Using neutral and outlier loci to uncover the genetic structure of

Mytilus galloprovincialis populations from the eastern Adriatic coast

using population genetic and seascape genetic approaches

Joanna Szondy Hamilton

A thesis submitted to Victoria University of Wellington

in fulfilment of the requirements for the degree of

Master of Science in Marine Biology

2019

Abstract

The genetic variation of the Mediterranean blue mussel, *Mytilus galloprovincialis* from the eastern coastline of the Adriatic Sea was investigated using polymorphic microsatellite markers. In total, 843 individuals were sampled from 18 populations representing a variety of coastal environments in Croatia, Bosnia-Herzegovina, Montenegro and Albania. Neutral loci revealed low levels of genetic structure in this continuously distributed species, while one outlier locus, MGE7, which may be under selection, provided strong evidence of genetic structure. The distribution of one of the alleles of this locus, MGE7^{243}, was significantly correlated with latitude. A genetic seascape analysis using 9 environmental and 3 geospatial variables revealed a strong association between MGE7^{243} and three highly correlated environmental variables, maximum sea surface temperature, minimum salinity and maximum chlorophyll-a. This association was maintained for homozygous genotypes for the MGE7^{243} allele, but not for heterozygotes, providing further evidence that the locus MGE7 may be under selection or closely linked to a gene under selection. These findings highlight how previously unrecognised genetic structure can be identified through the use of genetic seascape approaches.

Acknowledgements

Table of contents

Supplementary material.

List of Figures

List of Tables

Supplementary Figures and Tables

x

Chapter 1: Introduction

Study Species

Mytilus galloprovincialis Lamarck 1819, the Mediterranean blue mussel, is a species of marine mollusc inhabiting intertidal habitats in temperate and subtropical regions in many parts of the world. It is important both ecologically as an ecosystem engineer and commercially in aquaculture, and it is also on the list of 100 of the World's Worst Invasive Alien species (Fly et al. 2015, Kovačić et al. 2017, Global Invasive Species Database 2019). It is one member of the *Mytilus edulis* species complex made up of *M. edulis, M. galloprovincialis* and *M. trossulus* (McDonald et al. 1991, Hilbish et al. 2000). The species complex has an antitropical distribution with each species having its own distinct distribution, evolutionary history and genetic lineage (Quesada et al. 1995, Hilbish et al. 2000). The species interbreed and form hybrid zones where their distributions overlap (Braby & Somero 2006b, Rawson & Harper 2009, Roux et al. 2014). The native range of *M. galloprovincialis* extends from the Black Sea, along the northern coast of the Mediterranean Sea, along the northwestern coast of Africa both within the Mediterranean and in the Atlantic, and then north along the Atlantic coast of Spain to France and to the southwest of the British Isles, where it has interbred with *M. edulis* (Hilbish et al. 2000, Riginos & Cunningham, 2005). Of the three sibling species, *M. galloprovincialis* has been most widely dispersed to the Southern hemisphere, where two distinct lineages have been identified. The first lineage, Southern hemisphere *M. galloprovincialis*, is thought to have crossed the tropics during the Pleistocene, while the Northern hemisphere lineage is a more recent introduction, possibly anthropogenic, via ship hulls and ballast water (Hilbish et al. 2000, Westfall & Gardner 2010, Gardner et al. 2016).

Although Northern hemisphere *M. galloprovincialis* is present in the Southern hemisphere, there is no evidence of Southern hemisphere lineages of blue mussels in the Mediterranean Sea. In Croatia, however, there is evidence of very low levels of introgression from *M. edulis* and *M. trossulus* (Hamer et al. 2012). The nuclear DNA marker Me 15/16, which is not completely diagnostic between the three species, was examined in 110 samples from 22 Croatian sites. The majority of mussels sampled (108) were homozygous for the G (*M. galloprovincialis*) allele, but there were two individuals with a heterozygous genotype, one with the G and the E (*M. edulis*) allele, and one with the G and the T (*M. trossulus*) allele. While *M. galloprovincialis* and *M. edulis* are known to hybridise and exchange alleles, *M. galloprovincialis* and *M. trossulus* do not occur naturally in the same areas in Europe. In areas where *M. galloprovincialis* has been introduced into the range of *M. trossulus*, they do, however, hybridise although hybrids may have reduced reproductive viability (Braby & Somero 2006b, Lockwood & Somero 2011). The presence of the T allele is therefore surprising and was explained either via a two-step introgression process, or through human-mediated transport.

The existence of northern lineage mussels in the Southern hemisphere underlines its invasive potential (Westfall 2011, Global Invasive Species Database 2019). It has extended its range to North America, Chile, South Africa and Namibia, Japan, Hong Kong, New Zealand and Australia (McDonald et al. 1991, Larraín et al. 2015, Griffiths et al. 1992, Gardner et al. 2016). Its success as an invader is in part due to its tolerance of a wide variety of environmental conditions, and also human-mediated transfers, either unintended via, ships, barges and even oil rigs, or intended, for the purposes of aquaculture (Branch & Steffani 2004, Evans & Somero 2010, Tomanek 2012, McQuaid & Phillips 2000)

The main mechanism for exchange of individuals between populations is, as for many benthic marine invertebrates, via a pelagic larval stage which lasts for up to 8 weeks depending on environmental conditions (Beaumont et al. 2004). During this pelagic stage, larvae are not entirely passive particles floating in the water, since they display various swimming behaviours depending on the stage of development (Gosling 2003). Their ultimate fate, whether they are retained close to their natal origin or transported hundreds of kilometers away, is however determined by sea currents. Each female can produce up to 40 million eggs, only a fraction of which will be fertilized, survive predation and dispersal and reach habitat suitable for settlement (Gosling 2003, Cowen & Sponaugle 2009). In the Mediterranean Sea, their main habitat is the rocky intertidal zone, although in aquaculture areas they are grown on longlines and rafts (Michaelidis et al. 2014, Mandić et al. 2017). Environmental conditions in the summer in some areas of the Mediterranean Sea are close to, and in some instances have exceeded, the limits of their tolerance (Anestis et al. 2007, Ramón et al. 2007, Michaelidis et al. 2014). There have been increasing incidents of mass mortality during marine heat waves that have wiped out whole populations of mussels (Lejeusne et al. 2010, Di Camillo & Cerrano 2015, Galli et al. 2017).

Mytilus galloprovincialis is a widely studied model organism, and an ideal subject to investigate how genetic connectivity is affected by physical oceanography at different spatial-scales, from the main currents and gyres of the sea to local tides, eddies and currents (Lockwood & Somero 2011, Wilson-Sanders 2011, Michaelidis et al. 2014). The species has been studied extensively to understand the dynamics of hybrid zones, and to recognise and monitor its invasive potential, but there has been much less attention on populations within its native range (Gardner & Skibinski 1990a & b, Branch & Steffani 2004, Gardner et al. 2016). Seascape genetic approaches, which search for links between environmental features and

genetic patterns, are beginning to be employed and have used combinations of hydrographic modelling and environmental variables together with single- and multi-locus genetic data (Gilg & Hilbish 2003, Braby & Somero 2006b, Stuckas et al. 2017, Kijewski et al. 2019). These seascape studies have, however, mainly focussed on hybrid zone dynamics and genetic differences between sibling *Mytilus* species: *M. edulis* and *M. galloprovincialis* (Gilg & Hilbish 2003), *M. trossulus* and *M. galloprovincialis* (Braby & Somero 2006b) and *M. edulis* and *M. trossulus* (Stuckas et al. 2017, Kijewski et al. 2019) with less attention placed on how environmental features relate to the genetic patterns of a single species in its native range.

Microsatellite markers

Microsatellite markers are one of the most popular and widely employed genetic markers in population genetic studies to assess population structure and connectivity (Chambers & MacAvoy 2000, Selkoe & Toonen 2006). Microsatellites are ubiquitous throughout eukaryotic genomes and can occur anywhere in the genome from intergenic regions to exons (Tóth et al. 2000, Bagshaw 2017). Microsatellites, also known as short tandem repeats (STRs) and short sequence repeats (SSRs), are short sequences of DNA consisting of tandem repeats of between 2 – 6 bp (e.g., the motif CGT repeated multiple times). The repetitive nature of a microsatellite can lead to strand slippage and misalignment during DNA replication resulting in high rates of mutation and polymorphism. It is this polymorphism that is useful in genetic studies (Chambers & MacAvoy 2000, Vieira et al. 2016). The high mutation rate of microsatellites means that genetic analyses using these markers have a more recent scale of reference compared to mitochondrial or nuclear DNA sequence (Waits & Storfer 2016). Traditionally microsatellites have been used in population genetics as neutral markers, though their distribution throughout the genome means that they need not be selectively neutral (Vieira et al. 2016, Bagshaw 2017). They are otherwise known to be implicated in several human

diseases and although originally microsatellites were thought to be genomic aberrations that could cause disease, it is now recognised that they can have important functions in the cell, which may become pathological when the number of repeats has expanded above a certain threshold (Bagshaw et al. 2017, Sawaya et al. 2013).

Marine population genetics

Low genetic differentiation amongst populations is a characteristic of many marine species over large spatial scales and in early work was explained by a model of open seas, large effective populations and high fecundity, with high demographic connectivity mediated through the long-distance dispersal of a pelagic larval stage (Waples 1998, Cowen & Sponaugle 2009). Low genetic differentiation has been observed in *Mytilus* species over different spatial scales. For example, no genetic differentiation was found in *M. californianus* over its entire range > 4000 km (Addison et al. 2008). For *M. galloprovincialis* in the Mediterranean Sea, there were only low levels of differentiation within the Adriatic Sea (Štambuk et al. 2013), the Aegean and Ionian Seas (Giantsis et al. 2014a & b) and for sample sites around the Atlantic and Alboran coast of the Iberian peninsula (Diz & Presa, 2008), although the latter three studies did report differentiation between basins and between western and eastern populations in the Adriatic Sea.

Low genetic differentiation, measured by the F_{ST} statistic, may not, however, imply high genetic connectivity but may arise instead from large effective populations even when gene flow is low (Hedgecock et al. 2007, Gagnaire et al. 2015). Increasingly, marine population genetic studies are finding high levels of larval retention (Teske et al. 2015, Miller et al. 2018) including studies of *Mytilus* congeners, which have found the majority of mussel larvae settle within 40 km of their natal origin (McQuaid & Phillips, 2000, Gilg & Hilbish 2003, Becker et

al. 2007, Stuckas et al. 2017). Biophysical models of larval dispersal in the Adriatic Sea have also identified areas of predicted larval retention, especially along the Croatian coast (Dubois et al. 2016, Bray et al. 2017). A high level of larval retention infers low gene flow and would therefore be expected to lead to genetic structuring of populations. Thus, the low genetic structure so far reported for *M. galloprovincialis* in the Mediterranean Sea may be not result of high levels of connectivity (Diz & Presa, 2008, Štambuk et al. 2013, Giantsis et al. 2014a & b).

