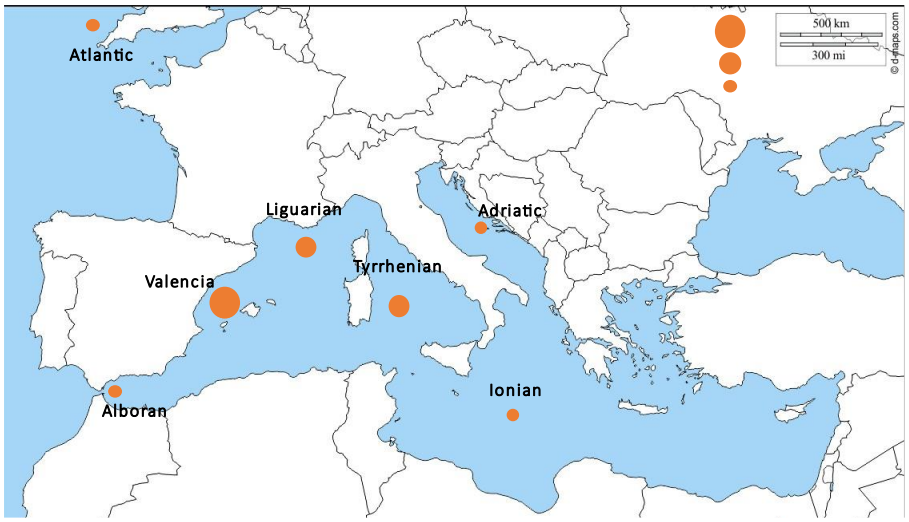


Figure 9 – Geographic areas of all sampled striped dolphin (*Stenella coeruleoalba*) individuals in the Mediterranean Sea and North Atlantic Ocean used in this study. Orange circles sizes are proportional to the number of samples available for each location.

Sample distribution per year and per period was done according to sample availability, quality and balance between periods (Table 1 – final sample selection). Samples from the Alboran Sea and Valencia were considered as the western end of the Western Mediterranean Basin, whereas samples from the Ligurian and Tyrrhenian Seas were the eastern end of the Western Mediterranean Basin, while the Ionian and Adriatic samples (three samples) were from the Eastern Mediterranean Basin. This is primarily a comparison between the western and eastern ends of the Western Mediterranean Basin, since there were only three samples from the eastern basin. There were also three samples from the eastern North Atlantic.

Table 1 - Final number of samples per location/year and respective retained reads for chapter 2. Atlantic samples (1 from year 1994 and 2 from year 1995) were used as control for the Mediterranean region. Valencia contained 70 samples from Gkafas et al. (2020).

Year	Number of Samples	Retained Reads
1990	24	101255073
1991	3	17605003
1992	3	5877582
1993	6	13526172
1994	1	1010528
1995	2	7420386
1996	5	21451201
1997	8	59580636
1998	2	9988443
1999	3	10778142
2000	4	19226882
2001	4	14433645
2002	4	13862127
2003	8	31379730
2004	9	63319823
2005	3	10500026
2006	7	41272547
2007	9	36408094
2008	13	72325149
Atlantic	3	15201618

Location	Number of Samples	Retained Reads
Ligurian	22	116502398
Tyrrhenian	10	34661780
Alboran	5	36279465
Ionian	2	3763381
Valencia	78	349100559
Adriatic	1	10913606
Atlantic	3	15201618
Total	121	566422807

DNA quantification was done first using Nanodrop, as a relatively quick and accurate way to quantify DNA concentration and levels of purity in the samples. However, for Next Generation Sequencing (NGS) purposes, nanodrop is known to sometimes overestimate the amount of DNA present in a sample, leading to sub-optimal results. Therefore, once a subset of suitable samples was selected, quantification was again done using the Qubit 2.0 high sensitivity kit.

DNA extraction was performed using QIAamp DNA mini kit (QIAGEN) and following standard manufacturer instructions. DNA quantification was estimated using a Qubit 2.0 fluorometer high sensitivity kit (Life Technologies) and integrity was checked using 1% agarose gel electrophoresis stained with ethidium bromide. Before proceeding to the next step, DNA concentration was normalized for each sample to estimated 15 ng/μL.

2.2.2 Genomic library preparation and ddRAD sequencing

A DNA library – a collection of sequencing-competent DNA fragments of 84 samples (7 pools of 12 samples each), was prepared for ddRADseq (Figure 10) following Peterson and collaborators (2012), with modifications.

The targeted genome was digested and fragmented using two restriction enzymes – MspI (4bp cutter) and HindIII (6bp cutter), and then attached to a series of adapters designed in-house (Appendix A, Table S1). Samples individually barcoded with a unique adapter, were then pooled into a multiplex of 12 individuals, and cleaned with Purelink PCR Micro Kit columns (Thermofisher). Size selection was performed with Pippin Prep electrophoresis gel (Sage Science), according to the machine protocol, and a base pair range selection of 273-372 bp. To uniquely identify amplified products from the reaction, each pool of samples was then double indexed with PCR, using a Phusion™ Polymerase kit (Thermofisher), followed by one more cleaning step with Invitrogen PureLink PCR Micro Kit columns (Thermofisher) to remove PCR by-products. A quality control stage followed, using TapeStation (Agilent Bioanalyzer) – to quantify molarity and library fragment size of the pools, as well as quantitative PCR (qPCR) using KAPA SYBR® FAST kit (Roche), as an accurate quantification step. After the verification of concentrations of each pool, a final library of 2 nM concentration was then produced from all pools, ready for sequencing.

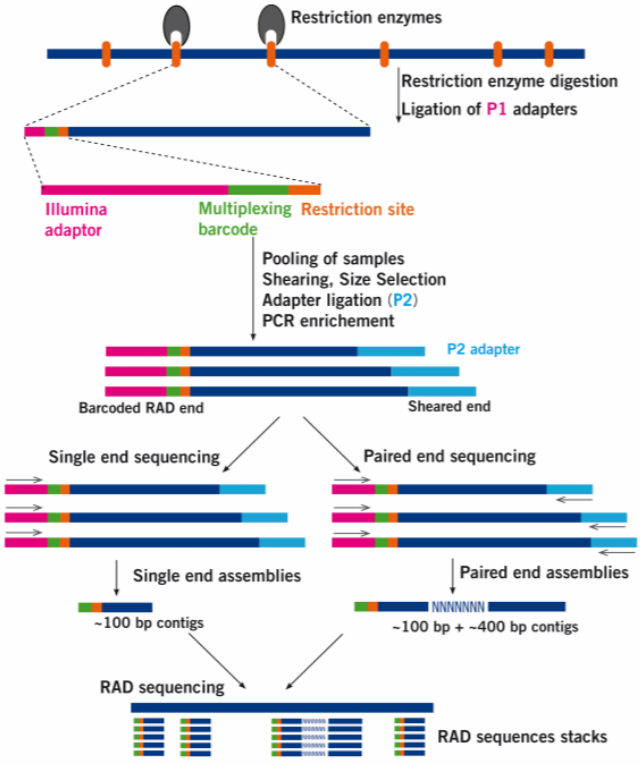


Figure 10 - RAD sequencing overview (adapted from Floragenex).

Sequencing was performed on an Illumina HiSeq2500 platform in paired-end (2X 125 bp) read mode, using one lane at DBS Genomics of Durham University.

2.2.3 Bioinformatic analysis

The demultiplex of the sequenced data and sorting into individual samples was performed using *process_radtags* with STACKS v.2.2 (Catchen et al., 2013). Samples retained (with more than 1M reads), were mapped against the *Tursiops truncatus* reference genome (mTurTru1.mat.Y RefSeq, GenBank Assembly ID: GCA_011762595.1) using BWA v.0.7.17 and BWA-mem2 v.2.2.1 (Li and Durbin, 2009), followed by sorting and data indexing with SAMTOOLS v.1.16.1 (Li et al., 2009). A mapping quality filter was applied, also with SAMTOOLS, right before SNP calling with STACKS v.2.2 (Catchen et al., 2013). *Refmap* and *populations* modules were applied to search for variants and remove loci present in fewer than 80% of individuals (Catchen et al., 2013). The output files of SNP data were later converted into PED and MAP formats, using PGDspider (Lischer and Excoffier, 2012) or PLINK (Purcell et al., 2007), and VCFTOOLS v.0.1.16 (Danecek et al., 2011). Following Gfakas et al. 2020, the VCF file was filtered to include a minimum of 80% of taxa for which a SNP was scored, a minor allele frequency of 0.05, a minimum depth value of 5, and a minimum genotype quality of 5. Read depth per individual and SNP was also calculated using VCFTOOLS, and binary files (BED, RAW and BIM) were generated using PLINK.

Population structure and diversity was then accessed in RStudio v4.2.2 (R Core Team, 2020) using the package *SambaR* (de Jong et al., 2021).

The use of two different datasets with distinct sample numbers was applied to allow structure comparisons between the Atlantic and Mediterranean individuals. For both, the same filtering steps were applied, in order to avoid low quality samples and SNPs, or an excess of rare alleles, thus enhancing the power of the analysis (de Jong et al., 2021). Two datasets were generated according to the inclusion/exclusion of Atlantic samples and differentiation between populations. One dataset was composed of 118 samples: 83 individuals from the western end of the western basin + 32 individuals from the eastern end of the western basin + 3 individuals of the eastern basin of the Mediterranean Sea. The other was composed of 121 samples, where 3 Atlantic individuals were included. After imported to R, the data were filtered in order to exclude samples and loci with more than 20% of missing data, and to select one SNP per 500 bp region in order to avoid linkage disequilibrium issues. Next, the function *findstructure* of *SambaR* was run, producing a wide variety of outputs, which included the principal component analyses (PCA) and the principal coordinate analyses (PCoA), as well as the *calcdistance*, *calcdiversity,* and *calckinship* functions to access F_{ST}, heterozygosity and inbreeding levels, respectively. F_{ST} was also assessed using Arlequin v.3.5 (Excoffier and Lischer, 2010) for the dataset without the Atlantic individuals.

The population dynamics associated with the historical demography of the Mediterranean samples was inferred using Stairway plot v2.1.1 analysis (Liu and Fu, 2020). The folded site frequency spectra (SFS), was calculated for each population, using the easySFS software (Gutenkunst et al., 2009). Following Taylor et al. (2007), the mutation rate was set to 2.5 x 10^{-8} substitutions per nucleotide per generation, and the generation time to 22 years. To estimate trajectories of historical effective population size (Ne) using a method based on linkage disequilibrium, SNeP software v1.11 (Barbato et al., 2015; Corbin et al., 2012) was used.

Genome-wide heterozygosity was also calculated as another way to indirectly estimate the effective population size. For that, BCFTOOLS v.1.15 (Danecek et al., 2021) was used, generating a file including all monomorphic and polymorphic sites. This procedure varies from the heterozygosity generated in SambaR which contains polymorphic sites (more specifically: biallelic SNPs), and hence heterozygosity values are typically higher. Based on the obtained results, the intraspecific variability was estimated with Kimura's (1991) formula:

$He = 4.Ne.\mu$

where Ne is the effective population size, and μ is the mutation rate per nucleotide site per generation.

2.3 Results

The ddRAD sequencing generated a total of 282 million reads from the original library of 84 striped dolphin samples. After quality control –

and loss of samples due to low quality, filtering steps and the addition of the sequencing data of the 70 samples from Valencia (Gkafas et al. 2020), there were 566 million reads (Table 1). Calling SNPs resulted in a total of 53702 SNPs for the dataset with Atlantic individuals, and 49810 SNPs for the dataset without them, of which 13590 and 14240 remained after filtering, respectively. The proportion of retained individuals and SNPs after filtering steps applied can be seen in Figure S1 of the supplementary material.

Correspondence analyses (CA), principal component analyses (PCA), multi-dimensional scaling (MDS) (Figure 11), principal coordinate analyses (PCoA) (Appendix A, Figure S3), suggested a weak genetic differentiation pattern between East and West regions of the western basin of the Mediterranean Sea (and including three samples from the eastern basin), with some overlap between them. A more pronounced differentiation between the Mediterranean and Atlantic individuals, was seen when the three samples from the North Atlantic were included (Figure 11 B, D, F).

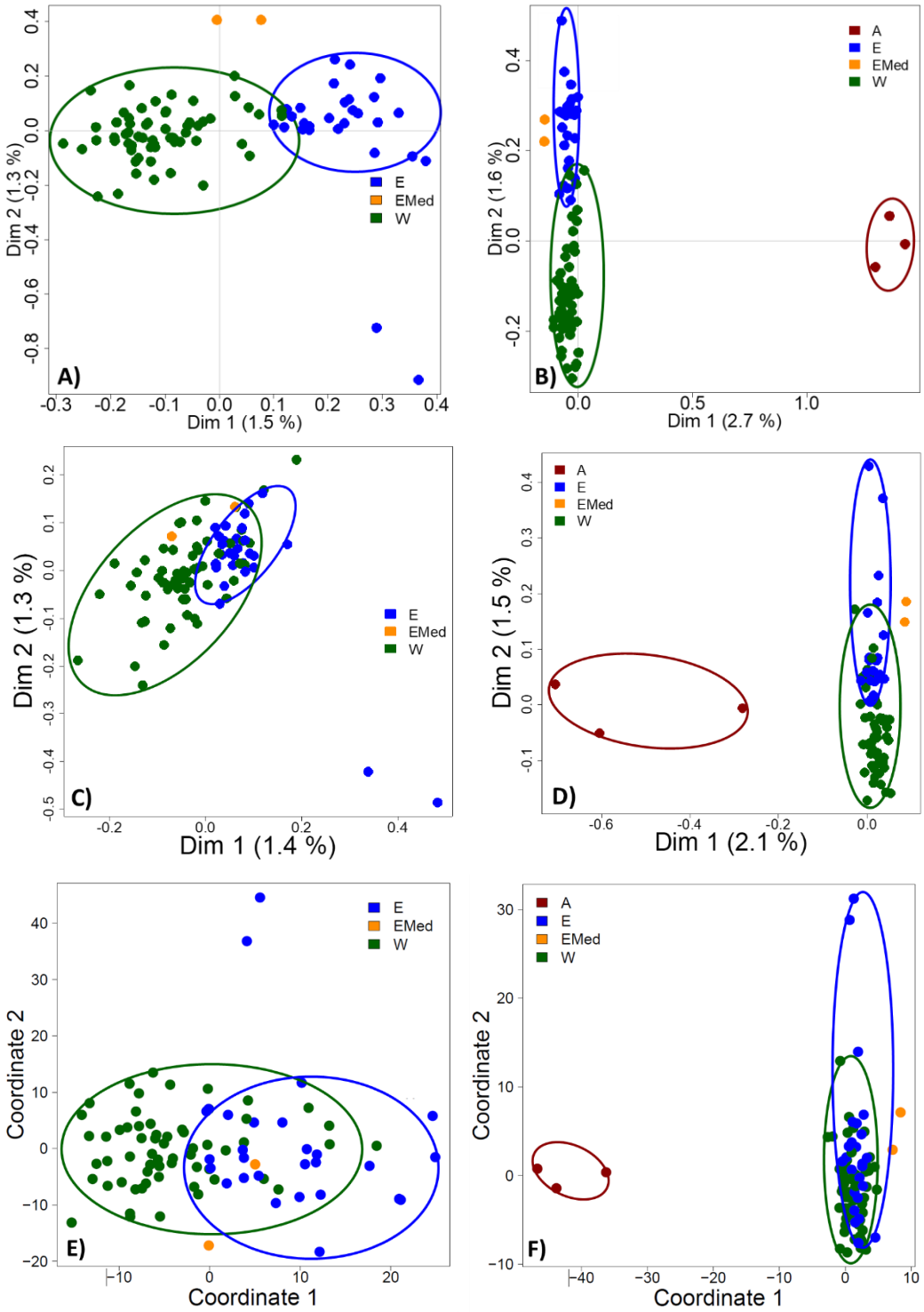


Figure 11 - Several analysis with the same pattern of structure for the dataset without (left) and with (right) Atlantic individuals, together with West (W) and East (E) of the western basin, and East basin (EMed) of the Mediterranean Sea. A,B) Correspondence analyses (CA); C,D) Principal component analyses (PCA); E,F) Multi-dimensional scaling (MDS).

Genetic differentiation between the Mediterranean groups was also assessed by pairwise F_{ST} analysis, which corroborates the weak differentiation between East and West regions of the western basin, with a value of 0.00056 (NS). Overall, heterozygosity estimates showed levels of genetic diversity consistent with other abundant species (see Discussion; Figure 12, Appendix A, Figure S3, Table S2).

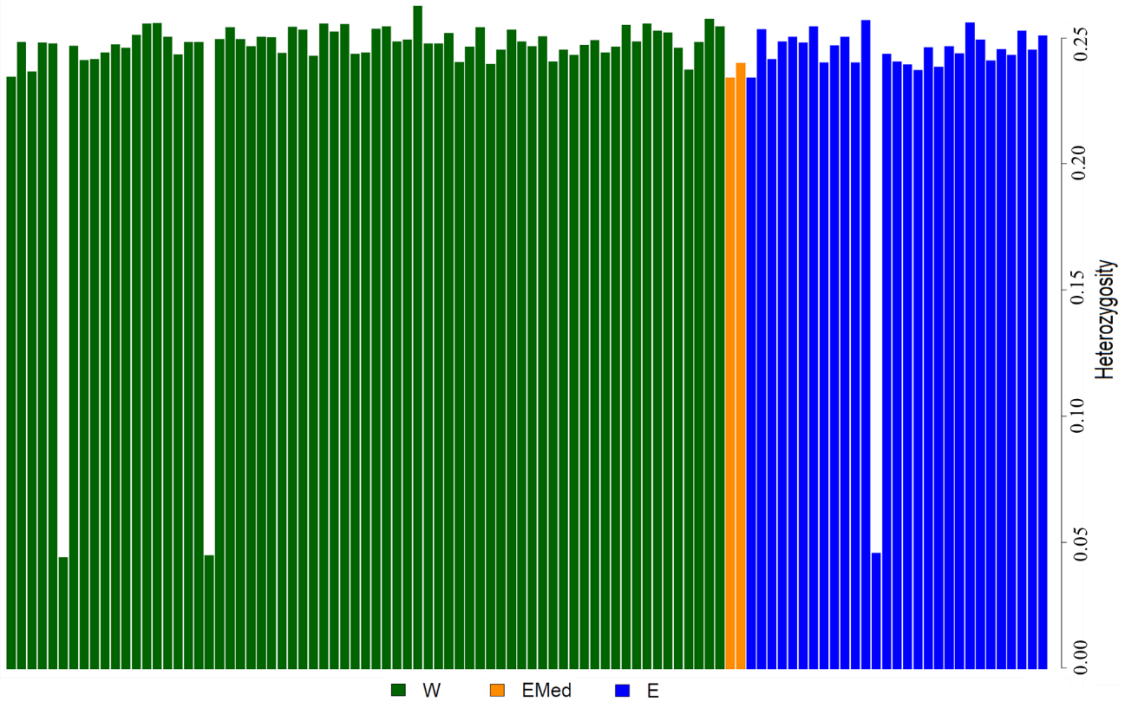


Figure 12 - Levels of heterozygosity per individual from the West (W) and East (E) of the western basin, and East basin (EMed) of the Mediterranean Sea.

Stairway plot profiles (Figure 13) for both regions of the Mediterranean Sea suggested a similar pattern of the effective population size (Ne) over the last 10k-20k years, to the present.

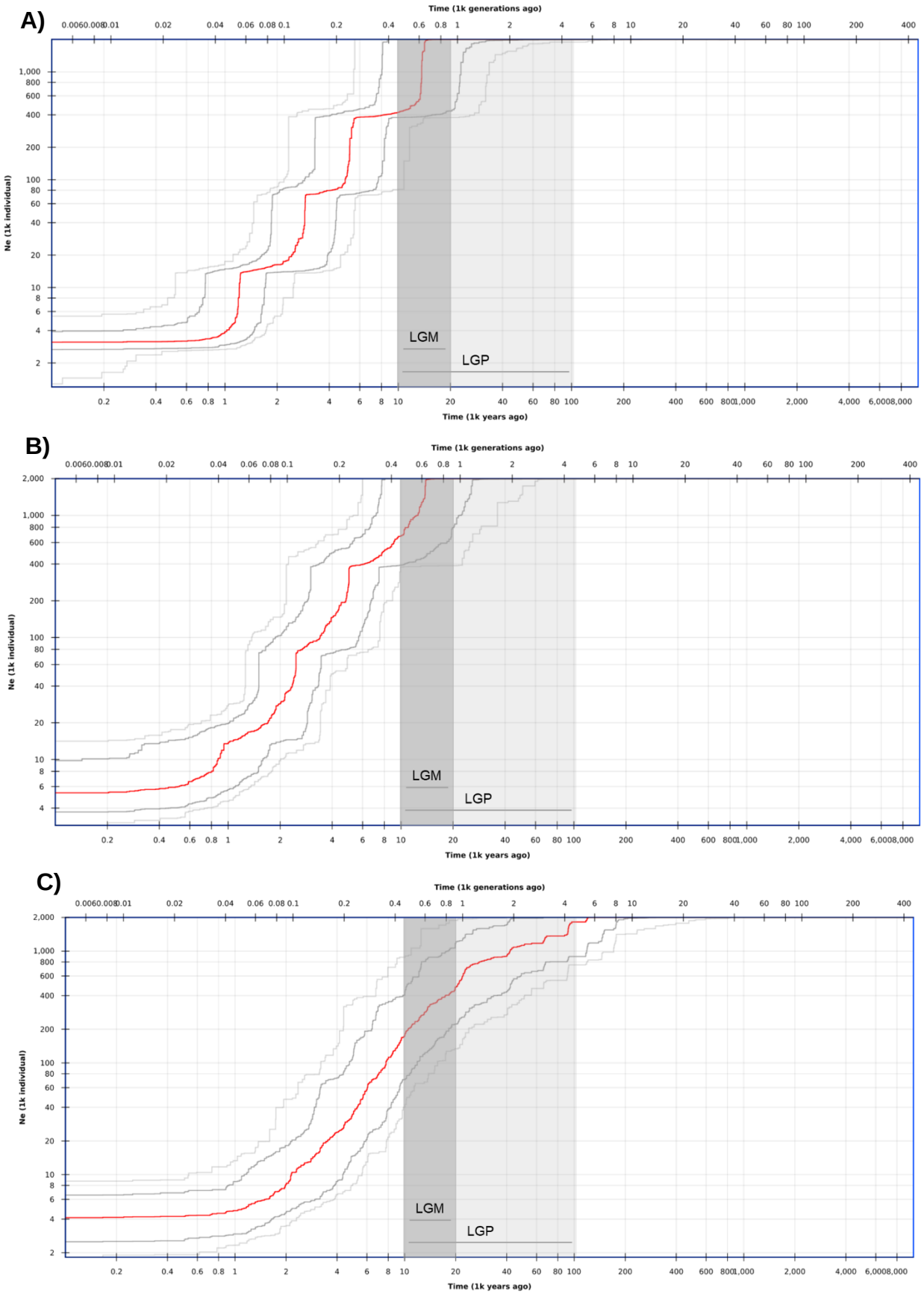

A)
B)
C)
Time (1k generations ago)
Time (1k years ago)
Ne (1k individual)
LGM
LGP

Figure 13 - Demographic history reconstruction by stairway plots for the A) West cluster of the western basin, B) East cluster of the western basin and east basin, and C) overall region of the Mediterranean Sea, up to 600 kya. The last glacial period (LGP) and last glacial maximum (LGM) are represented. Time is shown in thousands of years. Both axis are log scale.

Overall, the demographic history reconstruction starts within the last glacial period (LGP) and an abrupt population decline is evident, particularly after the last glacial maximum (LGM ~20kya), with an apparent plateau starting approximately 1-2kya. A more robust estimate of recent and current Ne can be gained using inference from linkage disequilibrium (using the program SNeP). This showed a similar pattern of abrupt decline (~7k) over the last glacial period in the western region, while a more steady dynamic could be observed for the eastern region over the same period of time (Figure 14).

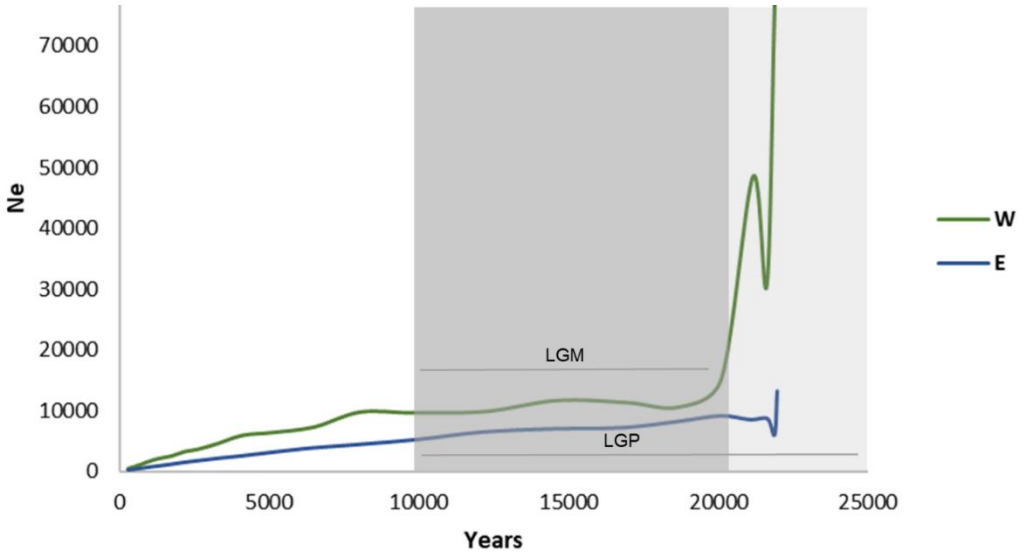


Figure 14 - Demographic reconstruction by SNeP, showing the demographic history of the of West (W) and East (E) clusters, up to 25 kya. Last glacial period (LGP) and last glacial maximum (LGM) are represented.

After 20ka, both revealed a more steady profile until the present. The stairway plots suggested an Ne of approximately 5000, as did SNeP until the very recent past when the SNeP estimates decline to zero – likely an artifact of the analysis. However, both methods seem to present low numbers of the current effective population size comparing with the recent census estimates. Although the cause of these low values is unknown, various factors can affect Ne such as variation in reproductive success, migration, and mortality rates. Current estimates of striped dolphin abundance and distribution, amount to 419,361 individuals, with the following distribution (ACCOBAMS, 2021): Western Mediterranean - 315,789; Central Mediterranean - 45,042; Adriatic Sea - 10,264; Eastern Mediterranean - 48,266.

Results from genome-wide analysis generated by BCFTOOLS (Appendix A – Table S2), revealed a high number of segregating sites, sites with low allele frequency, and a level of heterozygosity of 0.002. Using this value, together with the mutation rate per generation for this species ($2.5x10^{-8}$) in Kimura's (1991) formula, the result for the long-term average estimate of Ne was 20000 individuals. Long-term Ne reflects an average over the past and can be higher than current estimates.

2.4 Discussion

Genetic diversity and population structure can be influenced by environmental variables together with biotic factors, which over time can

lead to phylogeographical breaks and diversification shifts (Avise, 2000; Perry et al., 2005). In the present chapter, genomic-wide characteristics were explored to assess the levels of genetic diversity, geographic structure and historical demography of the Mediterranean striped dolphin.

For the population structure analysis, the present data suggested weak differentiation with overlap between the individuals from the west and east region of the western basin of the Mediterranean Sea, according to the PCoA, CA, PCA, and MDS analysis performed, though the F_{ST} comparison was low and non-significant. The genome-wide heterozygosity (0.002), was within the range expected for mammals (Teixeira and Huber, 2021). The low percentage levels of genetic variance explained by each coordinate, as well as the non-significant pairwise F_{ST} result, reinforce the suggestion of weak differentiation between these Mediterranean groups. However, using the same analysis, a more evident distinction was indicated from the ordination analyses between the Mediterranean and Atlantic individuals.

Other studies have found significant differentiation for the striped dolphin both between the Atlantic and the Mediterranean and within the Mediterranean region, though the regions compared and sample sizes typically differed (e.g. Bourret et al., 2007; Garcia-Martinez et al., 1999; Gaspari et al. 2007, 2019; Gkafas et al., 2017). For example, Gkafas et al. (2017) found a relatively high F_{ST} (0.03) but compared the western basin with samples from Greece and further east. Population genetic structure within the Mediterranean has also been described for various other

delphinid species. There is evidence for a relatively strong break between the eastern and western basins, but in other cases the same environmental conditions may imply different impacts on different taxa. For example, common dolphins (*Delphinus delphis*) showed no clear population division between the Mediterranean and the North Atlantic, but did between the eastern and western basins (Moura et al., 2013a), while bottlenose dolphins (*Tursiops truncatus*) showed multiple differentiated populations across the Mediterranean and including the boundary with the North Atlantic (Natoli et al., 2005; Louis et al., 2014).

Putative phylogeographical breaks have been described between the Mediterranean region and the North Atlantic for several taxa (e.g. bottlenose dolphin – *Tursiops truncatus* Natoli et al., 2005; common dolphin – *Delphinus* sp. Natoli et al., 2006; harbour porpoises – *Phocoena phocoena* Fontaine et al., 2014; sperm whale – *Physeter macrocephalus* Violi et al., 2023; common sole – *Solea solea* Rolland et al., 2007). In the open sea, no physical barriers are recognized, but certain areas may promote different levels of transition and isolation, due to their specific environmental variables potentially preventing gene flow (Limborg et al., 2012). An example of this is the Almeria Oran Front, considered as a natural marine barrier to gene flow possibly by vicariance, temperature, salinity, and hydrography (Patarnello et al., 2007; Galarza et al., 2009). This has been proposed for several species (e.g. Mediterranean mussels, *Mytilus galloprovincialis*, El Ayari et al., 2019; and bottlenose dolphins,

Tursiops truncatus, Natoli et al., 2005). The front is just east of the Alboran Sea, west of where most samples for this study were collected.

Historical context, including relevant biotic and abiotic factors, acts as a background of information in the differentiation process and dispersion pattern of species, which help us to better understand them. The demographic reconstruction applied to the Mediterranean regions, revealed a similar pattern between them, with a decline in effective population size (Ne) levels since the last glacial period. The Pleistocene epoch was marked by a series of environmental changes, with consequences for numerous relevant biotic interactions, though the specific drivers are not known (Hewitt, 2004; Banguera-Hinestroza et al., 2014; Moura et al., 2014a). Changes related to sea level during these glacial periods (Lambeck and Purcell, 2005), may also have played relevant roles on species dispersion and colonization within the Mediterranean. It is also known that this region functioned as a refuge for numerous species during the last glacial period (Patarnello et al., 2007). The change in oceanographic features may have contributed to the creation of isolated refugia and also promoted the increasing of barriers to gene flow. Post-glacial expansion re-distributed some populations (Xue et al., 2014), impacting the local biodiversity that exists today (Patarnello et al., 2007). Ultimately, this may have contributed to an increase of vicariance between this region and the Atlantic, and also between the basins within the Mediterranean Sea, as suggested by Gkafas et al. (2017). There are clear differences between the effective population size estimated here and the

census population size (Nc – expected to be 10x higher than Ne; Frankham, 1995). Recent estimations suggested a total of 419,361 striped dolphin individuals within the Mediterranean Sea (ACCOBAMS, 2021), closer to 100 times higher than Ne. It is possible that population cycling driven by the epizootics has depressed Ne (see Vucetich et al. 1997). Nonetheless, it is also worth acknowledging that the present analysis using stairway plots was based on the allele frequency spectrum from DNA sequenced data, previously mapped against a reference genome from another species. Differences between the bottlenose dolphins and striped dolphins go beyond the genetic level, also involving ecological and biological processes, which can also impact their demographic history. Thus, interpretation of these results should be taken cautiously.

The eastern and western Mediterranean basins are composed of Ionian, Aegean and Levantine Seas in the east, and Alboran, Balearic, Ligurian, Strait of Sicily, and Thyrrenian Seas in the west (Theocharis et al., 1993). All of these areas are considered to have unique oceanographic features, such as salinity, nutrient, and circulation levels, as well as dynamics (Theocharis et al., 1993; Kourafalou et al., 2004). It may be possible that this subdivision of habitat systems helps promote some level of population differentiation within the basins, though the evidence for this is weak for striped dolphin in the regions assessed. Further studies on these basins would be important to better understand the existence or absence of genetic structure in this region, targeting a considerable and balanced sample number of free ranging animals to avoid possible bias

related to oceanic currents and strandings, particularly when it involves highly mobile species. Additionally, complete information on important factors of each individual such as age, sex, cause of death and contaminant load would be essential to accurately interpret the present results.

Understanding how each species responds to environmental variables, subtle habitat differences, and particular ecological niches, and what drives population structuring over time, is of utmost importance, particularly when it comes to effective conservation measurements. The present study considers the most common cetacean species of the Mediterranean region, and finds relatively weak population structure within the western basin. However, taking into account the environmental threats (mostly anthropogenic) that it faces worldwide, its present classification in the IUCN red list – Vulnerable in the Mediterranean (Aguilar and Gaspari, 2012), and poor data for the rest of the world, it is critical to increase the available knowledge to support the effective conservation of this species. This is particularly relevant when, in the recent decades, this species is also considered to be the most susceptible cetacean to one of the major causes of death in marine mammals worldwide – morbillivirus infection, which will be considered in more detail in the next chapters.

Chapter 3

Genomic diversity associated with epizootic susceptibility in *Stenella coeruleoalba*

3.1 Introduction

Since 1980 (Hersh et al., 1990; Duignan et al., 1996), cetacean populations have been facing epizootic outbreaks worldwide caused by cetacean morbillivirus (CeMV) (genus *Morbillivirus*, family *Paramyxoviridae*). These outbreaks are often associated with high levels of transmission and mass mortality events leaving several thousands of dead animals (Aguilar and Raga, 1993). Cetacean morbillivirus has been described to include three major groups: porpoise morbillivirus (PMV) (Kennedy, 1998), dolphin morbillivirus (DMV) (Domingo et al., 1990), and pilot whale morbillivirus (PWMV) (Taubenberger et al., 2000; Bellière et al., 2011). An early detection was in *Delphinus delphis* in the Northwestern Atlantic along the northeast coast of the USA (Duignan et al., 1995). In the Mediterranean Sea, the first cases were reported in 1990 (Domingo et al., 1990; Di Guardo and Mazzariol, 2013). Over the past 33 years, three epidemic outbreaks have occurred, from 1990-1992, 1997-1999 and 2006-2008, suggesting a reoccurring cycle of 3-5 years (Raga et al., 2008; Van Bressem et al., 2009; Di Guardo et al., 2013; Di Guardo and Mazzariol, 2013). Although not associated with mass die-off events, mortality related to morbillivirus continues to be reported, namely in 2009-2011, 2013, 2016, 2018-2021 (Rubio-Guerri, et al., 2013; Casalone et al., 2014; Van Bressem et al., 2014; Mira et al., 2019; Vargas-Castro et al., 2023). Milder symptoms, together with high prevalence levels of CeMV but low infection rates, have been suggested to be not associated with epizootic outbreaks (Vargas-Castro et al., 2023). This persistence suggests the existence of an endemic infection

cycle in this region, which can cause systemic infections (Rubio-Guerri, et al., 2018). The infection resurgence, together with the close genetic proximity found between the multiple DMV strains circulating among the Mediterranean striped dolphins, suggests low levels of herd immunity against the virus in dolphin populations (Van Bressem et al., 2001, 2014; Keck et al., 2010; Bellière et al., 2011; Di Guardo and Mazzariol, 2013). It is known that chemical pollution has a considerable impact on the immune system function of cetaceans and contributes to the presence of virus and parasites (Kannan et al., 1993; Nakayama et al., 2009; Fossi et al., 2013; Desforges et al., 2016). In the particular case of the Mediterranean striped dolphin, pollutant contamination has been hypothesised as an important contributor to Morbillivirus susceptibility (Aguilar and Borrell, 2005; Profeta et al., 2015).

Viral adaptation to the environment and its hosts has been suggested, with the circulation of a new DMV strain that possibly came from the northeast Atlantic, in particular from the coasts of Galicia and Portugal, into the Mediterranean Sea. It was potentially introduced by pilot whales (*Globicephala melas*) or striped dolphins, as both have been proposed to be DMV carriers (Duignan et al., 1995; Morris et al., 2015; Mira et al., 2019; Pautasso et al., 2019; Cerutti et al., 2020). Although morbillivirus impacts a wide range of marine mammal species (Kennedy, 1998), the striped dolphin is apparently especially vulnerable (Cotté et al., 2010). This is one of the most cosmopolitan marine mammals (listed by the IUCN Red List of Threatened Species as 'least concern' globally), however

the population in the Mediterranean Sea is listed as “vulnerable” (Aguilar and Gaspari, 2012). It is known to be highly gregarious, and this characteristic is thought to have contributed significantly to the spread of the infection (Van Bressem et al., 2014). Although striped dolphins are not the only gregarious cetacean species and contaminant load, particularly in a semi-enclosed area, has been suggested to exacerbate their vulnerability to this virus, when it comes to susceptibility this is a worldwide scenario for this species.

Although there is evidence for some genetic structure for this species within the Mediterranean, the level of differentiation is relatively weak (Valsecchi et al., 2004; Gaspari et al., 2007; Gkafas et al., 2017; Gaspari et al., 2019; Chapter 2). Lower genetic diversity within the Mediterranean and restricted gene flow with the North Atlantic have also been suggested (Garcia-Martinez et al., 1999; Bourret et al., 2007; Gaspari et al., 2007; Chapter 2). Inbreeding has been correlated to impacts on the fitness, health and survival of many species (e.g. Coltman et al., 1998; Amos et al., 2001; Acevedo-Whitehouse et al., 2003, Gkafas et al. 2020). However, for striped dolphins, different outcomes have been suggested regarding inbreeding levels and their influence on mortality associated with morbillivirus epidemic events. During epizootic events, Valsecchi and co-workers (2004) reported not only elevated levels of inbreeding detected in stranded dolphins, being particularly high in the early outbreak, but also the return to pre-outbreak diversity levels occurring in 10 years after the end of the epidemic. In contrast, Batley et al. (2021) found that levels of

inbreeding in *Tursiops aduncus* populations had no detectable influence on the outcome for individuals. However, they found 50 genes that were good candidates for studying the complex cetacean immune system response towards the morbillivirus threat. To the best of my knowledge, this is the only study that has attempted to understand the vulnerability and susceptibility of dolphins to viral infection at the molecular level involving large genome datasets.