To try to uncover whether low differentiation is due to high gene flow or large effective populations, models such as hierarchical analysis of molecular variance (AMOVA) and isolation-by-distance, IBD, have often been employed. AMOVA partitions genetic variance into different levels of organisation from regional groupings of populations down to the individual, allowing indirect assessment of gene flow between regions compared to between populations. IBD indirectly assesses levels of gene flow by looking at the balance between genetic drift and dispersal over geographic distance. The original IBD model of Wright (1943), which was developed to specifically test for the influence of weak gene flow as a function of geographic distance, had individuals uniformly distributed through space with mating occurring only between neighbouring individuals and dispersal being finite, which leads to a pattern of genetic differentiation that increases with geographic distance. IBD, which is usually tested for by Mantel tests, is one of the simplest models of population structure and has been observed across diverse taxa in many settings (Guillot et al. 2009, Jenkins et al. 2010) including oysters, dolphins and reef fishes (Rose et al. 2006, Amaral et al. 2012, Puebla et al. 2012, Nanninga et al. 2014).

Although now widespread, the use of Mantel tests to infer patterns of IBD has been criticised for a number of reasons, some seemingly contradictory. With regards to sampling design, the number of populations, individuals per population and the scale of the sampling area, especially in terms of mean dispersal distance, are very important to the success of

detecting IBD. Early studies in the marine setting often failed to detect IBD and this is thought, in some cases, to be because an insufficient number of populations were sampled rather than an absence of IBD (Selkoe et al. 2016). A meta-analysis of 240 IBD studies that used Mantel tests found that the choice of marker and genetic measure had no effect on IBD results, but that the number of populations included was asymptotically related to IBD significance. A 9 population-study had a 50% probability of detecting IBD, while one with over 23 populations had a 90% probability (Jenkins et al. 2010). There is a contrasting view that Mantel tests have low statistical power to detect IBD because of high variance in the genetic measures used (which leads to a weaker correlation) and because the IBD slope may be very low due to large genetic patch size (the IBD slope is inversely related to the density of genes and the square of the mean dispersal distance). Moreover, estimations of correlation are biased when too few individuals and populations are collected over too small an area (Leblois et al. 2003), so when sampling scale is large enough to overcome these biases the statistical power of the Mantel test can be too low to detect IBD (Guillot et al. 2009). Therefore, the number and size of populations and the size of study area compared to dispersal distance have important effects on both the validity and sensitivity of Mantel tests and there may not be a sweet spot where a Mantel test is both valid and sensitive enough to detect IBD.

Thus, while both F_{ST} values and Mantel tests for IBD can be used to estimate connectivity indirectly by inferring the gene flow necessary to produce the observed pattern of genetic structure, there can be significant problems in their implementation (Guillot et al. 2009, Gagnaire et al. 2015). Migration can, however, also be estimated directly through clustering methods such as Structure (Pritchard et al. 2000), AWclust (Gao & Starmer 2008) and Discriminant Analysis of Principal Components (DAPC) (Jombart et al. 2010), which try to assign individuals to a population cluster and can hence identify recent migrants or admixed individuals. Each clustering method has its own assumptions, strengths and drawbacks

(discussed in Chapter 2) but a general drawback is that their power to detect clusters generally increases with genetic differentiation, and they are therefore better suited to inferring connectivity when dispersal is limited and effective populations sizes small (Gagnaire et al. 2015).

There has also been much attention focussed on the specificity of these approaches when the underlying (true) genetic structure does not conform to the model assumed explicitly in the test. For example, simulations have shown that the presence of hierarchical or clustered population structures can bias IBD results and, in the reverse situation, the presence of IBD can bias results of analyses specifically testing for hierarchical or clustering population genetic structure (Guillot et al. 2009, Meirmans 2012, Perez et al. 2018). The genetic signal from the underlying population structure may therefore lead to spurious results in tests formulated to detect a different model of structure. In particular, where regional structure is discovered, Mantel tests across regions are likely to return positive tests from the regional structure unrelated to the mechanisms underpinning IBD, so in this case, IBD should only be tested within regions. Clearly, results must be carefully interpreted and, especially in cases of weak or nuanced structure, conclusions cannot be made without considering the results from all analytical methods. Nevertheless, when results cannot be disentangled, signals from diverse tests do lend confidence to the existence of some structure even though the model of genetic differentiation cannot be assessed.

In the marine environment, it has frequently been found that the pattern of genetic homogeneity found over large distances breaks down at smaller distances, with surprising levels of genetic differentiation occurring over smaller scales, which may change over time (Hedgecock & Pudovkin 2011, Eldon et al. 2016). This apparent contradiction between little or no structure over large distances and fluctuating allele or genotype frequencies at finer spatial scales has been termed chaotic genetic patchiness (CGP, Johnson and Black 1982). CGP

was originally reported in an undescribed species of limpet (*Siphonaria* sp.) and has also been found in species with long pelagic larval phases such as the sea urchin, *Sterechinus nuemayeri*, with larval duration of about 4 months (Miller et al. 2018) and the spiny lobster, *Jasus edwardsii*, which has one of the longest larval durations at 12 – 24 months (Villacorta-Rath et al. 2018).

CGP is challenging to explain and is likely multifactorial in origin and although not generally found in *Mytilus* species, some of the mechanisms put forward to explain CGP may be relevant to the larval dispersal and genetic connectivity of *M. galloprovincialis* in the Adriatic Sea. Four important mechanisms for CGP have been described: (1) Sweepstake reproductive success, (2) collective dispersal, (3) temporal asynchrony, and (4) selection (Eldon et al. 2016). Sweepstake reproductive success predicts a very large variation in individual reproductive success, dependent on the coordination of biological development, spawning, larval maturation and settlement, with oceanographic conditions favourable to each stage. It was first used to describe the genetic patterns observed in the Pacific oyster, *Crassostrea gigas* (Hedgecock 1994). Collective dispersal describes any biophysical process where groups of individuals are drawn together into "packets" of water and transported together, so that immigrants to the same population have a higher than random chance of originating from the same natal population and dispersal is therefore not a mixing process. Temporal asynchrony can lead to genetic differentiation over fine spatial scales when the timing of spawning varies from patch to patch due to environmental and genetic differences between patches. Along the Californian coast, the difference in timing of reproduction in the broadly co-occurring *Mytilus* species, *M. galloprovincialis* and *M. californianus*, has led to very different patterns of larval dispersal and connectivity because of seasonal differences in near-shore currents (Carson et al. 2010). The two species also differ in fine spatial scale distributions, with *M. galloprovincialis* mainly bay-dwelling and *M. californianus* mostly

inhabiting open-coast sites. Shoreline configuration, such as headlands and embayments, affect local hydrodynamics by creating eddies and fronts which have been found to retain phytoplankton within bays, a result which presumably can be extended to mussel larvae (Archambault et al. 1999 and references therein). This was reflected in the levels of self-recruitment observed, 42% in the bay-dweller, *M. galloprovincialis*, compared 28% in the open-coast *M. californianus* (Carson et al. 2010). Within a species, genotypic and size/age differences could also influence the timing of spawning and the distribution of individuals over different habitats, so differences in larval dispersal and connectivity may occur (Gardner & Skibinski 1988, 1990a). Genetic differentiation due to selection is often found in association with an environmental gradient, with a resulting predictable cline in genetic structure (Hilbish & Koehn 1985), but can also be associated with small-scale differences in habitat, such as exposure to wave action and substrate type leading to a genetic pattern of CGP (e.g., Gardner & Skibinski 1991; Gérard et al. 2015).

Seascape genetics

Seascape genetics is a young discipline related to landscape genetics but focussed on marine habitats and associated ocean dynamics, and with analytical methods still in development. Both landscape and seascape genetics seek to investigate the connections between the environment and microevolutionary processes and therefore require both environmental and genetic datasets (Manel et al. 2003, Holderegger & Wagner 2006, Selkoe et al. 2016). Landscape and seascape genetics approaches can therefore be summarised into three steps, the first related to obtaining genetic data quantifying an aspect of microevolutionary processes, the second quantifying landscape or seascape heterogeneity, and thirdly a statistical method to link the two. Finding the right combinations of the three to answer specific research questions is not trivial (Selkoe et al. 2016). For example, if the research

question is related to genetic differentiation amongst populations, models of larval dispersal which incorporate rare hydrodynamic events are important because very low levels of migration are required to prevent the accumulation of genetic differences between populations (Palumbi 2003). Questions related to the vulnerability of populations after disease and mass-mortality events and the ability of meta-populations to repopulate an area after such events will be more concerned with average larval dispersal and the speed with which populations can be re-established.

One of the most challenging aspects of obtaining seascape environmental data is the fluid nature of the marine environment. Dispersal of individuals and connectivity is mediated by the sea, a dynamic and heterogeneous medium driven by oceanographic features such as ocean and nearshore currents, winds and tides, and is shaped by geologic features such as islands, continental shelves and shoreline configuration (Liggins et al. 2013, Selkoe et al. 2016). The sea is heterogeneous: it varies spatially in 3 dimensions and temporally with respect to many biologically relevant characteristics: temperature, salinity, pH, nutrient concentrations (Riginos & Liggins 2013). Increasingly, environmental data is becoming available from satellite data and in the Adriatic Sea, has been used to estimate sea surface temperature and chlorophyll-a concentrations at a basin-wide scale (Böhm et al. 2003). Estimates for coastal areas, which are critical for studies on intertidal organisms, are, however, more difficult to make partly because of high levels of automated cloud detection in coastal areas (Böhm et al. 2003) and partly because even at the 1 km spatial resolution available (the size of a pixel in the satellite image), the remote sensing of small bays and convoluted coastlines is not possible (Đurović et al. 2018). Finer-scale satellite data focussed on specific coastal areas is, however, beginning to become available. Satellite data from Landsat 8 has recently been used to monitor eutrophication in Kotor Bay in Montenegro, which is one of the sampling sites in this study (Đurović et al. 2018). Existing environmental data, which can be patchy both spatially and

temporally, can be extended through the use of hydrodynamic modelling. Lipizer et al. (2014) used data collected from the Adriatic Sea over almost 100 years to model sea temperature, salinity and dissolved oxygen concentration at different depths at a basin-wide scale. As with the satellite-derived data, information in coastal areas was more difficult to estimate due to the high complexity of coastlines and variability of coastal currents.

One important development in seascape genetics is the use of biophysical models of larval dispersal (Cowen & Sponaugle 2009, Selkoe et al. 2016). Larval dispersal is the driving process behind connectivity for benthic species that have a pelagic larval stage. It depends on the coordination of biological processes such as reproduction, spawning, larval development and settlement, with seascape factors such as local sea currents and environmental variables (e.g. temperature, salinity and food availability) within the mass of water in which larvae are transported (Gosling 2003). As its name suggests, biophysical modelling of larval dispersal requires an interdisciplinary approach to incorporate biological processes with physical properties of the sea such as oceanographic circulation and fine-scale hydrodynamics. So far, modelling has been able to replicate or validate genetic parameters at broad spatial scales but problems still exist with respect to fine-scale hydrodynamic modelling and the incorporation of the local-scale events at either end of dispersal, that is, during events surrounding spawning and settlement (Selkoe et al. 2016). Limitations in some modelling efforts include that species-specific biological processes are not included in the model (processes such as larval swimming behaviour or variation in the duration of the pelagic stage due to environmental factors) and assumptions of homogenous release of larvae from all regions which does not allow for habitat patchiness. Nevertheless, recent modelling of larval dispersal has allowed meaningful connections to be made with genetic data. Two recent studies have modelled larval dispersal along the eastern coast of the Adriatic, the first covering the whole Mediterranean Sea (Dubois et al. 2016) and the second covering the Adriatic Sea (Bray et al. 2017). Although neither study

was species-specific, both were able to make useful predictions about larval dispersal routes and larval retention, highlighting both the general hydrodynamic conditions which support larval retention or transport, and, specific areas predicted to have increased or decreased levels of larval retention. The general hydrodynamic conditions that favour larval retention include complex shoreline configurations, weak currents and extended continental shelves while energetic nearshore currents impede larval retention. Specific predictions included low levels of larval retention south of the Po river delta where the southward-flowing Western Adriatic current originates and increased larval retention along the Croatian coastline with the convoluted coastline highlighted as an important factor impeding larval transport. This prediction has important implications for this study since the Croatian coast, with its numerous island chains and highly indented bays, covers more than half of the study range.

This work is part of a collaboration between Victoria University of Wellington and the University of Zagreb, Croatia. Faculty members from both universities were involved in the collection and initial processing of *Mytilus galloprovincialis* samples from four different countries along the eastern coast of the Adriatic Sea. Tissue samples preserved in 99% ethanol were brought back to New Zealand, while the valves remained in Croatia and will be used in a companion study on shell morphometry.

Research aims and thesis structure

The overall objective of this Master's thesis was to investigate the genetic structure of the endemic mussel, *Mytilus galloprovincialis*, along the entire eastern coast of the Adriatic Sea, with hypotheses formulated from oceanographic and seascape considerations.

Specific aims of the thesis were:

1. To test the null hypothesis of panmixia over 800 km of coastline against the alternate hypotheses of regional differentiation based on the 3 gyres of the Adriatic Sea and of isolation by distance (IBD) of a gradient population model with limited dispersal;

2. To test the null hypothesis that environmental variables are not correlated with genetic variation and thereby to identify markers and environmental variables that explain genetic patterns.

The thesis is divided into four chapters. This present chapter, Chapter 1, is a general introduction to the study species and to marine population genetics and seascape genetics, and contains an overview of the theoretical framework that underpins this research. Chapter 2 describes the population genetic structure of *M. galloprovincialis* in the study area and tests the hypotheses of regional genetic differentiation and isolation by distance. Chapter 3 takes a seascape genetics approach to identify associations between environmental variables and genetic structure, and Chapter 4 examines the key findings in a broader context and indicates possible future directions for research.

Chapter 2: Genetic structure of *Mytilus galloprovincialis* populations from the eastern coastline of the Adriatic Sea

Introduction

The Mediterranean blue mussel, *Mytilus galloprovincialis*, is a model organism with a large body of research investigating many different aspects of its biology. Productive areas of investigation include its role as an ecosystem engineer, its physiological adaptations to the rocky intertidal zone, aquaculture, its potential as an invasive species, and the dynamics of hybrid zones (Arribas et al. 2014, Braby & Somero 2006b, Kovačić et al. 2017, Gardner et al. 2016, Gardner & Skibinski 1988). There is, however, surprisingly little information about the population genetic structure of this species within its native range. This chapter adds to this limited information by investigating the population structure of *M. galloprovincialis* along the eastern coastline of the Adriatic Sea, part of its native range.

Patterns of genetic structure in the Mediterranean Sea

In the Mediterranean Sea, oceanographic currents, fronts, channels and straits such as the Almería-Oran Oceanographic Front, the Sicily Channel and the Otranto Strait, at the entrance of the Adriatic Sea, have been shown to mediate genetic differentiation in a variety of marine organisms, including *M. galloprovincialis,* which showed strong differentiation between populations in the Atlantic and Alboran Seas and populations in the Mediterranean Sea (Diz & Presa 2008, Pascual et al. 2017). Biophysical approaches, which model larval or small particle transport, have shown that the currents and gyres that divide the Adriatic Sea

into northern, central and southern regions may act as barriers to larval transport, although the patterns of particle transport vary considerably seasonally and between years (Liubartseva et al. 2016, Bray et al. 2017). This northern, central and southern regional pattern of genetic differentiation has been observed in a shore crab (*Carcinus aestuarii*) on the western Adriatic coast using microsatellites (Schiavina et al. 2014) but so far, evidence of regional differentiation for blue mussels has been mixed. Two studies covering the central and eastern Mediterranean Sea using microsatellites and mitochondrial sequence variation were consistent with regional differentiation in the Adriatic Sea but only two of the study populations were situated in the Adriatic Sea, one from the western coast, Ravenna, and the other from Zadar on the eastern coast (Giantsis et al. 2014a & b), while another study based on the Croatian coast, which coincides with the northern and central regions, found only limited genetic differentiation (Štambuk et al. 2013).

A pattern of isolation by distance, IBD, was observed in the sea star *Astropecten aranciacus*, from seven locations in the Mediterranean and two in the eastern Atlantic Ocean. There was strong differentiation between one of the Atlantic populations, on the Island of Madeira, and the other eight populations. Mantel tests for IBD were applied to all nine populations and to the eight non-differentiated populations and were significant in both cases, although the Mantel R was higher when only eight populations were tested (Zulliger et al. 2009). Similarly, for the bivalve, *Ostrea edulis*, significant IBD was seen over 15 sampling sites from Norway to the Black Sea and was still significant when the three easternmost populations, which were genetically differentiated from the other populations, were excluded from the Mantel test (Launey et al. 2002). In contrast, for *M. galloprovincialis*, IBD was not observed in 17 populations around the Iberian Peninsula (Diz & Presa 2008), 9 populations in the Aegean and Ionian seas (Giantsis et al. 2014a), or 10 populations on the Croatian coast in the Adriatic Sea (Štambuk et al. 2013).

Study Range

The Adriatic Sea is a semi-enclosed, elongated arm of the Mediterranean Sea lying between the Balkan and Apennine peninsulas. It is almost 800 km long on a northwest to southeast axis and about 72 km wide at its narrowest point. This point, the Strait of Otranto, is commonly held to be the boundary between the Ionian and Adriatic Seas, although the International Hydrodynamic Organization places the boundary further south, running from the mouth of the Butrint River, across Corfu to Cape Santa Maria di Leuca in Italy (International Hydrographic Organization 1953, UNEP/MAP-RAC/SPA 2015, Lipizer et al. 2014, Tudor & Janeković 2016, Böhm et al. 2003). The study range covers almost the entire eastern coastline of the Adriatic from the Istrian Peninsula in the north to Butrint Lagoon which connects to the Butrint River on the southern boundary, and spans four countries, Croatia, Bosnia-Herzegovina, Montenegro and Albania.

(a)   (b)
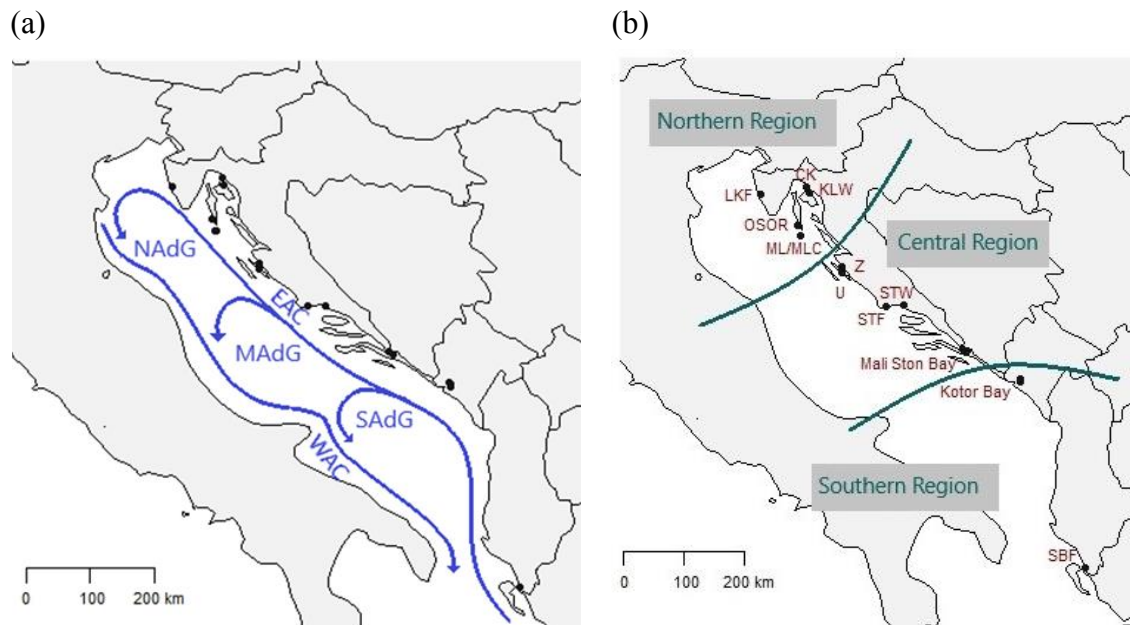
Figure 2.1 Maps of the Adriatic Sea showing (a) general currents: EAC Eastern Adriatic Current, WAC Western Adriatic Current, NAdG Northern Adriatic Gyre, MAdG Middle Adriatic Gyre, SAdG Southern Adriatic Gyre (adapted from Figure 1 Lipizer et al. 2014) and (b) the 3 regions used to test the hypothesis of genetic differentiation based on Adriatic Sea oceanography, and the locations of sampling sites. Abbreviations as per Table 2.1.

The Adriatic Sea is divided into 3 basins defined by bathymetry: the shallow northern basin, defined as the area bounded by the 100 m isobaths, the middle Adriatic basin which extends to the Palagruža Sill at 170 m and has a deepest point of 273 m in the Middle Adriatic Pit, and the deepest basin in the south which includes the 1,230 m deep South Adriatic Pit. From here the sea floor rises to the Otranto Sill at 780 m depth and then falls away again as the Adriatic Sea gives way to the Ionian Sea. The main circulation of the Adriatic Sea is cyclonic with the north-flowing Eastern Adriatic Current (EAC) originating in the Ionian Sea and the Western Adriatic Current (WAC) flowing southward from the mouth of the Po River. Three cyclonic gyres, which correspond to the northern, central and southern basins, operate with varying strength through the year and significant variation between years (Figure 2.1a) (Artegiani et al. 1997, Dubois et al. 2016, Bray et al. 2017, Zonn & Kostianoy 2017).