Recent technological advances in genetic sequencing have enabled the assessment of large amounts of data and made the generation of those datasets less time consuming and more affordable. In my study I chose double digest restriction-site associated DNA sequencing (ddRADseq), which identifies genetic markers across the genome. This strategy for genome sampling has become common in population genetics studies. It provides a high-resolution assessment of genetic variation among regions resulting from drift (Chapter 2), and a genome-wide assessment of inbreeding at the individual level. The identification of candidate adaptive genes related to relevant ecological functions can also be used to assess the impact on specific gene systems. Although the genotyped nucleotides might not necessarily be within a functional gene, they can be identified as experiencing a selective sweep if within linkage disequilibrium of a gene that is functional and has experienced selection as a result of morbillivirus events. In this chapter ddRAD data are used to investigate possible genomic signature for morbillivirus susceptibility and resistance.

3.2 Methods

3.2.1 Sample Selection and Preparation

Apart from the differentiated North Atlantic samples, all samples from Chapter 2 were pooled for the analyses in chapter 3 (see Figure 9, Table 2). This is justified because signals for structure were weak and F_{ST} values non-significant (see Chapter 2). Instead, samples were divided among epizootic and non-epizootic periods. Procedures for sample selection and preparation were the same as described in Chapter 2.

As mentioned in Chapter 2, the cause of stranding/death of the sampled dolphins, as well as the infectious status for morbillivirus per sampled dolphin was unknown, but it is assumed for the purposes of these analyses, that stranded animals during epizootic events have a higher probability of being infected. Epizootic and non-epizootic periods included were defined according to previous studies (Van Bressem et al., 2001, 2014; Bellière et al., 2011; Rubio-Guerri et al., 2013; Gaspari et al., 2019) (Table 2). Other information details on each individual were limited and important factors such as age, sex (in some cases), and contamination load were not available.

Table 2 - Final number of samples per location/year and respective retained reads for chapter 3. Grey areas represent epizootic periods, while the remaining are reported as non-epizootic period. M/F/? – represents individuals gender Male, Female or the number of non-identified individuals, respectively.

Period	Year	Number of Samples	M/F/?	Retained Reads
1	**1990**	24	16/7/1	101255073
	1991	3	0/2/1	17605003
	1992	3	2/1/0	5877582
2	**1993**	6	3/2/1	13526172
	1994	1	1/0/0	1010528
	1995	2	0/2/0	7420386
	1996	5	3/2/0	21451201
3	**1997**	8	4/4/0	59580636
	1998	2	2/0/0	9988443
	1999	3	3/0/0	10778142
4	**2000**	4	0/4/0	19226882
	2001	4	2/2/0	14433645
	2002	4	1/3/0	13862127
	2003	8	4/4/0	31379730
	2004	9	5/4/0	63319823
	2005	3	2/1/0	10500026
5	**2006**	7	2/5/0	41272547
	2007	9	6/3/0	36408094
	2008	13	9/4/0	72325149

Location	Number of Samples	Retained Reads
Ligurian	22	116502398
Tyrrhenian	10	34661780
Alboran	5	36279465
Ionian	2	3763381
Valencia	78	349100559
Adriatic	1	10913606
Total	118	551221189

A further subdivision was included that focused only on individuals from the peak epizootic and non-epizootic periods (Table 3).

Table 3 - Subdivision of samples from the peak periods, per location/year and respective retained reads for chapter 3. Grey areas represent epizootic periods, while the remaining are reported as non-epizootic period. M/F/? – represents individuals gender Male, Female or the number of non-identified individuals, respectively.

Year	Number of Samples	M/F/?	Retained Reads	Location	Number of Samples	Retained Reads
1991	3	0/2/1	17605003	**Ligurian**	3	10686630
1994	1	1/0/0	1010528	**Tyrrhenian**	4	12462008
1995	2	0/2/0	7420386	**Alboran**	0	0
1998	2	2/0/0	9988443	**Ionian**	0	0
2002	4	1/3/0	13862127	**Valencia**	22	94525673
2003	8	4/4/0	31379730	**Adriatic**	0	0
2007	9	6/3/0	36408094	Total	29	117674311

3.2.2 Genomic library preparation and ddRAD sequencing

The DNA library preparation, as well as ddRADseq protocol steps, were described in Chapter 2.

3.2.3 Bioinformatic analysis

From demultiplex process to VCF filtering, the steps applied were the same as in Chapter 2. To look for a possible relationship between SNPs and the epizootic periods, the dataset was analysed after filtered, using a generalized linear model (GLM) in TASSEL v.5.2.25, with a previous sorting required by TASSEL using the SortGenotypeFile plugin (Bradbury et al., 2007). For the GLM analysis samples were allocated to different categories, namely epizootic vs non-epizootic; epizootic peaks vs non-epizootic peaks (the peak period is defined as the middle years of an epizootic/non-epizootic event: Table 2), as well as the same categories within males and females. The filtered vcf file and the respective trait file (with the samples allocated to the different categories described above) were merged with the *-intersect* flag, and the number of permutations was set to 1000, with the flag *-FixedEffectLMPlugin*. This tested the associations between segregating sites and traits using a fixed effect linear model. Next, the qqman package (Turner, 2018) was applied to produce Manhattan plots – the plot of the negative logarithm of *p-value* for each SNP, allowing their visualization across the genome. A low threshold line ($-\log10 = 10^{-3}$) was chosen in case many loci of relatively low effect were responsible for the trait (other examples of arbitrary threshold line selection: Duncan et al., 2014; Lawal et al., 2018; Coolen et al., 2019, 2023). Dots above the line were selected and merged with the annotated file of the *Tursiops truncatus* reference genome (mTurTru1.mat.Y RefSeq, GenBank Assembly ID: GCA_011762595.1), using the *awk* command and

BEDTOOLS v.2.30 (Quinlan and Hall, 2010). The SnpEff program (Cingolani et al., 2012) was used to assess genomic (within or outside exons), as well as functional effect (missense vs. synonymous changes) information, regarding all SNPs associated with a gene. The VCF file for the main dataset was aligned to the *Tursiops truncatus* genome annotation (as used previously) to provide information on SNP location. Functions associated with the respective genes were explored using information provided by UniProtKB (The UniProt Consortium, 2023). Particular attention was given to genes that may be involved in pathogen resistance pathways, later screened in Ensemble and NCBI websites. The genotypes of each sample were extracted using TASSEL v.5.2.25, after the selection of SNPs of interest (outlier loci from GLM, and missense substitutions where loss of function were represented by stop gained variants) with VCFTOOLS. Statistical significance of genotype frequency differences was compared between epizootic vs non-epizootic periods using the 2X3 Fisher exact probability test (Lowry, 1998) and the p value adjusted to 0.002 (outlier loci), 0.00005 (missense variants), or 0.00250 (stop gained variants) applying the Bonferroni type 1 error correction to avoid false positives.

The frequency of heterozygotes and minor allele frequency (MAF) were analysed for each outlier of the GLM and Manhattan plot analyses, and then plotted for different time periods to uncover possible patterns among the years and epizootic periods. To further assess the existence of possible clusters between epizootic and non-epizootic periods within the

main dataset, as well as within a sample set restricted to the western region, a Factorial Correspondence Analysis was performed using the software Genetix v.4.05.2 (Belkhir et al., 2002). The SNPs were those restricted to missense mutations in coding regions, as identified using the program snpEff.

Genomic diversity metrics, particularly related to inbreeding and heterozygosity-fitness correlations (HFCs), were assessed using the R package InbreedR (Stoffel et al., 2016). In this analysis, a proxy for identity disequilibrium (g_2) was calculated, based on the formula of Hoffman et al. (2014), using 100 permutations and a confidence interval of 0.95 by bootstrapping over individuals (e.g. $g_2 = 0$ implies no variance in inbreeding in the sample). Heterozygosity-heterozygosity correlation coefficients (HHCs; Balloux et al., 2004) were also calculated as another way of estimating identity disequilibrium. However this method is considered to be less robust than g_2 because HHC sample distributions are not independent. Standardized multilocus heterozygosity (sMLH) and its variance (Coltman et al., 1999) were calculated to estimate the average heterozygosity across the genome. The inbreeding coefficient (F_{IS}), was also assessed according to the following function (Kardos et al., 2015):

$F_{IS} = 1-(H_{obs}/H_{exp})$.

The proportion of heterozygous sites (F_H) was calculated based on the following formula (Kardos et al., 2015):

$F_H = 1-(H/N)$

Where H refers to the number of homozygous sites (observed or expected), and N is the total number of sites per individual. Homozygosity information was retrieved from *SambaR* (de Jong et al., 2021) in RStudio v4.2.2 (R Core Team, 2020). sMHL and F_{IS} were later plotted by year to assess the trend over time. The sample size per year is given in Table 2 and 3.

To further assess the impact of inbreeding the correlation between the trait value (*W*, epizootic events) and inbreeding: r^2_{Wf}, and the correlation between the trait value and heterozygosity: r^2_{Wh} were calculated (Szulkin et al., 2010). All analysis were performed for six different SNP datasets (main dataset, peak dataset, females-only main dataset, females-only peak dataset, males-only main dataset and males-only peak dataset). Statistical significance was assessed using a chi-square analysis, and the critical value set to 0.05 (corrected for type 1 error using Bonferroni to avoid false positives).

3.3 Results

From all the scenarios analysed, the GLM and Manhattan plot analyses retrieved only one (comparing peak periods for both males and females) with SNP outliers – above the set threshold (Figure 15).

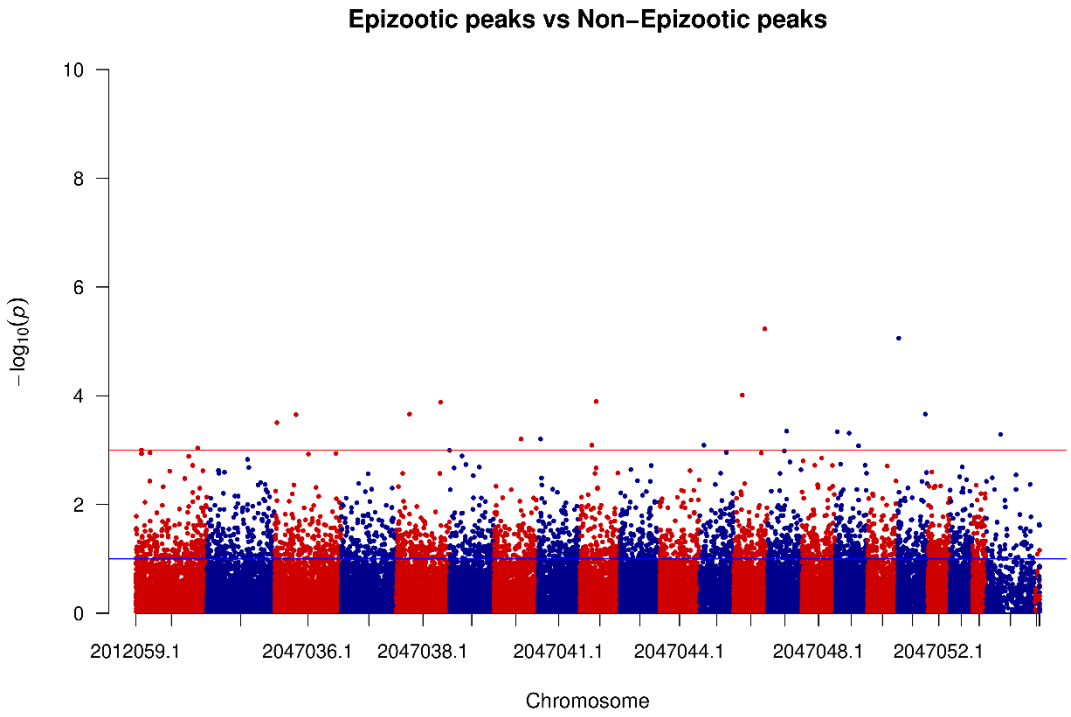


Figure 15 – Manhattan plot with outlier SNPs (above red line – arbitrarily set to 10^{-3}) when comparing peaks of epizootic and non-epizootic periods. Blue and red identify different chromosomes.

When comparing the peak periods between epizootic and non-epizootic events, a total of 23 SNP positions were detected as outliers, across 12 chromosomes. From those, 3 SNPs were in an intergenic region – variations present between the coding genes, within an unknown gene of uncharacterized function (C16H10orf90) or within 65,000 bp to the closest gene (SLC30A5, KIRREL3, 3 variants associated with NDUFA4, DOK6, LHCGR, EFNB2), 2 SNPs were associated with an upstream gene variant (USH2A, TANK), 7 were associated with intron variant (SMARCA5, CDCA7L, FILIP1, NF2, LIPA, UNC5B, SPATA13), and the remaining 3 were within intergenic regions with a range of 6 000 – 65 000 bp of distance

to the closest genes: TENT4A, PPA2, and FRMPD3. The biological functions of these genes are related to vision and auditory functions – USH2A (Chen et al., 2014; Perrino et al., 2020), cellular and energetic activities – SMARCA5, SPATA13, SLC30A5, PPA2 (Parker et al., 2013; Guimier et al., 2016; Kambe et al., 2002; Waseem et al., 2020; Bomber et al., 2023), inflammatory response and immune system processes – TANK, LIPA, FRMPD3, TENT4A (Lim et al., 2018; Yan et al., 2006; Guo and Cheng, 2007; Bauler et al., 2008; Lim et al., 2018), a viral receptor – EFNB2 (Bowden et al., 2008), brain and neural system activity – USH2A, NF2, CDCA7L, FILIP1, UNC5B, KIRREL3, NDUFA4, DOK6, EFNB2 (Nagano et al., 2002; Crowder et al., 2004; Grönholm et al., 2005; Ou et al., 2006; Bhalla et al., 2008; Chen et al., 2014; Bhat et al., 2020; Perrino et al., 2020; Piffko et al., 2022), DNA-damage regulation – TENT4A (Lim et al., 2018), and sexual development – LHCGR (Qiao et al., 2009).

The individual genotypes across these detected variants, were analysed comparing genotype frequencies between epizootic and non-epizootic periods including both males and females. In the main dataset, where all samples were included, only one gene – DOK6 revealed significant differences (p value = 0.000) between the different periods, with high levels of heterozygosity during epizootic periods. However, for the peak period dataset, most genes presented significant differences (p value < 0.002) (with the exception of USH2A, TANK, KIRREL3, CDCA7L, and LHCGR, although still significant if considering a p value of 0.05) and most had significantly greater homozygosity during epizootic periods (with the

exception of USH2A, and C16H10orf90). From the 23 variations extracted from the GLM analysis, only one – USH2A was also associated with a missense substitution.

The heterozygosity frequency and minor allele frequency (Figure 16, Table S3) revealed different patterns according to each locus in the context of the different epizootic periods. With the exception of DOK6, EFNB2, C16H10orf90, LIPA, and USH2A, all the remaining loci revealed a pattern of lower heterozygosity levels during epizootic periods.

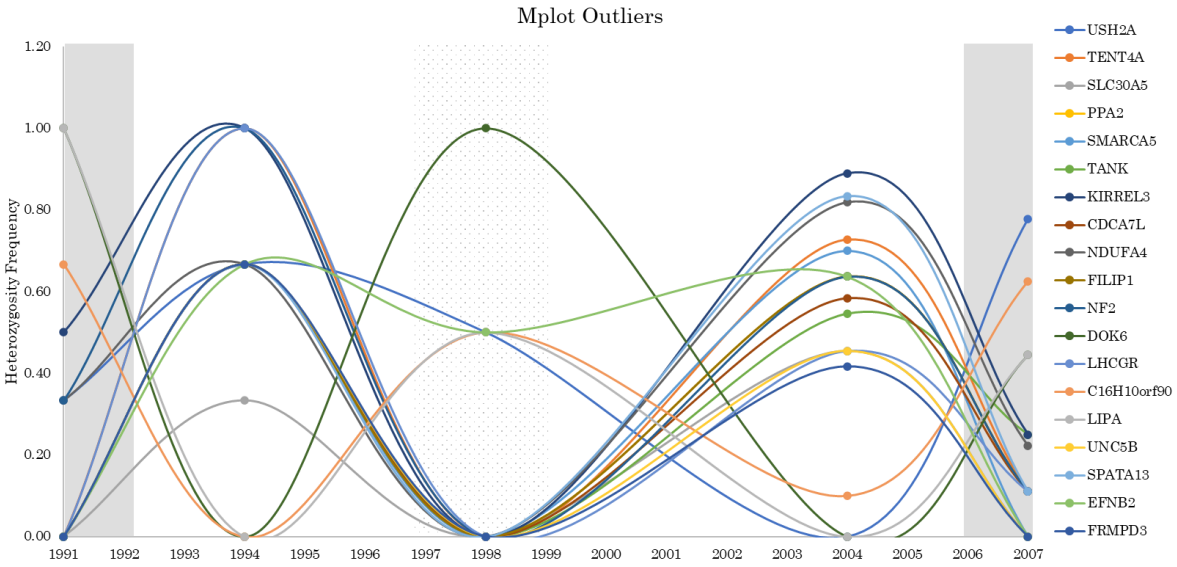


Figure 16 – Heterozygosity frequency of each outlier from the Manhattan plot analysis, through time, covering the peaks of epizootic and non-epizootic periods.

The individual genotypes across the detected 1059 missense variations were also analysed, between epizootic and non-epizootic periods. Within the main dataset, where all samples were included, no differences were found with *p* value adjusted to 0.00005 but, if considering a *p* value of 0.05, 7 genes revealed genotype differences between periods: 2 SNPs

within LOC101316607 like-PRAMEF8 (p = 0.001; p = 0.031) and one within UTS2 (p = 0.006), ABCD4 (p = 0.040), CLIP1 (p = 0.014), PIWIL2 (p = 0.031), LOC101319173 like-40S RPS21 (p = 0.039), and POP1 (p = 0.012). With the exception of PIWIL2, all of the above genes showed significant allele frequency differences ($p < 0.05$) between periods. For the peak period dataset, no differences were found when the p-value was adjusted to 0.00005, but for $p < 0.05$, 4 genes revealed genotype differences between periods: LOC101316607 like-PRAMEF8 (p = 0.0186), ENDOD1 (p = 0.016;), CIAO1 (p = 0.016), and TEKT4 (p = 0.028). With the exception of TEKT4, all of the above genes showed significant allele frequency differences ($p < 0.05$) between periods. The strongest inference may be from LOC101316607 like-PRAMEF8, since it was the only locus identified in both analyses.

These genes are mostly related to cellular, immunity and reproductive mechanisms – LOC101316607 like-PRAMEF8, TEKT4, PIWIL2 (Lee et al., 2006; Roy et al., 2007; Mistry et al., 2011), inflammatory response and immune system processes – ENDOD1 (Fenech et al., 2020), and cellular and metabolic activities – CIAO1, ABCD4, CLIP1 (Weisbrich et al., 2007; Kim et al., 2018; Kitai et al., 2021), cardiovascular activity – UTS2 (Johns et al., 2004), RNA mechanisms and inflammatory responses – RPS21, POP1 (Smirnova et al., 2020, Abdulhadi-Atwan et al., 2020).

The individual genotypes across the detected 20 stop gained variants (loss of function) were also analysed between epizootic and non-

epizootic periods. Considering the main dataset, only a weak difference was found within CER1 (p = 0.041). Allele frequency differences were also weak (p = 0.0436) and the function of this gene is associated with brain and neural system activity (Belo et al., 2009).

Genomic diversity was assessed over time in the context of the morbillivirus epizootic periods. There was no positive correlation with the pattern reported in Gaspari et al. (2019; see Figure 6 and 17), though some sample sizes per year are very small for the RAD dataset (Table 2).

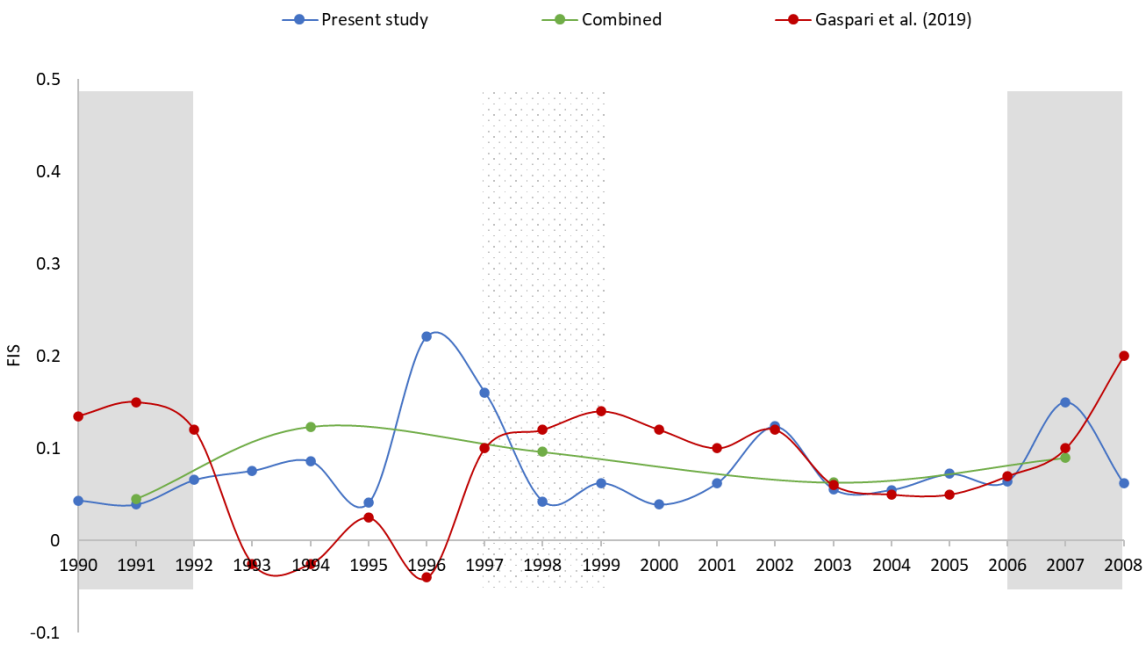


Figure 17 - Genetic variation of Mediterranean striped dolphin across time, assessed by the inbreeding coefficient (F_{IS}). Blue line represents the values obtained in the present study, green line represents the main period averages, and the red line is an estimation of Gaspari et al., (2019) results. Grey areas represent well documented morbillivirus epizootic years, while the dotted area represents an epizootic less described in the current literature.

To help address sample size issues, the values for each epizootic and non-epizootic period were averaged across all years (defining three epizootic and two non-epizootic periods), which again showed no consistent pattern and little variation over time (Figure 18).

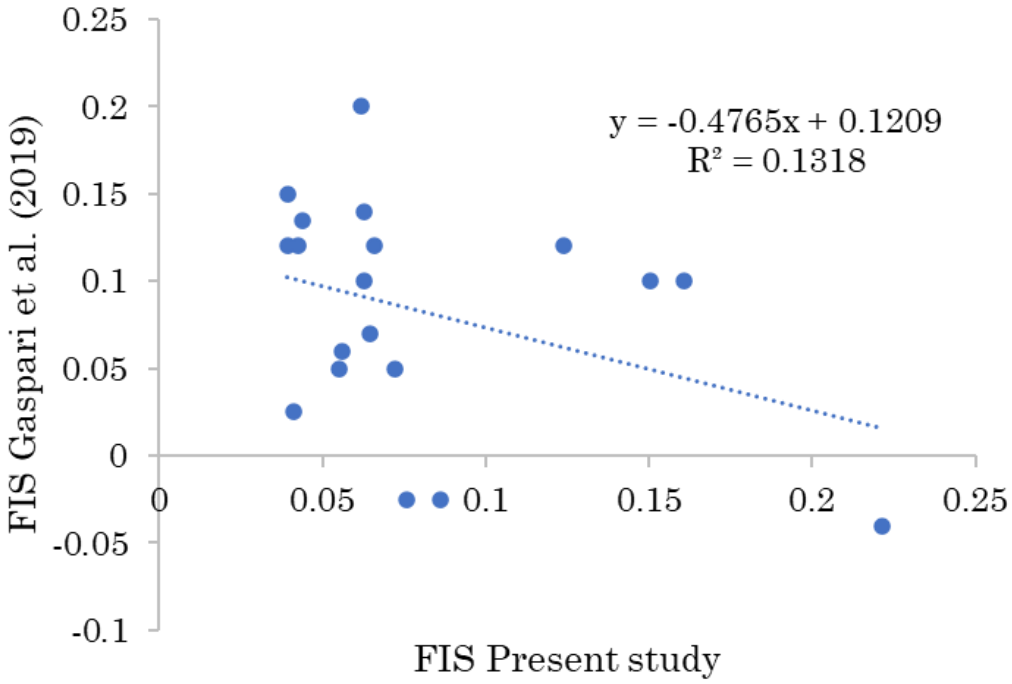


Figure 18 - Inbreeding coefficient (FIS) revealing a negative correlation between the data of the present study and an estimation of Gaspari et al. (2019).

When analysing associations between epizootic events and genetic diversity, low correlation levels, with the absence of statistical significance, were detected for all datasets (Table 4-6 and Figure 19).

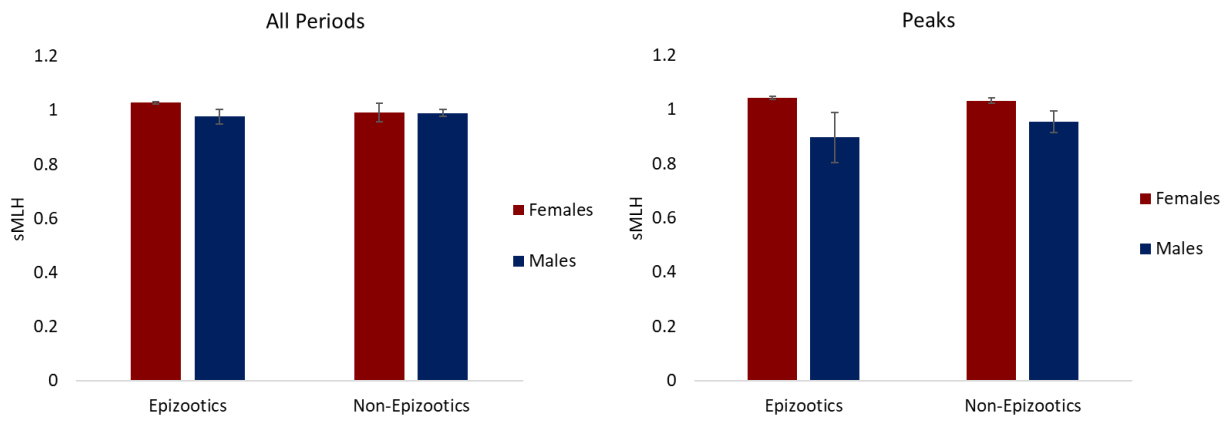


Figure 19 - Comparison of genetic variation of Mediterranean striped dolphin, between males and females, during epizootic and non-epizootic events of (left) all periods and (right) the peak periods of a considered epizootic/non-epizootic event. Error bars represent the standard error. No significant differences were found between groups, with 95% confidence interval chi-square test.

Table 4 - Genetic measures, for all periods, from InbreedR, based on 14 240 SNPs. $o^2(h)$ - Distribution variance of standardized multilocus heterozygosity; r^2_{Wh} - expected correlation between a fitness trait and heterozygosity; r^2_{Wf} - expected correlation between a fitness trait and inbreeding; g^2 - identity disequilibrium; HHC - distribution variance of heterozygosity–heterozygosity correlation coefficient, with confidence intervals (CI: 2.5–97.5%). Due to unidentified gender, three samples were not included in the gender analysis.

	$o^2(h)$	r^2_{Wh}	r^2_{Wf}	g^2	HHC
All (118)	0.018	0.0004	0.0004	0.022±0.01, p = 1 CI: [0.006-0.045]	0.95±0.034 CI: [0.852-0.979]
Females (50)	0.014	0.024	0.018	0.019±0.01, p = 1 CI: [0.004-0.047]	0.971±0.005 CI: [0.961-0.98]
Males (65)	0.021	0.002	0.002	0.026±0.01, p = *1* CI: [0.007-0.047]	0.932±0.071 CI: [0.731-0.982]

Table 5 - Genetic measures, for the peak periods, from InbreedR, based on 14 240 SNPs. $\sigma^2(h)$ -Distribution variance of standardized multilocus heterozygosity; r^2_{Wh} - expected correlation between a fitness trait and heterozygosity; r^2_{Wf} - expected correlation between a fitness trait and inbreeding; g^2 - identity disequilibrium; HHC - distribution variance of heterozygosity–heterozygosity correlation coefficient, with confidence intervals (CI: 2.5–97.5%). Due to unidentified gender, one sample was not included in the gender analysis.

	$\sigma^2(h)$	r^2_{Wh}	r^2_{Wf}	g^2	HHC
All (29)	0.026	0.014	0.011	0.036±0.03, p = *1* CI: [0.011-0.087]	0.923±0.07 CI: [0.735-0.987]
Females (14)	0.0006	0.055	0.007	0.004±0.003, p = *1* CI: [0.0008-0.011]	0.543±0.15 CI: [0.246-0.796]
Males (14)	0.053	0.019	0.012	0.084±0.05, p = *1* CI: [0.007-0.047]	0.918±0.09 CI: [0.68-0.992]

Table 6 - Chi-Square analysis of the genetic diversity between gender within the different epizootic periods from the main dataset (all periods) and the peaks of those periods. Critical value was set to 0.05.

Periods	Gender	Chi-Square (X^2)	Degrees of Freedom (*df*)	*p* value
All Periods	Males	8.070	42	1.000
	Females	1.537	25	1.000
Peaks	Males	4.430	7	0.729
	Females	1.335	7	0.987

Furthermore, levels of inbreeding among loci (g^2) were not different from 0 (p = 1), and the HHC distribution mean, with a general wide range of confidence intervals, does not support the presence of inbreeding in the present datasets (Table 4 and 5). Heterozygosity levels, assessed by sMLH, for males and females within each dataset, and between epizootic periods, also did not reveal statistical significance for either males or females (Figure 19, Table 6).

The factorial correspondence analysis between the epizootic periods based on missense mutations (Figure 20), did not reveal the existence of strong structure between the analysed groups. For both, only a subtle differentiation was found in the presence of factor 4.

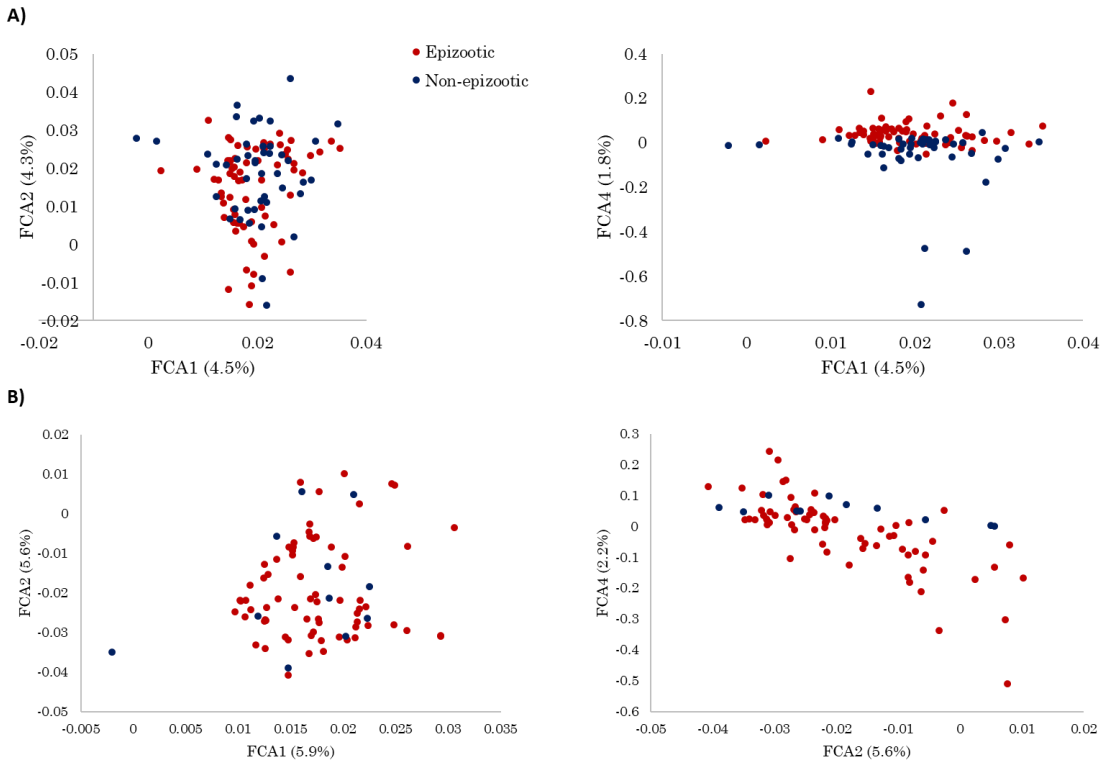


Figure 20 - Factorial correspondence analysis between epizootic periods within A) all dataset and B) putative west region of the Mediterranean Sea.

3.4 Discussion

Wildlife species face multiple challenges, including from infectious disease, which is a major conservation concern due to these periodic epizootics. Disease has a profound effect as a demographic driver. Susceptibility to environmental stressors, such as pathogens, can be facilitated by a range of genetic factors. The possible association between genome-wide heterozygosity of the Mediterranean striped dolphin, and morbillivirus epizootic events was explored, with a special emphasis on inbreeding levels and possible signs of selection. Although the disease status of individuals was not known, the timing was documented so that both the full epizootic periods and just the peak periods could be independently compared.

Previous studies have suggested an association between levels of inbreeding and a higher susceptibility and vulnerability to infectious diseases, as well as immune system impairment, including cases of morbillivirus infection in striped dolphins (Valsecchi et al., 2004; Smith et al., 2009; Gaspari et al., 2019). Both Valsecchi et al. (2004) and Gaspari et al. (2019) reported an association between inbreeding and the epizootic periods. Another study, that looked at the impact of lungworm infection, reported a relationship between genetic diversity and disease in striped dolphins (Gkafas et al., 2020). Contrary to morbillivirus, which has three main target organs, these nematodes are relatively common (Raga and Carbonell, 1985) in delphinid respiratory system, and can also cause almost total obstruction of bronchi and bronchioles (Raga et al., 1987),

leading to significant impacts on health of these organisms. While there was no effect overall, there was a strong correlation for females. This sex-specific pattern was suggested to be associated with female stress and higher energetic demand related to pregnancy but may also be affected by vertical transmission. The vertical transmission hypothesis was also proposed for morbillivirus infection, as well as several other parasite species, but with both transmission form and respective consequences dependent on DMV type (Soto et al., 2011).

In my study, no clear relationship was observed between epizootic and non-epizootic periods. Genetic variation across time, assessed by inbreeding coefficient, revealed no consistent pattern with genetic variation, nor a positive correlation with the results from Gaspari et al., (2019) – which reported higher levels of inbreeding during epizootic events (Figure 17 and 18). Also, no structure was found among missense mutations (Figure 20), nor was there a significant correlation when looking into genome-wide heterozygosity levels and fitness associated with morbillivirus epizootic infection survival (Table 4-6 and Figure 19). This is in accordance with Batley et al. (2019, 2021) who reported on the absence of high inbreeding levels in Indo-Pacific bottlenose dolphins (*Tursiops aduncus*) infected by morbillivirus, compared to uninfected individuals, rejecting a correlation between infection and inbreeding levels.

In my study an exception to this was found when the Manhattan plot analysis was used to identify loci showing an association with epizootics beyond a defined threshold. Only the peak periods showed outlier SNPs at

this threshold (10^{-3}). For the comparison of peak periods, genotype profiles at the 23 loci identified were generally significantly different within compared to outside the epizootics, as expected (since they were chosen based on having a correlation in the Manhattan plot). However, there was also a consistent pattern between them, with most showing a greater level of homozygosity during epizootics, and higher heterozygosity frequency during non-epizootic periods (Figure 16). Batley et al. (2019, 2021), in contrast to the lack of correlation overall, reported high levels of homozygosity present in at least 84% of all samples of *Tursiops aduncus* within 11 SNPs that were associated with immune system genes, previously suggested to be involved in morbillivirus immune responses. Nonetheless, comparisons and extrapolations between species with different ecologies and social structures should be taken cautiously. In the present study, the fluctuation patterns observed over time for heterozygosity and MAF, in dead stranded striped dolphin might be related to greater fitness/protective roles of certain alleles and also the function of the respective loci.