Materials and Methods

Sample Collection

In total, 843 individuals of *M. galloprovincialis* were collected from eighteen populations along the eastern coastline of the Adriatic Sea, a region spanning four countries (Croatia, Bosnia-Herzegovina, Montenegro, and Albania) and a variety of coastal environments (Figure 2.2 and Table 2.1). Seven of the populations were farmed and four of these were paired with a nearby wild population. The number of individuals sampled per site varied between 31 and 50. Samples of mantle tissue were preserved in 99% ethanol.
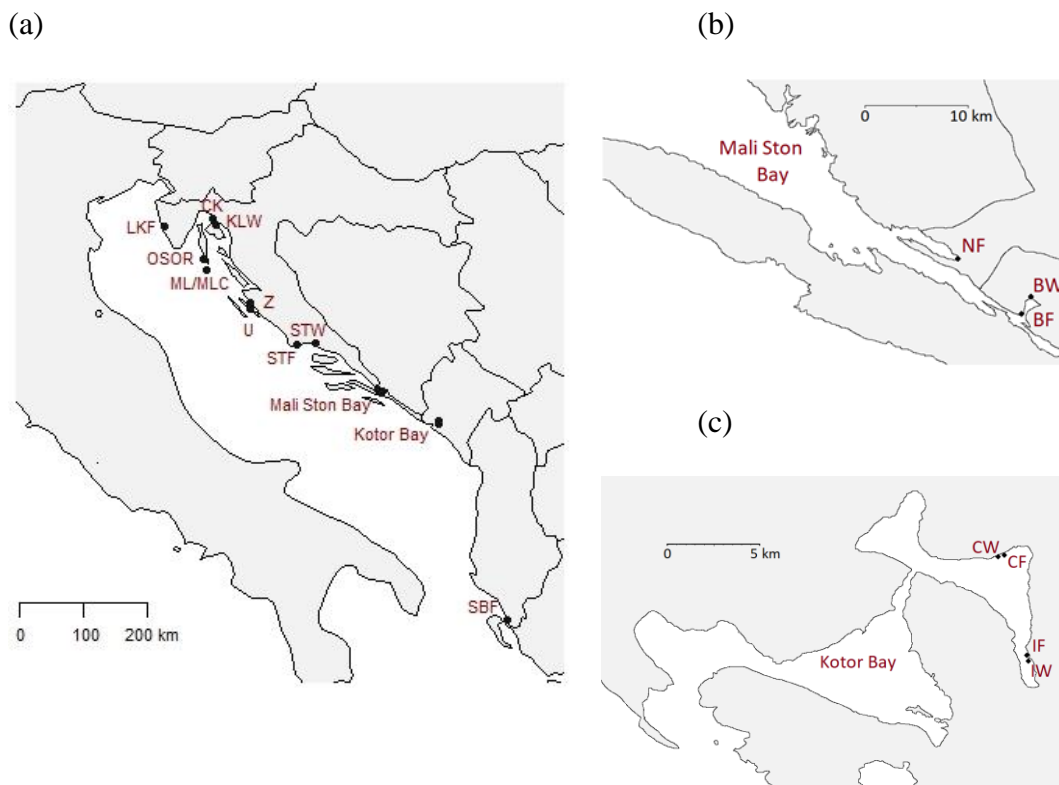
Figure 2.2 Maps of the Adriatic Sea showing (a) the sampling sites of *Mytilus galloprovincialis* collected from the eastern coast of the Adriatic Sea, (b) 3 sites in Mali Ston Bay, Bosnia & Herzegovina (NF) and Croatia (BF & BW) (c) 4 sites in Kotor Bay (CW, CF, IF, IW). Site abbreviations as per Table 2.1.

Table 2.1 *Mytilus galloprovincialis* sampling site description, sample size, collection date and geospatial coordinates.

Site code	Site description	Wild (W) Farmed (F)	Country	N	Collection date	Latitude	Longitude
LKF	Limski Channel	F	Croatia	31	09/2015	45.13334	13.66667
CK	Črišnjeva Krk bridge	W	Croatia	43	09/2015	45.24935	14.58243
KLW	Klimno, Soline Bay	W	Croatia	49	09/2015	45.15497	14.61889
OSOR	Cres-Lošinj bridge	W	Croatia	35	09/2015	44.69283	14.39231
ML	Mali Lošinj Harbour	W	Croatia	50	09/2015	44.53228	14.46835
MLC	Mali Lošinj Čikat	W	Croatia	50	09/2015	44.53181	14.45143
Z	Gaženica Harbour, Zadar	W	Croatia	50	09/2015	44.09786	15.25289
U	Ugljan-Pašman bridge	W	Croatia	50	09/2015	44.01455	15.25086
STF	Poljica, near Split	F	Croatia	48	09/2015	43.51543	16.13917
STW	Vranjic, Split	W	Croatia	48	09/2015	43.53180	16.46667
NF	Neum, Mali Ston Bay	F	Bosnia & Herzegovina	50	09/2015	42.91312	17.62381
BF	Bistrina, Mali Ston Bay	F	Croatia	50	09/2015	42.86880	17.69835
BW	Bistrina, Mali Ston Bay	W	Croatia	50	09/2015	42.87976	17.70917
CF	Bay of Kotor	F	Montenegro	50	09/2015	42.48539	18.74596
CW	Bay of Kotor	W	Montenegro	50	09/2015	42.48568	18.74349
IF	Bay of Kotor	F	Montenegro	50	09/2015	42.43705	18.76321
IW	Bay of Kotor	W	Montenegro	44	09/2015	42.43596	18.76367
SBF	Butrint Lagoon	F	Albania	45	05/2017	39.75474	20.03251
Total				843			

Laboratory Protocols

DNA had previously been extracted and amplified for eight microsatellite loci described in three published papers. One of these loci, Mgµ5 (Presa et al. 2002), exhibited extensive stuttering and multiple peaks and was dropped from the dataset. Details of the seven remaining loci are summarised in Table 2.2. PCR reactions for these loci had previously been optimised and the presence of the correct repeat motif confirmed by Westfall (2011). DNA extractions and amplifications were performed by three different operators. Total DNA was extracted using the Geneaid Genomic DNA Mini Kit following the manufacturer's instructions and DNA concentrations were quantified using a NanoDrop™ ND-1000 (Thermo Scientific). PCR cycling conditions followed Westfall (2011). Amplified fragments were scanned using an automated analyser ABI 3730XL by two different service providers (Macrogen, South Korea and Sangon Biotech, China).

Table 2.2 Details of microsatellite loci used in this study.

Locus	Repeat Motif	Size Range (bp)	Species used for development	Source	GenBank Accession no.	Reference
Mgµ1	$(TG)_n$	168 - 208	*M. galloprovincialis*	Not stated - probably northern Spain	AF445370	Presa et al. 2002
MGE3	$(GTG)_n$	450 - 456			AJ624208	
MGE4	$(CTA)_n$	168 - 180	*M. galloprovincialis*	EST sequences from GenBank	AJ624159	Yu & Li 2007
MGE5	$(TGA)_n$	220 - 244			AJ623869	
MGE7	$(CAA)_n$	242 - 260			AJ516476	
MT203	$(CA)_n$	161 - 197	*M. trossulus*	Baltic Sea	BV725482	Gardeström et al. 2008
MT282	$(GT)_n$	336 - 354			BV725484	

Note: The size range for each locus is the range found in the reference paper.

Genetic Analyses

The Adriatic Sea can be divided into three regions, northern, central and southern, as determined by the three gyres depicted in Figure 2.1a & b, which may form barriers to larval dispersal and therefore lead to genetic discontinuities. The 18 blue mussel populations along the eastern Adriatic coast were divided according to their location in these regions, giving 6, 7 and 5 populations in the northern, central and southern regions, respectively. The first null hypothesis to be tested was that there was no genetic differentiation amongst the regions defined by the three gyres with the alternate hypothesis being regional genetic differentiation. Blue mussels are almost continuously distributed along the eastern coast of the Adriatic Sea and may form a gradient population according to Wright's model of IBD (Wright 1943) where genetic differentiation increases with geographic distance because of mating between neighbouring individuals and finite larval dispersal. The second null hypothesis to be tested was that of panmixia (i.e. no genetic structure) within the study range with an alternate hypothesis of IBD.

Genetic data and construction of datasets

Geneious version 8.0.5 (http://www.geneious.com, Kearse et al. 2012) was used to score electropherograms with manual checking to standardise across genotyping providers and operators. The Microchecker application (Van Oosterhout *et al.* 2004) was used to check the microsatellite data for genotyping artefacts such as null alleles, large allele dropout and stuttering. The software Arlequin ver. 3.5.2.2 (Excoffier & Lischer 2010) was used to detect loci under selection using the Finite Island Model, 100 simulated demes and 20,000 coalescent simulations. Under neutral conditions loci exhibit a consistent relationship between

differentiation (commonly quantified by the F_{ST} statistic) and heterozygosity, a measure of diversity. Atypically high or low levels of differentiation can signify the presence of selection (Beaumont & Nichols 1996).

Because missing data can be problematic for some analyses, particularly the distance-based AMOVA, Mantel and spatial autocorrelation tests, several datasets have been used, detailed in Table 2.3. The first dataset, the 4 of 7 dataset, includes all 7 loci and consists of 837 individuals each having data for at least 4 loci. The second excludes the Mgµ1 locus, which has a large amount of missing data, and covers 6 loci for 835 individuals each again having data for at least 4 loci (the 4 of 6 dataset). A third complete dataset, with no missing data, of 577 individuals and 6 loci was used for analyses where missing data can have a large effect on the results. Finally, because the locus MGE7 was potentially under selection (see the outlier analysis in the Results section), analyses using the 4 of 6 dataset of 835 individuals were repeated using only 5 loci. This dataset, the 4 of 5 dataset, was essentially the same as the 4 of 6 dataset, with the same individuals included, but with the locus MGE7 excluded.

Table 2.3 Datasets used in this study, showing the loci included and the number and percentage of individuals scored.

Dataset Locus	LKF	CK	KLW	Osor	ML	MLC	Z	U	STF	STW	NF	BF	BW	CF	CW	IF	IW	SB	Total
4 of 7	31	43	48	34	49	50	49	50	48	48	50	50	50	50	50	50	44	43	837
																			100%
MGE3	31	43	48	34	49	50	49	50	47	47	49	49	49	48	49	46	40	41	826
	100%	100%	100%	100%	100%	100%	100%	100%	98%	98%	98%	98%	98%	96%	98%	92%	95%	98%	99%
MT282	29	34	43	28	38	46	36	42	45	43	45	44	46	34	39	42	37	41	712
	94%	79%	90%	82%	79%	92%	73%	84%	94%	90%	90%	88%	92%	69%	78%	84%	84%	95%	85%
MGE4	29	43	45	33	47	48	49	44	48	45	46	43	47	48	47	46	39	42	791
	94%	100%	94%	97%	96%	96%	100%	88%	100%	90%	92%	86%	94%	96%	94%	92%	91%	98%	95%
MGE5	31	43	48	34	49	50	47	49	47	47	48	49	48	42	48	49	43	42	813
	100%	100%	94%	97%	100%	100%	100%	100%	98%	98%	96%	98%	96%	84%	96%	98%	98%	98%	97%
MGE7	31	40	48	32	45	50	35	49	48	48	37	50	50	37	37	41	35	42	755
	100%	93%	100%	94%	92%	100%	71%	98%	100%	93%	74%	100%	100%	74%	74%	82%	80%	98%	90%
MT203	29	42	46	34	48	50	48	47	47	45	48	49	49	46	47	49	41	42	807
	94%	98%	96%	100%	98%	100%	98%	94%	98%	94%	96%	98%	98%	92%	94%	98%	93%	98%	96%
Mgu1	13	16	31	22	19	19	17	14	28	22	18	30	29	19	20	15	13	24	369
	42%	37%	65%	65%	39%	38%	35%	28%	58%	46%	36%	60%	58%	38%	40%	30%	30%	56%	44%
4 of 6	31	43	48	34	49	50	49	50	48	48	50	50	50	49	50	50	43	43	835
																			100%
MGE3	31	43	48	34	48	50	49	50	47	47	49	47	48	48	49	49	43	42	824
	100%	100%	100%	100%	98%	100%	100%	100%	98%	98%	98%	94%	96%	96%	98%	98%	98%	98%	99%
MT282	29	34	43	28	38	46	36	42	45	43	45	44	46	34	39	42	37	41	712
	94%	79%	90%	82%	78%	92%	73%	84%	94%	90%	90%	88%	92%	69%	78%	84%	84%	95%	85%
MGE7	31	43	48	34	49	50	47	49	47	47	48	49	48	41	48	49	42	42	811
	100%	100%	100%	100%	100%	100%	100%	100%	98%	98%	96%	98%	96%	84%	96%	98%	98%	98%	97%
MGE5	29	43	45	33	47	48	49	44	48	45	46	43	47	48	47	46	39	42	789
	94%	100%	94%	97%	96%	96%	100%	88%	100%	90%	92%	86%	94%	98%	94%	92%	91%	98%	94%
MGE4	31	40	48	32	45	50	35	49	48	48	37	50	50	37	37	41	35	42	755
	100%	93%	100%	94%	92%	100%	71%	98%	98%	93%	74%	100%	100%	76%	74%	82%	81%	98%	90%
MT203	29	42	46	34	48	50	48	47	47	45	48	49	49	46	47	49	41	42	807
	94%	98%	96%	100%	98%	100%	98%	94%	98%	94%	96%	98%	98%	92%	94%	98%	93%	98%	96%
6 of 6	25	31	39	25	33	44	26	33	42	37	26	35	41	21	25	30	25	39	577
	94%	98%	96%	100%	98%	100%	98%	94%	98%	94%	96%	98%	98%	94%	94%	98%	98%	98%	97%
All loci	25	31	39	25	33	44	26	33	42	37	26	35	41	21	25	30	25	39	577
	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

To assess whether sample sizes were sufficiently large to characterise the allelic variation of each population, the rarefaction procedure in the software package PopGenKit (Rioux Paquette 2011) in R 3.5.1 (R Core Team 2014) was used with 1000 jackknife replicates to estimate the total allelic diversity for each marker, each population and over all populations.