Some of the loci identified in the current study through looking at Manhattan Plot outliers were associated with neural functions. It is known that the brain is one of the main target organs of morbillivirus, particularly for central nervous system (CNS)-localised cases (Soto et al., 2011), with reports of disoriented animals coming close to shore, where most of the time the individuals end up dying (Kennedy 1998; Valsecchi et al., 2004). CNS-localised cases are characteristically associated with brain

inflammatory lesions and the absence of the virus in other organs, which is suggested to be related to a lack of excretion or transmission of the virus to other marine mammals (Soto et al., 2011). Similarly to what happens to humans with subacute sclerosing panencephalitis, caused by measles virus, it is suggested that after a primary non-fatal infection, fatal nervous lesions may actually end up occurring, associated with the virus persistence in a latent form in the CNS over a long period (Connolly et al. 1967, Garg 2008: Soto et al., 2011), after which the infection can reactivate. In the present study, several genes related to brain and neural system functions revealed significant differences between epizootic and non-epizootic periods, and most of them involved on relevant pathways. In particular, nervous system development/maintenance (DOK6, UNC5B; Crowder et al., 2004; Bhat et al., 2020), nervous system processes, skin tumours and ocular abnormalities (NF2; Grönholm et al., 2005), cortical neuron migration and dendritic morphology (FILIP1; Nagano et al., 2002), neurological and cognitive disorders (KIRREL3; Bhalla et al., 2008), and signaling receptor binding and viral receptor (EFNB2; Piffko et al., 2022) particularly for paramyxoviruses – the same family to which the morbillivirus belongs. DOK6, specifically expressed in the nervous system and with relevant roles in neurite growth, was suggested as a prognostic biomarker for acute myeloid leukemia (AML) (Sun et al., 2019). Missense variations on FILIP1, were suggested to cause neurological and muscular disorders, particularly *arthrogryposis multiplex congenita* (AMC) (Schnabel et al., 2023). Mutations in KIRREL3 were identified as risk

factors for autism spectrum disorder and intellectual disability (Taylor et al., 2020). Similarly, variations in EFNB2 were also related to intellectual disabilities, and other health problems such as hearing loss, seizures, and congenital heart defects (Lévy et al., 2018). As mentioned above, all these genes are related to relevant pathways and outcomes of neural system functions, highly impacted by morbillivirus infection but not only.

High levels of homozygosity in immune system genes may impair pathogen recognition (Blanchong et al., 2016; Smith et al., 2009). Inflammatory response and immune system processes were highlighted in the present study by the significant genotype variation found in/near various genes. These were associated with antiviral innate immunity (TANK; Guo and Cheng, 2007), immune response (LIPA; Yan et al., 2006), innate immune system (FRMPD3, ENDOD1; Bauler et al., 2008; Fenech et al., 2020), viral mRNA Translation (LOC101319173 like-40S RPS21, Smirnova et al., 2020), autoantigen (POP1; Abdulhadi-Atwan et al., 2020), and DNA-damage regulation and infectious diseases (TENT4A; Lim et al., 2018). Viral interaction with host cells was recently demonstrated with TENT4A, which is used by the hepatitis A virus (HAV) to replicate its own genome (Li et al., 2022). Knowing the mechanisms behind this interaction allowed the discovery of a promising antiviral therapy for this disease (Li et al., 2022).

FILIP1, NF2, UNC5B, and LIPA were all within intronic regions, which might influence gene function, for example through splicing dynamics (Seoighe and Korir, 2011; Guigó and Ullrich, 2020).

The great majority of the outlier SNPs were not associated with non-synonymous changes, with the exception of one – USH2A, which is associated with vision, auditory, and neural system functions (Perrino et al., 2020). This locus did not show a clear heterozygosity pattern across the epizootic periods (Figure 16). To the best of my knowledge, there is no information available about how morbillivirus infections could impact auditory and visual functions in particular. However these are neural system related processes, which can be highly impaired by this viral infection leading to degeneration at the sensorial and neuronal level in the inner ear and the retina (Reiners et al., 2005, 2006). In fact, Usher syndrome, which is characterized by a combined deaf-blindness in humans, is suggested to be associated with a USH-protein network, in which USH2A is present, being regulated in the synaptic terminals of both retinal photoreceptors and inner ear (Reiners et al., 2005, 2006).

Because the Manhattan plot analysis was based on few individuals per category, as well as a limited number of SNPs (mapped against the reference genome available for the most related species: *Tursiops truncatus*), it is possible that the results also reflect that limitation. Furthermore, the majority of SNPs found were not within coding regions, and may or may not be associated with regulation effects. However, three of the loci found (USH2A, LIPA, and NF2 in its inverted form – INF2) were suggested as putative candidates under selection for morbillivirus resistance in Batley et al. (2021), although not considered as immune-related candidates.

Overall, contrary to the suggestions of earlier studies for this species, there was no association between genomic indicators of inbreeding and viral susceptibility in this study. When all missense or loss of function mutations were assessed, there were no strong effects. Strong differences were found when using a Manhattan plot comparison of peak epizootic and non-epizootic periods. This seems unlikely to be by chance, because the pattern of genotype difference was the same for most loci (more homozygous during the epizootic) and the functions of the identified loci were consistent with what could be expected in the context of morbillivirus infection. This is consistent with heterozygosity providing some protection, but the extent of the effect would depend on various variables, including the strength of selection, none of which is known. Nonetheless, and because strandings can occur from a variety of reasons, further studies with a more balanced and larger sample set, as well as data on the infection status, age, sex, and contaminant load of individual dolphins, should be carried out to provide a more accurate interpretation. Also, investigating specific candidate gene systems expected to be related to viral resistance is expected to be useful, rather than looking for correlations from random genome scans. This was done for Chapter 4.

Chapter 4

Susceptibility-related loci to morbillivirus epizootics in *Stenella coeruleoalba*

4.1 Introduction

As significant evolutionary drivers for wildlife populations, the emergence, and re-emergence of pathogens and consequent infectious diseases, have become a major concern for the conservation of many species (Acevedo-Whitehouse and Cunningham, 2006; Van Bressem et al., 2009; Stejskalova et al., 2017). The rapid evolution and transmission dynamics of pathogens, as well as their ability to cause mass morbidity and mortality events, leads to genetic diversity loss, declines in population numbers and increase of extinction risks, posing an effective threat to species adaptation, survival, and conservation measurements (Van Bressem et al., 2014; Stejskalova et al., 2017).

For cetaceans, disease outbreaks followed by mass mortality events are increasing worldwide, becoming a cause for concern. This is especially true for species that have high potential for transmission (highly social species) and populations that are small or threatened (Van Bressem et al., 2009). As discussed in Chapter 1, cetacean morbillivirus (CeMV) has been considered a major threat to cetacean populations, being highly contagious and virulent (Sacristán et al., 2015). Belonging to the family of RNA viruses Paramyxoviridae, morbilliviruses comprise seven different genera that infect mammals: canine distemper virus (CDV) in carnivores, feline morbillivirus in felids, rinderpest in cattle, peste des petits ruminants virus in ruminants, measles virus (MV) in humans, and there have been various outbreaks in marine mammals (see Chapter 1, Figure 3). Strains of morbillivirus share a phylogenetic origin, which might be associated

with the frequent cross-species transmissions (Kennedy et al., 1991; da Fontoura Budaszewski and von Messling, 2016; Geoghegan et al., 2017; Jo et al., 2018a, 2018b). This has been considered a possible advantage in the study of the immune system response to this virus, allowing for extrapolation among host species.

Over the past 33 years, the Mediterranean Sea has been the stage of several morbillivirus epizootics, being associated with high mortality rates and frequent strandings, particularly with the striped dolphin species (*Stenella coerueoalba*) (Domingo et al., 1990; Valsecchi et al., 2004; Raga et al., 2008, Di Guardo and Mazzariol, 2013). Although this species susceptibility has been suggested to be associated with several factors such as high levels of inbreeding, low genetic diversity, vertical transmission, high pollutant load (Aguilar and Borrell, 1994, 2005; Valsecchi et al., 2004, Di Guardo and Mazzariol, 2016), it is still not clear what makes this species so prone to this virus pathology, compared to other cetaceans.

Host-pathogen dynamics are known to be determinant drivers in the adaptation and plasticity of immune responses and, ultimately, the susceptibility and evolution of the individual (Stejskalova et al., 2017). Yet, studies addressing the host genetic basis for disease susceptibility or resistance are limited, and have mostly focused on a small number of functionally important genomic regions. Candidate genes/molecules suggested to be involved in the host immune response against morbillivirus include the major histocompatibility complex – MHC (Acevedo-Whitehouse and Cunningham, 2006; Cammen et al., 2015;

Manlik et al., 2019) and the homologue to human leukocyte antigen – HLA (Martin and Carrington, 2005). The signalling lymphocyte activation molecule – SLAMF1 (Hashiguchi et al., 2011; Shimizu et al., 2013; Melia et al., 2014; Ohishi et al., 2019) is relevant as a viral receptor on lymphoid cells, and CD46 (cluster of differentiation 46) as a binder of viral genes. Poliovirus-related receptor-4 – NECTIN4/PVRL4 (Stejskalova et al., 2017) functions as and epithelial cell receptor, and Basigin – BSG/CD147 (Watanabe et al., 2010, Stejskalova et al., 2017) functions as a cell entry receptor on epithelial and neuronal cells not expressing SLAMF1. Vitamin A and D are viral receptors and interleukin 10 (IL10), and tumor necrosis factor (TNFA) are cytokine receptor genes (Sato et al., 2012). Pathogen-associated sensing genes include toll-like receptors – TLRs (Stejskalova et al., 2017) and cluster of differentiation 209 (CD209). Solute carrier family 11 member 1 – SLC11A1 (Archer et al., 2015; Stejskalova et al., 2017) works as mediator of natural resistance by macrophage transport. Natural killer cell receptors – NKRs and the natural cytotoxicity triggering receptor 1 – NCR1 (Carrillo-Bustamante et al., 2016) provide crucial elements of innate immune response.

Compared to the MHC, which is the most studied and well characterized immune system region, less is known about other immunity-related genes and their roles in cetacean defence against morbillivirus. Some of these are interconnected with the innate and adaptive immune systems. For example, while cytokine receptors are recognized as important in morbillivirus host defence, other elements/mechanisms such

as their protein activators, might influence the overall outcome for the individual as well (Gelain and Bonsembiante, 2019, Batley et al., 2021). In humans, Fc receptor proteins (FcRs), which bind to antibodies and stimulate cellular/humoral responses (Takai, 2002, 2005), have been suggested as a point of interaction for measles virus proteins. This interaction leads to a general immunosuppression caused by the impairment of several factors such as the antigen specific T cell proliferation and interleukin production (Marie et al., 2001).

Other candidates reflecting susceptibility and resistance to morbillivirus were linked to relevant pathways involved with the target organs of the virus. Although neuro-inflammation is known to be the main cause of death in stranded cetaceans infected with morbillivirus (Pintore et al., 2018; Sierra et al., 2020; Wessels et al., 2021), research on the associated neuropathological changes, is still limited. In fact, neurological lesions, such as demyelination, have been mostly associated with canine and phocine distemper virus infection (Vandevelde and Zurbriggen, 2005; Mazzariol et al., 2013; Duignan et al., 2014; Ulrich et al., 2014), but also with CeMV infection (Duignan et al., 1992; Domingo et al., 1995; Soto et al., 2011; Pautasso et al., 2019). The most recent findings, and the first on neuropathological characterization in cetaceans infected with morbillivirus, reported neurological lesions in the central nervous system (CNS) of these animals (Giorda et al. 2022). This was mostly related to myelin reduction and demyelination, but also astro-microgliosis, neuronal necrosis, spongiosis, malacia, and non-suppurative meningoencephalitis.

Viral meningoencephalitis is frequently seen in cetaceans (Sierra et al., 2020; Wessels et al., 2021). Besides the central nervous system, the cardiovascular, and respiratory tract, are affected by morbillivirus infection, with leukocytes thought to play an important role on viral antigen dissemination within the individual (Groch et al., 2020).

Taking advantage of next-generation sequencing advances, the present study focuses on precise, and high-quality target enrichment, being more cost-effective than whole genome sequencing (Jones and Good, 2016). Using specific genomic regions of interest as a target, hybridization-based capture was applied to investigate the genetic basis of susceptibility to morbillivirus in striped dolphins, using an 18-year sample archive, covering different regions across the Mediterranean Sea, as well as epizootic and non-epizootic periods.

Limited research on the genetic variation of the immune system response in cetaceans has been conducted, particularly with a focus on ecologically relevant functional traits, and even less while this species experienced morbillivirus epizootics (Van Bressem et al., 2014; Stejskalova et al., 2017; Batley et al., 2019, 2021). To the best of my knowledge Batley et al. (2021) – which suggested 50 genes as relevant candidates in *Tursiops aduncus*, is the only study that attempted to better understand the complex interaction between the cetacean immune system and morbilliviruses infection, based on whole genome data sets. Information on genetic footprints of selection can lead us to a better understanding of local

adaptation and susceptibility patterns, but also of the long-term viability and evolutionary potential of wild species.

4.2 Methods

4.2.1 Selection of immune system genes and probe design

Due to logistic constraints and the incomplete functional annotation of the dolphin genomes available to date, only a portion of the immune system related genes were possible to select and genotype. Genes whose functions are more likely to be associated with immune response and viral resistance to morbillivirus and other potential opportunistic infections, were thus selected. These were identified by querying the Gene Ontology Consortium database at http://www.geneontology.org/ using all annotation sources available in the search engine. For this purpose, the Gene Ontology (GO) terms associated with SLAMF1 and NECTIN4 were first used to identify other genes that could have similar immune functions. Furthermore, other terms associated with immune function were also queried in the GO database. Initial queries were run to identify terms that were too generic and avoid targeting an excessive number of genes of unclear function, or whose function might be only tangentially involved with the immune response. The final query thus contained the following terms: “innate immune receptor”; “immune response”; “immune suppression”; “nectin4”; “slamf”; “immune complex formation”; “virus

antigen"; and "recombination-activating gene". Only genes annotated from mammal species were selected, as they are the most closely related to cetaceans. Information regarding each gene name, class type, taxonomic group, database and the functional details of each gene were extracted from the GO database.

Duplicates were removed by using the 'remove duplicates' function in Microsoft Excel v2304, based on the gene name. The final database was then inspected by eye, to ensure functionally relevant genes were identified. For immune system genes, gene classes known to be important in the immune response were searched for in previous papers. This included Toll-Like receptors, CD genes, immunoglobulin genes, MHC, Interleukins and Recombination Activation Genes (RAG). At the same time, genes with functions or characteristics that were considered too generic and whose biological significance might be difficult to interpret were also removed (e.g. ADM and ADM2, both involved in blood vessel morphogenesis). Finally, RNA genes, protein mitochondrial genes, Y chromosome genes, olfactory receptor genes (known to be pseudogenized in odontocetes), pseudogenes, and genes marked as non-functional were also removed from the dataset. Non-functional genes were removed at this stage, to maximize capture of functional genes. The final list (Appendix C, Gene list) included genes from the following gene annotation sources: UniProt KnowledgeBase, Rat Genome Database, Mouse Genome Informatics, IntAct Molecular Interaction Database and RNAcentral Database.

The respective exon sequences for the final gene list were then extracted using the ENSEMBL BioMart tool from the dolphin Ensembl Genes database (turTru1). The list of gene names was used as the query list, and the option "exon sequences" was selected. Each sequence name was designed to include the following information: Gene name > Chromosome/scaffold name > Gene start (bp) > Gene end (bp) > Constitutive exon > Exon region start (bp) > Exon region end (bp) > Genomic coding start > Genomic coding end > Exon stable ID > Gene stable ID. After the count was performed and the results obtained, data were exported as a "FASTA" file.

Some genes from the initial list were not yet annotated in dolphins, and their exons were therefore not downloaded. The final list contained 7307 individual exons belonging to 1408 unique genes, corresponding to a total of 1 244 564 bp. This list was then used as a template for probe design, in collaboration with Arbor Biosciences to produce the MyBaits target enrichment kit, and different filtering strategies were applied by the company. A trial-and-error approach was used to identify filtering levels that would still produce good quality data at the sequencing stage (according to Arbor biosciences experience) but still provide coverage for key genes required to achieve the scientific objectives of the project and its logistical constraints (e.g. ensuring that all exons from SLAMF1 were included in the final probe design). Final filtering steps applied were as follow: first, a design of 100 nucleotides baits with a 50% overlap between successive baits was chosen, meaning that each exon will be fully covered

by different baits. Comparing the genome of Artiodactyls and whales (given the limited information available for cetaceans), the target sequences from repetitive regions were identified and marked, resulting in 0.9% of problematic baits. Using the BLAST (Basic Local Alignment Search Tool) algorithm, the resulting baits were compared to the *Tursiops truncatus* reference genome, and only baits that were ≤ 25% complementary to repetitive regions were kept; baits that were complementary to the mitochondrial genome sequence were also excluded. Only baits with a GC content between 35-60% were selected, avoiding the risk of bait binding to non-complementary sequences (by company experience). Furthermore, only target loci that were > 96% covered by the baits as well as 99% within 100 bp of a bait were kept, to guarantee that each probe matched to its respective region only. The total number of baits obtained was thus 19,836.

4.2.2 Sample selection and Preparation

A total number of 96 samples were selected as this was assessed in previous trials to provide the best per sample coverage on a single Illumina lane sequencing. Because dolphins are diploid, they contain two copies of each gene, and thus for each sample genotyped, it was expected to obtain the sequence data for two gene copies. Procedures of sample selection and preparation were the same as detailed in the previous chapter.

4.2.3 Illumina sequencing library preparation and exome capture kit for Illumina sequencing

After probe design was completed, Illumina libraries were prepared, and exomes were isolated from these libraries using enrichment kits (Liu et al., 2012; Sestak et al., 2012) to be then sequenced on the Illumina HiSeq platform. DNA concentration was first normalized for each sample to 16 ng/μL, and then fragmented to 200 base pairs using a Covaris sonication system to facilitate downstream reactions and reading during Illumina sequencing process. A single-tube library preparation protocol, as described in Carøe et al. (2018), was then performed using three sequential reactions:

End-repair – to convert the ends of the sheared DNA fragments to blunt-ends required for the subsequent ligation of adaptors. The reaction mix included T4 DNA polymerase (0.03 U/μL), T4 PNK (0.25 U/μL), dNTP (0.25 mM), T4 DNA ligase buffer (1x), and the reaction booster buffer. Samples were then incubated for 30 minutes at 20ºC, followed by 30 minutes at 65ºC, and then cooling down to 4ºC.

Adaptor ligation – The adapters, PEG-4000 (6.25%), and T4 DNA ligase (8 U/μL) were added and incubated for 30 minutes at 20ºC, followed by 10 minutes at 65ºC, and then cooling down to 4ºC.

Fill-in – The reaction included buffer (0.33x), dNTP (0.33 mM), Bst 2.0 Warmstart polymerase (0.21 U/μL), and molecular grade water

with the sample. It was incubated for 15 minutes at 65ºC, followed by 15 minutes at 80ºC, and then cooling down to 4ºC.

After the final fill-in reaction, a library purification was performed to remove reagents used in these steps, as well as adapter dimers and DNA molecules that were too small/long relative to the target of 200 bp. For this step, speed-beads (carboxyl-modified Sera-Mag magnetic speed-beads in a PEG/NaCl buffer) were used following Rohland and Reich (2012) protocol. After washing, samples were eluted in EBT buffer and incubated for 10 minutes at 37ºC. Quantitative PCR (qPCR) was performed using CFX Connect Real-Time PCR Detection System from Bio-Rad. This qPCR step estimates the ideal number of cycles required for each sample in the downstream index PCR. For this step, KAPA SYBR FAST qPCR (KAPA BIOSYSTEMS) kit was used, following standard manufacturer instructions. The PCR cycle conditions were 95°C for 1 minute followed by 35-40 cycles of 95°C for 15 seconds and 63°C for 60 seconds, from which a dissociation curve was produced. To evaluate success, a few random samples were chosen to measure DNA quantification using Qubit.

Index PCR was then performed with P7 and P5 primers in combinations (Appendix C, Table S4 and S5) taking into account a reasonable balance for the Illumina platform. Each sample received a unique combination of P5 and P7 from a set of 8 unique P5 primers and 12 unique P7. This meant that 8 sets of samples had the same P5 but different P7. Sample-primer assignment was performed to avoid inclusion of samples of the same year/location sharing the same P5 or P7. This was

done to prevent any possible reading or procedural error from affecting an entire year or population. The reaction conditions were an initial denaturation for 45 seconds at 98°C, followed by the pre-defined number of cycles of 20 seconds at 98°C, 30 seconds at 60°C and 30 seconds at 72°C. Final extension was for 60 seconds at 72°C, and then cooling down to 4ºC. For this step, PfuTurbo Cx Hotstart DNA Polymerase (Agilent Technologies) was used. The success of library building preparation was then evaluated by measuring 6 randomly chosen samples on Qubit.

The capture protocol followed the myBaits manual version 4.01 (Arbor Biosciences) in three main steps (Figure 21):

Hybridization setup – Libraries were mixed with several blocking nucleic acids, denatured and mixed with hybridization proprietary reagents (Hyb N, Hyb D, Hyb S, Hyb R, baits produced for this samples, Block A, Block C, Block O). Hybridization was allowed to occur during 16h of incubation at 65ºC, which facilitates the encounter of the baits with target library molecules.

Bind and wash – Streptavidin-coated magnetic beads were used to bind bait-target hybrids. Non-target DNA was thus removed with several washing rounds using a proprietary washing buffer.

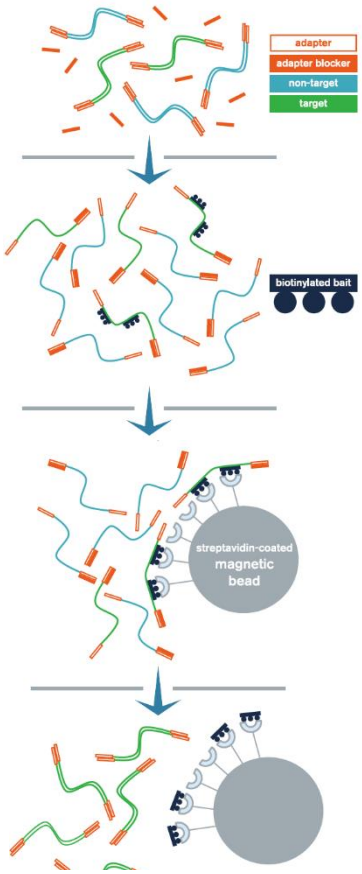


Figure 21 - Procedure overview of capture protocol. Adapted from myBaits manual version 4.01 (arbor biosciences).

Library resuspension and amplification – The bead-bound enriched library was resuspended in Tris-TWEEN solution (EB buffer), heat-denatured from the baits through incubation for 5 minutes at 95ºC , separated from the beads with pelleting on a magnetic plate, and then amplified. Amplification was done using Phusion Hot Start II High-Fidelity DNA Polymerase (Thermo Scientific) kit, following the myBaits manual.

After amplification, SPRI beads were used to purify the samples one last time. Samples were then resuspended in EB buffer after washing with 80% ethanol, and beads were pelleted and discarded. After the final cleaning, samples were measured on Qubit to evaluate final DNA concentration and further dilutions needed before performing one last qPCR. qPCR was conducted using KAPA SYBR FAST qPCR (KAPA BIOSYSTEMS). Six samples chosen randomly were analysed in a TapeStation System 2200 (Agilent Technologies), to further assess the average fragment size. Each pool was adjusted to 2nM, to which 5 μL were then collected and combined in the same tube, pooling 80 of the 96 samples at the same concentration. The library was sequenced on an Illumina HiSeq2500 platform in paired-end 125bp read mode, using one lane (at Durham Genomics).

4.2.3 Bioinformatic analysis

From demultiplex process to conversion to BAM files, the steps applied were the same as in previous chapters. Only 72 samples (with more

than 1 million reads, Table 7) were retained after sequencing and alignment to the reference genome.

Table 7 - Final number of samples per location/year and respective retained reads for chapter 4. Grey areas represent epizootic periods, while the remaining are reported as non-epizootic period.

Period	Year	Number of Samples	Retained Reads
1	**1990**	6	9012563
	1991	5	12508224
	1992	3	3058722
2	**1993**	11	8422153
	1994	1	660883
	1996	1	851749
3	**1997**	10	13206113
	1998	3	2556095
	1999	2	1438815
4	**2003**	4	4577456
	2004	5	9050878
	2005	5	7313710
5	**2006**	7	5526994
	2007	2	2137938
	2008	7	8773595

Location	Number of Samples	Retained Reads
Ligurian	30	43019378
Tyrrhenian	15	17377631
Alboran	5	4296199
Ionian	8	5408182
Spain	11	12711092
Adriatic	3	6283406
Total	72	89095888

Conversion from BAM to VCF, and the following compression and index of all VCF files, were performed by BCFtools using the plugins

mpileup, view and index (Danecek et al., 2021). VCF files were then merged into one, as described in the previous chapters. The SnpEff program (Cingolani et al., 2012) was used to assess functional effect (missense vs. synonymous changes) information, regarding all SNPs associated with a gene. This process was performed as in chapter 3. Bonferroni type 1 error correction was applied with *p* value adjusted to 0.00024, to avoid false positives. A further subdivision of the main dataset was included that focused only on individuals from the peak epizootic and non-epizootic periods, as in chapter 3 (Table 8). Due to the limited number of samples, in this chapter, "peaks" were considered as the whole period with the exception of the first year.

Table 8 - Subdivision of samples from the peak periods, per location/year and respective retained reads for chapter 4. Grey areas represent epizootic periods, while the remaining are reported as non-epizootic period.

Period	Year	Number of Samples	Retained Reads
1	**1991**	5	12508224
	1992	3	3058722
2	**1994**	1	660883
	1996	1	851749
3	**1998**	3	2556095
	1999	2	1438815
4	**2004**	5	9050878
	2005	5	7313710

Location	Number of Samples	Retained Reads
Ligurian	19	29639724
Tyrrhenian	10	12261617
Alboran	0	0
Ionian	1	660883
Spain	2	1554339
Adriatic	2	4234046
Total	34	48350609

5	**2007**	2	2137938
	2008	7	8773595

To assess signs of geographical variation, samples were categorized into Eastern or Western region, as in chapter 2, or into epizootic vs non-epizootic periods to assess signs on functional changes. All missense variations (210) were grouped (Appendix C, Table S6) into two functional groups – immune system-related genes (146) or other genes (64), based on the associated terms/details provided by GOnet online tool (Pomaznoy et al., 2018) or UniProt (The UniProt Consortium, 2023).

Using each group as a different dataset, ordination (PCA and PCoA), Nei's genetic distance, and F_{ST}, were then assessed with SambaR (de Jong et al., 2021) in RStudio v4.2.2 (R Core Team, 2020) to assess signs of geographical variation or functional changes. The ancestry coefficients were calculated using the software Admixture-1.3 (Alexander et al., 2009), with posterior plotting produced in SambaR, to allow the allocation of individuals to putative populations.

4.3 Results

The exome sequencing through hybridization-based capture generated a total of 1228598.211 reads from 80 Mediterranean striped dolphins. After filtering, 89095888 reads remained. Calling SNPs resulted in a total of 8257279 possible SNPs, of which 1426 remained after quality

filtering – the final dataset for analysis. The proportion of retained individuals and SNPs after filtering steps were applied can be seen in Tables 7 and 8.

The individual genotypes across the detected 210 missense variations were analysed, between epizootic and non-epizootic periods. Within the main dataset, where all samples were included, only one SNP located within HAVCR2 was significantly different ($p = 0.00009905$) with the adjusted p value of 0.00024, between the different periods, revealing the absence of heterozygosity during non-epizootic periods. However if considering a p value of 0.05, 3 genes also revealed genotype differences between periods CUBN ($p = 0.026$), TMOD3 ($p = 0.031$), and TMPRSS15 ($p = 0.016$), and greater levels of heterozygosity during epizootic periods (with the exception of CUBN). All of the above genes showed significant allele frequency differences ($p < 0.05$) between periods (with the exception of TMOD3). For the peak period dataset, no differences were found when the p-value was adjusted to 0.00024, but for $p < 0.05$, 8 genes revealed genotype differences between periods: TMOD3 ($p = 0.015$), HAVCR2 ($p = 0.033$), TMPRSS15 ($p = 0.041$), PDE5A ($p = 0.020$), AIMP1 ($p = 0.033$), CRTAM ($p = 0.042$), IL26 ($p = 0.021$), LOC117310972 (like-PKM) ($p = 0.034$). The tendency for greater levels of heterozygosity during epizootic periods was not evident in this data set, and with the exception of PDE5A and LOC117310972 (like-PKM), all of the above genes showed no significant allele frequency differences ($p < 0.05$) between periods. The

strongest inference may be from HAVCR2, since it was the only locus identified in both analyses.

Overall, MAF and heterozygosity patterns of variation across time revealed irregular patterns (Figure 22, Appendix C, Table S6), but there was a tendency for greater levels of heterozygosity during epizootic periods. HAVCR2 showed higher MAF both outside and within the peak periods, while other loci such as TMOD3 and TMPRSS15 showed higher MAF only within the peaks dataset. PDE5A and IL26 revealed an opposite pattern, consistent between datasets. SLAMF1, although also included among the missense substitutions detected, did not show any clear pattern associated with epizootic periods.

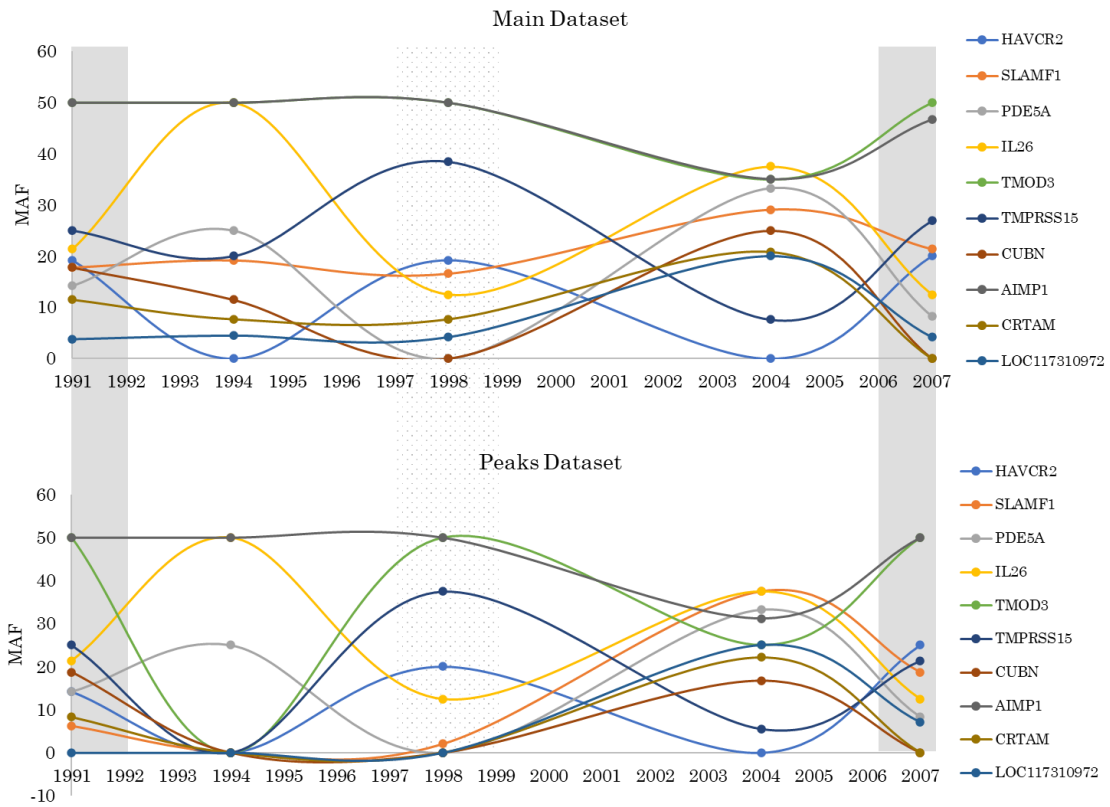


Figure 22 – Minor allele frequency (MAF) of the significant missense variations detected and SLAMF1, through time, for the main and peaks datasets.

HAVCR2, CRTAM, LOC117310972 (like-PKM) are mostly related to innate and adaptive immune response (Monney et al., 2002; Sánchez-Fueyo et al., 2003; Boles et al., 2005; Buschow et al., 2010; Gorman and Colgan, 2014; Liu et al., 2022). CUBN, AIMP1, and IL26 are associated with inflammatory responses, wound repair and lipoprotein metabolism (Kozyraki et al., 1999; Shalak et al., 2001; Crider-Pirkle et al., 2002; Hör et al., 2004; Han et al., 2006; Commins et al., 2008). TMOD3 functions in cytoskeleton organization and erythrocyte development (Kumari et al., 2020), TMPRSS15 is associated with viral proliferation (Hayashi, et al., 2018), and PDE5A is associated with smooth muscle relaxation in the

cardiovascular system (Lin et al., 2006). When looking for signs of geographical or epizootic influence on each functional group (Table 9) – immune-related loci or other functional loci, no differentiation pattern was observed (Figures 24-28).

Table 9 - Missense loci grouped by related function. Immune system loci are represented without background colour, and loci not directly related with the immune system are represented in grey background colour.

Loci	Chromosome	snp Position			
ABCC2	NC_047049.1	28101427	CD19	NC_047048.1	16533653
ADA2	NC_047044.1	101439541	CD2	NC_047034.1	81235174
ADAMTS12	NC_047036.1	27667577	CD207	NC_047047.1	29372174
ADAMTS3	NC_047038.1	88369041	CD207	NC_047047.1	29372187
AIMP1	NC_047038.1	33069439	CD207	NC_047047.1	29372233
ANKRD1	NC_047049.1	35093587	CD300LF	NC_047053.1	51955509
ANPEP	NC_047035.1	139006288	CD44	NC_047041.1	85012804
ANPEP	NC_047035.1	139006299	CD6	NC_047041.1	98143804
ANPEP	NC_047035.1	139006310	CD7	NC_047053.1	46066345
ANPEP	NC_047035.1	139016253	CD81	NC_047041.1	106684355
APLF	NC_047047.1	31505819	CDH17	NC_047050.1	37611290
APP	NC_047037.1	125578044	CDH17	NC_047050.1	37611319
BPIFA1	NC_047048.1	60915316	CDH17	NC_047050.1	37611513
BPIFB1	NC_047048.1	60965310	CDH17	NC_047050.1	37657830
C3	NC_047036.1	166279881	CDH17	NC_047050.1	37658401
C8A	NC_047034.1	136096969	CDH17	NC_047050.1	37662597
C8A	NC_047034.1	136121533	CFI	NC_047038.1	30057129
C8B	NC_047034.1	136065647	CHIA	NC_047034.1	86342382
C9	NC_047036.1	32294215	COMMD9	NC_047041.1	85967687
C9	NC_047036.1	32296482	CRTAM	NC_047041.1	10371895
C9	NC_047036.1	32296517	CRTC3	NC_047035.1	139746394
C9	NC_047036.1	32296523	CX3CR1	NC_047043.1	59764746
CARD11	NC_047048.1	2085818	DEFB128	NC_047048.1	59491432
CCL24	NC_047048.1	8748972	DTX3L	NC_047037.1	76170169
CCR2	NC_047043.1	61809712	EIF2AK4	NC_047035.1	83099312
CCR8	NC_047043.1	59710716	EIF2AK4	NC_047035.1	83112114
CCR8	NC_047043.1	59711555	F2RL1	NC_047036.1	60252082
CD109	NC_047045.1	10819910	FBXO9	NC_047043.1	24364183
CD19	NC_047048.1	16533605	FCRL3	NC_047034.1	70303714
			FCRL3	NC_047034.1	70305769

FCRL5	NC_047034.1	70417745
FCRL5	NC_047034.1	70421297
FCRL5	NC_047034.1	70421300
FCRL5	NC_047034.1	70421309
FCRL5	NC_047034.1	70421327
FGA	NC_047038.1	104531782
FGA	NC_047038.1	104531991
FUT10	NC_047054.1	25616918
FZD4	NC_047041.1	45203182
GBA	NC_047034.1	72151912
HAVCR2	NC_047036.1	132314371
HNMT	NC_047040.1	91135194
IFI6	NC_047034.1	162365902
IFNGR1	NC_047045.1	63216313
IFNGR1	NC_047045.1	63216394
IFNGR1	NC_047045.1	63216618
IL13	NC_047036.1	109716305
IL13	NC_047036.1	109716315
IL18RAP	NC_047047.1	4363887
IL2	NC_047038.1	19040347
IL24	NC_047034.1	35404576
IL26	NC_047044.1	48177312
IL9	NC_047036.1	112763755
KLRB1	NC_047044.1	94672601
KLRD1	NC_047044.1	94315557
KLRD1	NC_047044.1	94315562
KLRD1	NC_047044.1	94315605
KLRD1	NC_047044.1	94315612
KLRD1	NC_047044.1	94315623
KLRD1	NC_047044.1	94316474
KLRD1	NC_047044.1	94316521
KLRD1	NC_047044.1	94316533
KLRD1	NC_047044.1	94316535
LCP1	NC_047051.1	20586390
LGR4	NC_047041.1	78017770
LOC101316502	NC_047040.1	71478148
LOC101330057	NC_047041.1	96406790
LOC101336089	NC_047043.1	23989861
LOC101336089	NC_047043.1	23990451
LOC101336672	NC_047043.1	40251451
LOC109548948	NC_047041.1	105650378
LOC117307709	NC_047046.1	76500592
LOC117310972	NW_022983279.1	44615
LY86	NC_047043.1	4561547
LYST	NC_047049.1	77227150
MALT1	NC_047046.1	72765640
MDM2	NC_047044.1	47687681
MFSD6	NC_047040.1	43751155
MIA3	NC_047034.1	20334219
MIA3	NC_047034.1	20334743
MIA3	NC_047034.1	20336120
MMP8	NC_047041.1	30326966
MRC1	NC_047035.1	172007410
NLRP3	NC_047036.1	156782649
PANX1	NC_047041.1	39120443
PDE5A	NC_047038.1	21763502
PKHD1L1	NC_047050.1	51376544
PKHD1L1	NC_047050.1	51395395
PKHD1L1	NC_047050.1	51400265
PKHD1L1	NC_047050.1	51415016
PLAU	NC_047049.1	58415713
PTPRB	NC_047044.1	46236784
RTKN2	NC_047049.1	68634576
RTKN2	NC_047049.1	68634748
RTKN2	NC_047049.1	68664606
SERPINB1	NC_047043.1	1892051
SETD2	NC_047043.1	62393197
SETD2	NC_047043.1	62393526
SETD2	NC_047043.1	62393676
SETD2	NC_047043.1	62393779
SETD2	NC_047043.1	62393826
SETD2	NC_047043.1	62394269
SETD2	NC_047043.1	62394318
SLAMF1	NC_047034.1	68128635
SPINK5	NC_047036.1	124034024
STK10	NC_047036.1	146940490
TAB3	NC_047055.1	102651120
TAP1	NC_047043.1	40194419
TAP1	NC_047043.1	40196619
TCN1	NC_047041.1	97721155
TCN1	NC_047041.1	97723131
TMEM131L	NC_047038.1	105531935
TMEM63A	NC_047034.1	22703574
TMOD3	NC_047035.1	94711194
TRAF6	NC_047041.1	86163999
XDH	NC_047047.1	64081805
XPR1	NC_047034.1	49095311
AHI1	NC_047045.1	61775199
ALMS1	NC_047047.1	26878467
ALMS1	NC_047047.1	26878901
BBS7	NC_047038.1	19639815
BBS7	NC_047038.1	19642847
CCDC146	NC_047042.1	20233669
CCNH	NC_047036.1	69364735

CLEC4F	NC_047047.1	29383876	RAPGEF4	NC_047040.1	59538410
CLEC4F	NC_047047.1	29384273	RELN	NC_047042.1	19635931
CLEC9A	NC_047044.1	94494661	RNASE4	NC_047035.1	75834234
CTC1	NC_047053.1	8244520	RNASE4	NC_047035.1	75834236
CUBN	NC_047035.1	171046708	RNASE4	NC_047035.1	75834237
CUBN	NC_047035.1	171082995	RNASE4	NC_047035.1	75834260
CUBN	NC_047035.1	171115501	RNASE4	NC_047035.1	75834271
CUBN	NC_047035.1	171124118	RNASE4	NC_047035.1	75834288
CUBN	NC_047035.1	171161782	RNASE4	NC_047035.1	75834289
CUBN	NC_047035.1	171282812	RNASE4	NC_047035.1	75834304
FSTL4	NC_047036.1	110312074	RNASE4	NC_047035.1	75834459
GLP2R	NC_047053.1	7043281	RNASE4	NC_047035.1	75834559
HSD17B4	NC_047036.1	97995841	RNASE4	NC_047035.1	75834569
IL20RA	NC_047045.1	63125623	SLC17A2	NC_047043.1	20789008
INPP4B	NC_047038.1	115486930	SORBS1	NC_047049.1	31787135
LECT2	NC_047036.1	112821000	STIL	NC_047034.1	145192829
LGR5	NC_047044.1	45410953	TMPRSS15	NC_047037.1	118749862
LOC101320643	NC_047041.1	80614522	TMPRSS15	NC_047037.1	118751353
LOC101324829	NC_047035.1	94761236	TMPRSS15	NC_047037.1	118769983
LOC101325883	NC_047050.1	61145026	TMPRSS15	NC_047037.1	118785956
LOC101325883	NC_047050.1	61145050	TMPRSS15	NC_047037.1	118845446
LOC101326785	NC_047042.1	67202553	TOM1L1	NC_047053.1	24997997
LOC101330624	NC_047036.1	69836996	TRAF3IP1	NC_047040.1	2643598
LOC101330624	NC_047036.1	69837032	TRAF3IP1	NC_047040.1	2643686
LOC101330624	NC_047036.1	69837167	TUB	NC_047041.1	61540345
LRIT2	NC_047049.1	50077722	VIPR1	NC_047043.1	57626052
MYO3B	NC_047040.1	61508334	WDR77	NC_047034.1	86207593
MYO3B	NC_047040.1	61585837	WNT8A	NC_047036.1	114863112
MYO3B	NC_047040.1	61506120			

PCA and PCoA for both datasets do not suggest an influence involving geography or a general effect associated with functional loci (Figures 23-24).

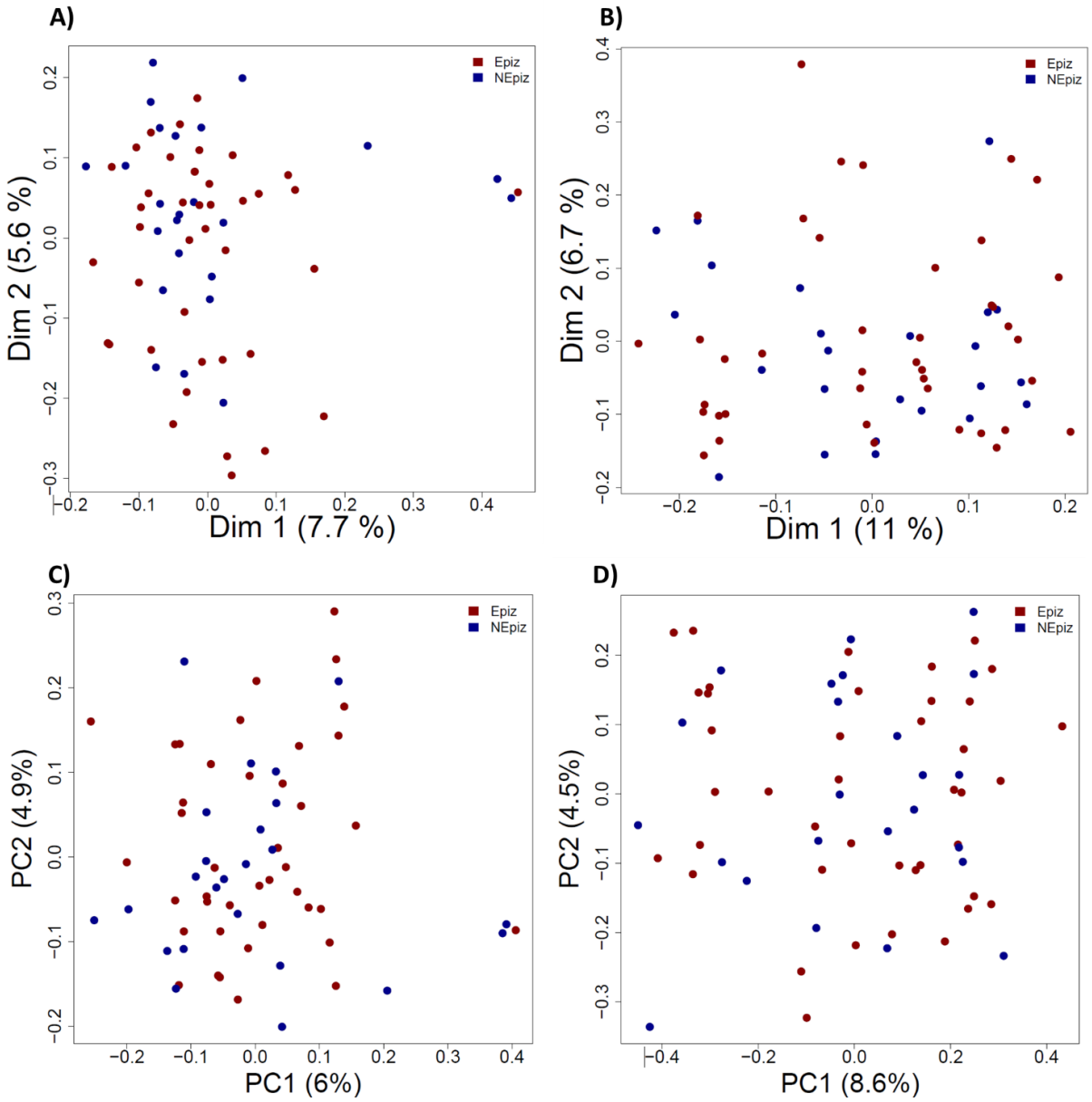


Figure 23 – Ordination analyses with A,B) principal component analysis (PCA), and C,D) principal coordinate analysis (PCoA), of the epizootic (Epiz) and non-epizootic (NEpiz) periods in the Mediterranean Sea, using the immune system group/dataset.

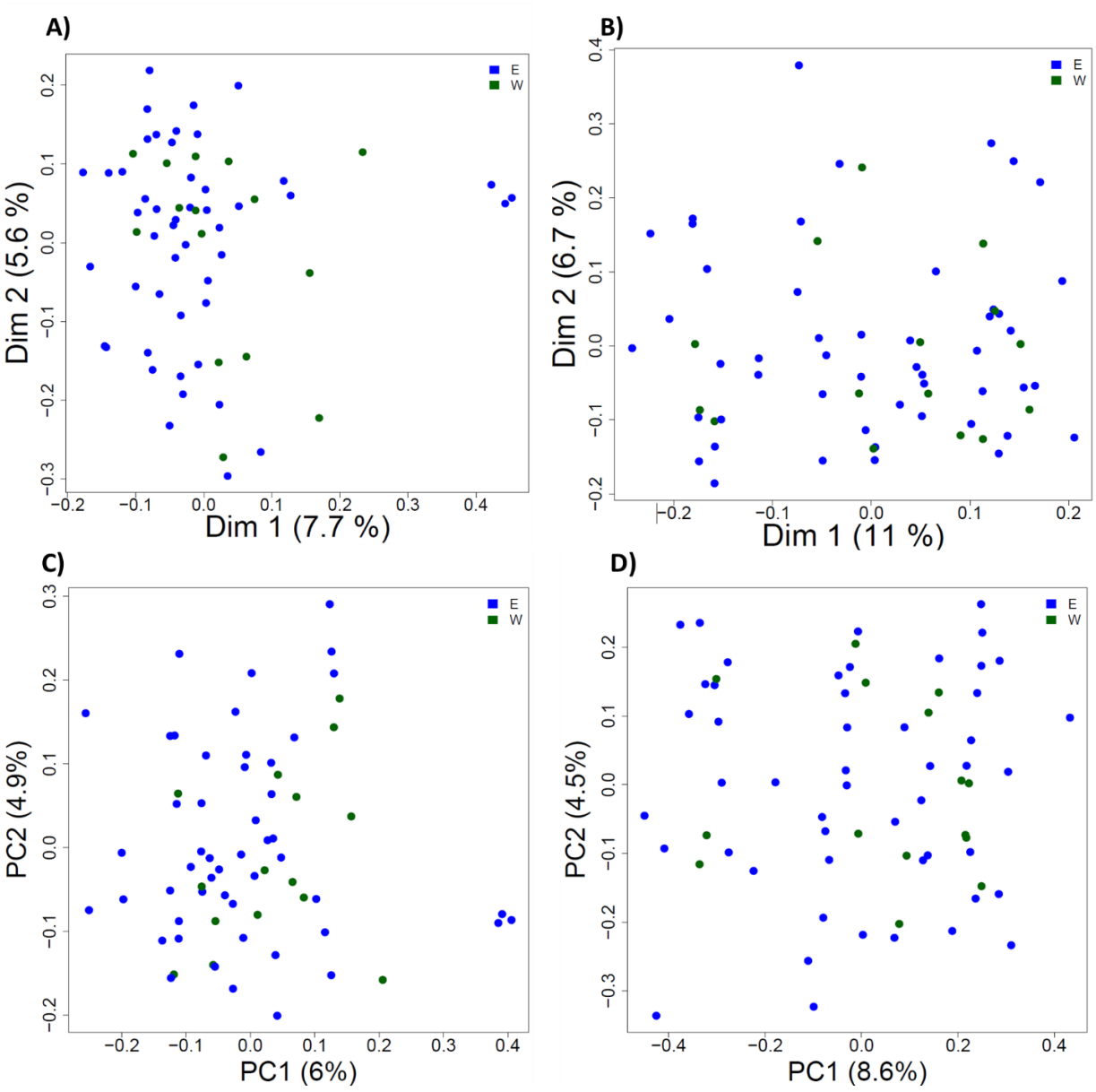


Figure 24 - Ordination analyses with A,B) principal component analysis (PCA), and C,D) principal coordinate analysis (PCoA), of the West (W) and East (E) regions of the Mediterranean Sea, using the other functional loci group/dataset.

Genetic differentiation between the Mediterranean regions or the epizootic periods, by pairwise F_{ST} analysis, corroborates the absence of differentiation for both datasets involved, with values varying between 0.005 (epizootic vs non-epizootic, within both datasets), 0.007 (East vs

West region, within the immune system dataset), and 0.005 (East vs West region, within the other functional loci group/dataset). Genetic distance by Nei supports the previous results of no differentiation, and high variability in genetic distance within epizootic periods and the Mediterranean regions (Figure 25).

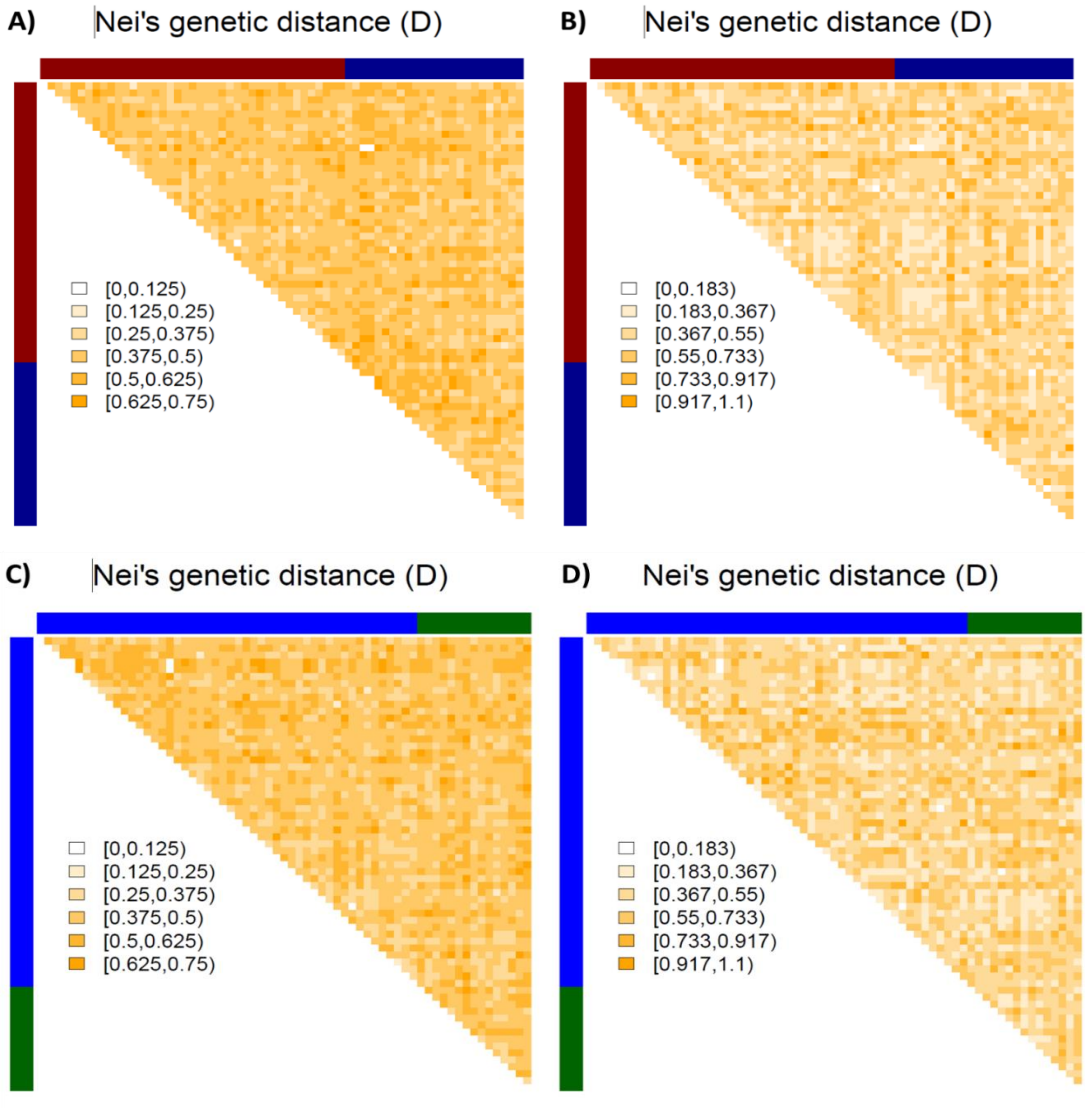


Figure 25 - Nei's genetic distance representing genetic differentiation between individuals for the immune system group/dataset (A, C), and other functional loci group/dataset (B, D). Individuals from the epizootic (red) and non-epizootic (dark blue) periods are represented in A and B, and West (green) and East (blue) region of the Mediterranean Sea are represented in C and D.

The ADMIXTURE plot also supported the absence of a clear differentiation between epizootic periods or geographic regions (Figure 26-27).

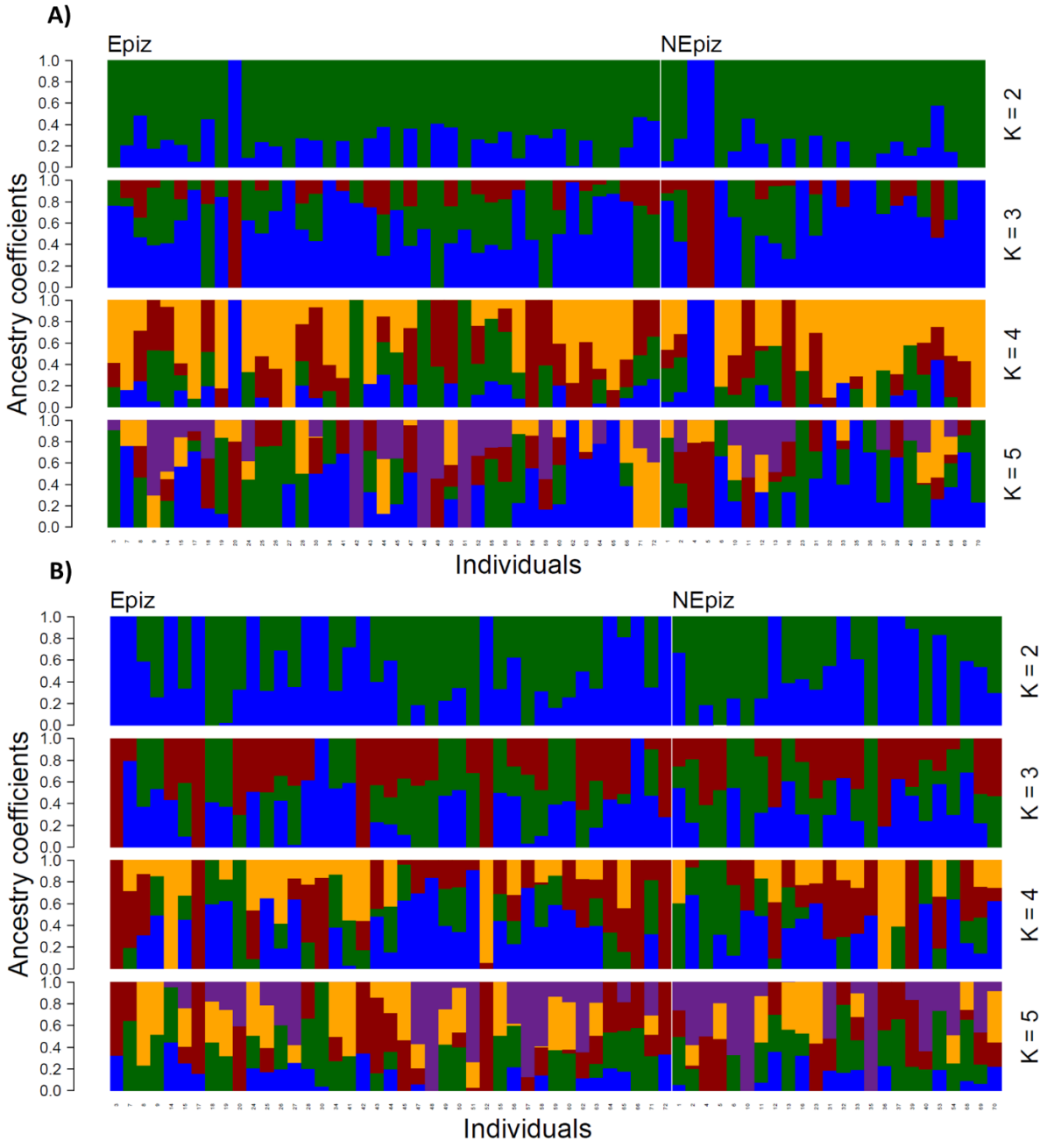


Figure 26 - Admixture plot and ancestry coefficients based on the A) immune system group/dataset, B) other functional loci group/dataset, comparing epizootic (Epiz) and non-epizootic periods (NEpiz). K represents the number of clusters for structured populations.

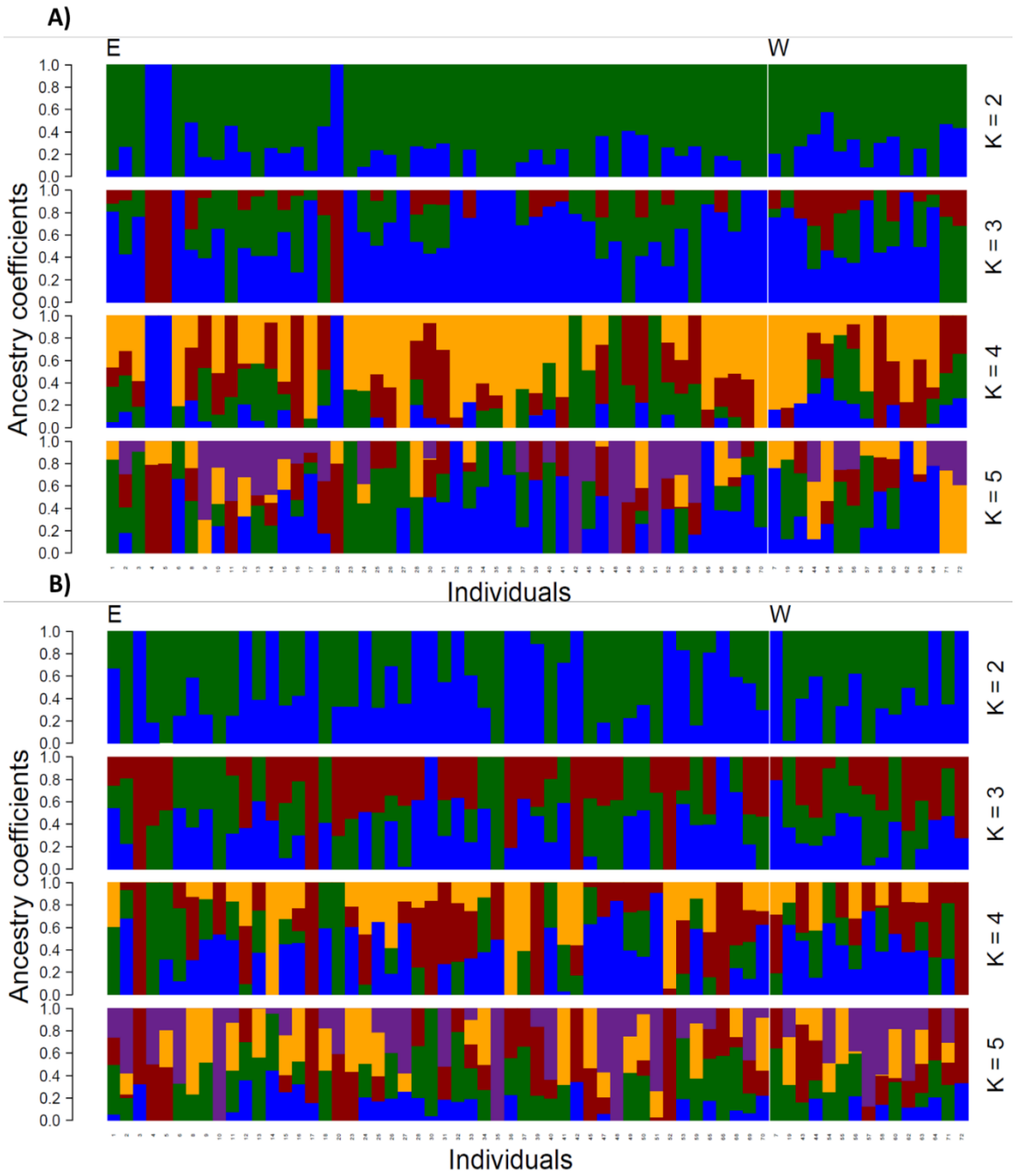


Figure 27 - Admixture plot and ancestry coefficients based on the A) immune system group/dataset, B) other functional loci group/dataset, comparing the West (W) and East (E) regions of the Mediterranean Sea. K represents the number of clusters for structured populations.

4.4 Discussion

Host-pathogen dynamics are considered as key drivers of evolution, with genetic factors of both as the most important determinants of vulnerability and survival outcomes. Although the immunogenetic diversity is the basis of the host defence mechanisms – influencing its resistance, tolerance or susceptibility to infection, few studies have focused on a deeper understanding of these complex networks in wildlife disease (Bossart et al., 2019; Smith et al., 2009). The identification of immune system pathways associated with the morbillivirus disease in wild cetaceans' species, has been considered a potential key to understand the highly contagious and fatal outcomes in these animals, especially those that seems to be particularly vulnerable to this virus, like the striped dolphin. The possible association between genome-wide variation within candidate genes of the Mediterranean striped dolphin, expected to be related to viral resistance, and morbillivirus epizootic events, were explored.

Previous studies have attempted to uncover the relationship between genetic variants and immune response to this infectious disease in cetaceans. However, these were studies that generally targeted a small number of functionally important genomic regions, such as MHC, SLAMF1, NECTIN4, HLA, and thus, with reduced representation (Hashiguchi et al., 2011; Stejskalova et al., 2017; Batley et al., 2019). Batley et al. (2021) aimed to look for signs of selection in a dolphin species – *Tursiops aduncus*, using a non-target approach and whole genome-based

information, which retrieved 50 candidate genes related to the immune system response against morbillivirus, with high levels of homozygosity.

In contrast with that result and the results from chapter 3, the missense variations identified here revealed high overall levels of heterozygosity during epizootic periods. This difference, however, was significant for only one locus (HAVCR2). Low levels of genetic diversity, with higher homozygosity in immune-related genes, tend to be related to reduced fitness and adaptive potential, which can be determinant in, for example, pathogen recognition (Blanchong et al., 2016; Smith et al., 2009). No loss of function was observed from snpEff analyses, and the majority of missense variations identified in the present study were related to the immune system, as expected. The strongest effect was for HAVCR2, important with the regulation of macrophage activation, it is a cell surface viral receptor that modulates innate and adaptive immune responses, being associated with cytokine signaling in immune system, and interleukin-2 family signaling pathways (Monney et al., 2002; Gorman and Colgan, 2014). It is thought to also have a dramatic impact acting as an inhibitor of T-helper type 1 lymphocyte when upregulated, which leads to immunological tolerance and immunopathology (Sánchez-Fueyo et al., 2003) by overexpression, hyperactivation, exhaustion, and paralysis of T-cells, also observed in severe COVID19 cases (Kalfaoglu et al., 2020).

In the sample set presented here, the minor allele (A) was found only in heterozygotes, and the minor allele frequency increased from zero outside of the epizootic periods to 19.5% among those that were found dead

during the epizootic. It is possible that the homozygous condition (GG) at this residue in this locus is highly protective. Recent literature available suggests that heterozygous mutations in this gene are associated with severe diseases, such as hemophagocytic syndrome (HLH) in subcutaneous panniculitis-like T-cell lymphoma (SPTCL) (Polprasert et al., 2019; Zhang et al., 2022). These were considered pathogenic mutations in germline TIM-3, which is responsible for amplifying inflammatory signals (Zhang et al., 2022). However, the precise role of this gene in the cetacean-morbillivirus interaction is unknown.

TMOD3 and TMPRSS15 revealed the same variation pattern for MAF as HAVCR2. TMPRSS15, responsible for the activation of pancreatic proteolytic proenzymes, is known to play a role in viral proliferation for the influenza A virus (Hayashi et al., 2018). Both morbillivirus and the influenza A virus have negative-sense single-stranded RNA genomes. TMOD3 has been associated with hepatic cancer cell proliferation, invasion and migration in humans (Zheng et al., 2019). Among the other loci that showed some genotypic differentiation between epizootic and non-epizootic periods, AIMP1 is known for its critical role in antitumor and antiviral immune responses. It is usually upregulated in cases of viral infection (Liang et al., 2018). With related pathways associated with viral mRNA translation, and tRNA aminoacylation, this gene also possesses inflammatory cytokine activity, wound healing/repair, and its downregulation was suggested to be associated with SARS-CoV-2 infection (Feng et al., 2022; Feng and Zhang, 2022). CRTAM is part of the relevant

innate immune system pathways, and associated with class I MHC mediated antigen activity. This locus regulates the activation, differentiation and tissue retention of various T-cells, and exhibits cytotoxic activity (Boles et al., 2005; Takeuchi et al., 2016). This is also a commonly identified gene among COVID-19 patients (Alqutami et al., 2021). IL26, highly expressed in chronic inflammatory diseases, is involved in the production of inflammatory cytokines and chemokines, which in turn recruit neutrophils to inflammation sites, suggested to be a potent pro-inflammatory mediator (Paiva et al., 2021). It is also associated with mucosal immunity, presenting an overexpression in herpesvirus saimiri-transformed T cells (Hör et al., 2004; Commins et al., 2008). Besides its crucial role in the regulation of glycolysis, LOC117310972 (like-PKM) is also associated with the innate immune system pathway, and has been suggested to interact with and induce the replication of the influenza virus (Buschow et al., 2010; Miyake et al., 2017; Liu et al., 2022). Although proposed to have relevance to morbillivirus infection in other studies (see Introduction), viral receptor SLAMF1 showed no clear association in this study.

In some cases, the immune system role of a certain gene could be indirectly expressed and thus its function may be harder to detect. For example, although CUBN is associated with lipoprotein, vitamin, and iron metabolism (Kozyraki et al., 1999), it is thought to have an important function on the maternal-fetal interface by binding to LGALS3. Associated with acute inflammatory activity, this complex is involved in a cascade of

immune responses including neutrophil and mast cell functions, and chemoattraction of monocytes and macrophages (Crider-Pirkle et al., 2002). Additionally, in coronavirus infection, it is suggested to be involved in remyelination (the generation of new myelin sheaths when upregulated), which is a determinant for the correct function of the CNS (Sariol et al., 2020). As mentioned in the previous chapter, one of the main affected organs in cetaceans with morbillivirus infection is the brain. Although this gene was not within the differentiated missense variations, it is indirectly related to one (though not one significant beyond the false discovery threshold). PDE5A is associated with muscle cell relaxation, and vasodilation, being particularly important in the regulation of the vascular tone (Lin et al., 2006), which is another system critically targeted and impacted by morbillivirus. In fact, PDE5A was suggested to be related to pulmonary hypertension pathology, being highly upregulated in the pulmonary endothelium of patients suffering from this disease (Gebska et al., 2011). This gene was also suggested to be a potentially good biomarker for COVID19 and ischemic stroke (Wu and Han, 2023).

Taken together, the present results with the ones from the previous chapter, it is evident that several SNPs were not necessarily linked with the immune system pathway in a direct way, but with processes of the main target organs for this virus. Looking deeper into those pathways and respective functional loci, might also be relevant to better understand this susceptibility.

Although there is no information available related to the role and related pathways of the above mentioned genes in cetaceans facing morbillivirus pressure, the present results together with the general information available in other mammals, as well as the fact that none of these same loci were reported in Batley et al. (2021) candidate gene list, supports the need for future studies to further investigate and better characterize these genes in the context of morbillivirus infection. Also, no clear pattern was observed when multiple loci were grouped by functionality, suggesting the absence of genetic differentiation patterns and the need to focus on specific loci. Because different genotypes can be introduced in a population by mutation (including recombination) but can also be transferred among distinct populations by migration (gene flow), further studies, with larger and balanced sample set including samples from Atlantic individuals, would be important to identify the process through which relevant variations were introduced. Additionally, and because strandings can occur from a variety of reasons, data on the infection status of individual dolphins, as well as other important factors such as age, sex, and contaminant load, would be crucial for accurate interpretations.

Chapter 5

General Conclusions

The objective of this thesis was to better understand the genomic basis of striped dolphin susceptibility and resistance to morbillivirus disease, as well as the genetic diversity and structure of this species in the Mediterranean Sea. The approach was to assess aspects that could be influencing the vulnerability of striped dolphin to morbillivirus infection, including association of death during epizootics with inbreeding and genotype at susceptibility-related candidate genes. The present chapter aims to integrate the findings presented in the previous chapters and consider them in the context of the literature.

5.1 Genomic diversity of the Mediterranean striped dolphin reflects weak geographical structure in the western basin and no correlation with epizootic period

Striped dolphin has a world-wide distribution, being adapted to a wide range of habitats, and is the most abundant cetacean species in the Mediterranean Sea. In Chapter 2, the most prominent genetic distinction observed was between the Atlantic and Mediterranean individuals, but at a level that would be consistent with some level of connectivity (and exposure to interaction with Atlantic populations). However, the sample size from the Atlantic was very small. Previous studies using lower resolution genetic markers have shown the same geographic structure (Bourret et al., 2007; Garcia-Martinez et al., 1999; Gaspari et al. 2007,

2019; Gkafas et al., 2017). It is suggested that this may be associated with phylogeographical boundaries also highlighted for other species (e.g. bottlenose dolphin – *Tursiops truncatus* Natoli et al., 2005; common dolphin – *Delphinus* sp. Natoli et al., 2006; harbour porpoises – *Phocoena phocoena* Fontaine et al., 2014; sperm whale – *Physeter macrocephalus* Violi et al., 2023; common sole – *Solea solea* Rolland et al., 2007).

Possible structure among samples in the Mediterranean Sea was tested to help determine the potential for connectivity (disease spread) and whether the assumptions of panmixia would be violated when including all samples in further analyses. All but a few samples were from the western basin of the Mediterranean Sea (see chapter 2), and although some evidence of structure was found between eastern and western regions within the basin using ordination analyses, high levels of connectivity seem likely, and combining sample sets did not appear to violate the assumptions of panmixia.

Previous studies can be informative about the potential for isolation and connectivity throughout the region. Using microsatellite loci, Gaspari et al. (2019) found low levels of differentiation between the Ligurian and Adriatic ($F_{ST} = 0.008$), and the Ligurian and Balearic ($F_{ST} = 0.01$). Valsecchi et al. (2004) also found a significant difference either side of Italy ($F_{ST} = 0.012$), as did Gaspari et al. (2004). When comparing 29 samples from Greece and Israel to 111 samples from Gibraltar to the Ionian Sea, Gkafas et al. (2017) found significant differentiation ($F_{ST} = 0.03$), but also estimated contemporary gene flow and found a strong directional effect

from the west to the east. Population genetic structure within the Mediterranean has been reported for various cetacean species (Natoli et al., 2005), particularly between the eastern and western basins (Moura et al., 2013a). Particular areas with specific hydrographic characteristics and oceanographic features may promote habitat differentiation, and distinct levels of transition and isolation (Theocharis et al., 1993; Kourafalou et al., 2004), which in turn, can impact gene flow (Limborg et al., 2012).

The historical demographic profile of a species can also influence their present distribution and pattern of population differentiation. Numerous refugia were generated during the last glacial period, which affected distribution patterns that are still influencing some species today. For example, orca (*Orcinus orca*) in the North Pacific near Nemuro Strait may be descendants of ancestors that settled there around 20,000 years ago (Filatova et al., 2023). This inference was based on a pattern of genetic diversity consistent with continued occupation of a refugial region rather than post-refugial expansion, together with unique cultural traits.

Population structure is highly related to effective population size, with smaller populations having the potential to differentiate more quickly by genetic drift. Changes in genetic variation and group composition can be influenced by various factors, such as external pressures and environmental dynamics, which in turn can potentially influence demographic patterns and thus, genetic changes. When analysing genetic variations through time, Gaspari et al. (2019) found the most significant changes across a temporal rather than a spatial scale, suggesting a

correlation between the cyclical fluctuations encountered with previous morbillivirus epizootics. It had been reported that individuals found stranded during the epizootic outbreaks had higher levels of inbreeding, which is associated with higher vulnerability to infectious diseases and immune system impairment (Valsecchi et al., 2004; Smith et al., 2009; Gaspari et al., 2019). Gkafas et al. (2020) also reported a relationship between genetic diversity and disease, with lungworm infection in striped dolphins. Although there was no effect overall, a sex-specific pattern was detected for females, probably associated with pregnancy stress, which can also be affected by vertical transmission – a hypothesis that has also been proposed for morbillivirus infection (Soto et al., 2011).

Gaspari et al. (2019) proposed that fluctuations of population F_{IS} levels within and between morbillivirus epizootic periods were significantly different to those expected by chance. However, that pattern was not found in the data presented here. In fact, not only was there no associated pattern of genetic variation across time with the morbillivirus outbreaks, but the levels of inbreeding were, in general, low. Batley et al. (2019, 2021) rejected a correlation between morbillivirus infection and high levels of inbreeding for Indo-Pacific bottlenose dolphins (*Tursiops aduncus*) infected with the same virus.

Heterozygosity–fitness correlation analyses have been investigated in several studies associated with different proxies for fitness, such as parasite load (Rijks et al., 2008; Gkafas et al., 2020), reproductive success (Slate et al., 2000), and longevity (Coltman et al., 1999). Gkafas et al.

(2020) also reported an absence of correlation between parasite load and levels of genetic diversity for common dolphins, suggesting this may be associated with sample size or virulence level of the pathogen in that species. From the data presented in chapter 3, genome-wide heterozygosity for the striped dolphin was found to be within the average values reported for mammals (see Teixeira and Huber, 2021). No significant correlation was found between diversity levels and fitness associated with surviving morbillivirus infection. However, the disease status of the sampled individuals used in this study was not known, and only the timing of stranding and respective sample collection was documented. Although the majority of samples were collected within well-known epizootic outbreaks, that does not guarantee a cause of death attributed/exempt to morbillivirus infection. As necropsies were not conducted, other relevant factors, such as health status, body condition, and age, were not taken into account and, together with infection status, could have biased the present study. In fact, although there is already some indication of a correlation between age and infection level, with greater age having higher infection burden (Gaspari et al., 2019), the current literature is still limited about these mechanisms and how they can directly impact demographic and population structure dynamics.