Analyses of Hardy-Weinberg equilibrium (HWE) and Linkage Disequilibrium (LD) were performed using Genepop 4.7.0 (Rousset 2008). Departures from HWE were assessed by two methods, the probability test and the U-test with the alternate hypothesis of heterozygote deficiency. Both tests employ a Markov Chain (MC) algorithm to estimate an unbiased exact p-value using 10,000 dememorization steps, 20 batches and 5,000 iterations per batch. LD was assessed using a probability test, with the default MC parameters of 10,000 dememorization steps, 100 batches and 5,000 iterations per batch. False discovery rate control for multiple statistical testing was applied to p-values (Verhoeven et al. 2005).

The effective population size, Ne, was estimated using the linkage disequilibrium method of Hill (1981), Waples (2006) and Waples & Do (2010) as implemented in NeEstimator V2.1 (Do et al. 2014). Rare alleles were excluded from the calculation when their frequency was less than the critical value, Pcrit, which was set above $1/(2S)$, as recommended in Waples & Do (2010).

Allele frequency estimation

The frequency of alleles, number of private alleles (PA), observed and expected heterozygosities (H_O and H_E) and the fixation index (F_{IS}) were calculated using GenAlEx 6.503 (Peakall & Smouse 2012). Allelic richness (A_R) and private allelic richness (PA_R) were calculated using the software HP-RARE 1.1 (Kalinowski 2005) which uses rarefaction to

compensate for different population sizes. These calculations were performed with a rarefaction sample size of 56 genes for the 4 of 6 dataset and 26 genes for the Mgμ1 locus alone since this locus has a large proportion of missing values and also a very large number of different alleles which unduly affects the comparisons made between other populations. Linear regression was performed between latitude and both F_{IS} and A_R, and Kruskal-Wallis tests were employed to test whether F_{IS} and A_R from farmed and wild populations were drawn from the same distribution.

To test the hypothesis that alleles are drawn from the same distribution in the eighteen populations exact *G*-tests for population differentiation were performed in Genepop version 4.7.0 for all populations and for all pairs of populations and using the default Markov chain parameters.

Since the frequencies of the most common alleles for two loci, MGE7 and Mgμ1, were higher/lower in the 5 southern populations than in the other regions, correlations of these allele frequencies with geospatial position were tested. Also, because any correlation could be driven by the allele frequencies in the southern region, the correlation was also tested for 6 subsets of the study range to check whether any correlation was consistent over the study range.

Differentiation between regions and populations

To test whether there were genetic differences between populations in the three regions determined by the three gyres, which divide the Adriatic Sea (Figure 2.1a & b), a hierarchical analysis of molecular variance (AMOVA, Excoffier et al. 1992) was performed using GenAlEx 6.503 (Peakall & Smouse 2012) for 3 regions and 18 populations with 999 permutations to determine the statistical significance of the calculated F_{RT} values for total and pairwise options. The *F* statistics were also calculated for individual loci to find the contribution of each locus

to the overall differentiation indices. The p-values were corrected for family-wise error using the False Discovery Rate formula. Distance-based analyses such as AMOVAs are susceptible to bias when there is substantial missing data, so analyses were performed on both the 4 of 7 dataset, which has substantial amounts of missing data due to the Mgµ1 locus and the 4 of 6 dataset which excludes this locus. Analyses were also repeated without the MGE7 locus since the outlier analysis indicated it might be under selection. Pairwise population F_{ST} values were also calculated to measure genetic differentiation between populations and for use in the Barrier and Mantel tests (see later section).

To test the hypothesis that genotypes are drawn from the same distribution in the three regions exact G-tests for population differentiation were performed in Genepop version 4.7.0 for all regions and for all pairs of regions and using the default Markov chain parameters.

Clustering Analyses

Three different cluster analysis programs, Structure (Pritchard et al. 2000), AWclust (Gao & Starmer 2008) and Discriminant Analysis of Principal Components (DAPC) (Jombart et al. 2010) were used to assess genetic structure over the eighteen populations. The three programs, Structure, AWclust and DAPC all attempt to group genetic data into clusters, but use different frameworks, methodologies and assumptions, which perform differently to uncover underlying genetic structure and which may be of particular use when the genetic structure is weak or nuanced. Structure uses Bayesian analysis carried out through Markov Chain Monte-Carlo algorithms, which assumes conformation to HWE and LD assumptions, while AWclust and DAPC are non-parametric analyses and therefore do not require conformity to HWE and LD assumptions. All analyses were performed using three datasets, the 4 of 7 dataset, the 4 of

6 dataset which excludes the locus Mgμ1, and the 4 of 5 dataset of neutral loci which also excludes locus MGE7.

Non-spatial Bayesian Structure analyses were performed with a burn-in length of 50,000 steps and run length of 100,000 steps, with 10 iterations for each K = 1 to 18, where K is the number of clusters assumed in the model. The runs used the correlated allele frequency model and sampling locations as priors, which can improve clustering for low levels of genetic differentiation but can also lead to overestimation of the number of clusters (Structure manual). Both the Admixture and No admixture models were used for comparison although the Admixture model which allows for mixed ancestry from the K clusters was considered a more accurate representation of the population dynamics of *Mytilus galloprovincialis* in the Adriatic Sea. The Evanno method (Evanno et al. 2005) as implemented in the program Structure Harvester (Earl & vonHoldt 2012) was used to detect the most appropriate number of clusters by calculating the greatest value for ΔK from the Structure results and linear correlations were performed between latitude and the proportion of cluster membership for each population.

The non-parametric program AWclust was designed for SNP analysis where each SNP is coded as 0 for no variant alleles, 1 for heterozygous and 2 for homozygous variants. It has been adapted for use with microsatellite data (Gruber et al. 2013, Wei et al. 2013a) by assigning a 'SNP' to each allele of each locus with 0 coded when the allele is not present in an individual, 1 when one copy of the allele is present and 2 when both copies of the allele is present. To determine the most appropriate number of clusters, K, the gap statistic was calculated for K = 1 to 8 using 100 null simulations. Analyses were completed on the 4 of 7, 4 of 6 and 4 of 5 datasets. AWclust assigns each individual of each population to a cluster. To see if cluster membership followed a latitudinal cline similar to MGE7 allele frequency, the proportion of

each population belonging to each cluster was calculated and linear correlations carried out between latitude and the proportion of cluster membership for each population.

Non-parametric DAPC analyses were performed using the DAPC web server using R package adegenet 3.5.1 (Jombart 2008, Jombart and Ahmed 2011) using 28 PCA eigenvalues. The DAPC procedure uses a combination of Principal Components Analysis (PCA) and Discriminant Analysis (DA). PCA is used first to transform genetic data into a format suitable for DA, which partitions total variance into between-group and within-group components and allows between-group variance to be maximized while within-group variance is minimized. The best-supported number of clusters is found using the Bayesian Information criterion which rates models using the likelihood function, but also adds a penalty term for each additional parameter used in a model to avoid overfitting.

Relationship between genetic and geographic distance

Mantel tests and spatial autocorrelation analyses were employed as implemented in GenAlEx 6.503 (Peakall & Smouse 2012) to examine the relationship between genetic and geographic distance. After initial analyses using the 4 of 6 dataset showed inconsistent results depending on whether the "Interpolate missing values" option was used, all tests were performed on the 6 of 6 dataset with no missing values. The genetic measures used were F_{ST} and Linearised F_{ST} (LinF_{ST}) calculated by AMOVA in GenAlEx 6.503. The geographic measures used were CD and $\log(1 + CD)$, where CD is the coastal distance, the shortest distance by sea between each population pair as measured in Google Maps. CD is a more appropriate measure for geographic distance than the default geographic distance calculated in GenAlEx 6.503 from geospatial coordinates, GGD, since it is more closely related to the distance larvae disperse. Because of the highly indented nature of the eastern Adriatic coast,

CD is more than twice GGD for some population pairs. For long approximately one-dimensional habitats such as coastlines, modelling has shown that an approximately linear relationship between $\text{Lin}F_{ST}$ and CD exists, but in two-dimensional habitats, the linear relationship is between $\text{Lin}F_{ST}$ and $\log(1 + \text{CD})$. When the distance between populations is less than half the width of the habitat, as occurs in the northern part of the sampling range, the 2-dimensional model applies (Rousset 1997). The Mantel tests were performed with 9999 permutations to determine the statistical significance of the correlation coefficient.

Spatial autocorrelation correlation coefficients, r, were calculated to gauge the genetic similarity/dissimilarity between individuals whose geographical separation falls within specified distance classes. In total, 9999 permutations were performed to define the 95% confidence interval of the null hypothesis for each distance class and overall correlogram significance. Bootstrap trials (10,000) were completed to calculate the 95% confidence intervals for the observed r in each distance class. Distance classes were chosen to give the most even number of pairwise comparisons per class. The Mgµ1 locus was excluded from analyses because of the high level of missing alleles and analyses were performed both with and without the option of interpolating missing value. The 6 of 6 dataset with no missing data was also used to find the contribution of the MGE7 locus to spatial autocorrelation and to gauge the extent over which positive spatial autocorrelation operates.

Results

Performance of microsatellite loci and construction of datasets

Six of seven microsatellite loci amplified were able to be scored satisfactorily with 85.1-98.7% of individuals scored. Locus Mgμ1 showed stuttering and multiple peaks for a proportion of samples, with 44.1% of individuals able to be reliably scored overall and the percentage scored per population ranging between 28.0% and 64.7%. There was no relationship between percentage scored per population and population latitude ($R^2 = 0.0004$, $p = 0.94$). The percentage of each locus scored per population was broadly consistent across operators and the two service providers.

Missing data can be problematic for some analyses, so several datasets were constructed as described in the methods section and detailed in Table 2.3. The third dataset of 577 individuals and 6 loci with no missing data was used for some analyses where missing data appeared to have a large effect on the results. This dataset has the disadvantage of having small sample numbers for some populations and a corresponding reduction in the number of alleles represented, but it has the advantage of being complete (balanced). In this third dataset, seven populations were reduced to less than 30 individuals, the smallest being 21 at CF in Kotor Bay which lost representation of 25% of alleles present in the 4 of 6 dataset.

Microchecker identified an excess of homozygote genotypes across all eighteen populations for five of the seven loci (MT282, MGE4, MGE7, MT203 and Mgμ1). For the remaining two loci, MGE3 and MGE5, only 1 and 3 populations, respectively, had excess homozygotes. For the loci MGE7, MGE4 and MT282, 10, 13 and 16 populations, respectively, showed a deficiency of heterozygote genotypes with alleles of one repeat unit difference an indicator for scoring error due to stuttering. The traces for these loci did not show stuttering so

the deficiency was put down to the existence of one very common allele together with the homozygote excess. There was no evidence for large allele dropout.

Outlier analysis using the 4 of 7 dataset did not detect any significant deviations in the F_{ST} *vs.* heterozygosity distribution although the p-value for the locus MGE7, which had a high observed F_{ST} value (Figure 2.3), was 0.057. When the tests were repeated without the Mgμ1 locus using the 4 of 6 dataset, the p-value for MGE7 was 0.040, suggesting that this locus may be under selection, or close to a gene under selection. Since many of the analyses undertaken have an assumption that loci are not under selection, most analyses have been repeated using a dataset excluding the MGE7 locus (the 4 of 5 dataset consisting of 835 individuals).
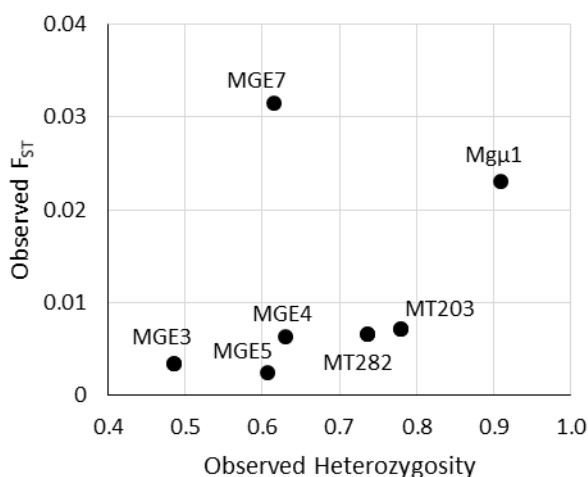
Figure 2.3 Outlier Analysis to detect loci under selection.

For five loci, rarefaction curves showed that at least 80% of allelic variation in each population was represented in the samples (see Supplementary Figure 1). For the locus MT203, which had 24 different alleles, 12 of the 18 populations were estimated to have 80% of their allelic variation represented in the samples, the remaining 6 having 70-80% of the variation.

Only one population appeared to make the 80% threshold of allelic representation for the Mgµ1 locus (48 different alleles) with samples for two populations exhibiting approximately 75% representation. The rarefaction curves for the remaining 15 populations were still increasing at the actual sample size.

All eighteen populations showed significant departures from HWE expectations with five individual loci being out of HWE at every population. Loci MGE3 and MGE5 met expectations at 15 and 11 populations, respectively. There was no evidence of linkage disequilibrium for any pair of loci, either at population level or across all populations.

Estimates of the effective population size, N_e using the 4 of 6 dataset and the 4 of 5 dataset which excludes the locus potentially under selection, showed that the effective sizes of populations in the study range were large (Table 2.4). The upper limit of the 95% confidence interval was "infinite" for 16 of the 18 populations using the 4 of 5 dataset and for 15 of the 18 populations using the 4 of 6 dataset, where an "infinite" population size means that there is no evidence of genetic variation caused by genetic drift due to a finite number of parents. Only one population, IW, had a finite upper limit for the 95% confidence interval for both datasets.

Table 2.4 Estimates of effective population size, N_e, (NeEstimator) with the 95% confidence intervals. Pcrit is the critical value for allele frequencies below which rare alleles were excluded from the calculation.

| Population | Pcrit | 4 of 6 dataset including MGE7 | | | 4 of 5 dataset excluding MGE7 | | |
		Ne	95% Confidence Interval Lower	Upper	Ne	95% Confidence Interval Lower	Upper
LKF	0.02	46.2	20.4	618.3	42.9	17.3	∞
CK	0.02	∞	386.8	∞	∞	202.2	∞
KLW	0.02	∞	444.5	∞	∞	195.5	∞
OSOR	0.02	∞	190.1	∞	∞	675.4	∞
ML	0.02	120.6	49.0	3465.5	394.7	65.1	∞
MLC	0.02	∞	∞	∞	∞	∞	∞
Z	0.02	108.8	44.8	∞	82.3	36.9	1301.4
U	0.02	148.5	50.2	∞	189.0	49.3	∞
STF	0.02	∞	115.9	∞	∞	125.7	∞
STW	0.02	210.1	59.3	∞	224.0	52.2	∞
NF	0.02	∞	220.6	∞	∞	314.4	∞
BF	0.02	∞	360.4	∞	∞	138.2	∞
BW	0.02	157.9	60.9	∞	237.8	62.1	∞
CF	0.025	88.8	35.8	∞	384.8	53.9	∞
CW	0.02	346.4	69.4	∞	198.9	55.5	∞
IF	0.02	∞	252.6	∞	∞	150.7	∞
IW	0.02	76.4	35.6	774.6	50.0	26.2	161.1
SBF	0.02	∞	108.4	∞	∞	∞	∞
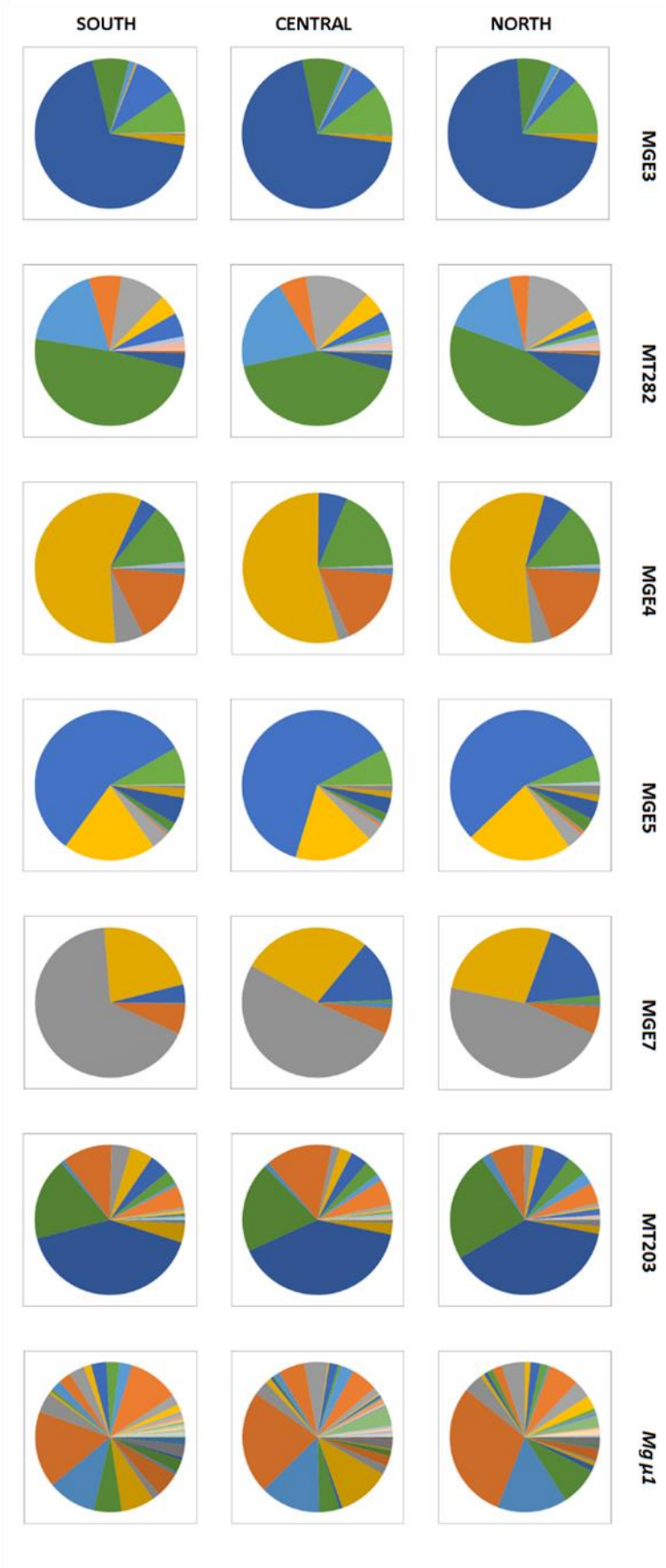
Figure 2.4 Allelic frequencies for 7 loci for each of the three regions.

Figure 2.5 Allelic frequencies for MGE7 for each population, each region and across all data.

Population and regional allelic differentiation

Allele frequencies by population are listed in Supplementary Table 2 and the regional frequencies are illustrated in Figure 2.4. The probability that alleles from each population were drawn from the same distribution, calculated by exact G-tests in the genic differentiation option in Genepop, was less than 1×10^{10} across all loci, and less than 1×10^{5} for individual loci MT282, MGE4, MGE7, MT203 and Mgµ1. The probability that alleles of the MGE5 locus were drawn from the same distribution was much higher at 0.014 and for MGE3 the probability was not significant at 0.175. The number of significant pairwise population exact G-tests for individual loci is shown in Table 2.5. The MGE7 locus, which has only 6 alleles, had 103 significant comparisons out of 153, and the Mgµ1 locus, with 48 alleles, had 109 significant comparisons. Together, these two loci accounted for just over half of all significant comparisons. The ratio of number of significant tests to the number of alleles ranged from 1.2 − 4.9 for the loci not under selection, but was 17.2 for MGE7. The frequencies of MGE7 alleles by population and by region are shown in Figure 2.5.

Table 2.5 Exact G-tests for population differentiation showing the number of alleles for each locus and the number of significant tests for each pairwise population comparison.

Locus	Number of alleles	Number of significant tests (out of 153 pairwise population comparisons)
MGE3	13	16
MT282	16	78
MGE4	8	38
MGE5	15	28
MGE7	6	103
MT203	24	47
Mgμ1	48	109

The most common alleles for loci MGE7 and Mgμ1 were markedly different between regions and closer examination revealed a significant linear relationship with latitude for the most common allele for both loci, which was not the case for any other locus. The most common allele of the MGE7 locus, sized 243 bp, MGE7[243], showed an increasing southward trend with regional frequencies of 46.5%, 51.4% and 66.9% for the northern, central and southern regions (Figure 2.6a). The population frequencies ranged from 39.1-54.4% for populations in the North, 45.8-59.5% in the Central region and 54.9-77.0% in the South. Linear regression analysis of the frequency of the most common allele at each population against latitude showed a significant negative relationship ($R^2 = 0.476$, $p = 0.0015$, and gradient, $m = -0.054$) (Figure 2.6b). Correlation tests on six subsets of the study range (North & Central, North & South and Central & South, North, Central, South) indicated that the north-south cline was fairly consistent across the study range with gradients of -0.034 (North & Central), -0.056 (North & South), -0.020 (Central & South), -0.029 (North), -0.026 (Central), -0.019 (South)

(Figure 2.6c). However, only the correlation for the North & South subset (with 11 populations) was statistically significant ($R^2 = 0.551$, p = 0.009).