5.2 Genome-wide signatures and potential candidate genes of morbillivirus pressure

Overall, contrary to the suggestions of earlier studies for this species, there was no association between genomic indicators of inbreeding and viral susceptibility in this study, nor strong effects associated with missense or loss of function mutations.

However, when looking for correlations from random genome scans, the Manhattan plot analysis allowed the identification of 23 loci beyond an assigned threshold (though well below the Bonferroni correction) when the epizootic and non-epizootic peak periods were compared. As expected, genotype differences were mostly significant (beyond correction for type one error), however they also showed greater levels of homozygosity during peak epizootic periods. This pattern was also observed in 84% of samples of *Tursiops aduncus* within 11 SNPs in immune system genes suggested to be involved in morbillivirus immune responses in Batley et al. (2021). High levels of homozygosity in immune system genes tend to be associated with pathogen recognition impairment (Blanchong et al., 2016; Smith et al., 2009). Identified SNPs in/close to immune system genes were related to antiviral innate immunity (TANK; Guo and Cheng, 2007), immune response (LIPA; Yan et al., 2006), innate immune system (FRMPD3, ENDOD1; Bauler et al., 2008; Fenech et al., 2020), viral mRNA Translation (LOC101319173 like-40S RPS21, Smirnova et al., 2020), autoantigen (POP1; Abdulhadi-Atwan et al., 2020), and DNA-damage

regulation and infectious diseases (TENT4A; Lim et al., 2018). Other SNPs were also associated with loci that have relevant functions, such as neural and nervous system development/ maintenance (DOK6, UNC5B; Crowder et al., 2004; Bhat et al., 2020), nervous system processes, skin tumours and ocular abnormalities (NF2; Grönholm et al., 2005), cortical neuron migration and dendritic morphology (FILIP1; Nagano et al., 2002), neurological and cognitive disorders (KIRREL3; Bhalla et al., 2008), and signaling receptor binding and viral receptor (EFNB2; Piffko et al., 2022), particularly for paramyxoviruses – the same family to which the morbillivirus belongs to.

From all SNPs identified, only one represented a non-synonymous change – USH2A, which is associated with vision, auditory, and neural system functions (Perrino et al., 2020). This locus, together with LIPA and NF2 in its inverted form – INF2, although as not immune-related candidates, were all suggested as putative candidates under selection for morbillivirus resistance in Batley et al. (2021). USH2A was also the only locus with an irregular pattern of heterozygosity frequency among periods, with a tendency for increased heterozygosity during the epizootics. Although, to the best of my knowledge, there is no information available about how morbillivirus infections could impact auditory and visual functions, it is known that in general neural system related processes suffer from viral infection. Such impacts might lead to degeneration at the sensorial and neuronal level in the inner ear and the retina (Reiners et al., 2005, 2006). In humans, Usher syndrome – characterized by a combined

deaf-blindness, is suggested to be associated with a USH-protein network, in which USH2A is present (Reiners et al., 2005, 2006).

Although some SNPs identified using the Manhattan plot were not within coding regions, and may or may not be associated with regulation effects, there were nevertheless strong differences in genotype frequency found between epizootic and non-epizootic peak periods. For a given locus this could be due to linkage disequilibrium with a functional region or drift. Overall, because the pattern of genotype differences was the same for most loci (higher homozygosity during the epizootics), and related functions of nearby loci when known were consistent with what could be expected in the context of morbillivirus infection, this pattern seems unlikely to be by chance. However, the Manhattan plot could pick out SNPs based on differences in the context of morbillivirus epizootic periods, and there were few that stood out. When over 1059 SNPs associated with missense mutations were picked out from the RAD data, there was little if any correlation with the epizootic periods. The only patterns found were insignificant trends (not significant after Bonferroni correction) seen at ENDOD1, which has relevant functions for innate immune signalling through the cGAS-STING pathway (Fenech et al., 2020), and LOC101316607 like-PRAMEF8, predicted to be particularly relevant in gametogenesis and oncogenesis (Kern et al., 2021). Consistent with the results based on genome-wide inbreeding, the data suggest that specific loci will be important rather than a general effect.

The search for signs of selection in wild cetacean species associated with morbillivirus disease, has been highly focused on immune system genes. At the same time, it is important to not disregard other mechanisms and pathways that might be related to resistance and adaptation strategies, even when not directly considered as part of the immunological process. In general, investigating specific candidate gene systems expected to be related to viral resistance, is considered to be a more precise and direct approach, especially in those rare cases when the gene system is well understood. This is potentially useful to help uncover susceptibility determinants to morbillivirus using informed inference, rather than only looking for correlations from random genome scans.

This was the approach undertaken in Chapter 4, with a hybridization-based capture procedure, having specific genomic regions as targets of interest. From this analysis, 210 SNPs were identified as missense variations, none of which reflected loss of function. Within 9 loci that presented variations between the outbreaks (TMPRSS15, AIMP1, CRTAM, HAVCR2, IL26, PKM, CUBN, PDE5A, TMOD3), the strongest effect was found for HAVCR2, a cell surface viral receptor that modulates innate and adaptive immune responses, being associated with cytokine signaling in the immune system, and interleukin-2 family signalling pathways (Monney et al., 2002; Gorman and Colgan, 2014).

To assess possible signs of geographic variation and a general effect in and out of epizootic periods, all missense variations were grouped according to loci having either immune system or other functions. No clear

patterns were observed, suggesting the absence of genetic differentiation either geographically or over general functions between epizootic periods. This reinforced the importance of also considering specific loci. When doing so, those that showed a significant pattern revealed high overall levels of heterozygosity during those periods, instead of the homozygosity seen at the outlier loci in chapter 3. This was also the opposite from Batley et al. (2021) who found an excess of homozygotes at 50 candidate immune system-related genes in *Tursiops aduncus* using a non-target approach and whole genome-based information. Homozygosity detected in random regions of the genome, but heterozygosity preserved in those related with immune system loci, is a pattern already reported in other studies with different taxa even during pandemic events (Smith et al., 2009). Because genetic diversity tends to be associated with higher fitness and so more chances of survival, this might be related with molecular trade-offs that enable the prioritization of this diversity on considered crucial regions of the genome. However, despite the widely reported tendency described above, heterozygosity might sometimes be related with not so beneficial effects. For the Mediterranean striped dolphins, HAVCR2 showed the strongest effect, where the minor allele frequency increase from 0% outside of epizootics, to 19.2% within the epizootics. Increased frequency of the minor allele will increase heterozygosity under the Hardy-Weinberg equilibrium (HWE) - maximum when p=q=0.5. This, together with the literature available suggesting severe diseases associated with heterozygotes in this gene (please see below), may suggest that the major

allele is protective at this locus, but more research is required to better understand the biological mechanisms involved in this case. The same pattern was seen for TMOD3, TMPRSS15. The opposite pattern was seen for PDE5A and IL26 which had greater homozygosity during the epizootics (as seen for the outlier loci in chapter 3). The same was observed for SLAMF1, one of the most studied genes associated with morbillivirus susceptibility in cetaceans, although with very weak variations in the present study.

These different variation patterns reflecting cyclical fluctuations in the genetic composition of individuals found dead over time, could be related to the function of those loci in most cases. It is possible that this may be related to the gene-for-gene model (Flor, 1942; Flor, 1955; Woolhouse et al., 2002), where there is matching between host and pathogen genotypes, though this cannot be fully tested with the current dataset. If the major allele at HAVCR2 has a protective role, that could suggest greater vulnerability for individuals that by chance have the minor allele. Although like SLAMF1, HAVCR2 is also a cell surface viral receptor, which encodes T cell immunoglobulin and mucin domain-containing protein 3 (TIM-3), the role of this gene in cetacean-morbillivirus interaction is unknown. In fact, to my knowledge, this study is the first report of missense variations on HAVCR2 in dolphins. However, the literature available suggests that heterozygotes in this gene are associated with severe diseases, such as hemophagocytic lymphohistiocytosis (HLH) in subcutaneous panniculitis-like T-cell lymphoma (SPTCL) (Polprasert et

al., 2019; Wegehaupt et al., 2020; Zhang et al., 2022; Planas et al., 2023). These were considered inflammatory and pathogenic mutations in germline TIM-3 (Gayden et al., 2018; Wolf et al., 2020; Zhang et al., 2022). This molecule acts as a checkpoint receptor, being a critical regulator of innate immunity and inflammatory responses, including monocyte and macrophage activation (Wolf et al., 2020; Planas et al., 2023). When suppressed or inactive, it over-activates T cells, amplifying the inflammatory signals (Gayden et al., 2018; Wolf et al., 2020; Zhang et al., 2022). These mutations, responsible for a rare form of lymphoma in humans, were considered to be a result of vertical transmission (Gayden et al., 2018). TIM-3 coded by HAVCR2 gene, might be a promising molecule in immune system regulation and a potential target for immunotherapy. HAVCR2 was also considered to be involved in acute T cell lymphoblastic leukemia (T-ALL), as part of the regulatory circuit mechanism (Zhu et al., 2018). Recent literature has also suggested that HAVCR2 might influence immunological tolerance scenarios, also observed in severe COVID19 cases (Sánchez-Fueyo et al., 2003; Kalfaoglu et al., 2020; see Chapter 4). TMOD3 DNA site 38 changed from G to A causing amino acid site 13 to change from Glycine to Aspartic acid – c.38G>A (p.Gly13Asp) and TMPRSS15 c.1997A>C (p.Gln666Pro) where 'c' is for coding nucleotide number and 'p' is for protein amino acid number. These changes shared the same pattern and are both related to relevant immune system functions. While TMPRSS15 plays an important role in viral proliferation for the influenza A virus, which like morbillivirus has a negative-sense single-stranded

RNA genome (Hayashi et al., 2018), TMOD3 has been associated with hepatic cancer cell proliferation, invasion and migration in humans (Zheng et al., 2019).

Also worth noting, PDE5A c.1965C>A (p.Asp655Glu) and IL26 c.34C>A (p.Leu12Met) had higher heterozygosity outside epizootic outbreaks, which may suggest again a relationship between MAF and morbillivirus susceptibility. IL26 is associated with mucosal immunity, presenting an overexpression in herpesvirus saimiri-transformed T cells (Hör et al., 2004; Commins et al., 2008), and is also suggested to be a potent pro-inflammatory mediator (Paiva et al., 2021). PDE5A was suggested to be associated with pulmonary hypertension pathology (Gebska et al., 2011), regulation of the vascular tone (Lin et al., 2006), and also a potential biomarker for COVID19 and ischemic stroke (Wu and Han, 2023). Both of these systems are the main targets of morbillivirus disease.

Because the mutation sites reported in the literature available do not match the ones revealed in this study, and none of the mentioned loci were reported in Batley et al. (2021) candidate gene list, deeper research at this level would be important to better understand the precise interaction mechanisms between the virus and these potential candidate genes in cetaceans, as well as the true nature of these mutations.

5.3 Final remarks and future perspectives

Investigations into population structure for this study were primarily to confirm panmixia for the sample set to be used to assess susceptibility

to morbillivirus, and to consider the historical demographic profile. The latter suggested a decline, but perhaps associated primarily with the climatic transition out of the Pleistocene. Only very weak population structure was seen within the western basin, evident in ordination analysis, but not significant for F_{ST}. However, it may suggest a broader pattern of isolation by distance. To better support adequate conservation management and provide a complete understanding of the real population geographic structure and historical demography, further studies should be carried out with high resolution genetic data and robust sample sizes across the species range, including stranded and free ranging individuals to diminish possible influences of local hydrodynamics and currents.

Genome-wide scans revealed low levels of inbreeding, as well as no fitness/heterozygosity correlations, even when comparing males and females, with morbillivirus pressure as the fitness trait. However, the identification of particular loci with functions that were consistent with morbillivirus infection/disease suggests that these findings are unlikely to be by chance and could be related to morbillivirus susceptibility. Previous studies have found fitness/heterozygosity correlations in the same population and other pathogens (e.g. Gkafas et al., 2020). Due to the unknown infection status and sample number for certain years, the possibility of biased results in this present study cannot be excluded.

The hybridization-based capture approach reinforced the indication from other analyses undertaken in this thesis, that there was no clear general effect, even among multiple loci with shared immune system

functions. Instead, one highly relevant locus, HAVCR2, showed the strongest correlation with morbillivirus exposure. Because the majority of loci detected in this present study, with associated missense variations, were not previously reported and have limited information published in the available literature, more research is needed on their role in viral susceptibility. In fact, and because dolphins are considered as good models of human disease studies, several identified loci have been associated with COVID19 disease, which may permit some extrapolated conclusions.

Mortality rate as well as the number of strandings associated with epizootic outbreaks, have been characterized by a progressive reduction in population size in recent years (Rubio-Guerri, et al., 2013; Casalone et al., 2014; Van Bressem et al., 2014; Mira et al., 2019). Some authors suggest a possible endemic viral circulation and infection cycle in the region, able to cause systemic infections (Rubio-Guerri, et al., 2018; Vargas-Castro et al., 2023) with milder symptoms. Supporting this, Vargas-Castro et al. (2023) recently reported one of the highest prevalence levels of CeMV in Mediterranean stranded cetaceans (mostly striped dolphins), not associated with an epizootic outbreak. Of 354 stranded dolphins collected between 2018 and 2021, 113 were infected (while nearly all would be expected to be infected during an outbreak; Domingo et al., 1990; Aguilar and Raga, 1993; Raga et al., 2008; Keck et al., 2010; Casalone et al., 2013; Rubio-Guerri et al., 2013; Mazzariol et al., 2016, 2017; Pautasso et al., 2019). Therefore, this also suggests that Mediterranean striped dolphin could be a reservoir for morbillivirus in the region, and also an ideal model

for future research that aims to better understand the adaptive process and mechanisms against pathogenic infection. This, together with the fact that morbillivirus infection has been increasingly observed in other marine mammals, including endangered species, emphasizes the urgency of future research.

Overall, further studies also based on a time-series analysis, with a more balanced and larger sample set, including stranded and free ranging individuals, should be carried out to provide a more accurate interpretation of these data. Also, and because strandings can occur from a variety of reasons, data on the infection status of individual dolphins, as well as other important factors such as age, sex, and contaminant load, would be crucial to accurately interpret the present results. In some cases, the analyses were based on few individuals per category, and sequence mapping was against the reference genome available for the most related species: *Tursiops truncatus*. It is possible that the results also reflect these limitations. Mapping against the exact species reference genome (when available), before reanalysing the pipeline, could also contribute to obtaining a more complete set of relevant loci. Also, because different genotypes can be introduced in a population by mutation but can also be transferred among distinct populations by migration, further studies, with robust sample sets from Atlantic individuals would be important to identify the process through which relevant variations were introduced.

Successive epizootics are suggested to be related to changes not only at the genetic level, but possibly with age structure too. The influence of

demographic and ecological factors should be considered more fully in future studies. Although both sexes seem to be affected in a similar way, sexually mature individuals tend to suffer the highest mortality rates, and pollutant load has been suggested as one of the possible explanations (Calzada et al., 1994). As mentioned previously, several studies have reported high levels of inbreeding associated with epizootic, but also the recovery of those values to pre-outbreak levels within 10 years of the end of the epidemic (Valsecchi et al., 2004). The available literature suggests that demographic patterns might be correlated with epizootic events, by population dynamic mechanisms (Gkafas et al., 2017; Gaspari et al., 2019), such as density-dependent recruitment, and immigration factors, which could be associated with a higher susceptibility to morbillivirus infection.

On the other hand, and given the correlations found between genetic variation at genotype level and epizootic events, the detection of selective sweeps could unveil additional genetic adaptation mechanisms involved in epizootic susceptibility/survival to morbillivirus. This may represent processes beyond the epizootic recovery derived from population growth and immigration.

Taking into account the complexity of the natural environment, it may be relevant to look into factors that may exacerbate dolphin vulnerability. For example, chemical pollution is known to be one of the main threats to cetaceans, having a considerable impact on their immune system function. High levels of contamination in the environment, which is typical in the Mediterranean Sea, are considered to contribute for the

presence of virus, parasites and the severity of its impacts on hosts (Kannan et al., 1993; Nakayama et al., 2009; Fossi et al., 2013; Desforges et al., 2016). In the particular case of the Mediterranean striped dolphin, pollutant contamination has been hypothesised as an important contributor to Morbillivirus susceptibility (Aguilar and Borrell, 2005; Profeta et al., 2015). However, most studies of contamination effects on dolphin immune function have focused on analyses of contaminant load in stranded animals showing signs of infection, with limited analyses on the effects such pollutants can have on gene expression (Desforges et al., 2016). From the myriad contaminants released into the environment by human activity, xenobiotics (e.g. POPs - persistent organic pollutants such as polychlorinated biphenyl – PCB, and polycyclic aromatic hydrocarbons – PAH; and heavy metals) have received particular attention in the literature dedicated to marine environment, given their high toxicity and environmental persistence (Nomiyama et al., 2014; Desforges et al., 2016). Given that cetaceans are top predators, they are particularly prone to environmental contamination due to bioaccumulation and biomagnification processes through food webs. Furthermore, PCBs, known to have immunodepressive effects (Kannan et al., 1993; Aguilar and Borrell, 1994), are not readily metabolized by odontocetes (Tanabe et al., 1988; Norstrom et al., 1992; Meyer et al., 2018), and therefore will potentially accumulate more than in other marine top predators, contributing to increase striped dolphin susceptibility to Morbillivirus. A recent study has also found that marine mammals may be particularly

vulnerable to pollution from man-made organophosphorus compounds, due to an ancestral molecular change in a gene (PON1) encoding an enzyme which reduces oxidative damage and defends against neurotoxins found in such compounds (Meyer et al., 2018). It is thought that this change might have been somehow beneficial in ancestral oceans, but result in detrimental fitness effects in the presence of anthropogenic pollutants. Due to the highly industrialized coasts of the Mediterranean Sea and given the above evidence suggesting a higher vulnerability of cetaceans to pollutants such as PCB's, the higher contaminant levels found in this region (Fowler, 1987), may interact synergistically with intrinsic genetic factors to increase striped dolphin susceptibility to Morbillivirus in this region. Identifying reliable molecular and biochemical fitness markers is an important step in the development of biomonitoring tools, which are of utmost importance to gain insights into the overall impact pollution can have on disease susceptibility in wild marine organisms. Understanding such processes for large marine top predators such as cetaceans can be crucial for biomonitoring purposes, given that such umbrella species and considered environmental sentinels, can provide important information on whole ecosystem health (Ozaki et al., 2006).

Taken together, the present time-series data provide a contribution to the scientific understanding of the relationship between genetic variation and susceptibility of striped dolphin to morbillivirus disease, as well as its geographical diversity, conferring relevant information that could be useful to future studies and conservation measures for this species.

References

ACCOBAMS. (2021). Estimates of abundance and distribution of cetaceans, marine mega-fauna and marine litter in the Mediterranean Sea from 2018-2019 surveys. In: Panigada, S., Boisseau, O., Canadas, A., Lambert, C., Laran, S., McLanaghan, R. and Moscrop, A. (eds). Agreement on the Conservation of Cetaceans of the Black Sea, Mediterranean Sea and contiguous Atlantic area (ACCOBAMS) - ACCOBAMS Survey Initiative Project, Monaco.

Abdulhadi-Atwan, M., Klopstock, T., Sharaf, M., Weinberg-Shukron, A., Renbaum, P., Levy-Lahad, E., Zangen, D. (2020). The novel R211Q POP1 homozygous mutation causes different pathogenesis and skeletal changes from those of previously reported POP1-associated anauxetic dysplasia. American Journal of Medical Genetics. Part A, 182, 5, 1268–1272.

Acevedo-Whitehouse, K., Cunningham, A. (2006). Is MHC enough for understanding wildlife immunogenetics? Trends in Ecology and Evolution, 21, 433–438.

Acevedo-Whitehouse, K., Gulland, F., Greig, D., Amos, W. (2003). Inbreeding-dependent pathogen susceptibility in California sea lions. Nature, 422, 35.

Aguilar, A. (2000). Population biology, conservation threats, and status of Mediterranean striped dolphins (*Stenella coeruleoalba*). Journal of Cetacean Research and Management, 2, 17-26.

Aguilar, A., Borrell, A. (1994). Abnormally high polychlorinated biphenyl levels in striped dolphins (*Stenella coeruleoalba*) affected by the 1990–1992 Mediterranean epizootic. Science of the Total Environment, 154, 237-247.

Aguilar, A., Borrell, A. (2005). DDT and PCB reduction in the western Mediterranean from 1987 to 2002, as shown by levels in striped dolphins (*Stenella coeruleoalba*). Marine Environmental Research, 59, 391-404.

Aguilar, A., Gaspari, S. (2012). *Stenella coeruleoalba*. The IUCN Red List of Threatened Species 2012: e.T20731A2773889: https://www.iucnredlist.org/species/20731/2773889

Aguilar, A., Raga, J. A. (1993). The striped dolphin epizootic in the Mediterranean Sea. Ambio, 22, 524-528.

Agrawal, A., Lively, C. M. (2002). Infection genetics: gene-for-gene versus matching-alleles models and all points in between. Evolutionary Ecology Research, 4, 79–90.

Alexander, D. H., Novembre, J., Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Research, 19, 9, 1655–1664.

Alqutami, F., Senok, A., Hachim, M. (2021). COVID-19 Transcriptomic Atlas: A Comprehensive Analysis of COVID-19 Related Transcriptomics Datasets. Frontiers in genetics, 12, 755222.

Amaral, A. R., Beheregaray, L. B., Bilgmann, K., Boutov, D., Freitas, L., Robertson, K. M., et al. (2012). Seascape genetics of a globally distributed, highly mobile marine mammal: The shortbeaked common dolphin (genus *Delphinus*). PLoS One, 7, e31482.

Amarasinghe, G. K., Ayllón, M. A., Bào, Y., et al. (2019). Taxonomy of the order Mononegavirales: update 2019. Archives of virology, 164, 7, 1967–1980.

Amos, W., Harwood, J. (1998). Factors affecting levels of genetic diversity in natural populations. Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 353, 1366, 177–186.

Amos, W., Worthington Wilmer, J., Fullard, K., Burg, T. M., Croxall,J. P., Bloch, D., Coulson, T. (2001). The influence of paternal relatedness on reproductive success. Proceedings of the Royal Society of London Series B, 268, 2021–2027.

Archer F. I., Perrin W. F. (1999) *Stenella coeruleoalba*. Mammalian Species, 603, 1–9.

Archer, N. S., Nassif, N. T., O'Brien, B. A. (2015). Genetic variants of SLC11A1 are associated with both autoimmune and infectious diseases: systematic review and meta-analysis. Genes & Immunity, 16, 275– 283.

Avise, J. C. (2000). Phylogeography: The history and formation of species. Cambridge, MA: Harvard University Press

Bahri-Sfar, L., Lemaire, C., Hassine, O. K. B., Bonhomme, F. (2000). Fragmentation of sea bass populations in the western and Eastern Mediterranean as revealed by microsatellite polymorphism. Proceeding of the Royal Society of London B, 267, 929–935.

Balloux, F., Amos, W., Coulson, T. (2004). Does Heterozygosity Estimate Inbreeding in Real Populations? Molecular Ecology, 13, 10, 3021–31.

Banguera-Hinestroza, E., Evans, P. H. G., Mirimin, L. et al. (2014). Phylogeography and population dynamics of the white-sided dolphin (*Lagenorhynchus acutus*) in the North Atlantic. Conservation Genetics, 15, 789–802.

Barbato, M., Orozco-ter Wengel, P., Tapio, M., Bruford, M.W. (2015). SNeP: a tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. Frontiers in Genetics, 6, 109.

Batley, K. C., Sandoval-Castillo, J., Kemper, C. M., et al. (2019). Genome-wide association study of an unusual dolphin mortality event reveals candidate genes for susceptibility and resistance to cetacean morbillivirus. Evolutionary Applications, 1-15.

Batley, K. C., Sandoval-Castillo, J., Kemper, C. M., Zanardo, N., Tomo, I., Beheregaray, L. B., Möller, L. M. (2021). Whole genomes reveal multiple candidate genes and pathways involved in the immune response of dolphins to a highly infectious virus. Molecular Ecology, 30, 6434-6448.

Bauler, T. J., Hendriks, W. J. A. J., King, P. D. (2008). The FERM and PDZ Domain-Containing Protein Tyrosine Phosphatases, PTPN4 and PTPN3, Are Both Dispensable for T Cell Receptor Signal Transduction. PLOS ONE, 3, 12, e4014.

Bellière, E. N., Esperon, F., Sanchez-Vizcaino, J. M. (2011). Genetic comparison among dolphin morbillivirus in the 1990-1992 and 2006-2008 Mediterranean outbreaks. Infection, Genetics and Evolution, 11, 1913-20.

Belkhir, K., Borsa, P., Chikhi, L., Raufaste, N., Bonhomme, F. (2002). GENETIX 4.04, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions: CNRS UMR 5000, Université de Montpellier II, Montpellier, France.

Belo, J. A., Silva, A. C., Borges, A-C, et al. (2009) Generating asymmetries in the early vertebrate embryo: the role of the Cerberus-like family. The International Journal of Developmental Biology, 53, 1399–1407.

Benton, M. J. (2009). The Red Queen and the Court Jester: species diversity and the role of biotic and abiotic factors through time. Science (New York, N.Y.), 323, 728–732.

Bhalla, K., Luo, Y., Buchan, T., et al. (2008). Alterations in CDH15 and KIRREL3 in patients with mild to severe intellectual disability. American journal of human genetics, 83, 6, 703–713.

Bhat, S. A., Sarwar, Z., Gillani, S. Q., et al. (2020). Polyomavirus Small T Antigen Induces Apoptosis in Mammalian Cells through the UNC5B Pathway in a PP2A-Dependent Manner. Journal of Virology, 94, 14, e02187-19.

Bierne, N., Bonhomme, F., David, P. (2003). Habitat preference and the marine-speciation paradox. Proceeding of the Royal Society of London B, 270, 1399–1406.

Blanchong, J. A., Robinson, S. J., Samuel, M. D., Foster, J. T. (2016). Application of genetics and genomics to wildlife epidemiology. The Journal of Wildlife Management, 80, 593 – 608.

Boles, K. S., Barchet, W., Diacovo, T., Cella, M., Colonna, M. (2005). The tumor suppressor TSLC1/NECL-2 triggers NK-cell and CD8+ T-cell responses through the cell-surface receptor CRTAM. Blood, 106, 3, 779–786.

Bomber, M. L., Wang, J., Liu, Q., Barnett, K. R., Layden, H. M., Hodges, E., Stengel, K. R., Hiebert, S. W. (2023). Human SMARCA5 is continuously required to maintain nucleosome spacing. Molecular cell, 83, 4, 507–522.e6.

Bossart, G. D., Romano, T. A., Peden-Adams, M. M., Schaefer, A. M., Rice, C. D., Fair, P. A., Reif, J. S. (2019). Comparative innate and adaptive immune responses in atlantic bottlenose dolphins (*Tursiops truncatus*) with viral, bacterial, and fungal infections. Frontiers in Immunology, 10, 1125.

Bourret, V., Macé, M., Crouau-Roy, B. (2007). Genetic variation and population structure of western Mediterranean and northern Atlantic *Stenella coeruleoalba* populations inferred from microsatellite data. Journal of the Marine Biological Association of the UK, 87, 265–269.

Bowden, T. A., Aricescu, A. R., Gilbert, R. J., Grimes, J. M., Jones, E. Y., Stuart, D. I. (2008). Structural basis of Nipah and Hendra virus attachment to their cell-surface receptor ephrin-B2. Nature structural & molecular biology, 15, 6, 567–572.

Bowen, L., Aldridge, B. M., Gulland, F., Woo, J., Van Bonn, W., DeLong, R., Stott, J. L., Johnson, M. L. (2002). Molecular characterization of expressed DQA and DQB genes in the California sea lion (*Zalophus californianus*). Immunogenetics, 54, 332-347.

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., Buckler, E. S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics, 23, 2633-2635.

Braun, B. A., Marcovitz, A., Camp, J. G., Jia, R., Bejerano, G. (2015). Mx1 and Mx2 key antiviral proteins are surprisingly lost in toothed whales. Proceedings of the National Academy of Sciences of the United States of America, 112, 8036–40.

Brockhurst, M. A., Champan, T., King, K. C., Mank, J. E., Paterson, S., Hurst, G. D. D. (2014). Running with the Red Queen: The role of biotic conflicts in evolution. Proceedings of the Royal Society B: Biological Sciences, 281, 20141382.

Buschow, S. I., van Balkom, B. W., Aalberts, M., Heck, A. J., Wauben, M., Stoorvogel, W. (2010). MHC class II-associated proteins in B-cell exosomes and potential functional implications for exosome biogenesis. Immunology and Cell Biology, 88, 8, 851–856.

Calzada, N., Lockyer, C. H., Aguilar, A. (1994). Age and sex composition of the striped dolphin die-off in the western mediterranean. Marine Mammal Science, 10, 299-310.

Cammen, K. M., Wilcox, L. A., Rosel, P. E., Wells, R. S., Read, A. J. (2015). From genome-wide to candidate gene: An investigation of variation at the major histocompatibility complex in common bottlenose dolphins exposed to harmful algal blooms. Immunogenetics, 67, 2, 125– 133.

Carøe, C., Gopalakrishnan, S., Vinner, L. et al. (2018). Single-tube library preparation for degraded DNA. Methods in Ecology and Evolution, 9, 410–419.

Carreras-Carbonell, J., Macpherson, E., Pascual, M. (2006). Population structure within and between subspecies of the Mediterranean triplefin fish *Tripterygion delaisi* revealed by highly polymorphic microsatellite loci. Molecular Ecology, 15, 3527–3539.

Carreras, C., Pascual, M., Cardona, L., Aguilar, A., Margaritoulis, D., Rees, A., et al. (2006). The genetic structure of the loggerhead sea turtle (*Caretta caretta*) in the Mediterranean as revealed by nuclear and mitochondrial DNA and its conservation implications. Conservation Genetics, 8, 761–775.

Carrillo-Bustamante, P., Keşmir, C., de Boer, R. J. (2016). The evolution of natural killer cell receptors. Immunogenetics, 68, 3–18.

Carroll, L. (1871). "2: The Garden of Live Flowers". Through the Looking-Glass (The Millennium Fulcrum Edition 1.7 ed.). Project Gutenberg.

Casado-Amezúa, P., Goffredo, S., Templado, J., Machordom, A. (2012). Genetic assessment of population structure and connectivity in the threatened Mediterranean coral *Astroides calycularis* (Scleractinia, Dendrophylliidae) at different spatial scales. Molecular Ecology, 21, 3671–3685.

Casalone, C., Mazzariol, S., Pautasso, A. et al. (2014). Cetacean strandings in Italy: an unusual mortality event along the Tyrrhenian Sea coast in 2013. Diseases of Aquatic Organisms, 109, 81-86.

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., and Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. Molecular Ecology, 22, 3124–3140.

Centelleghe, C., Da Dalt, L., Marsili, L., et al. (2019). Insights into dolphins' immunology: immuno-phenotypic study on Mediterranean and Atlantic stranded cetaceans. Frontiers in Immunology, 10, 888.

Cerutti, F., Giorda, F., Grattarola, C. et al. (2020). Specific capture and whole-genome phylogeography of Dolphin morbillivirus. Scientific Reports, 10, 20831.

Chapman, J. R., Nakagawa, S., Coltman, D. W., Slate, J., Sheldon, B. C. (2009). A quantitative review of heterozygosity–fitness correlations in animal populations. Molecular Ecology, 18, 2746-2765.

Chen, X., Sheng, X., Liu, X., et al. (2014). Targeted next-generation sequencing reveals novel USH2A mutations associated with diverse disease phenotypes: implications for clinical and molecular diagnosis. PLoS One, Aug 18, 9, 8, e105439.

Cingolani, P., Platts, A., Wang, leL., et al. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. Fly, 6, 2, 80–92.

Coltman, D. W., Bowen, W. D., Wright, J. M. (1998). Birth weight andneonatal survival of harbour seal pups are positively correlated withgenetic variation measured by microsatellites. Proceedings of the Royal Society of London Series B, 265, 803–809.

Coltman, D. W, Pilkington, J. G., Smith, J. A., Pemberton, J. M. (1999). Parasite-Mediated Selection Against Inbred Soay Sheep in a Free-Living, Island Population. Evolution, 53, 4, 1259–67.

Commins, S., Steinke, J. W., Borish, L. (2008). The extended IL-10 superfamily: IL-10, IL-19, IL-20, IL-22, IL-24, IL-26, IL-28, and IL-29. The Journal of allergy and clinical immunology, 121, 5, 1108–1111.

Connolly, J. H., Allen, I. V., Hurwitz, L. J., Millar, J. H. D. (1967). Measles virus antibody and antigen in subacute sclerosing panencephalitis. Lancet, 289, 542−544.

Coolen, S., Van Dijen, M., Van Pelt, J. A., Van Loon, J. J. A., Pieterse, C. M. J., Van Wees, S. C. M. (2023). Genome-wide association study reveals WRKY42 as a novel plant transcription factor that influences oviposition preference of Pieris butterflies. Journal of Experimental Botany, 74, 1690–1704.

Coolen, S., Van Pelt, J. A., Van Wees, S. C. M., et al. (2019). Mining the natural genetic variation in *Arabidopsis thaliana* for adaptation to sequential abiotic and biotic stresses. Planta, 249, 1087–1105.

Corbin, L. J., Liu, A. Y. H., Bishop, S. C., Woolliams, J. A. (2012). Estimation of historical effective population size using linkage disequilibria with marker data. Journal of Animal Breeding and Genetics = Zeitschrift Für Tierzüchtung Und Züchtungsbiologie, 129, 4, 257–70.

Cotté, C., Guinet, C., Taupier-Letage, I., Petiau, E. (2010), Habitat use and abundance of striped dolphins in the western Mediterranean Sea prior to the morbillivirus epizootic resurgence. Endangered Species Research, 12, 203-214.

Crider-Pirkle, S., Billingsley, P., Faust, C., Hardy, D. M., Lee, V., Weitlauf, H. (2002). Cubilin, a binding partner for galectin-3 in the murine utero-placental complex. The Journal of Biological Chemistry, 277, 18, 15904–15912.

Crowder, R. J., Enomoto, H., Yang, M., Johnson, E. M., Jr, Milbrandt, J. (2004). Dok-6, a Novel p62 Dok family member, promotes Ret-mediated neurite outgrowth. The Journal of Biological Chemistry, 279, 40, 42072–42081.

Danecek, P., Auton, A., Abecasis, G. et al. (2011). The variant call format and VCFtools. Bioinformatics 27, 2156–2158.

Danecek, P., Bonfield, J. K., Liddle, J. et al. (2021). Twelve years of SAMtools and BCFtools. GigaScience, 10, 2, giab008.

David, P. (1998). Heterozygosity–fitness correlations: new perspectives on old problems. Heredity, 80, 531-537.

de Bruyn, P. J. N., Tosh, C. A., Terauds, A. (2013). Killer whale eco- types: is there a global model? Biological Reviews, 88, 62–80.

Deecke, V. B., Ford, J. K. B., Slater, P. J. B. (2005) .The vocal behaviour of mammal-eating killer whales: communicating with costly calls. Animal Behaviour, 69, 395–405.

de Jong, M. J., de Jong, J. F., Hoelzel, A. R., Janke, A. (2021). SambaR: An R package for fast, easy and reproducible population-genetic analyses of biallelic SNP data sets. Molecular Ecology Resources, 21, 4, 1369–1379.

da Fontoura Budaszewski, R., von Messling, V. (2016). Morbillivirus experimental animal models: Measles virus pathogenesis insights from canine distemper virus. Viruses, 8, 10, E274.

Delpeut, S., Noyce, R. S., Richardson, C. D. (2014). The tumor-associated marker, PVRL4 (Nectin-4), is the epithelial receptor for morbilliviruses. Viruses, 6, 2268–2286.

Dengjel, J., Schoor, O., Fischer, R. et al. (2005). Autophagy promotes MHC class II presentation of peptides from intracellular source proteins. Proceedings of the National Academy of Sciences USA, 102, 7922–7927.

Desforges, J. P., Sonne, C., Levin, M., Siebert, U., De Guise, S., Dietz, R. (2016). Immunotoxic effects of environmental pollutants in marine mammals. Environment International, 86, 126-139.

Di Guardo, G., Mazzariol, S. (2013). Dolphin Morbillivirus: a lethal but valuable infection model. Emerging Microbes and Infections, 2, e74.

Di Guardo, G., Mazzariol, S. (2016). Cetacean Morbillivirus-Associated Pathology: Knowns and Unknowns. Frontiers in Microbiology, 7, 112.

Di Guardo, G., Mazzariol, S. (2019). Cetacean morbillivirus: a land-to-sea journey and back? Virologica Sinica, 34, 240–242.

Domingo, M., Ferrer, I., Pumarola, M., Marco, A., Plana, J., Kennedy, S., Mcaliskey, M., Rima, B. K. (1990). Morbillivirus in dolphins. Nature, 336, 21.

Domingo, M., Vilafranca, M., Visa, J., Prats, N., Trudgett, A., Visser, I. (1995). Evidence for chronic morbillivirus infection in the Mediterranean striped dolphin (*Stenella coeruleoalba*). Veterinary Microbiology, 44, 229–239.

Domingues, V. S., Santos, R. S., Brito, A., Alexandrou, M., Almada, V. C. (2007). Mitochondrial and nuclear markers reveal isolation by distance and effects of pleistocene glaciations in the northeastern Atlantic and Mediterranean populations of the white seabream (*Diplodus sargus*, l.). Journal of Experimental Marine Biology and Ecology, 346, 102–113.

Dronamraju, K. R. (2004). Infectious Disease and Host-Pathogen Evolution. New York: Cambridge University Press, 370 p.

Duignan, P. J., Geraci, J. R., Raga, J. A., Calzada, N. (1992). Pathology of morbillivirus infection in striped dolphins (*Stenella coeruleoalba*) from Valencia and Murcia, Spain. Canadian Journal of Veterinary Research, 56, 242–248.

Duignan, P. J., House, C., Geraci, J. R., et al. (1995). Morbillivirus infection in two species of pilot whale (*Globicephala* sp.) from the Western Atlantic. Marine Mammal Science, 11, 150–162.

Duignan, P. J., House, C., Ode, D. K., et al. (1996). Morbillivirus in bottlenose dolphins: Evidence for recurrent epizootics in the western Atlantic and Gulf of Mexico. Marine Mammal Science, 12, 499–515.

Duignan, P. J., Van Bressem, M. F., Baker, J. D., et al. (2014). Phocine distemper Virus: Current knowledge and future directions. Viruses, 6, 5093–5134.

Duncan, L. E., Holmans, P. A., Lee, P. H. et al. (2014). Pathway analyses implicate glial cells in schizophrenia. PLoS ONE 9, e89441.

Ebert, D. (2008). Host-parasite coevolution: Insights from the Daphnia-parasite model system. Current Opinion in Microbiology, 11, 290–301.

Edwards, S. V., Potts, W. K. (1996). Polymorphism of genes in the major histocompatibility complex (MHC): implications for conservation genetics of vertebrates. Molecular Genetic Approaches in Conservation, Oxford University Press, 214–237.

El Ayari, T., Trigui El Menif, N., Hamer, B., Cahill, A. E., Bierne, N. (2019). The hidden side of a major marine biogeographic boundary: a wide mosaic hybrid zone at the Atlantic-Mediterranean divide reveals the complex interaction between natural and genetic barriers in mussels. Heredity, 122, 6, 770–784.

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

Fenech, E. J., Lari, F., Charles, P. D. et al. (2020). Interaction mapping of endoplasmic reticulum ubiquitin ligases identifies modulators of innate immune signalling. eLife, 9, e57306.

Feng, M., Zhang, H. (2022). Aminoacyl-tRNA Synthetase: A Non-Negligible Molecule in RNA Viral Infection. Viruses, 14, 3, 613.

Feng, N., Liu, Y., Wang, J. et al. (2016). Canine distemper virus isolated from a monkey efficiently replicates on Vero cells expressing non-human primate SLAM receptors but not human SLAM receptor BMC Veterinary Research, 12, p. 160.

Feng, Y., Tang, K., Lai, Q., et al. (2022). The Landscape of Aminoacyl-tRNA Synthetases Involved in Severe Acute Respiratory Syndrome Coronavirus 2 Infection. Frontiers in Physiology, 12, 818297.

Filatova, O. A., Fedutin, I. D., Borisova, E. A., Meschersky, I. G., Hoyt, E. (2023). Genetic and cultural evidence suggests a refugium for killer whales off Japan during the Last Glacial Maximum. Marine Mammal Science, 1–11.

Flor, H. H. (1942). Inheritance of pathogenicity in a cross between physiologic races 22 and 24 of *Melampsora lini*. Phytopathology, 32, 5.

Flor, H. H. (1955). Host-parasite interaction in flax rust-its genetics and other implications. Phytopathology, 45, 680-85.

Flor, H. H. (1971). Current status of the gene-fob-gene concept. Annual Review of Phytopathology, 3531, 275-296.

Fontaine, M. C., Roland, K., Calves, I. et al. (2014). Postglacial climate changes and rise of three ecotypes of harbour porpoises, *Phocoena phocoena*, in western Palearctic waters. Molecular Ecology, 23, 3306–3321.

Ford, M. J. (2002). Applications of selective neutrality tests to molecular ecology. Molecular Ecology, 11, 1245–1262.

Fossi, M. C., Panti, C., Marsili, L. et al. (2013). The Pelagos Sanctuary for Mediterranean marine mammals: Marine Protected Area (MPA) or marine polluted area? The case study of the striped dolphin (*Stenella coeruleoalba*). Marine Pollution Bulletin, 70, 64-72.

Fowler, S. W. (1987). PCBs and the environment: the Mediterranean marine ecosystem. Waid, J.S. (Ed.), PCBs and the Environment, 3, CRC Press, Boca Raton, 210-239.

Frankham, R. (1995). Effective population size - adult size ratios in wildlife - a review. Genetical research, 66, 95–107.

Frankham, R. (2005). Genetics and extinction. Biological Conservation, 126, 131-40.

Galarza, J. A., Carreras-Carbonell, J., Macpherson, E., Pascual, M., Roques, S., Turner, G. F., et al. (2009). The influence of oceanographic fronts and early-life-history traits on connectivity among littoral fish species. Proceeding of the National Academy of Sciences USA, 106, 1473–1478.

García-Martínez, J., Moya, A., Raga, J. A., Latorre, A. (1999). Genetic differentiation in the striped dolphin *Stenella coeruleoalba* from European waters according to mitochondrial DNA (mtDNA) restriction analysis. Molecular Ecology, 8, 1069-1073.

Garg, R. K. (2008). Subacute sclerosing panencephalitis. Journal of Neurology, 255, 1861−1871.

Gaspari, S., Azzelino, A., Airoldi, S., Hoelzel, A. R. (2007). Social kin associations and genetic structuring of striped dolphin populations (*Stenella coeruleoalba*) in the Mediterranean Sea. Molecular Ecology, 16, 2922-2933.

Gaspari, S., Holcer, D., Mackelworth, P. et al. (2013). Population genetic structure of common bottlenose dolphins (*Tursiops truncatus*) in the

Adriatic Sea and contiguous regions: Implications for international conservation. Aquatic Conservation: Marine Freshwater Ecosystem, 25, 212-222.

Gaspari, S., Marsili, L., Natali, C., Airoldi, S., Lanfredi, C., Deeming, C., Moura A. E. (2019). Spatio-temporal patterns of genetic diversity in the Mediterranean striped dolphin (*Stenella coeruleoalba*). Journal of Zoological Systematics and Evolutionary Research, 00, 1–14.

Gaspari, S., Scheinin, A., Holcer, D., et al. (2015). Drivers of population structure of the bottlenose dolphin (*Tursiops truncatus*) in the eastern Mediterranean Sea. Evolutionary Biology, 42, 177-90.

Gaudieri, S., Dawkins, R. L., Habara, K., Kulski, J. K., Gojobori, T. (2000). SNP profile within the human major histocompatibility complex reveals an extreme and interrupted level of nucleotide diversity. Genome Research, 10, 1579–86.

Gayden, T., Sepulveda, F. E., Khuong-Quang, D. A. et al. (2018). Germline HAVCR2 mutations altering TIM-3 characterize subcutaneous panniculitis-like T cell lymphomas with hemophagocytic lymphohistiocytic syndrome. Nature Genetics, 50, 1650–1657.

Gebska, M. A., Stevenson, B. K., Hemnes, A. R., et al. (2011). Phosphodiesterase-5A (PDE5A) is localized to the endothelial caveolae and modulates NOS3 activity. Cardiovascular research, 90, 2, 353–363.

Gelain, M. E., Bonsembiante, F. (2019). Acute phase proteins in marine mammals: State of art, perspectives and challenges. Frontiers in Immunology, 10, 1220.

Geoghegan, J. L., Duchene, S., Holmes, E. C. (2017). Comparative analysis estimates the relative frequencies of co-divergence and cross-species transmission within viral families. PLoS Pathogens, 13(2), e1006215.

Giorda, F., Crociara, P., Iulini, B., et al. (2022). Neuropathological Characterization of Dolphin Morbillivirus Infection in Cetaceans Stranded in Italy. Animals: an open access journal from MDPI, 12, 4, 452.

Gkafas, G. A., de Jong, M., Exadactylos, A., Raga. J. A., Aznar, F. J., Hoelzel, A. R. (2020). Sex-specific impact of inbreeding on pathogen load in the striped dolphin. Proceedings of the Royal Society B: Biological Sciences, 2872020019520200195.

Gkafas, G. A., Exadactylos, A., Rogan, E., Raga, J. A., Reid, R., Hoelzel, A.R. (2017). Biogeography and temporal progression during the evolution of striped dolphin population structure in European waters. Journal of Biogeography, 44, 2681–2691.

Gonçalves-Carneiro, D., McKeating, J. A., Bailey, D. (2017). The Measles Virus Receptor SLAMF1 Can Mediate Particle Endocytosis. Journal of Virology, 91, e02255-16.

Gorman, J. V., Colgan, J. D. (2014). Regulation of T cell responses by the receptor molecule Tim-3. Immunologic research, 59, 1-3, 56–65.

Groch, K. R., Díaz-Delgado, J., Santos-Neto, E. B., et al. (2020). The Pathology of Cetacean Morbillivirus Infection and Comorbidities in Guiana Dolphins During an Unusual Mortality Event (Brazil, 2017–2018). Veterinary Pathology, 57, 6, 845-857.

Grönholm, M., Teesalu, T., Tyynelä, J., et al. (2005). Characterization of the NF2 protein merlin and the ERM protein ezrin in human, rat, and mouse central nervous system. Molecular and Cellular Neurosciences, 28, 4, 683–693.

Guigó, R., Ullrich, S. (2020). Dynamic changes in intron retention are tightly associated with regulation of splicing factors and proliferative activity during B-cell development. Nucleic Acids Research, 48, 1327–1340.

Guimier, A., Gordon, C. T., Godard, F. et al. (2016). Biallelic PPA2 Mutations Cause Sudden Unexpected Cardiac Arrest in Infancy. American Journal of Human Genetics, 99, 3, 666–673.

Guo, B., Cheng, G. (2007). Modulation of the interferon antiviral response by the TBK1/IKKi adaptor protein TANK. The Journal of biological chemistry, 282, 16, 11817–1182.

Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., Bustamante, C. D. (2009). Inferring the joint demographic history of multiple populations from multidimensional SNP data. PLoS Genetics, 5, e1000695.

Han, J. M., Park, S. G., Lee, Y., Kim, S. (2006). Structural separation of different extracellular activities in aminoacyl-tRNA synthetase-interacting multi-functional protein, p43/AIMP1. Biochemical and Biophysical Research Communications, 342, 1, 113–118.

Hashiguchi, T., Ose, T., Kubota, M., Maita, N., Kamishikiryo, J., Maenaka, K., Yanagi, Y. (2011). Structure of the measles virus hemagglutinin bound to its cellular receptor SLAM. Nature Structural & Molecular Biology, 18, 135–141.

Hauffe, A. Rizzoli, C. Vernesi, eds), pp. 294– 318. Cambridge University Press, Cambridge.

Hayashi, H., Kubo, Y., Izumida, M., et al. (2018). Enterokinase Enhances Influenza A Virus Infection by Activating Trypsinogen in Human Cell Lines. Frontiers in Cellular and Infection Microbiology, 8, 91.

Hersh, S. L., Odell, D. K., Asper, E. D. (1990). Bottlenose dolphin mortality patterns in the Indian/Banana river system in Florida. In The Bottlenose Dolphin; Leatherwood, S.P., Reeves, R.R., Eds.; Academic Press: San Diego, CA, USA, 155–164.

Hewitt, G. M. (2004). Genetic consequences of climatic oscillations in the Quaternary. Philosophical Transactions of the Royal Society B London, 359, 183–195.

Hill, A. V. S., Allosopp, C. E., Kwaitowski, D. et al. (1991). Common West African HLA antigens are associated with protection from severe malaria. Nature, 352, 595–600.

Hoelzel, A. R., Hey, J., Dahlheim, M. E., Nicholson, C., Burkanov, V., Black, N. (2007). Evolution of population structure in a highly social top predator, the Killer Whale. Molecular Biology and Evolution, 24, 1407–1415.

Hoelzel, A. R. (2009). Evolution of population structure in marine mammals. In: Population Genetics for Animal Conservation (G. Bertorelle, M.W. Bruford, H.C.

Hoffman, J. I., Simpson, F., David, P. et al. (2014). High-throughput sequencing reveals inbreeding depression in a natural population. Proceedings of the National Academy of Sciences, 111, 10, 3775–80.

Hör, S., Pirzer, H., Dumoutier, L., et al. (2004). The T-cell lymphokine interleukin-26 targets epithelial cells through the interleukin-20 receptor 1 and interleukin-10 receptor 2 chains. The Journal of Biological Chemistry, 279, 32, 33343–33351.

Huelsmann, M., Hecker, N., Springer, M. S., Gatesy, J., Sharma, V., Hiller, M. (2019). Genes lost during the transition from land to water in cetaceans highlight genomic changes associated with aquatic adaptations. Science Advances, 5, 9, eaaw6671.

Hughes, A. L., Yeager, M. (1998). Natural selection at major histocompatibility complex loci of vertebrates. Annual Review of Genetics, 32, 415-435.

Hughes, A. L., Nei, M. (1988). Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature, 335, 167–170.

Hughes, A. L., Nei, M. (1992). Maintenance of MHC polymorphism. Nature, 355, 402–403.

Janeway, C. A., Travers, P., Walport, M., Shlomchik, M. (2001). Immunobiology 5: the immune system in health and disease: Garland Publishing, 883 p.

Jo, W. K., Osterhaus, A. D., Ludlow, M. (2018a). Transmission of morbilliviruses within and among marine mammal species. Current Opinion in Virology, 28, 133–141.

Jo, W. K., Kruppa, J., Habierski, A., van de Bildt, M., Mazzariol, S., Di Guardo, G., Siebert, U., Kuiken, T., et al. (2018b). Evolutionary evidence for multi-host transmission of cetacean morbillivirus. Emerging Microbes and Infections, 7, 201.

Johns, D. G., Ao, Z., Naselsky, D., Herold, C. L., Maniscalco, K., Sarov-Blat, L., Steplewski, K., Aiyar, N., Douglas, S. A. (2004). Urotensin-II-mediated cardiomyocyte hypertrophy: effect of receptor antagonism and role of inflammatory mediators. Naunyn-Schmiedeberg's Archives of Pharmacology, 370, 4, 238–250.

Jones, M. R., Good, J. M. (2016). Targeted capture in evolutionary and ecological genomics. Molecular Ecology, 25, 185-202.

Kalfaoglu, B., Almeida-Santos, J., Tye, C.A., Satou, Y., Ono, M. (2020). T-Cell Hyperactivation and Paralysis in Severe COVID-19 Infection Revealed by Single-Cell Analysis. Frontiers in Immunology, 11, 589380.

Kambe, T., Narita, H., Yamaguchi-Iwai, Y., et al. (2002). Cloning and characterization of a novel mammalian zinc transporter, zinc transporter 5, abundantly expressed in pancreatic beta cells. The Journal of biological chemistry, 277, 21, 19049–19055.

Kannan, K., Tanabe, S., Borrell, A., Aguilar, A., Focardi, S., Tatsukawa, R. (1993). Isomerspecific analysis and toxic evaluation of polychlorinated biphenyls in striped dolphins affected by epizootic in the western Mediterranean Sea. Archives of Environmental Contamination and Toxicology, 25, 227-233.

Kardos, M., Luikart, G., Allendorf, F. (2015). Measuring individual inbreeding in the age of genomics: marker-based measures are better than pedigrees. Heredity, 115, 63–72.

Kaur, B., Bhatia, D., Mavi, G. S. (2021). Eighty years of gene-for-gene relationship and its applications in identification and utilization of R genes. Journal of Genetics, 100, 50.

Keck, N., Kwiatek, O., Dhermain, F. et al. (2010). Resurgence of Morbillivirus infection in Mediterranean dolphins off the French coast. Veterinary Record, 166, 654-655.

Kennedy, S. (1998). Morbillivirus infections in aquatic mammals. Journal of Comparative Pathology, 119, 201-225.

Kennedy, S., Smyth, J. A., Cush, P. F., McAliskey, M., McCulloch, S. J., Rima, B. K. (1991). Histopathologic and immunocytochemical studies of distemper in harbor porpoises. Veterinary Pathology, 28, 1, 718– 7.

Kern, C.H., Yang, M., Liu, W-S. (2021). The PRAME family of cancer testis antigens is essential for germline development and gametogenesis. Biology of Reproduction, 105, 2, 290–304.

Kim, K. S., Maio, N., Singh, A., Rouault, T. A. (2018). Cytosolic HSC20 integrates de novo iron-sulfur cluster biogenesis with the CIAO1-mediated transfer to recipients. Human Molecular Genetics, 27, 5, 837–852.

Kimura, M. (1991). The neutral theory of molecular evolution: a review of recent evidence. Idengaku zasshi, 66, 4, 367–386.

Kitai, K., Kawaguchi, K., Tomohiro, T., Morita, M., So, T., Imanaka, T. (2021). The lysosomal protein ABCD4 can transport vitamin B12 across liposomal membranes in vitro. The Journal of Biological Chemistry, 296, 100654.

Klein, J., Sato. A. (1998). Birth of the major histocompatibility complex. Scandinavian Journal of Immunology, 47, 199-209.

Kourafalou, V. H., Savvidis, Y. G., Koutitas, C. G., Krestenitis, Y. N. (2004). Modeling studies on the processes that influence matter transfer on the Gulf of Thermaikos (NW Aegean Sea). Continental Shelf Research, 24, 203–222.

Kozyraki, R., Fyfe, J., Kristiansen, M., et al. (1999). The intrinsic factor-vitamin B12 receptor, cubilin, is a high-affinity apolipoprotein A-I receptor facilitating endocytosis of high-density lipoprotein. Nature Medicine, 5, 6, 656–661.

Kumari, R., Jiu, Y., Carman, P. J., et al. (2020). Tropomodulins Control the Balance between Protrusive and Contractile Structures by Stabilizing Actin-Tropomyosin Filaments. Current Biology: CB, 30, 5, 767–778.e5.

Lambeck, K. Purcell, A. (2005). Sea-level change in the Mediterranean Sea since the LGM: Model predictions for tectonically stable areas. Quaternary Science Reviews, 24, 1969–1988.

Lawal, R. A., Al-Atiyat, R. M., Aljumaah, R. S., Silva, P., Mwacharo, J. M., Hanotte, O. (2018). Whole-genome resequencing of red junglefowl and indigenous village chicken reveal new insights on the genome dynamics of the species. Frontiers in Genetics, 9, 264.

Lee, J. H., Schütte, D., Wulf, G., Füzesi, L., Radzun, H. J., Schweyer, S., Engel, W., Nayernia, K. (2006). Stem-cell protein Piwil2 is widely expressed in tumors and inhibits apoptosis through activation of Stat3/Bcl-XL pathway. Human molecular genetics, 15, 2, 201–211.

Lejeusne, C., Chevaldonné, P., Pergent-Martini, C., Boudouresque, C. F., Perez, T. (2010) Climate change efects on a miniature ocean: the highly diverse, highly impacted Mediterranean Sea. Trends in Ecology and Evolution, 25, 250–260.

Lévy, J., Haye, D., Marziliano, N., et al. (2018). EFNB2 haploinsufficiency causes a syndromic neurodevelopmental disorder. Clinical Genetics, 93, 1141– 1147.

Liang, D., Tian, L., You, R., et al. (2018). AIMp1 Potentiates TH1 Polarization and Is Critical for Effective Antitumor and Antiviral Immunity. Frontiers in Immunology, 8, 1801.

Li, H., Durbin, R. (2009). Fast and accurate short read alignment with burrows-wheeler transform. Bioinformatics, 25, 1754–1760.

Li, H., Handsaker, B., Wysoker, A., et al. (2009). 1000 Genome Project Data Processing Subgroup. The sequence alignment/map (SAM) format and SAMtools. Bioinformatics, 25, 2078–2079.

Li, Y., Misumi, I., Shiota, T., et al. (2022). The ZCCHC14/TENT4 complex is required for hepatitis A virus RNA synthesis. Proceedings of the National Academy of Sciences of the United States of America, 119, 28, e2204511119.

Limborg, M. T., Hanel, R., Debes, P. V., Ring, A. K., Andre, C., Tsigenopoulos, C. S., Bekkevold, D. (2012). Imprints from genetic drift and mutation imply relative divergence times across marine transition zones in a pan-European small pelagic fish (*Sprattus sprattus*). Heredity, 109, 96–107.

Lim, J., Kim, D., Lee, Y. S. et al. (2018). Mixed tailing by TENT4A and TENT4B shields mRNA from rapid deadenylation. Science, 361, 6403, 701–704.

Lin, C. S., Lin, G., Xin, Z. C., Lue, T. F. (2006). Expression, distribution and regulation of phosphodiesterase 5. Current Pharmaceutical Design, 12, 27, 3439–3457.

Lischer, H. E. L., Excoffier, L. (2012). PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. Bioinformatics (Oxford, England), 28, 2, 298–299.

Liu, C., Liu, C., Fu, R. (2022). Research progress on the role of PKM2 in the immune response. Frontiers in Immunology, 13, 936967.

Liu, F., Wu, X., Li, L., Zou, Y., Liu, S., Wang, Z. (2016). Evolutionary characteristics of morbilliviruses during serial passages in vitro: Gradual attenuation of virus virulence. Comparative Immunology, Microbiology and Infectious Diseases, 47, 7–18.

Liu, K., Wang, A., Ran, L. et al. (2019). ARHGEF38 as a novel biomarker to predict aggressive prostate cancer. Genes & diseases, 7, 2, 217–224.

Liu, X., Fu, Y. X. (2015). Exploring population size changes using SNP frequency spectra. Nature Genetics, 47, 555-559.

Liu, X., Fu, Y. X. (2020). Stairway Plot 2: demographic history inference with folded SNP frequency spectra. Genome Biology, 21, 1, 1-9.

Liu, Z., Kozarewa, I., Rosa-Rosa, J. M. et al. (2012). A modified method for whole exome resequencing from minimal amounts of starting DNA. PLoS One, 7, e32617.

Lively, C. M. (2010). A review of Red Queen models for the persistence of obligate sexual reproduction. Journal of Heredity, 101, S13-S20.

Louis, M., Viricel, A., Lucas, T. et al. (2014). Habitat-driven population structure of bottlenose dolphins, *Tursiops truncatus*, in the North-East Atlantic. Molecular Ecology, 23, 857–874.

Lowry, R. (1998) VassarStats: http://www.vassarstats.net Assessed on 2023.

Luijckx, P., Fienberg, H., Duneau, D., Ebert, D. (2013). A Matching-Allele Model Explains Host Resistance to Parasites. Current Biology, 23, 1085–1088.

Marie, J. C., Kehren, J., Trescol-Biemont, M. C., et al. (2001). Mechanism of measles virus-induced suppression of inflammatory immune responses. Immunity, 14, 69– 79.

Martin, M. P., Carrington, M. (2005). Immunogenetics of viral infections. Current Opinion in Immunology, 17, 510–516.

Manlik, O., Krützen, M., Kopps, A. M., et al (2019). Is MHC diversity a better marker for conservation than neutral genetic diversity? A case study

of two contrasting dolphin populations. Ecology and Evolution, 9, 6986–6998.

Mazzariol, S., Peletto, S., Mondin, A., Centelleghe, C., Di Guardo, G., Di Francesco, C. E., Casalone, C., Acutis, P. L. (2013). Dolphin morbillivirus infection in a captive harbor seal (*Phoca vitulina*). Journal of Clinical Microbiology, 51, 708–711.

McKnight, D. T., Schwarzkopf, L., Alford, R. A., Bower, D. S., Zenger, K. R. (2017). Effects of emerging 677 infectious diseases on host population genetics: a review. Conservation Genetics, 1-11.

Melia, M. M., Earle, J. P., Abdullah, H., Reaney, K., Tangy, F., Cosby, S. L. (2014). Use of SLAM and PVRL4 and identification of pro-HB-EGF as cell entry receptors for wild type phocine distemper virus. PLoS One, 9, e106281.

Mendez, M., Subramaniam, A., Collins, T., et al. (2011). Molecular ecology meets remote sensing: Environmental drivers to population structure of humpback dolphins in the Western Indian Ocean. Heredity, 107, 349–361.

Meyer, W. K., Jamison, J., Richter, R. et al. (2018). Ancient convergent losses of Paraoxonase 1 yield potential risks for modern marine mammals. Science, 361, 591–594.

Mira, F., Rubio-Guerri, C., Purpari, G., et al. (2019). Circulation of a novel strain of dolphin morbillivirus (DMV) in stranded cetaceans in the Mediterranean Sea. Scientific Reports, 9, 9792.

Mirimin, L., Westgate, A., Rogan, E., Rosel, P., Read, A. J., Coughlan, J., Cross, T. (2009). Population structure of short-beaked common dolphins (*Delphinus delphis*) in the North Atlantic Ocean as revealed by mitochondrial and nuclear genetic markers. Marine Biology, 156, 821–834.

Mistry, B. V. Chang, T. C., Yasue, H., Kim, D., Oatley, J., Liu, W. S. (2011). Roles of PRAME in Spermatogenesis., Biology of Reproduction, Volume 85, Issue Suppl_1, 574.

Miyake, Y., Ishii, K., Honda, A. (2017). Influenza virus infection induces host pyruvate kinase m which interacts with viral RNA-dependent RNA polymerase. Frontiers in Microbiology, 8, 162.

Monney, L., Sabatos, C. A., Gaglia, J. L., et al. (2002). Th1-specific cell surface protein Tim-3 regulates macrophage activation and severity of an autoimmune disease. Nature, 415, 6871, 536–541.

Montgelard, C., FranCois Catzejfis, F. M., Douzeryt, E. (1997). Phylogenetic relationships of artiodactyls and cetaceans as deduced from

the comparison of cytochrome b and 12S rRNA mitochondrial sequences. Molecular Biology and Evolution, 14, 550-559.

Morin, P. A., Archer, F. I., Foote, A. D. et al. (2010). Complete mitochondrial genome phylogeographic analysis of killer whales (*Orcinus orca*) indicates multiple species. Genome Research, 20, 908–916.

Morris, S. E., Zelner, J. L., Fauquier, D. A., Rowles, T. K., Rosel, P. E., Gulland, F., Grenfell, B. T. (2015). Partially observed epidemics in wildlife hosts: Modelling an outbreak of Dolphin morbillivirus in the northwestern Atlantic. Journal of The Royal Society Interface, 12, 1.

Moura, A. E., Natoli, A., Rogan, E., Hoelzel, A. R. (2013a). Atypical panmixia in a European dolphin species (*Delphinus delphis*): Implications for the evolution of diversity across oceanic boundaries. Journal of Evolution Biology, 26, 63–75.

Moura, A. E., Natoli, A., Rogan, E., Hoelzel, A. R. (2013). Evolution of functional genes in cetaceans driven by natural selection on a phylogenetic and population level. Evolutionary Biology, 40, 341–354.

Moura, A. E., Kenny, J. G., Chaudhuri, R., et al. (2014). Population genomics of the killer whale indicates ecotype evolution in sympatry involving both selection and drift. Molecular Ecology, 23, 21, 5179–5192.

Moura, A. E., van Rensburg, C. J., Pilot, M. et al. (2014a). Killer whale nuclear genome and mtDNA reveal widespread population bottleneck during the Last Glacial Maximum. Molecular Biology and Evolution, 31, 1121–1131.

Nakayama, K., Matsudaira, C., Tajima, Y. et al. (2009). Temporal and spatial trends of organotin contamination in the livers of finless porpoises (*Neophocaena phocaenoides*) and their association with parasitic infection status. Science of the Total Environment, 407, 61736178.

Nagano, T., Yoneda, T., Hatanaka, Y., Kubota, C., Murakami, F., Sato, M. (2002). Filamin A-interacting protein (FILIP) regulates cortical cell migration out of the ventricular zone. Nature cell biology, 4, 7, 495–501.

Natoli, A., Birkun, A., Aguilar, A., Lopez, A., Hoelzel, A. R. (2005). Habitat structure and the dispersal of male and female bottlenose dolphins (*Tursiops truncatus*). Proceeding of the Royal Society of London B, 272, 1217–1226.

Natoli, A., Canadas, A., Peddemors, V. M., Aguilar, A., Vaquero, C., Fernandez-Piqueras, P., Hoelzel, A. R. (2006). Phylogeography and alpha taxonomy of the common dolphin (*Delphinus* sp.). Journal of Evolutionary Biology, 19, 943–954.

Nomiyama, K., Kanbara, C., Ochiai, M. et al. (2014). Halogenated phenolic contaminants in the blood of marine mammals from Japanese coastal waters. Marine Environmental Research, 93,15-22.

Norris, R. D. (2000). Pelagic species diversity, biogeography, and evolution. Paleobiology, 26, 236–258.

Norstrom, R. J., Muir, D. C .G., Ford, C. A., Simon, M., Macdonald, C. R., Béland, P. (1992). Indications of P450 Monooxygenase activities in beluga (*Delphinapterus leucas*) and narwhal (*Monodon monoceros*) from patterns of PCBs, PCDD and PCDF accumulation. Marine Environmental Research, 34, 267-272.

Noyce, R. S., Bondre, D. G., Ha, M. N., Lin, L. T., Sisson, G., Tsao, M. S., Richardson, C. D. (2011). Tumor cell marker PVRL4 (Nectin 4) is an epithelial cell receptor for measles virus. PLoS Pathogens, 7, e1002240.

O'Brien, S. J., Yuhki, N. (1999). Comparative genome organization of the major histocompatibility complex: lessons from the Felidae. Immunological Reviews, 167, 133–144.

Ohishi, K., Maruyama, T., Seki, F., Takeda, M. (2019). Marine morbilliviruses: Diversity and interaction with signaling lymphocyte activation molecules. Viruses, 11, 606.

Ou, X. M., Chen, K., Shih, J. C. (2006). Monoamine oxidase A and repressor R1 are involved in apoptotic signaling pathway. Proceedings of the National Academy of Sciences of the United States of America, 103, 29, 10923–10928.

Ozaki, K., Isono, M., Kawahara, T., Iida, S., Kudo, T. Kukuyama, F. (2006). A Mechanistic Approach to evaluation of umbrella species as conservation surrogates. Conservation Biology, 20, 1507-1515.

Paiva, I. A., Badolato-Corrêa, J., Familiar-Macedo, D., de-Oliveira-Pinto, L. M. (2021). Th17 Cells in Viral Infections—Friend or Foe? Cells, 10, 5, 1159.

Parker, J., Tsagkogeorga, G., Cotton, J. et al. (2013). Genome-wide signatures of convergent evolution in echolocating mammals. Nature, 502, 228–231.

Patarnello, T., Volckaert, F. A., Castilho, R. (2007). Pillars of Hercules: is the Atlantic–Mediterranean transition a phylogeographical break? Molecular Ecology, 16, 21, 4426-4444.

Pautasso, A., Iulini, B., Grattarola, C., et al. (2019). Novel dolphin morbillivirus (DMV) outbreak among Mediterranean striped dolphins

Stenella coeruleoalba in Italian waters. Diseases of Aquatic Organisms, 132, 215-220.

Penczykowski, R. M., Laine, A. L., Koskella, B. (2016). Understanding the ecology and evolution of host-parasite interactions across scales. Evolutionary Applications, 9, 37-52.

Person, C., Samborski, D. J., Rohringer, R. (1962). The gene-for-gene concept. Nature, 194, 561-62.

Perrino, P. A., Talbot, L., Kirkland, R. et al. (2020). Multi-level evidence of an allelic hierarchy of USH2A variants in hearing, auditory processing and speech/language outcomes. Communications Biology, 3, 180.

Perry, A. L., Low, P. J., Ellis, J. R., Reynolds, J. D. (2005). Climate change and distribution shifts in marine fishes. Science, 308, 1912–1915.

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., Hoekstra, H. E. (2012). Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS ONE, 7(5), e37135.

Piertney, S. B., Oliver, M. K. (2006). The evolutionary ecology of the major histocompatibility complex. Heredity, 96, 7–21.

Piffko, A., Uhl, C., Vajkoczy, P., Czabanka, M., Broggini, T. (2022). EphrinB2-EphB4 Signaling in Neurooncological Disease. International Journal of Molecular Sciences, 23, 3, 1679.

Pintore, M. D., Mignone, W., Di Guardo, G., et al. (2018). Neuropathologic findings in cetaceans stranded in Italy (2002–14). Journal of Wildlife Diseases, 54, 295–303.

Pitceathly, R. D., Rahman, S., Wedatilake, Y., et al (2013). NDUFA4 mutations underlie dysfunction of a cytochrome c oxidase subunit linked to human neurological disease. Cell Reports, 3, 6, 1795–1805.

Planas, R., Felber, M., Vavassori, S., Pachlopnik Schmid, J. (2023). The hyperinflammatory spectrum: from defects in cytotoxicity to cytokine control. Frontiers in Immunology, 14, 1163316.

Polprasert, C., Takeuchi, Y., Kakiuchi, N., et al. (2019). Frequent germline mutations of HAVCR2 in sporadic subcutaneous panniculitis-like T-cell lymphoma. Blood Advances, 3(4), 588–595.

Pomaznoy, M., Ha, B., Peters, B. (2018). GOnet: a tool for interactive Gene Ontology analysis. BMC Bioinformatics, 19, 470.

Purcell, S., Neale, B., Todd-Brown, K., et al. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. American Journal of Human Genetics, 81, 3, 559–575.

Profeta, F., Di Francesco, C.E., Marsilio, F. et al. (2015). Retrospective seroepidemiological investigations against Morbillivirus, *Toxoplasma gondii* and *Brucella* spp. in cetaceans stranded along the Italian coastline (1998–2014). Research in Veterinary Science, 101, 89-92.

Qiao, J., Han, B., Liu, B.L., et al. (2009). A splice site mutation combined with a novel missense mutation of LHCGR cause male pseudohermaphroditism. Human Mutation, 30, 9, E855-65.

Quinlan, A. R., Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics (Oxford, England), 26, 6, 841–842.

Raga, J. A., Abril, F., Almor, P. (1987). *Skrjabinalius guevaraii* Galego and Selva 1979 (Nematoda: Pseudaliidae), a lungworm parasitizing dolphins (Cetacea: Delphinidae) in the western Mediterranean Sea. Rivista di Parassitologia, 4, 27–32.

Raga, J. A, Banyard, A., Domingo, M. et al. (2008). Dolphin Morbillivirus epizootic resurgence, Mediterranean Sea. Emerging Infectious Diseases, 14, 471-473.

Raga, J. A., Carbonell, E. (1985). New data about parasites on *Stenella coeruleoalba* (Meyen, 1833) (Cetacea: Delphinidae) in the western Mediterranean Sea. Investigations on Cetacea, 14, 337–338.

Reiners, J., Nagel-Wolfrum, K., Jürgens, K., Märker, T., Wolfrum, U. (2006). Molecular basis of human Usher syndrome: deciphering the meshes of the Usher protein network provides insights into the pathomechanisms of the Usher disease. Experimental Eye Research, 83, 1, 97–119.

Reiners, J., van Wijk, E., Märker, T., et al. (2005). Scaffold protein harmonin (USH1C) provides molecular links between Usher syndrome type 1 and type 2. Human molecular genetics, 14, 24, 3933–3943.

Rijks, J. M., Hoffman, J. I., Kuiken, T., Osterhaus, A. D. M. E., Amos, W. (2008). Heterozygosity and lungworm burden in harbour seals (*Phoca vitulina*). Heredity, 100, 587 – 593.

Rohland, N., Reich, D. (2012). Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. Genome Research, 22, 939–946.

Rolland, J. L., Bonhomme, F., Lagardère, F., Hassan, M., Guinand, B. (2007). Population structure of the common sole (*Solea solea*) in the

Northeastern Atlantic and the Mediterranean Sea: Revisiting the divide with epic markers. Marine Biology, 151, 327–341.

Roy, A., Lin, Y. N., Agno, J. E., DeMayo, F. J., Matzuk, M. M. (2007). Absence of tektin 4 causes asthenozoospermia and subfertility in male mice. FASEB journal: official publication of the Federation of American Societies for Experimental Biology, 21, 4, 1013–1025.

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

Rubio-Guerri, C., Jiménez, M.Á., Melero, M. et al. (2018). Genetic heterogeneity of dolphin morbilliviruses detected in the Spanish Mediterranean in inter-epizootic period. BMC Veterinary Research, 14, 248.

Rubio-Guerri, C., Melero, M., Esperón, F., et al. (2013). Unusual striped dolphin mass mortality episode related to cetacean morbillivirus in the Spanish Mediterranean Sea. BMC Veterinary Research, 9, 106.

Sacristán, C., Carballo, M., Muñoz, M. J., Bellière, E. N., Neves, E., Nogal, V., Esperón, F. (2015). Diagnosis of Cetacean morbillivirus: A sensitive one step real time RT fast-PCR method based on SYBR® Green. Journal of Virological Methods, 226, 25– 30

Sánchez-Fueyo, A., Tian, J., Picarella, D., et al. (2003). Tim-3 inhibits T helper type 1-mediated auto- and alloimmune responses and promotes immunological tolerance. Nature Immunology, 4, 11, 1093–1101.

Sariol, A., Mackin, S., Allred, M. G., et al. (2020). Microglia depletion exacerbates demyelination and impairs remyelination in a neurotropic coronavirus infection. Proceedings of the National Academy of Sciences of the United States of America, 117, 39, 24464–24474.

Sato, H., Yoneda, M., Honda, T., Kai, C. (2012). Morbillivirus receptors and tropism: Multiple pathways for infection. Frontiers in Microbiology, 3, 75.

Seoighe, C., Korir, P. K. (2011). Evidence for intron length conservation in a set of mammalian genes associated with embryonic development. BMC Bioinformatics, 12, 9, S16.

Schnabel, F., Schuler, E., Al-Maawali, A., et al. (2023). Homozygous loss-of-function variants in FILIP1 cause autosomal recessive arthrogryposis multiplex congenita with microcephaly. Human genetics, 142, 4, 543–552.

Sestak, K., Jin, X., He, M. et al. (2012). An effort to use human-based exome capture methods to analyze chimpanzee and macaque exomes. PLoS One, 7, e40637.