The most common allele for locus Mgμ1 was 160 bp and occurred at regional frequencies 30.0%, 22.2% and 16.5% in the northern, central and southern regions respectively and the ranges in each region were 13.6-41.9%, 5.6-31.8% and 8.3-23.3%. There was a significant positive relationship between population frequency and latitude for this allele ($R^2 = 0.392$, p = 0.0055, $m = 0.044$). Because of the large amount of missing data at this locus, further analysis was not performed.

(a)

(b)

(c)

(d)
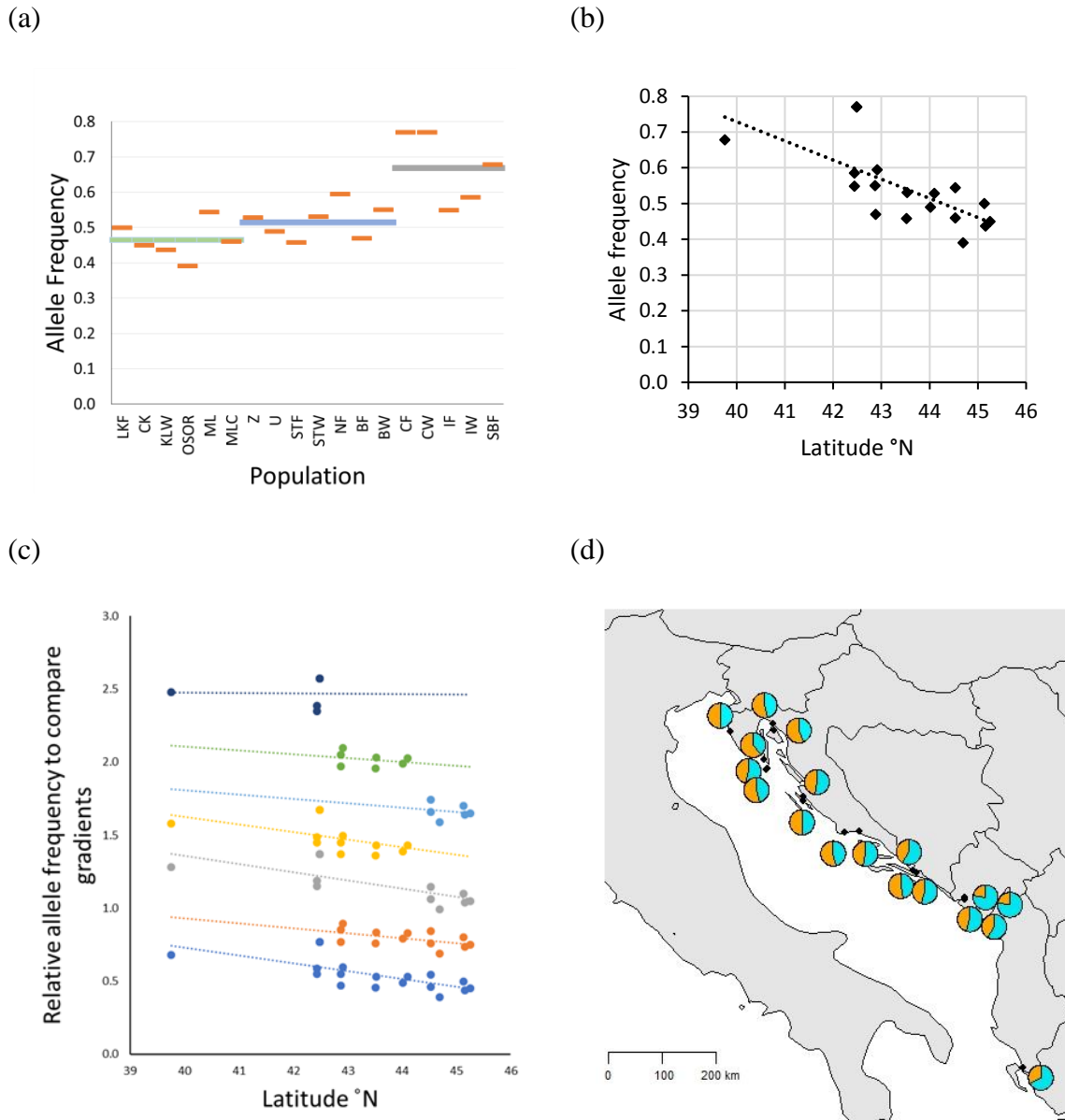
Figure 2.6 Frequency of the most common allele for MGE7 (243 bp) (a) by individual population (—), and by region (— northern region, — central region, — southern region); linear regression analysis of frequency of 243 bp allele as a function of latitude for, (b) the whole study range ($R^2 = 0.476$, p = 0.0015), (c) different subsections of the study range to check the consistency of the latitudinal cline, from bottom to top: mid blue – all 18 populations, $m = -0.054$; orange - northern and central regions, $m = -0.034$; grey – northern and southern regions, $m = -0.056$; yellow – central and southern regions, $m = -0.020$; light blue – northern region, $m = -0.029$; green – central region, $m = -0.026$; dark blue – southern region, $m = -0.019$; (d) map of sampled populations showing the proportion of the MGE7 allele size 243 bp (cyan) and all other alleles (orange).

Average allelic richness, A_R, excluding the Mgµ1 locus, ranged from 6.16 in NF to 7.90 in CK, and private allelic richness, PA_R, ranged from 0.00 in SBF to 0.40 in STF. In total, there were 28 private alleles, 13 of which were associated with the locus Mgµ1. Observed heterozygosity, H_O, was lower than the expected heterozygosity, H_E, at every population, with overall observed and expected heterozygosity across all populations and all loci of $H_O = 0.383$ and $H_E = 0.663$. This is reflected in the high inbreeding coefficient, F_{IS}, for which values ranged from 0.336 to 0.460 across all populations and was 0.399 overall (Table 2.6). The individual F_{IS} for each locus averaged over all populations varied considerably: 0.052, 0.115, 0.365, 0.471, 0.585, 0.599 and 0.602 for MGE3, MGE5, MT203, MGE4, MT282, MGE7 and Mgµ1 respectively. There was no relationship between A_R (calculated without the Mgµ1 locus) or F_{IS} and latitude ($R^2 = 0.030$, p = 0.49 and $R^2 = 0.154$, p = 0.11, respectively). Neither were there differences in A_R or F_{IS} between farmed and wild populations (Kruskal-Wallis H statistic = 0.740, p = 0.39 and H statistic = 1.386, p = 0.24 for A_R and F_{IS} respectively).

Table 2.6 Measures of genetic diversity for 18 populations of *Mytilus galloprovincialis* for 7 loci: A_R allelic richness, PA_R private allelic richness (excluding and including the Mgμ1 locus (-Mgμ1 and + Mgμ1 respectively), P_a number of private alleles, H_O observed heterozygosity, H_E expected heterozygosity, F_{IS} inbreeding coefficient.

Population	A_R (-Mgμ1)	PA_R (-Mgμ1)	A_R (+Mgμ1)	PA_R (+Mgμ1)	P_a	H_O	H_E	F_{IS}
LKF	6.56	0.12	11.00	0.17	0	0.408	0.656	0.349
CK	7.90	0.05	11.09	0.03	0	0.410	0.685	0.386
KLW	6.95	0.10	10.92	1.14	2	0.364	0.652	0.419
OSOR	7.06	0.01	11.24	0.00	0	0.401	0.691	0.374
ML	6.76	0.01	9.65	0.75	1	0.392	0.639	0.336
MLC	7.35	0.11	8.74	0.30	1	0.411	0.683	0.381
Z	6.76	0.02	11.94	1.01	1	0.369	0.671	0.419
U	6.51	0.22	11.63	2.79	5	0.387	0.663	0.381
STF	7.52	0.40	11.36	0.90	3	0.382	0.692	0.430
STW	7.05	0.12	9.34	0.00	1	0.387	0.663	0.390
NF	6.16	0.10	15.30	1.45	3	0.364	0.609	0.393
BF	6.94	0.11	13.36	0.69	2	0.392	0.688	0.403
BW	6.89	0.10	11.54	0.16	1	0.423	0.686	0.353
CF	6.60	0.01	13.27	1.24	1	0.354	0.617	0.426
CW	6.75	0.33	11.83	0.37	2	0.336	0.623	0.426
IF	6.98	0.15	9.44	0.00	1	0.347	0.669	0.456
IW	7.12	0.15	15.00	3.00	3	0.350	0.668	0.460
SBF	7.03	0.00	14.66	0.85	1	0.411	0.682	0.395

Results of the hierarchical AMOVA conducted on the 4 of 7 and 4 of 6 datasets were very similar; the 4 of 7 results are reported here (Tables 2.7 & 2.8). The regional differentiation index, which is the ratio of variation explained between regions to total variation, was small but significant ($F_{RT} = 0.0029$, $p < 0.001$). When only neutral loci were analysed using the 4 of 5 dataset, F_{RT} was no longer significant ($F_{RT} = 0.0007$, $p = 0.085$), while the F_{ST} index, which in the hierarchical AMOVA is an index of variation explained between regions and populations

to total variation, was significant, $F_{ST} = 0.002$, p = 0.024, suggesting that the smaller spatial scale between populations is more relevant to the processes through which genetic differentiation occurs compared to the larger distances between regions.

Table 2.7 Hierarchical AMOVA of 7 loci for 3 regions and 18 populations: df, degrees of freedom; SS, sum of squares; Est Var, estimated variation and % Var, Percentage variation.

Source of variation	7 loci				5 neutral loci			
	df	SS	Est Var	% Var	df	SS	Est Var	% Var
Among Regions	2	17.7	0.007	0.29%	2	6.3	0.001	0.07%
Among populations	15	76.3	0.017	0.72%	15	37.8	0.002	0.12%
Among Individuals	819	2872.9	1.164	49.2%	817	1850.4	0.665	39.69%
Within Individuals	837	987.5	1.180	49.8%	835	815.5	1.007	60.12%
			2.368	100.0%			1.675	100.00%

Table 2.8 Differentiation Indices from Hierarchical AMOVA.

Differentiation Index	7 loci		Neutral loci	
	Index	p-value	Index	p-value
F_{RT}	0.0029	<0.001	0.0007	0.085
F_{SR}	0.0072	<0.001	0.001	0.108
F_{ST}	0.010	<0.001	0.002	0.024
F_{IS}	0.497	<0.001	0.398	<0.001
F_{IT}	0.502	<0.001	0.398	<0.001

The southern region showed the greatest overall differentiation with pairwise F_{RT} of 0.008 and 0.004 (p $\leq$ 0.0001 for both) for comparisons with the North and Central regions, respectively (Table 2.9). The F_{RT} for the North-Central comparison was much lower, but still significant ($F_{RT} = 0.001$, p $= 0.012$). These results were supported by inter-region comparisons of pairwise population F_{ST}-values, with the mean F_{ST} for North-Central comparisons, 0.006, being similar to mean within-region comparisons of 0.005, 0.008 and 0.007 for North, Central, and South, respectively, and the North-South and Central-South means being larger at 0.011 and 0.009, respectively (see Table 2.10 for pairwise population F_{ST}-values). This pattern changed when only neutral loci were used: only the North-Central F_{RT} value was significant indicating that the large North-South and Central-South values were due to MGE7 and also showing that the subtle structure detected between the northern and central regions was mostly due to neutral loci. The pairwise F_{RT} values for individual loci were statistically significant (p ≤ 0.05) after FDR correction for the MGE7 locus between the northern and southern regions ($F_{RT} = 0.0393$, p ≤ 0.0001) and between the central and southern regions ($F_{RT} = 0.0195$, p $\leq$ 0.0001).