Shalak, V., Kaminska, M., Mitnacht-Kraus, R., Vandenabeele, P., Clauss, M., Mirande, M. (2001). The EMAPII cytokine is released from the mammalian multisynthetase complex after cleavage of its p43/proEMAPII component. The Journal of Biological Chemistry, 276, 26, 23769–23776.

Shimizu, Y., Ohishi, K., Suzuki, R. et al. (2013). Amino acid sequence variation of signaling lymphocyte activation molecule and mortality caused by morbillivirus infection in cetaceans. Microbiology and Immunology, 57, 624-632.

Sierra, E., Fernández, A., Felipe-Jiménez, I. et al. (2020). Histopathological differential diagnosis of meningoencephalitis in cetaceans: Morbillivirus, Herpesvirus, *Toxoplasma gondii*, *Brucella* sp., and *Nasitrema* sp. Frontiers in Veterinary Science, 7, 650.

Slate, J., Kruuk, L. E. B., Marshall, T. C., Pemberton, J. M., Clutton-Brock, T. H. (2000). Inbreeding depression influences lifetime breeding success in a wild population of red deer (*Cervus elaphus*). Proceedings of the Royal Society of London Series B — Biological Science, 267, 1657 – 1662.

Smirnova, E. V., Rakitina, T. V., Evtodienko, A. I.u, Kostanian, I. A., Lipkin, V. M. (2000). Klonirovanie i kharakteristika gena ribosomnogo belka S21 cheloveka [(letter)] [Cloning and characterization of the human ribosomal protein S21 gene]. Bioorganicheskaia khimiia, 26, 5, 392–396.

Smith, K. F., Acevedo-Whitehouse, K., Pedersen, A. B. (2009). The role of infectious diseases in biological conservation. Animal Conservation, 12, 1–12.

Soto, S., Alba, A., Ganges, L., et al. (2011). Post-epizootic chronic dolphin morbillivirus infection in Mediterranean striped dolphins *Stenella coeruleoalba*. Diseases of Aquatic Organisms, 96, 187-194.

Spurgin, L. G., Richardson, D. S. (2010). How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. Proceedings of the Royal Society B, 277, 979–88.

Stejskalova, K., Bayerova, Z., Futas, J. et al. (2017). Candidate gene molecular markers as tools for analyzing genetic susceptibility to Morbillivirus infection in stranded Cetaceans. HLA, 1– 11.

Stoffel, M. A., Esser, M., Kardos, M., Humble, E., Nichols, H., David, P., Hoffman, J. I. (2016). inbreedR: an R package for the analysis of inbreeding based on genetic markers. Methods in Ecology & Evolution, 7, 1331-1339.

Sun, G. K., Tang, L. J., Zhou, J. D., et al. (2019). DOK6 promoter methylation serves as a potential biomarker affecting prognosis in de novo acute myeloid leukemia. Cancer Medicine, 8, 14, 6393–6402.

Szulkin, M., Bierne, N., David, P. (2010). Heterozygosity-fitness correlations: a time for reappraisal. Evolution, 64, 5, 1202–17.

Takai, T. (2002). Roles of Fc receptors in autoimmunity. Nature Reviews Immunology, 2, 580– 592.

Takai, T. (2005). Fc receptors and their role in immune regulation and autoimmunity. Journal of Clinical Immunology, 25.

Takeuchi, A., Badr, M. elS., Miyauchi, K., et al. (2016). CRTAM determines the CD4+ cytotoxic T lymphocyte lineage. The Journal of Experimental Medicine, 213, 1, 123–138.

Tanabe, S., Watanabe, S., Kan, H., Tatsukawa, R. (1988). Capacity and mode of PCBs metabolism in small cetaceans. Marine Mammal Science, 4, 103-124.

Taubenberger, J. K., Tsai, M. M., Atkin, T. J., et al., (2000). Molecular genetic evidence of a novel morbillivirus in a long-finned pilot whale (*Globicephalus melas*). Emerging Infectious Diseases, 6, 42–45.

Taylor, B. L., Chivers, S. J., Larese, J., Perrin, W. F. (2007). Generation length and percent mature estimates for IUCN assessments of cetaceans. NOAA, Southwest Fisheries Science Center Administrative Report LJ-07-01. La Jolla, California.

Taylor, M. R., Martin, E. A., Sinnen, B., Trilokekar, R., Ranza, E., Antonarakis, S. E., Williams, M. E. (2020). Kirrel3-mediated synapse formation is attenuated by disease-associated missense variants. The Journal of Neuroscience: the official journal of the Society for Neuroscience, 40, 28, 5376–5388.

Temple, H. J., Terry, A. (Compilers) (2007). The Status and Distribution of European Mammals. Luxembourg: Office for Official Publications of the European Communities. viii + 48pp.

Teixeira, J. C., Huber, C. D. (2021). The inflated significance of neutral genetic diversity in conservation genetics. Proceedings of the National Academy of Sciences of the United States of America, 118, 10, e2015096118.

Theocharis, A., Georgopoulos, D., Lascaratos, A., Nittis, K. (1993). Water masses and circulation in the central region of the Eastern Mediterranean (E. Ionian, S. Aegean and NW Levantine). Deep-Sea Research II, 40, 1121–1142.

The UniProt Consortium. UniProt: the Universal Protein Knowledgebase in 2023. Nucleic Acids Research, 51, D523–D531.

Tschirren, B., Andersson, M., Scherman, K., Westerdahl, H., Mittl, P. R., Råberg, L. (2013). Polymorphisms at the innate immune receptor TLR2 are associated with Borrelia infection in a wild rodent population. Proceedings Biological Sciences, 280, 20130364.

Tsitrone, A., Rousset, F., David, P. (2001). Heterosis, marker mutational processes and population inbreeding history. Genetics, 159, 1845-1859.

Turner, A. K., Begon, M., Jackson, J. A., Bradley, J. E., Paterson, S. (2011). Genetic Diversity in Cytokines Associated with Immune Variation and Resistance to Multiple Pathogens in a Natural Rodent Population. PLoS Genetics, 7, e1002343.

Turner, S. D. (2018). qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. Journal of Open Source Software, 3, 25, 731.

Ulrich, R., Puff, C., Wewetzer, K., Kalkuhl, A., Deschl, U., Baumgärtner, W. (2014). Transcriptional changes in canine distemper virus-induced demyelinating leukoencephalitis favor a biphasic mode of demyelination. PLoS ONE, 9.

Valsecchi, E., Amos, W., Raga, J.A., Podesta, M., Sherwin, W. (2004). The effects of inbreeding on mortality during a morbillivirus outbreak in the Mediterranean striped dolphin (*Stenella coeruleoalba*). Animal Conservation, 7, 139-146.

Van Bressem, M. F., Waerebeek, K. Van, et al. (2001). An insight into the epidemiology of dolphin morbillivirus worldwide. Veterinary Microbiology, 81, 287–304.

Van Bressem, M. F., Duignan, P., Banyard, A. et al. (2014). Cetacean morbillivirus: current knowledge and future directions. Viruses, 6, 514581.

Van Bressem, M., Raga, A., Di Guardo, G. et al. (2009). Emerging infectious diseases in cetaceans worldwide and the possible role of environmental stressors. Diseases of Aquatic Organisms, 86, 143–157.

Van de Bildt, M. W., Martina, B. E., Vedder, E. J., et al. (2000). Identification of morbilliviruses of probable cetacean origin in carcases of Mediterranean monk seals (*Monachus monachus*). Veterinary Record, 146, 691–694.

Van Der Biezen, E. A., Jones, J. D. (1998). Plant disease-resistance proteins and the gene-for-gene concept. Trends in Biochemical Sciences, 23, 454–456.

Vandevelde, M., Zurbriggen, A. (2005). Demyelination in canine distemper virus infection: A review. Acta Neuropathologica, 109, 56–68.

Vargas-Castro, I., Peletto, S., Mattioda, V., et al. (2023). Epidemiological and genetic analysis of Cetacean Morbillivirus circulating on the Italian coast between 2018 and 2021. Frontiers in Veterinary Science, 10, 1216838.

Violi, B., de Jong, M. J., Frantzis, A., et al. (2023). Genomics reveals the role of admixture in the evolution of structure among sperm whale populations within the Mediterranean Sea. Molecular Ecology, 32, 2715–2731.

Vucetich, J. A., Waite, T. A., Nunney, L. (1997). Fluctuating population size and the ratio of effective to census population size. Evolution, 51, 2017-2021.

Waseem, N. H., Low, S., Shah, A. Z., et al. (2020). Mutations in SPATA13/ASEF2 cause primary angle closure glaucoma. PLoS genetics, 16(4), e1008721.

Watanabe, A., Yoneda, M., Ikeda, F., Terao-Muto, Y., Sato, H., Kai, C. (2010). CD147/EMMPRIN acts as a functional entry receptor for measles virus on epithelial cells. Journal of Virology, 84, 4183– 4193.

Webster, l. M., Mello, l. V., Mougeot, F., Martinez-Padilla, J., Paterson, S., Piertney, S. B. (2011). Identification of genes responding to nematode infection in red grouse. Molecular Ecology Resources, 11, 305-313.

Wegehaupt, O., Groß, M., Wehr, C., et al. (2020). TIM-3 deficiency presenting with two clonally unrelated episodes of mesenteric and subcutaneous panniculitis-like T-cell lymphoma and hemophagocytic lymphohistiocytosis. Pediatric Blood & Cancer, 67, 6, e28302.

Weisbrich, A., Honnappa, S., Jaussi, R., Okhrimenko, O., Frey, D., Jelesarov, I., Akhmanova, A., Steinmetz, M. O. (2007). Structure-function relationship of CAP-Gly domains. Nature Structural & Molecular Biology, 14, 10, 959–967.

Wessels, M. E.; Deaville, R.; Perkins, M. W. et al. (2021). Novel presentation of DMV-associated encephalitis in a long-finned pilot whale (*Globicephala melas*). Journal of Comparative Pathology, 183, 51–56.

Wolf, Y., Anderson, A. C., Kuchroo, V. K. (2020). TIM3 comes of age as an inhibitory receptor. Nature Reviews Immunology, 20, 173–185.

Woolhouse, M. E. J., Webster, J. P, Domingo, E., Charlesworth, B., Levin, B. R. (2002). Biological and biomedical implications of the co-evolution of pathogens and their hosts. Nature Genetics, 32, 569-577.

Wu, H., Han, F. (2023). Investigation of shared genes and regulatory mechanisms associated with coronavirus disease 2019 and ischemic stroke. Frontiers in Neurology, 14, 1151946.

Xue, D. X., Wang, H. Y., Zhang, T., Liu, J. X. (2014). Population genetic structure and demographic history of *Atrina pectinata* based on mitochondrial DNA and microsatellite markers. PLoS ONE, 9, e95436.

Yan, C., Lian, X., Li, Y. et al. (2006). Macrophage-specific expression of human lysosomal acid lipase corrects inflammation and pathogenic phenotypes in lal-/- mice. The American journal of pathology, 169, 3, 916–926.

Yuhki, N., O'Brien, S. J. (1990). DNA variation of the mammalian major histocompatibility complex reflects genomic diversity and population history. Proceedings of the National Academy of Sciences of the United States of America, 87, 836-840.

Zhang, Y., Wang, Z., Hu, G., et al. (2022). A novel germline HAVCR2 (TIM-3) compound heterozygous mutation is related to hemophagocytic lymphohistiocytic syndrome in EBV-positive peripheral T-cell lymphoma (NOS) with down-regulated TIM-3 signaling. Frontiers in Oncology, 12:870676.

Zheng, H., Yang, Y., Hong, Y. G., et al. (2019). Tropomodulin 3 modulates EGFR-PI3K-AKT signaling to drive hepatocellular carcinoma metastasis. Molecular carcinogenesis, 58, 10, 1897–1907.

Zhu, H., Zhang, L., Wu, Y., et al. (2018). T-ALL leukemia stem cell 'stemness' is epigenetically controlled by the master regulator SPI1. eLife, 7, e38314.

Zinzula, L., Mazzariol, S., Di Guardo, G. (2022). Molecular signatures in cetacean morbillivirus and host species proteome: Unveiling the evolutionary dynamics of an enigmatic pathogen? Microbiology and Immunology, 66, 52– 58.

Zueva, K. J., Lumme, J., Veselov, A. E., Kent, M. P., Primmer, C. R. (2018). Genomic signatures of parasite-driven natural selection in north European Atlantic salmon (*Salmo salar*). Marine Genomics, 39, 26–38.

Appendices

Appendix A: Supplementary material for chapter 2

Table S1 – Plate design of ddRADseq. Each of the following tables represents a 96 well plate for 84 samples. "Adapters" represents the adapter ligation mix, "Adapters/PCR Primers/POOLS" represents the combination of adapters, PCR primers and pool planning applied per well, and "Barcode/Index" represents the short DNA sequence used to identify the individuals, while also balancing the distribution between red (A and C) and green (T and G) laser on the Illumina platform.

Adapters

	01	02	03	04	05	06	07	08	09	10	11	12
A	01P1.P2	02P1.P2	03P1.P2	04P1.P2	05P1.P2	06P1.P2	07P1.P2	08P1.P2	09P1.P2	10P1.P2	11P1.P2	12P1.P2
B	01P1.P2	02P1.P2	03P1.P2	04P1.P2	05P1.P2	06P1.P2	07P1.P2	08P1.P2	09P1.P2	10P1.P2	11P1.P2	12P1.P2
C	01P1.P2	02P1.P2	03P1.P2	04P1.P2	05P1.P2	06P1.P2	07P1.P2	08P1.P2	09P1.P2	10P1.P2	11P1.P2	12P1.P2
D	01P1.P2	02P1.P2	03P1.P2	04P1.P2	05P1.P2	06P1.P2	07P1.P2	08P1.P2	09P1.P2	10P1.P2	11P1.P2	12P1.P2
E	01P1.P2	02P1.P2	03P1.P2	04P1.P2	05P1.P2	06P1.P2	07P1.P2	08P1.P2	09P1.P2	10P1.P2	11P1.P2	12P1.P2
F	01P1.P2	02P1.P2	03P1.P2	04P1.P2	05P1.P2	06P1.P2	07P1.P2	08P1.P2	09P1.P2	10P1.P2	11P1.P2	12P1.P2
G	01P1.P2	02P1.P2	03P1.P2	04P1.P2	05P1.P2	06P1.P2	07P1.P2	08P1.P2	09P1.P2	10P1.P2	11P1.P2	12P1.P2

Adapters/PCR Primers/POOLS

	01	02	03	04	05	06	07	08	09	10	11	12
A	01P1.02 PCR2	02P1.02 PCR2	03P1.02 PCR2	04P1.02 PCR2	05P1.02 PCR2	06P1.02 PCR2	07P1.02P CR2	08P1.02P CR2	09P1.02P CR2	10P1.02P CR2	11P1.02P CR2	12P1.02P CR2
B	01P1.04 PCR2	02P1.04 PCR2	03P1.04 PCR2	04P1.04 PCR2	05P1.04 PCR2	06P1.04 PCR2	07P1.04P CR2	08P1.04P CR2	09P1.04P CR2	10P1.04P CR2	11P1.04P CR2	12P1.04P CR2
C	01P1.05 PCR2	02P1.05 PCR2	03P1.05 PCR2	04P1.05 PCR2	05P1.05 PCR2	06P1.05 PCR2	07P1.05P CR2	08P1.05P CR2	09P1.05P CR2	10P1.05P CR2	11P1.05P CR2	12P1.05P CR2
D	01P1.06 PCR2	02P1.06 PCR2	03P1.06 PCR2	04P1.06 PCR2	05P1.06 PCR2	06P1.06 PCR2	07P1.06P CR2	08P1.06P CR2	09P1.06P CR2	10P1.06P CR2	11P1.06P CR2	12P1.06P CR2
E	01P1.07 PCR2	02P1.07 PCR2	03P1.07 PCR2	04P1.07 PCR2	05P1.07 PCR2	06P1.07 PCR2	07P1.07P CR2	08P1.07P CR2	09P1.07P CR2	10P1.07P CR2	11P1.07P CR2	12P1.07P CR2
F	01P1.12 PCR2	02P1.12 PCR2	03P1.12 PCR2	04P1.12 PCR2	05P1.12 PCR2	06P1.12 PCR2	07P1.12P CR2	08P1.12P CR2	09P1.12P CR2	10P1.12P CR2	11P1.12P CR2	12P1.12P CR2
G	01P1.16 PCR2	02P1.16 PCR2	03P1.16 PCR2	04P1.16 PCR2	05P1.16 PCR2	06P1.16 PCR2	07P1.16P CR2	08P1.16P CR2	09P1.16P CR2	10P1.16P CR2	11P1.16P CR2	12P1.16P CR2

	Barcode/Index											
	01	02	03	04	05	06	07	08	09	10	11	12
A	GCATG/ACATCG	AATCG/ACATCG	ACAGA/ACATCG	AAGTG/ACATCG	ATTAC/ACATCG	CAGGC/ACATCG	AGAATG/ACATCG	AGTTAA/ACATCG	CCACTG/ACATCG	AGTCAA/ACATCG	AGTGTT/ACATCG	CACGAC/ACATCG
B	GCATG/TGGTCA	AATCG/TGGTCA	ACAGA/TGGTCA	AAGTG/TGGTCA	ATTAC/TGGTCA	CAGGC/TGGTCA	AGAATG/TGGTCA	AGTTAA/TGGTCA	CCACTG/TGGTCA	AGTCAA/TGGTCA	AGTGTT/TGGTCA	CACGAC/TGGTCA
C	GCATG/CACTGT	AATCG/CACTGT	ACAGA/CACTGT	AAGTG/CACTGT	ATTAC/CACTGT	CAGGC/CACTGT	AGAATG/CACTGT	AGTTAA/CACTGT	CCACTG/CACTGT	AGTCAA/CACTGT	AGTGTT/CACTGT	CACGAC/CACTGT
D	GCATG/GATCTG	AATCG/GATCTG	ACAGA/GATCTG	AAGTG/GATCTG	ATTAC/GATCTG	CAGGC/GATCTG	AGAATG/GATCTG	AGTTAA/GATCTG	CCACTG/GATCTG	AGTCAA/GATCTG	AGTGTT/GATCTG	CACGAC/GATCTG
E	GCATG/TACAAG	AATCG/TACAAG	ACAGA/TACAAG	AAGTG/TACAAG	ATTAC/TACAAG	CAGGC/TACAAG	AGAATG/TACAAG	AGTTAA/TACAAG	CCACTG/TACAAG	AGTCAA/TACAAG	AGTGTT/TACAAG	CACGAC/TACAAG
F	GCATG/GGACGG	AATCG/GGACGG	ACAGA/GGACGG	AAGTG/GGACGG	ATTAC/GGACGG	CAGGC/GGACGG	AGAATG/GGACGG	AGTTAA/GGACGG	CCACTG/GGACGG	AGTCAA/GGACGG	AGTGTT/GGACGG	CACGAC/GGACGG
G	GCATG/TTTCAC	AATCG/TTTCAC	ACAGA/TTTCAC	AAGTG/TTTCAC	ATTAC/TTTCAC	CAGGC/TTTCAC	AGAATG/TTTCAC	AGTTAA/TTTCAC	CCACTG/TTTCAC	AGTCAA/TTTCAC	AGTGTT/TTTCAC	CACGAC/TTTCAC

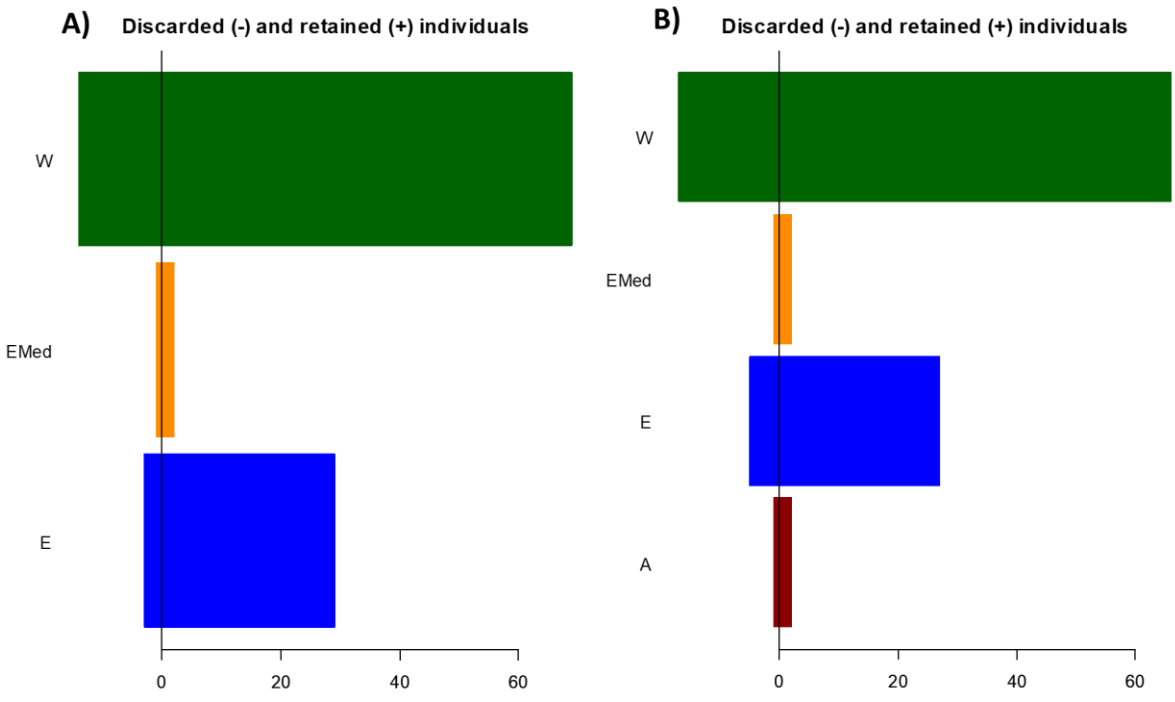


Figure S1 - Proportion of discarded/retained individuals for the dataset without (A) and with (B) Atlantic individuals, together with West (W) and East (E) of the western basin, and East basin (EMed) of the Mediterranean Sea. Bars before 0 represent the individuals that did not pass the quality filters.

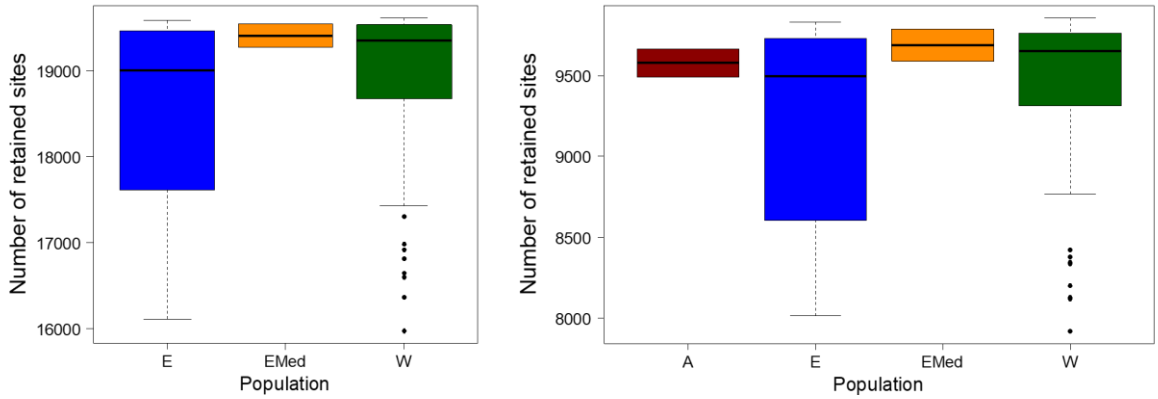


Figure S2 - Number of retained SNPs per population for the dataset without (A) and with (B) Atlantic individuals, together with West (W) and East (E) of the western basin, and East basin (EMed) of the Mediterranean Sea. Boxes represent the average of sites presented in the individuals.

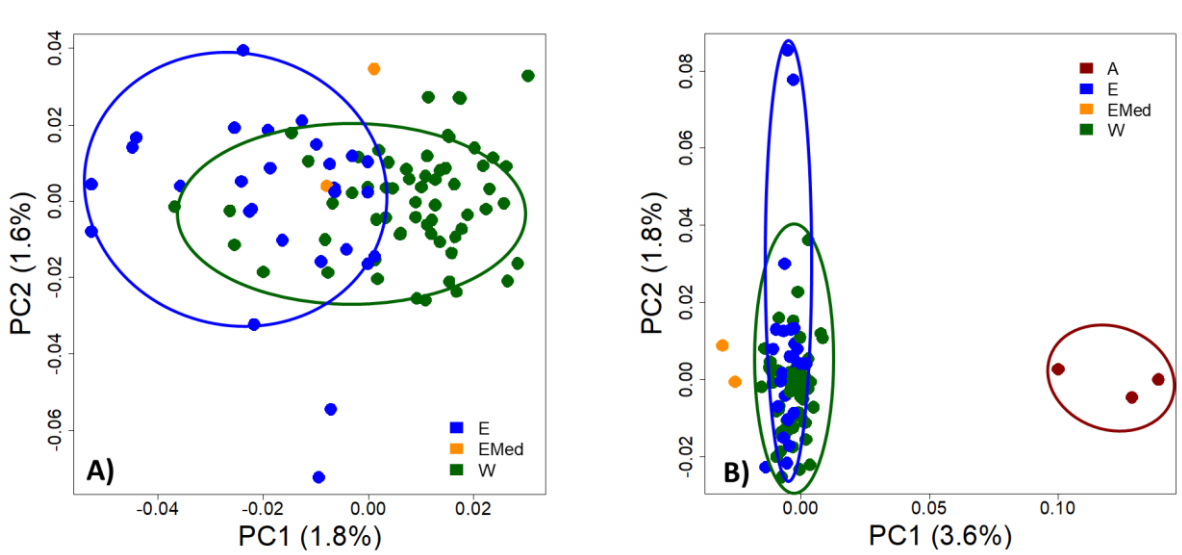


Figure S3 - Principal Coordinate Analysis (PCoA) plot of standardized data based on Pi (pairwise sequence dissimilarity), of the A) West (W) and East (E) of the western basin, and East basin (EMed) of the Mediterranean Sea, and B) with the Atlantic (A) individuals.

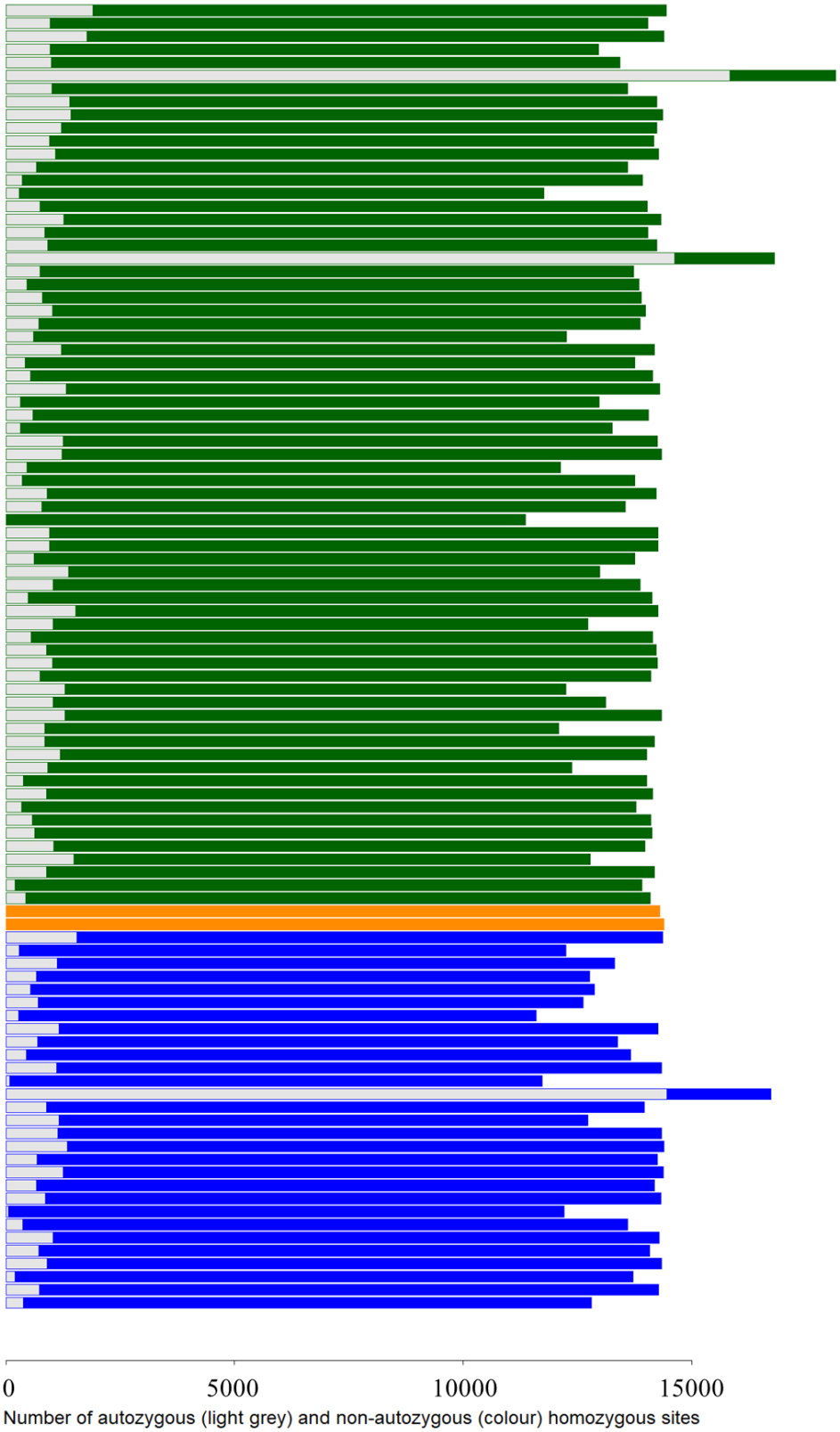


Figure S3 - Proportion of inbreeding per individual from the West (W) and East (E) of the western basin, and East basin (EMed) of the Mediterranean Sea.

Table S2 – Estimation results of genome-wide heterozygosity analysis, by bcftools, for Mediterranean individuals. n_AA - number of dominant homozygous allele; n_aa - number of homozygous recessive alleles; n_Aa - number of heterozygous alleles; n_missing - number of missing data points; total – total number of alleles; heterozygosity – estimates of heterozygosity; he_percent – percentage of heterozygosity; West (W) and East (E) of the western basin, and East basin (EMed) of the Mediterranean Sea.

sample	region	n_AA	n_aa	n_Aa	n_missing	total	heterozygosity	he_percent
79	EMed	1146764	480	2111	340483	1489838	0.001836682	0.1836682
98	EMed	994735	389	1665	493049	1489838	0.001670364	0.1670364
103	EMed	1092617	495	2070	394656	1489838	0.001890097	0.1890097
84	E	1067443	348	2103	419944	1489838	0.001965615	0.1965615
111	E	1018370	357	1982	469129	1489838	0.001941788	0.1941788
87	E	977433	255	1700	510450	1489838	0.001735778	0.1735778
108	E	1096470	453	2171	390744	1489838	0.001975263	0.1975263
106	E	1081587	387	2137	405727	1489838	0.0019712	0.19712
72	E	985852	261	1702	502023	1489838	0.001722995	0.1722995
92	E	1151939	491	2256	335152	1489838	0.001953778	0.1953778
114	E	1030702	407	1991	456738	1489838	0.001927209	0.1927209
97	E	1018870	1581	1089	468298	1489838	0.001066038	0.1066038
99	E	1035425	359	2032	452022	1489838	0.001957958	0.1957958
71	E	1095410	412	2170	391846	1489838	0.001976335	0.1976335
94	E	1140660	472	2335	346371	1489838	0.002042035	0.2042035
100	E	1135170	506	2321	351841	1489838	0.002039548	0.2039548
115	E	1013736	385	2053	473664	1489838	0.002020323	0.2020323
101	E	1099115	424	2198	388101	1489838	0.001995031	0.1995031
74	E	1102926	405	2232	384275	1489838	0.002018881	0.2018881
95	E	1029542	378	2045	457873	1489838	0.001981656	0.1981656
89	E	1098256	496	2219	388867	1489838	0.002015494	0.2015494
96	E	1099233	446	2305	387854	1489838	0.002091682	0.2091682
116	E	1025721	396	2054	461667	1489838	0.001997722	0.1997722
76	E	1117641	436	2281	369480	1489838	0.002035956	0.2035956
117	E	994038	387	1934	493479	1489838	0.001941067	0.1941067
75	E	1156381	456	2337	330664	1489838	0.002016091	0.2016091

90	E	1110883	445	2297	376213	1489838	0.002062633	0.2062633
73	E	1125275	455	2248	361860	1489838	0.001992947	0.1992947
83	E	1107240	429	2183	379986	1489838	0.001966929	0.1966929
88	E	1095073	450	2289	392026	1489838	0.002085056	0.2085056
109	E	948914	345	1787	538792	1489838	0.001878984	0.1878984
110	E	990459	363	2007	497009	1489838	0.002021496	0.2021496
118	E	1118830	467	2195	368346	1489838	0.001957214	0.1957214
78	E	1141546	493	2216	345583	1489838	0.001936631	0.1936631
85	E	1121638	477	2311	365412	1489838	0.002055271	0.2055271
82	W	1121243	493	2268	365834	1489838	0.002017786	0.2017786
104	W	994211	267	1812	493548	1489838	0.001818748	0.1818748
80	W	1008261	325	1847	479405	1489838	0.001827929	0.1827929
112	W	1074786	411	2302	412339	1489838	0.002136429	0.2136429
91	W	1072223	433	2125	415057	1489838	0.001977147	0.1977147
93	W	1112555	1867	1387	374029	1489838	0.001243044	0.1243044
107	W	1108827	450	2263	378298	1489838	0.002035914	0.2035914
77	W	1137175	474	2292	349897	1489838	0.00201063	0.201063
41	W	1043456	251	1706	444425	1489838	0.001631891	0.1631891
1	W	1215048	463	2344	271983	1489838	0.001924695	0.1924695
2	W	1172262	449	2330	314797	1489838	0.00198291	0.198291
3	W	1060049	344	1853	427592	1489838	0.001744417	0.1744417
4	W	1185452	432	2352	301602	1489838	0.001979405	0.1979405
54	W	1135839	367	2084	351548	1489838	0.001830816	0.1830816
34	W	1227449	476	2411	259502	1489838	0.001959627	0.1959627
35	W	1225413	503	2395	261527	1489838	0.001949832	0.1949832
36	W	1126008	353	2159	361318	1489838	0.001913125	0.1913125
37	W	1001695	216	1493	486434	1489838	0.001487935	0.1487935
5	W	1060835	345	1814	426844	1489838	0.001706501	0.1706501
6	W	1113361	371	1997	374109	1489838	0.001789861	0.1789861
9	W	569568	22	402	919846	1489838	0.000705273	0.0705273
10	W	1193254	444	2374	293766	1489838	0.00198483	0.198483
11	W	1048501	298	1703	439336	1489838	0.00162113	0.162113
12	W	1171535	417	2356	315530	1489838	0.002006288	0.2006288
13	W	1184538	377	2414	302509	1489838	0.002033135	0.2033135
38	W	1167722	412	2297	319407	1489838	0.001962525	0.1962525

39	W	1202282	476	2299	284781	1489838	0.001907794	0.1907794
40	W	1193345	432	2257	293804	1489838	0.00188707	0.188707
42	W	1124178	408	2037	363215	1489838	0.001808058	0.1808058
14	W	1127991	372	2279	359196	1489838	0.002015669	0.2015669
15	W	1032217	388	1890	455343	1489838	0.001826978	0.1826978
17	W	1052743	337	1905	434853	1489838	0.001805713	0.1805713
18	W	1135428	446	2253	351711	1489838	0.001979568	0.1979568
19	W	924281	160	1325	564072	1489838	0.001431247	0.1431247
20	W	1003746	366	1942	483784	1489838	0.001930314	0.1930314
21	W	1167627	487	2229	319495	1489838	0.00190457	0.190457
43	W	1170683	515	2311	316329	1489838	0.001969307	0.1969307
44	W	1149891	426	2268	337253	1489838	0.001967751	0.1967751
45	W	1120332	394	2162	366950	1489838	0.001925392	0.1925392
46	W	1026869	312	1842	460815	1489838	0.001790047	0.1790047
22	W	1102912	400	2100	384426	1489838	0.001899744	0.1899744
23	W	1021497	361	1853	466127	1489838	0.001810081	0.1810081
24	W	1126253	468	2182	360935	1489838	0.00193285	0.193285
25	W	1157151	467	2264	329956	1489838	0.001951923	0.1951923
26	W	1140370	444	2232	346792	1489838	0.001952677	0.1952677
28	W	1034762	362	1949	452765	1489838	0.001879328	0.1879328
29	W	1122950	461	2156	364271	1489838	0.001915479	0.1915479
30	W	1177804	485	2409	309140	1489838	0.002040319	0.2040319
53	W	1150310	458	2252	336818	1489838	0.001953132	0.1953132
47	W	1055487	309	1933	432109	1489838	0.0018275	0.18275
48	W	1150090	462	2351	336935	1489838	0.0020392	0.20392
31	W	1099999	442	2038	387359	1489838	0.001848561	0.1848561
32	W	1006883	404	1863	480688	1489838	0.001846108	0.1846108
33	W	1079896	421	2091	407430	1489838	0.001931804	0.1931804
52	W	1092799	377	2180	394482	1489838	0.001990221	0.1990221
56	W	1093092	460	2151	394135	1489838	0.001963123	0.1963123
57	W	923408	319	1612	564499	1489838	0.001742064	0.1742064
58	W	1091806	392	2140	395500	1489838	0.00195552	0.195552
59	W	1141515	450	2300	345573	1489838	0.002010024	0.2010024
49	W	1115639	378	2200	371621	1489838	0.001967418	0.1967418
50	W	1139731	479	2183	347445	1489838	0.001910901	0.1910901

51	W	1167805	486	2355	319192	1489838	0.00201171	0.201171
55	W	1090597	387	2118	396736	1489838	0.001937605	0.1937605
60	W	1097209	439	2106	390084	1489838	0.001914974	0.1914974
61	W	1016225	1618	1124	470871	1489838	0.001103078	0.1103078
62	W	979180	259	1607	508792	1489838	0.001638048	0.1638048
63	W	1185893	492	2301	301152	1489838	0.001935751	0.1935751
64	W	1141889	395	2202	345352	1489838	0.001924008	0.1924008
65	W	1159471	445	2212	327710	1489838	0.001903405	0.1903405
67	W	1018682	309	1952	468895	1489838	0.001911958	0.1911958
68	W	1153958	396	2326	333158	1489838	0.002010928	0.2010928
66	W	1150987	393	2325	336133	1489838	0.002015247	0.2015247
7	W	1131924	438	2131	355345	1489838	0.001878372	0.1878372
8	W	1168473	428	2301	318636	1489838	0.001964648	0.1964648
16	W	1118798	450	2132	368458	1489838	0.001901229	0.1901229
27	W	1194115	425	2371	292927	1489838	0.001980933	0.1980933
69	W	1145182	375	2222	342059	1489838	0.001935913	0.1935913
70	W	1176135	481	2267	310955	1489838	0.001923007	0.1923007
113	W	1122491	494	2216	364637	1489838	0.001969426	0.1969426
105	W	1056848	378	2147	430465	1489838	0.00202667	0.202667
81	W	1106030	452	2325	381031	1489838	0.002096848	0.2096848
102	W	1063752	388	2221	423477	1489838	0.002082784	0.2082784
86	W	1144838	473	2249	342278	1489838	0.00195981	0.195981

Appendix B: Supplementary material for chapter 3

Table S3 - Minor allele frequency (MAF) and heterozygosity frequency of each outlier from the Manhattan plot analysis, through time, covering the peaks of epizootic and non-epizootic periods.

Loci	MAF	Heterozygosity	Years
USH2A	16.7% (1/6)	1/3	1991
	33.3% (2/6)	2/3	1994-1995
	25% (1/4)	1/2	1998
	0%	0	2002-2003
	38.9% (7/18)	7/9	2007
TENT4A	0%	0	1991
	50% (3/6)	3/3	1994-1995
	0%	0	1998
	36.4% (8/22)	8/11	2002-2003
	5.5% (1/18)	1/9	2007
SLC30A5	0%	0	1991
	16.7% (1/6)	1/3	1994-1995
	0%	0	1998
	22.7% (5/22)	5/11	2002-2003
	0%	0	2007
PPA2	0%	0	1991
	33.3% (2/6)	2/3	1994-1995
	0%	0	1998
	31.8% (7/22)	7/11	2002-2003
	5.5% (1/18)	1/9	2007
SMARCA5	0%	0	1991
	33.3% (2/6)	2/3	1994-1995
	0%	0	1998
	35% (7/20)	7/10	2002-2003

	0%	0	2007
	0%	0	1991
	33.3% (2/6)	2/3	1994-1995
TANK	0%	0	1998
	35% (7/20)	6/11	2002-2003
	12.5% (2/16)	2/8	2007
	25% (1/4)	1/2	1991
	50% (2/4)	2/2	1994-1995
KIRREL3	0%	0	1998
	44.4% (8/18)	8/9	2002-2003
	12.5% (2/16)	2/8	2007
	0%	0	1991
	33.3% (2/6)	2/3	1994-1995
CDCA7L	0%	0	1998
	29.2% (7/24)	7/12	2002-2003
	5.5% (1/18)	1/9	2007
	16.7% (1/6)	1/3	1991
	33.3% (2/6)	2/3	1994-1995
NDUFA4	0%	0	1998
	40.9% (9/22)	9/11	2002-2003
	11.1% (2/18)	2/9	2007
	0%	0	1991
	33.3% (2/6)	2/3	1994-1995
FILIP1	0%	0	1998
	31.8% (7/22)	7/11	2002-2003
	5.5% (1/18)	1/9	2007
	16.7% (1/6)	1/3	1991
	50% (3/6)	3/3	1994-1995
NF2	0%	0	1998
	31.8% (7/22)	7/11	2002-2003
	5.5% (1/18)	1/9	2007

	50% (3/6)	3/3	1991
	0%	0	1994-1995
DOK6	50% (2/4)	2/2	1998
	0%	0	2002-2003
	22.2% (4/18)	4/9	2007
	0%	0	1991
	50% (2/4)	2/2	1994-1995
LHCGR	0%	0	1998
	22.7% (5/22)	5/11	2002-2003
	5.5% (1/18)	1/9	2007
	33.3% (2/6)	2/3	1991
	0%	0	1994-1995
C16H10orf90	25% (1/4)	1/2	1998
	5% (1/20)	1/10	2002-2003
	31.2% (1/16)	5/8	2007
	50% (3/6)	3/3	1991
	0%	0	1994-1995
LIPA	25% (1/4)	1/2	1998
	0%	0	2002-2003
	22.2% (4/18)	4/9	2007
	0%	0	1991
	33.3% (2/6)	2/3	1994-1995
UNC5B	0%	0	1998
	22.7% (5/22)	5/11	2002-2003
	0%	0	2007
	0%	0	1991
	33.3% (2/6)	2/3	1994-1995
SPATA13	0%	0	1998
	41.6% (10/24)	10/12	2002-2003
	5.5% (1/18)	1/9	2007

	0%	0	1991
	33.3% (2/6)	2/3	1994-1995
EFNB2	25% (1/4)	1/2	1998
	31.8% (7/22)	7/11	2002-2003
	0%	0	2007
	0%	0	1991
	33.3% (2/6)	2/3	1994-1995
FRMPD3	0%	0	1998
	20.8% (5/24)	5/12	2002-2003
	0%	0	2007

Appendix C: Supplementary material for chapter 4

Gene list - Final list of genes related with the immune system selected through Gene Ontology search.

ABCC2	AGBL4	ANGPT4	ARRDC1	BCL11A
ABI1	AGBL5	ANGPTL1	ARRDC3	BCL2L11
ABL1	AGL	ANGPTL3	ARSB	BCR
ACE2	AGR2	ANGPTL4	ASAH1	BECN1
ACKR4	AGTR2	ANG	ASGR1	BET1L
ACOD1	AHCYL1	ANK3	ASH2L	BIN2
ACPP	AHCY	ANKLE1	ASPM	BIRC2
ACSL4	AHI1	ANKRA2	ASS1	BIRC6
ACTR1B	AHSG	ANKRD17	ATG5	BLK
ACVR1B	AIF1	ANKRD1	ATG9A	BLM
ACVR1	AIMP1	ANPEP	ATOX1	BMP2
ACVR2A	AIPL1	APBB1IP	ATP6AP2	BMPER
ACVR2B	AIP	APBB1	ATP7A	BMPR1A
ACVRL1	AK7	APEH	ATRNL1	BMPR1B
ADA2	AKAP8	APLF	ATXN1	BMPR2
ADAM17	AKIRIN2	APLP2	ATXN7L3	BMX
ADAMDEC1	AKT1	APP	ATXN7	BNIP3L
ADAMTS12	AKT2	AREG	AXIN2	BPIFA1
ADAMTS3	ALCAM	ARG1	AXL	BPIFB1
ADAMTS9	ALDOB	ARHGAP35	BAG4	BPI
ADARB1	ALDOC	ARHGAP45	BAG6	BRI3
ADAR	ALKBH1	ARHGAP9	BAIAP2	BRS3
ADA	ALMS1	ARHGEF26	BAK1	BRWD1
ADCY6	ALOX15B	ARHGEF2	BANK1	BST1
ADD1	ALOX5	ARHGEF7	BAP1	BTBD11
ADIPOQ	ALYREF	ARMC8	BBS2	BTK
ADK	AMFR	ARNTL2	BBS4	BTNL2
ADORA1	AMOT	ARNTL	BBS7	BYSL
ADSS	AMPD3	ARNT	BCAP31	C1RL
AGAP2	ANGPT1	ARRB1	BCAS3	C1R

C1S	CCR2	CD81	CHRM2	CREBBP
C2	CCR5	CD84	CHRNB2	CRHR2
C3	CCR8	CD86	CHST3	CRKL
C5	CCR9	CD96	CHUK	CRLF3
C6	CCRL2	CDA	CIITA	CRMP1
C7	CCT2	CDC37	CILP	CRTAM
C8A	CCT8	CDH13	CISH	CRTC3
C8B	CD109	CDH17	CKLF	CRY1
C9	CD19	CDH1	CLEC1A	CRY2
CA1	CD207	CDH2	CLEC4E	CSE1L
CA2	CD226	CDH3	CLEC4F	CSF1R
CALCA	CD244	CDH4	CLEC7A	CSF2
CALCRL	CD27	CDH5	CLEC9A	CSF3R
CALR	CD28	CDHR3	CLIC5	CSK
CAND1	CD2	CDK13	CLNK	CSNK2B
CANX	CD300LF	CDK1	CLTC	CSTB
CAPG	CD302	CDK4	CLUAP1	CS
CAPN1	CD320	CDK5RAP3	CLU	CTC1
CAPZA1	CD34	CDK6	CMA1	CTDSPL2
CARD11	CD3D	CDK7	CNMD	CTF1
CAV1	CD3G	CDNF	CNOT7	CTGF
CAV2	CD40LG	CELA1	CNTN2	CTLA4
CAV3	CD44	CELSR1	CNTN4	CTNNA1
CBFB	CD47	CELSR3	COCH	CTNNB1
CBL	CD48	CEP290	COLEC10	CTNNBL1
CCL11	CD4	CFI	COLEC12	CTNND1
CCL13	CD53	CFL1	COMMD3	CTPS1
CCL1	CD58	CFP	COMMD9	CTSA
CCL20	CD59	CFTR	COPB1	CTSB
CCL24	CD5L	CHD7	CORO1A	CTSG
CCL25	CD63	CHEK1	CPB2	CTSH
CCL26	CD68	CHI3L1	CPD	CTSK
CCL27	CD69	CHIA	CPN1	CTSS
CCL28	CD6	CHID1	CPNE3	CTSV
CCL2	CD7	CHIT1	CPPED1	CTSW
CCNH	CD80	CHRDL1	CR2	CTSZ

CUBN	DHX58	ELMOD2	F7	FGG
CUL1	DMD	ELOC	FABP5	FGL2
CUL4A	DMRT1	EML2	FAF2	FGR
CUL5	DMTN	ENG	FAM120A	FIP1L1
CUX1	DNAAF4	EOMES	FAM120C	FKBP1B
CX3CL1	DNAI1	EP300	FAM210B	FLNB
CX3CR1	DNM1	EPAS1	FAM213A	FLOT1
CXCL10	DNM2	EPRS	FAP	FLT3
CXCL11	DRD1	EPX	FARP2	FN1
CXCL12	DRD2	ERAP2	FASN	FNBP1L
CXCL13	DRD3	ERBB2	FAT4	FNIP1
CXCL17	DRGX	ERBB3	FBLN1	FOLR1
CXCL9	DROSHA	ERBB4	FBN1	FRMPD3
CXCR6	DSP	ERBIN	FBXO7	FRRS1L
CYFIP1	DTX3L	ERCC2	FBXO9	FRZB
CYFIP2	DTX4	EREG	FBXW11	FSHB
CYLD	DYNC1LI1	ERN1	FBXW4	FSHR
CYTL1	DYNC1LI2	ERP44	FBXW7	FSTL4
DAB2	DYNLL2	ESM1	FCAR	FST
DAPK3	DYNLT1	ETV1	FCER1A	FUCA2
DBNL	ECEL1	ETV2	FCGRT	FUT10
DCAF1	ECM1	ETV4	FCMR	FYB1
DCC	EDA	ETV5	FCN3	FYB2
DCLK1	EDIL3	EVL	FCRL1	FYN
DCLRE1C	EDN1	EXOC1	FCRL3	FZD3
DCN	EDNRA	EXOSC3	FCRL4	FZD4
DCST1	EDNRB	EXOSC5	FCRLA	FZD6
DDOST	EEF2	EXOSC9	FCRLB	G6PD
DDRGK1	EFEMP1	EXT1	FECH	GABPA
DDX41	EGFR	EXTL3	FERMT1	GAK
DDX58	EGF	EZH2	FER	GALNT2
DEFB123	EIF2AK4	EZR	FEZ1	GAP43
DEFB127	ELF2	F12	FEZ2	GAPT
DEFB128	ELF4	F2RL1	FEZF2	GAS2L1
DERL1	ELMO1	F2RL3	FGA	GATA3
DERL2	ELMO2	F2	FGB	GBA

GCA	HACD3	HUWE1	IL1B	IMPDH2
GCG	HAMP	ICOS	IL1F10	INHBA
GCLC	HAS2	IFI35	IL1R2	INPP4B
GDI2	HAVCR2	IFI6	IL1RAP	INPP5D
GFRA1	HAX1	IFIT5	IL1RL2	INSR
GFRA2	HBEGF	IFNAR1	IL20RA	IP6K2
GFRA3	HCFC1	IFNAR2	IL20RB	IPO7
GFRAL	HCK	IFNB1	IL21R	IQGAP1
GHDC	HDGFL3	IFNGR1	IL21	IQGAP2
GHRH	HECW1	IFNG	IL22RA1	IQUB
GIGYF2	HELLS	IFNLR1	IL22RA2	IRAK1BP1
GIPC1	HFE	IGF1R	IL22	IRAK1
GJA1	HGF	IGF2R	IL23R	IRAK2
GLA	HGSNAT	IGSF6	IL24	IRAK3
GLB1	HGS	IKBKB	IL26	IRAK4
GLIPR1	HHEX	IKBKE	IL2RA	IRF1
GLMN	HHIPL1	IL10RA	IL2RG	IRF6
GLP2R	HIF1A	IL11RA	IL2	IRF8
GM2A	HK1	IL12A	IL33	IRF9
GP2	HK3	IL12B	IL36G	ISPD
GPC3	HLA	IL12RB2	IL4I1	IST1
GPI	HNF4A	IL13RA1	IL4R	ITCH
GPNMB	HNMT	IL13RA2	IL4	ITK
GPR18	HNRNPA2B1	IL13	IL5RA	ITPR1
GRHL3	HNRNPDL	IL15	IL5	ITPR3
GRIK2	HPN	IL16	IL6R	JADE2
GRIN2B	HPS1	IL17A	IL6ST	JAK1
GRM5	HPSE	IL17RB	IL6	JAK2
GRXCR1	HRG	IL17RC	IL7R	JAK3
GSK3B	HS1BP3	IL17RD	IL7	JAM3
GSN	HSD17B4	IL17RE	IL9	JCHAIN
GSTO1	HSP90B1	IL18R1	ILDR1	KARS
GSTP1	HSPG2	IL18RAP	ILDR2	KDELR1
GTF2F1	HTRA1	IL18	ILF2	KDM1A
GUSB	HTRA3	IL19	ILF3	KDM6A
GYG1	HTRA4	IL1A	IMPDH1	KDM6B

KEAP1	LGR5	MASP1	MMRN2	NABP1
KIRREL1	LINGO2	MBL2	MOG	NABP2
KIT	LMNB1	MCC	MOSPD2	NAV3
KLHL12	LMTK2	MCOLN1	MPO	NBEAL2
KLHL24	LNPEP	MCOLN3	MPZL1	NBN
KLHL6	LOXL2	MDM2	MRC1	NCAM1
KLK5	LOXL3	MEF2A	MRC2	NCBP3
KLRB1	LOXL4	MEF2C	MS4A1	NCF2
KLRD1	LPAR1	MEFV	MSH3	NCK1
KPNB1	LRCH1	MEGF8	MSH6	NCKAP1
KYNU	LRIG2	MEIS1	MSN	NCL
L1CAM	LRIT2	MEIS2	MST1	NCR1
LAMA2	LRIT3	MELK	MTHFD1	NCR3
LAMB2	LRMP	MEN1	MTHFR	NCSTN
LAMP1	LRRC17	MET	MTOR	NDEL1
LAMP2	LRRC3B	MFN2	MTSS1	NDFIP1
LAMP3	LRRC7	MFSD6	MUC15	NDP
LAMTOR3	LRRFIP2	MGRN1	MUC1	NEDD4
LANCL1	LRRIQ1	MIA3	MVP	NECTIN2
LAX1	LRRK2	MICALL1	MX1	NECTIN4
LBP	LTA4H	MID1	MYCBP2	NEK6
LCK	LTBR	MILR1	MYD88	NEO1
LCP1	LY86	MINK1	MYH10	NETO1
LEAP2	LY9	MKKS	MYH2	NETO2
LECT2	LYN	MKNK2	MYL9	NF1
LEF1	LYST	MKS1	MYLIP	NF2
LEMD3	LYZ	MLEC	MYLK3	NFASC
LEO1	MAGED1	MLH1	MYLPF	NFATC1
LEPR	MAGT1	MME	MYO18A	NFATC2
LEP	MALT1	MMP14	MYO1F	NFATC3
LGALS1	MANF	MMP1	MYO3A	NFATC4
LGALS3	MAP2K1	MMP28	MYO3B	NFE2L1
LGALS8	MAP3K7CL	MMP2	MYO6	NFE2L2
LGI1	MAPKAPK3	MMP3	MYO7A	NFIB
LGMN	MARCO	MMP8	MYO9B	NFIL3
LGR4	MARK2	MMP9	MYOF	NFKB1

NFKBIA	OPRK1	PGR	PPARGC1A	PTGIS
NFKBID	OPRM1	PHB2	PPARGC1B	PTGS1
NFKBIL1	OPTN	PHB	PPARG	PTGS2
NFRKB	ORMDL3	PHIP	PPIE	PTK6
NHLRC3	OTUB1	PIAS1	PPP1CA	PTN
NIT2	OTUD5	PIBF1	PPP2CA	PTPN11
NKAPL	OTUD7B	PID1	PPP2R3C	PTPN1
NLRP3	OTULIN	PIGR	PPP3CB	PTPN22
NMT2	OTX2	PIK3CG	PPP4R2	PTPRA
NOD1	OVOL2	PIK3R6	PPP6C	PTPRB
NOD2	P2RY12	PIM1	PQBP1	PTPRG
NOS3	PACRG	PINK1	PRCP	PTPRN
NOTCH1	PADI2	PKD1	PRDM16	PTX3
NPC1	PAFAH1B1	PKD2L1	PRDM1	PUM1
NPHP3	PAFAH1B2	PKD2L2	PRDX2	PUM2
NPHS1	PAK1	PKD2	PRDX4	PYGB
NPNT	PAK2	PKHD1L1	PRDX6	PYGO1
NR3C1	PAK3	PKM	PRG2	QPCT
NR5A2	PANX1	PLA2G7	PRG3	QSOX1
NRCAM	PARK7	PLAUR	PRG4	RAB10
NRG3	PARVA	PLAU	PRICKLE1	RABGEF1
NRP1	PBLD	PLEKHA1	PRICKLE2	RACGAP1
NRP2	PBX1	PLEKHG5	PRKD1	RACK1
NRROS	PDCL3	PLEK	PRKN	RAD21
NRXN1	PDE2A	PLGRKT	PRKRA	RAG1
NRXN3	PDE5A	PLP2	PRLR	RAG2
NTN4	PDGFD	PLS3	PROCR	RALBP1
NTRK1	PDGFRA	PLXNA1	PROK2	RALGAPA1
NTRK3	PDGFRB	PLXNA2	PROM1	RAMP2
NUB1	PDGFRL	PLXNA3	PROM2	RANBP9
NUMB	PDXK	PLXNB2	PRPF6	RAP1GAP
NUPL2	PELI1	PLXND1	PSEN1	RAPGEF1
OCLN	PELI2	PMS2	PSIP1	RAPGEF2
OFD1	PELI3	PNPLA8	PTCH2	RAPGEF4
ONECUT1	PEX5L	PPARA	PTGER4	RAPGEFL1
OPHN1	PGC	PPARD	PTGES2	RARG

RB1	RUVBL2	SERINC3	SLA2	SPP1
RBCK1	RYK	SERPINA1	SLAMF1	SPRED1
RBPJL	RYR2	SERPINA3	SLAMF6	SPSB1
RBPJ	SACM1L	SERPINB10	SLC11A1	SQSTM1
REEP1	SAMSN1	SERPINB1	SLC16A1	SRC
REEP2	SARM1	SERPINB2	SLC27A2	SRI
REEP3	SART1	SERPINB6	SLC44A2	SRPK1
REEP4	SASH1	SERPINC1	SLC50A1	SRPK2
REEP5	SASH3	SERPINE1	SLC7A10	SRPX2
RELN	SCAMP1	SETD2	SLC7A11	SRSF4
REN	SCAP	SFPQ	SLIT1	ST3GAL6
RETN	SCARB2	SFR1	SLIT3	ST6GAL1
RFTN1	SCEL	SFRP1	SLPI	STAM
RFTN2	SCRIB	SFRP4	SMARCAD1	STARD10
RGS1	SCYL2	SFTPB	SMO	STC1
RHAG	SDHAF4	SFTPC	SMPD3	STIL
RIF1	SEC14L1	SFTPD	SMPDL3B	STK10
RIOK3	SELENOS	SGIP1	SMURF1	STK36
RIPK3	SELE	SGPL1	SNAPC1	STK4
RNF111	SELL	SGSM1	SNAPC4	STMN1
RNF138	SELP	SH2B3	SNCA	STOML1
RNF168	SEMA3B	SH3BP1	SNX10	STOM
RNF220	SEMA3C	SH3GL2	SNX27	SUB1
RNF34	SEMA3D	SH3GLB1	SNX4	SUFU
RNF4	SEMA3E	SH3KBP1	SNX9	SULF1
RNF8	SEMA3F	SH3PXD2A	SOCS2	SWAP70
ROCK1	SEMA4A	SHANK2	SOCS5	SYK
ROCK2	SEMA4B	SHC1	SOD1	TAB1
RPA1	SEMA4D	SHC3	SORBS1	TAB2
RPGRIP1L	SEMA4G	SIAH1	SORT1	TAB3
RPS27A	SEMA5A	SIGIRR	SOSTDC1	TAC1
RSAD2	SEMA5B	SIN3A	SP1	TACR1
RSPO2	SEMA6A	SIRT1	SPARC	TAGAP
RTKN2	SEMA6C	SIT1	SPART	TAL1
RTN4	SENP2	SKAP1	SPG21	TALDO1
RUVBL1	SENP7	SKAP2	SPINK5	TANK

TAP1	TINAG	TRAT1	UBE2B	VPS33B
TAZ	TKFC	TREML1	UBE2D1	WASF2
TBC1D10C	TLR3	TRIB1	UBE2I	WASL
TBC1D23	TLR8	TRIB2	UBE2K	WDFY1
TBK1	TMBIM6	TRIM25	UBE2M	WDPCP
TBKBP1	TMC1	TRIM33	UBE2O	WDR1
TBR1	TMC6	TRIM37	UBR2	WDR61
TBX18	TMC8	TRIM38	UBR4	WDR77
TBX20	TMEM131L	TRIM4	UBR5	WIPF1
TBX21	TMEM173	TRIM63	UBXN2A	WIPF2
TBX2	TMEM179B	TRIM68	UCHL1	WIPF3
TBX3	TMEM63A	TRIM71	UCHL5	WISP3
TCF7	TMOD3	TRIP12	UFD1	WNT16
TCN1	TMPRSS15	TROVE2	UFL1	WNT2B
TCP1	TNFAIP1	TSC1	UMOD	WNT2
TDGF1	TNFAIP3	TSC2	USH1C	WNT5A
TEC	TNFRSF11A	TSG101	USP10	WNT8A
TEK	TNFRSF17	TSHR	USP12	WWOX
TENM2	TNFRSF8	TSPAN2	USP14	WWTR1
TESPA1	TNFRSF9	TSPAN6	USP15	XDH
TFF1	TNFSF10	TTC8	USP20	XPO1
TGFB3	TNFSF11	TUBB2B	USP33	XPR1
TGFBR1	TNFSF15	TUBB	USP34	XRCC4
TGFBR2	TNR	TUB	USP46	XRCC6
THBS1	TOM1L1	TULP1	USP4	XYLT1
THBS4	TOMM20L	TULP2	USP8	YARS
THEMIS2	TRAF1	TULP3	USP9X	YPEL5
THNSL2	TRAF3IP1	TXK	VANGL1	YTHDC2
THOC5	TRAF3IP2	TXNDC5	VANGL2	YTHDF2
THPO	TRAF3	TXNIP	VCL	ZAP70
TIA1	TRAF5	TYK2	VCP	ZNF451
TIAM1	TRAF6	TYR	VIL1	ZRANB1
TIE1	TRAF7	T	VIM	
TIGIT	TRAFD1	UBA3	VIPR1	
TIMP2	TRAIP	UBA52	VIP	
TINAGL1	TRAPPC1	UBA5	VLDLR	

Table S4 - List of indexed P7 and indexed P5 primers for the Illumina platform. Index is represented with lowercase letters.

Oligo ID	Oligo sequence (5'-3')
BEST_501	AATGATACGGCGACCACCGAGATCTACACcctgcgaACACTCTTTCCCTACACGACGCTCTT
BEST_503	AATGATACGGCGACCACCGAGATCTACACacctaggACACTCTTTCCCTACACGACGCTCTT
BEST_505	AATGATACGGCGACCACCGAGATCTACACatcttgcACACTCTTTCCCTACACGACGCTCTT
BEST_507	AATGATACGGCGACCACCGAGATCTACACttgcgaaACACTCTTTCCCTACACGACGCTCTT
BEST_515	AATGATACGGCGACCACCGAGATCTACACcaggtcgACACTCTTTCCCTACACGACGCTCTT
BEST_517	AATGATACGGCGACCACCGAGATCTACACggtacctACACTCTTTCCCTACACGACGCTCTT
BEST_519	AATGATACGGCGACCACCGAGATCTACACgagattcACACTCTTTCCCTACACGACGCTCTT
BEST_523	AATGATACGGCGACCACCGAGATCTACACcgagatcACACTCTTTCCCTACACGACGCTCTT
BEST_702	CAAGCAGAAGACGGCATACGAGATtgcagagGTGACTGGAGTTCAGACGTGT
BEST_704	CAAGCAGAAGACGGCATACGAGATttgatccGTGACTGGAGTTCAGACGTGT
BEST_706	CAAGCAGAAGACGGCATACGAGATtctccatGTGACTGGAGTTCAGACGTGT
BEST_708	CAAGCAGAAGACGGCATACGAGATttcgagcGTGACTGGAGTTCAGACGTGT
BEST_710	CAAGCAGAAGACGGCATACGAGATgtaccggGTGACTGGAGTTCAGACGTGT
BEST_716	CAAGCAGAAGACGGCATACGAGATcgcattaGTGACTGGAGTTCAGACGTGT
BEST_718	CAAGCAGAAGACGGCATACGAGATggacgcaGTGACTGGAGTTCAGACGTGT
BEST_720	CAAGCAGAAGACGGCATACGAGATgagcatgGTGACTGGAGTTCAGACGTGT
BEST_724	CAAGCAGAAGACGGCATACGAGATcatcttgGTGACTGGAGTTCAGACGTGT
BEST_730	CAAGCAGAAGACGGCATACGAGATgcggcatGTGACTGGAGTTCAGACGTGT
BEST_732	CAAGCAGAAGACGGCATACGAGATtactattGTGACTGGAGTTCAGACGTGT
BEST_736	CAAGCAGAAGACGGCATACGAGATtattccaGTGACTGGAGTTCAGACGTGT

Table S5 - Double indexing samples. Each of the following tables represents a 96 well plate for 96 samples. "P7/P5 index name" represents the index combination by oligo name, "P7/P5 index sequence" represents the index combination by the index sequence, and "P7/P5 index AC/TG" balance represents the balanced distribution between red (A and C) and green (T and G) laser on the Illumina platform.

P7/P5 Index Name

	01	02	03	04	05	06	07	08	09	10	11	12
A	501/702	501/704	501/706	501/708	501/710	501/716	501/718	501/720	501/724	501/730	501/732	501/736
B	503/702	503/704	503/706	503/708	503/710	503/716	503/718	503/720	503/724	503/730	503/732	503/736
C	505/702	505/704	505/706	505/708	505/710	505/716	505/718	505/720	505/724	505/730	505/732	505/736
D	507/702	507/704	507/706	507/708	507/710	507/716	507/718	507/720	507/724	507/730	507/732	507/736
E	515/702	515/704	515/706	515/708	515/710	515/716	515/718	515/720	515/724	515/730	515/732	515/736
F	517/702	517/704	517/706	517/708	517/710	517/716	517/718	517/720	517/724	517/730	517/732	517/736
G	519/702	519/704	519/706	519/708	519/710	519/716	519/718	519/720	519/724	519/730	519/732	519/736
H	523/702	523/704	523/706	523/708	523/710	523/716	523/718	523/720	523/724	523/730	523/732	523/736

P7/P5 Index Sequence

	01	02	03	04	05	06	07	08	09	10	11	12
A	CCTGCG A/TGCA GAG	CCTGCG A/TTGA TCC	CCTGCG A/TCTC CAT	CCTGCG A/TTCG AGC	CCTGCG A/GTAC CGG	CCTGCG A/CGCA TTA	CCTGCG A/GGAC GCA	CCTGCG A/GAGC ATG	CCTGCG A/CATC TTG	CCTGCG A/GCGG CAT	CCTGC GA/TAC TATT	CCTGCG A/TATT CCA
B	ACCTAG G/TGCA GAG	ACCTAG G/TTGA TCC	ACCTAG G/TCTC CAT	ACCTAG G/TTCG AGC	ACCTAG G/GTAC CGG	ACCTAG G/CGCA TTA	ACCTAG G/GGAC GCA	ACCTAG G/GAGC ATG	ACCTAG G/CATC TTG	ACCTAG G/GCGG CAT	ACCTAG G/TACT ATT	ACCTAG G/TATT CCA
C	ATCTTG C/TGCA GAG	ATCTTG C/TTGA TCC	ATCTTG C/TCTC CAT	ATCTTG C/TTCG AGC	ATCTTG C/GTAC CGG	ATCTTG C/CGCA TTA	ATCTTG C/GGAC GCA	ATCTTG C/GAGC ATG	ATCTTG C/CATC TTG	ATCTTG C/GCGG CAT	ATCTTG C/TACT ATT	ATCTTG C/TATT CCA
D	TTGCGA A/TGCA GAG	TTGCGA A/TTGA TCC	TTGCGA A/TCTC CAT	TTGCGA A/TTCG AGC	TTGCGA A/GTAC CGG	TTGCGA A/CGCA TTA	TTGCGA A/GGAC GCA	TTGCGA A/GAGC ATG	TTGCGA A/CATC TTG	TTGCGA A/GCGG CAT	TTGCGA A/TACT ATT	TTGCGA A/TATT CCA
E	CAGGTC G/TGCA GAG	CAGGTC G/TTGA TCC	CAGGTC G/TCTC CAT	CAGGTC G/TTCG AGC	CAGGTC G/GTAC CGG	CAGGTC G/CGCA TTA	CAGGTC G/GGAC GCA	CAGGTC G/GAGC ATG	CAGGTC G/CATC TTG	CAGGTC G/GCGG CAT	CAGGT CG/TAC TATT	CAGGTC G/TATT CCA
F	GGTACC T/TGCA GAG	GGTACC T/TTGA TCC	GGTACC T/TCTC CAT	GGTACC T/TTCG AGC	GGTACC T/GTAC CGG	GGTACC T/CGCA TTA	GGTACC T/GGAC GCA	GGTACC T/GAGC ATG	GGTACC T/CATC TTG	GGTACC T/GCGG CAT	GGTAC CT/TAC TATT	GGTACC T/TATT CCA
G	GAGATT C/TGCA GAG	GAGATT C/TTGA TCC	GAGATT C/TCTC CAT	GAGATT C/TTCG AGC	GAGATT C/GTAC CGG	GAGATT C/CGCA TTA	GAGATT C/GGAC GCA	GAGATT C/GAGC ATG	GAGATT C/CATC TTG	GAGATT C/GCGG CAT	GAGATT C/TACT ATT	GAGATT C/TATT CCA
H	CGAGAT C/TGCA GAG	CGAGAT C/TTGA TCC	CGAGAT C/TCTC CAT	CGAGAT C/TTCG AGC	CGAGAT C/GTAC CGG	CGAGAT C/CGCA TTA	CGAGAT C/GGAC GCA	CGAGAT C/GAGC ATG	CGAGAT C/CATC TTG	CGAGAT C/GCGG CAT	CGAGA TC/TAC TATT	CGAGAT C/TATT CCA

	P7/P5 Index AC/TG Balance											
	01	02	03	04	05	06	07	08	09	10	11	12
A	0.500	0.500	0.571	0.500	0.500	0.571	0.571	0.500	0.500	0.500	0.500	0.571
B	0.500	0.500	0.571	0.500	0.500	0.571	0.571	0.500	0.500	0.500	0.500	0.571
C	0.429	0.429	0.500	0.429	0.429	0.500	0.500	0.429	0.429	0.429	0.429	0.500
D	0.429	0.429	0.500	0.429	0.429	0.500	0.500	0.429	0.429	0.429	0.429	0.500
E	0.429	0.429	0.500	0.429	0.429	0.500	0.500	0.429	0.429	0.429	0.429	0.500
F	0.429	0.429	0.500	0.429	0.429	0.500	0.500	0.429	0.429	0.429	0.429	0.500
G	0.429	0.429	0.500	0.429	0.429	0.500	0.500	0.429	0.429	0.429	0.429	0.500
H	0.500	0.500	0.571	0.500	0.500	0.571	0.571	0.500	0.500	0.500	0.500	0.571

Table S6 - Minor allele frequency (MAF) and heterozygosity frequency of the significant missense variations detected and SLAMF1, through time, for the main and peaks datasets. Significant differences for * - adjusted p value of 0.00024; # - p value of 0.05.

	Main dataset		Peaks dataset		
Loci	**MAF**	**Heterozygosity**	**MAF**	**Heterozygosity**	**Years**
	19.2% (5/26)	5/13	14.2% (2/14)	2/7	1990-1992
	0%	0	0%	0	1993-1996
HAVCR2*	19.2% (5/26)	5/13	20% (2/10)	2/5	1997-1999
	0%	0	0%	0	2003-2005
	20% (6/30)	6/15	25% (4/16)	3/8	2006-2008
	50% (10/20)	10/10	50% (5/10)	5/5	1990-1992
	50% (7/14)	7/7	0%	0	1993-1996
TMOD3#	50% (12/24)	12/12	50% (4/8)	4/4	1997-1999
	35% (7/20)	7/10	25% (3/12)	3/6	2003-2005
	50% (13/26)	13/13	50% (6/12)	6/6	2006-2008
	25% (6/24)	6/12	25% (3/12)	3/6	1990-1992
	20% (4/20)	4/10	0%	0	1993-1996
TMPRSS15#	38.4% (10/26)	10/13	37.5% (3/8)	3/4	1997-1999
	7.6% (2/26)	2/13	5.5% (1/18)	1/9	2003-2005
	26.9% (7/26)	7/13	21.4% (3/14)	3/7	2006-2008

	17.8% (5/28)	5/14	18.7% (3/16)	3/8	1990-1992
	11.5% (3/26)	3/13	0	0	1993-1996
CUBN#	0%	0	0	0	1997-1999
	25% (6/24)	6/12	16.7% (3/18)	3/9	2003-2005
	0%	0	0	0	2006-2008
	8.3% (2/24)	2/12	14.2% (2/14)	2/7	1990-1992
	22.2% (4/18)	4/9	25% (1/4)	1/2	1993-1996
PDE5A#	15% (3/20)	3/10	0%	0	1997-1999
	26.9% (7/26)	7/13	33.3% (6/18)	6/9	2003-2005
	12.5% (3/24)	3/12	8.3% (1/12)	1/6	2006-2008
	50% (13/26)	13/13	50% (7/14)	7/7	1990-1992
	50% (2/24)	12/12	50% (2/4)	2/2	1993-1996
AIMP1#	50% (2/24)	12/12	50% (4/8)	4/4	1997-1999
	35% (7/20)	7/10	31.2% (5/16)	5/8	2003-2005
	46.7% (14/30)	14/15	50% (8/16)	8/8	2006-2008
	11.5% (3/26)	3/13	8.3% (1/12)	1/6	1990-1992
	7.7% (2/26)	2/13	0%	0	1993-1996
CRTAM#	7.7% (2/26)	2/13	0%	0	1997-1999
	20.8% (5/24)	5/12	22.2% (4/18)	4/9	2003-2005
	0	0	0%	0	2006-2008
	19.2% (5/26)	5/13	21.4% (3/14)	3/7	1990-1992
	34.6% (9/26)	9/13	50% (2/4)	2/2	1993-1996
IL26#	29.1% (7/24)	7/12	12.5% (1/8)	1/4	1997-1999
	29.1% (7/24)	7/12	37.5% (3/8)	3/4	2003-2005
	13.3% (4/30)	4/15	12.5% (2/16)	2/8	2006-2008
	3.8% (1/26)	1/13	0%	0	1990-1992
	4.5% (1/22)	1/11	0%	0	1993-1996
LOC117310972#	4.2% (1/24)	1/12	0%	0	1997-1999
	20% (4/20)	4/10	25% (4/16)	4/8	2003-2005
	4.2% (1/24)	1/12	7.1% (1/14)	1/7	2006-2008
SLAMF1	17.8% (5/28)	5/14	6.2% (1/16)	1/8	1990-1992

	19.2% (5/26)	5/13	0%	0	1993-1996
	16.6% (4/24)	4/12	2% (2/10)	2/5	1997-1999
	29.1% (7/24)	7/12	37.5% (6/16)	6/8	2003-2005
	21.4% (6/28)	6/14	18.7% (3/16)	3/8	2006-2008