Table 2.9 Pairwise F_{RT} for comparisons between the three regions are below the diagonal with p-values following 9999 permutations above the diagonal. Results for six loci are on the left and for the five neutral markers on the right.

6 loci

Region	North	Central	South
North		0.012	≤ 0.0001
Central	0.001		≤ 0.0001
South	0.008	0.004	

5 loci

Region	North	Central	South
North		0.013	0.146
Central	0.002		0.376
South	0.001	0.000	

The two populations Zadar (Z) and Ugljan-Pašman (U) were included in the central region, but because of the position of several large seaward islands, larval dispersal could primarily be into the northern region. Under the hypothesis of regional differentiation, F_{RT} values would be expected to increase if populations close to the boundary between regions were correctly assigned. Repeat AMOVA analyses with these two populations in the northern region were almost unchanged supporting the low pairwise differentiation found between the northern and central regions.

Table 2.10 Pairwise F_{ST} calculated by AMOVA for 18 populations of *Mytilus galloprovincialis* based on 6 loci below the diagonal and on 5 neutral loci above. F_{ST} values with significant p-values (after FDR correction) in bold.

	LKF	CK	KLW	OSOR	ML	MLC	Z	U	STF	STW	NF	BF	BW	CF	CW	IF	IW	SBF	Mean F_{ST}
LKF		0.001	0.000	0.004	**0.015**	0.000	0.008	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.003	0.000	0.000	0.000	0.002
CK	0.010		0.000	0.000	**0.009**	0.002	**0.010**	0.007	0.001	0.000	0.004	0.000	0.001	0.000	0.000	0.000	0.000	0.001	0.002
KLW	**0.011**	0.003		0.000	0.006	0.000	**0.010**	**0.014**	0.002	0.000	0.007	0.002	0.000	0.001	0.000	0.000	0.006	0.000	0.003
OSOR	**0.015**	0.000	0.000		0.000	0.000	0.004	0.004	0.000	0.000	0.008	0.004	0.000	0.000	0.000	0.000	0.000	0.002	0.001
ML	**0.016**	0.007	**0.010**	0.000		**0.009**	0.005	**0.012**	**0.009**	0.000	**0.014**	**0.016**	0.008	0.001	0.007	**0.010**	**0.009**	**0.014**	0.008
MLC	0.000	0.001	0.000	0.000	0.007		0.004	0.003	0.000	0.000	0.005	0.000	0.000	0.004	0.004	0.000	0.000	0.000	0.002
Z	**0.029**	**0.012**	**0.019**	0.008	**0.011**	**0.015**		0.007	**0.014**	**0.011**	**0.017**	**0.009**	0.002	0.001	**0.011**	0.005	**0.010**	**0.014**	0.008
U	0.006	0.003	**0.011**	0.002	**0.009**	0.003	**0.013**		0.004	**0.009**	0.000	0.002	0.002	0.007	0.000	0.000	0.003	**0.009**	0.004
STF	0.000	0.001	0.005	0.002	0.007	0.000	**0.025**	0.004		0.000	0.000	0.002	0.000	0.002	0.000	0.001	**0.010**	0.000	0.002
STW	0.006	0.000	0.000	0.000	0.005	0.000	**0.017**	0.000	0.003		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
NF	**0.019**	0.006	**0.016**	**0.011**	**0.016**	**0.015**	**0.011**	0.007	**0.014**	**0.009**		0.002	0.003	0.000	0.000	0.001	0.000	0.007	0.004
BF	0.000	0.000	0.005	0.005	**0.013**	0.000	**0.020**	0.001	0.000	0.000	**0.013**		0.000	0.000	0.001	0.000	0.000	0.000	0.002
BW	0.000	0.001	0.000	0.003	0.006	0.000	**0.015**	0.001	0.000	0.000	**0.012**	0.000		0.000	0.000	0.001	0.000	0.000	0.001
CF	**0.028**	**0.011**	**0.030**	**0.017**	**0.009**	**0.025**	**0.011**	**0.022**	**0.023**	**0.014**	**0.013**	**0.017**	**0.016**		0.000	0.000	0.000	0.002	0.001
CW	**0.023**	0.004	**0.026**	**0.013**	**0.011**	**0.019**	**0.015**	**0.012**	**0.014**	0.007	0.002	**0.014**	**0.011**	0.000		0.000	0.002	0.003	0.002
IF	**0.016**	0.000	0.007	0.000	**0.012**	0.004	0.004	0.000	**0.010**	0.001	0.000	0.003	0.004	**0.009**	0.000		0.000	0.003	0.001
IW	**0.014**	0.000	**0.009**	0.001	**0.009**	0.005	0.006	0.005	0.008	0.000	0.000	0.006	0.005	**0.009**	0.002	0.000		0.005	0.002
SBF	0.008	0.005	0.008	**0.010**	**0.012**	0.001	**0.027**	**0.013**	0.003	0.000	**0.019**	0.001	0.000	**0.015**	**0.009**	**0.011**	**0.010**		0.003
Mean F_{ST}	0.011	0.004	0.009	0.005	0.009	0.005	0.014	0.006	0.006	0.003	0.01	0.005	0.004	0.015	0.01	0.005	0.005	0.008	0.003

Pairwise population F_{ST} values for the 4 of 6 and 4 of 5 datasets were patchy with 64 of 153 comparisons significant after FDR correction for the 4 of 6 dataset and only 1 for the neutral dataset (Table 2.10). Two populations stood out with unusually high F_{ST} values for the neutral dataset, ML and Z, both with mean F_{ST} of 0.008 compared to mean values of 0.001 – 0.004 for all other population pairwise comparisons.

The probability that genotypes from each region were drawn from the same distribution was less than 1.1×10^{-6} for the 4 of 7 dataset. Pairwise regional differences were statistically significant for all three comparisons, North vs. Central, North vs. South and Central vs. South, with the latter two comparisons more significant statistically (Table 2.11). When this test was repeated on the 4 of 6 dataset, without Mgμ1, the overall probability increased slightly to 2.2×10^{-6} over all loci and pairwise regional differences were no longer significant for the North vs. Central comparison. When this test was repeated on the 5 neutral loci only, the overall p-value was 0.056 and only the North vs. South pairwise comparison was significant.

Table 2.11 Exact G-tests for regional differentiation showing the probabilities that genotypes are drawn from the same distribution.

Dataset	Loci omitted	Global Differentiation	Pairwise Regional Differentiation		
			North vs. Central	North vs. South	Central vs. South
4 of 7		1.1×10^{-6}	0.027	0.010	0.012
4 of 6	Mgμ1	2.2×10^{-6}	0.119	0.044	< 0.001
4 of 5	Mgμ1 MGE7	0.056	0.113	0.008	0.472

Clustering Analyses

The Structure runs using the Admixture model had a maximum likelihood statistic, ΔK, at K = 5, with a local maximum at K = 2 for both the 6 and 7 loci datasets. When the No Admixture model was used, the maximum likelihood was attained at K = 2 (Supplementary Figure 2a & b). The Structure runs for K = 5 did not show any obvious genetic structure but for K = 2, using both models, the bar plot suggested a north-south cline in the proportion of ancestry from each cluster, something that was confirmed by linear correlation tests of proportion of cluster membership as a function of latitude, $R^2 = 0.459$, p = 0.002 and gradient, *m* = -0.058 for the 4 of 6 dataset (Figure 2.7a and Supplementary Figure 2c, d & e). Structure runs on the neutral dataset did not show this north-south cline ($R^2 = 0.111$, p = 0.178 and *m* = -0.0143, Figure 2.7a) and there was no evidence of any genetic differentiation (K = 1). Figure 2.7b is a map showing the proportion of cluster membership for each population.

(a)   (b)
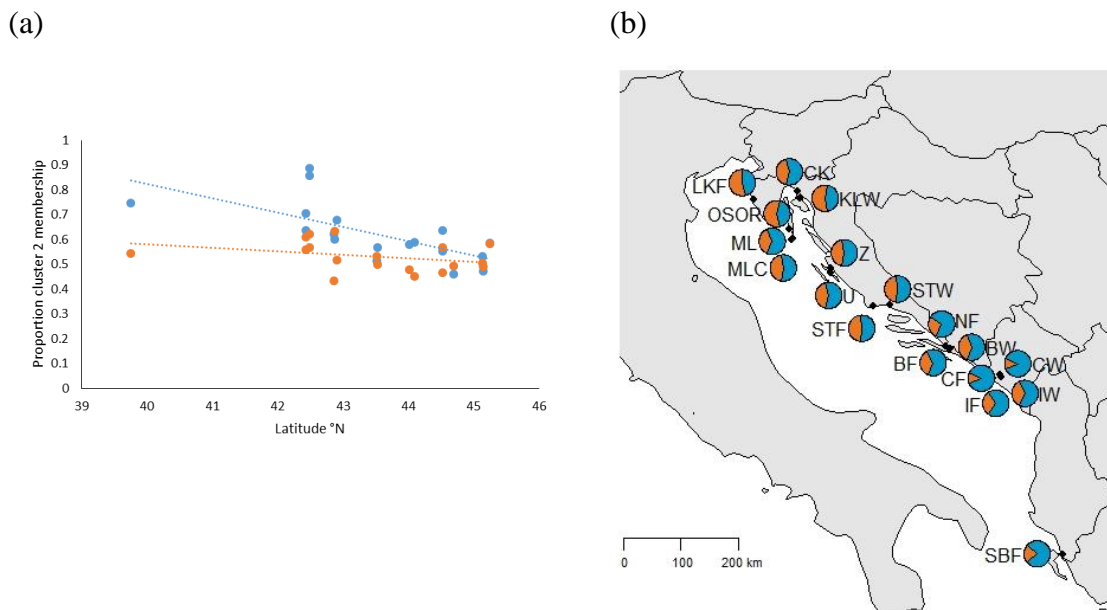
Figure 2.7 (a) Linear regression analysis of proportion of cluster 2 membership from Structure analysis as a function of latitude for 6 loci (blue), $R^2 = 0.459$, p = 0.002 and 5 neutral loci (orange), $R^2 = 0.111$, p = 0.178; (b) map of proportion of cluster membership by population for 6 loci.

The 2- and 3-dimensional scaling plots of the allele sharing distance from the non-parametric AWclust analyses of the 4 of 6 loci dataset showed loose clustering of individuals into two or three observable groups (Figure 2.8b). The gap statistic, which indicates the optimal number of clusters, had local maxima at K = 2 and 4 and an absolute maximum at K = 7 with the confidence intervals overlapping to a large degree, indicating only very weakly defined structure (Figure 2.8a). For K = 2, the correlation between cluster 2 membership and latitude was significant ($R^2 = 0.308$, p = 0.017) but weaker than the corresponding correlation from the Structure program (Figure 2.8c). Results for the 4 of 7 dataset were similar but for the 5 neutral loci, only one cluster was evident (Figure 2.8d) and when K = 2, the correlation between cluster 2 membership and latitude was not significant ($R^2 = 0.073$, p = 0.279).

(a)

(b)

(c)

(d)
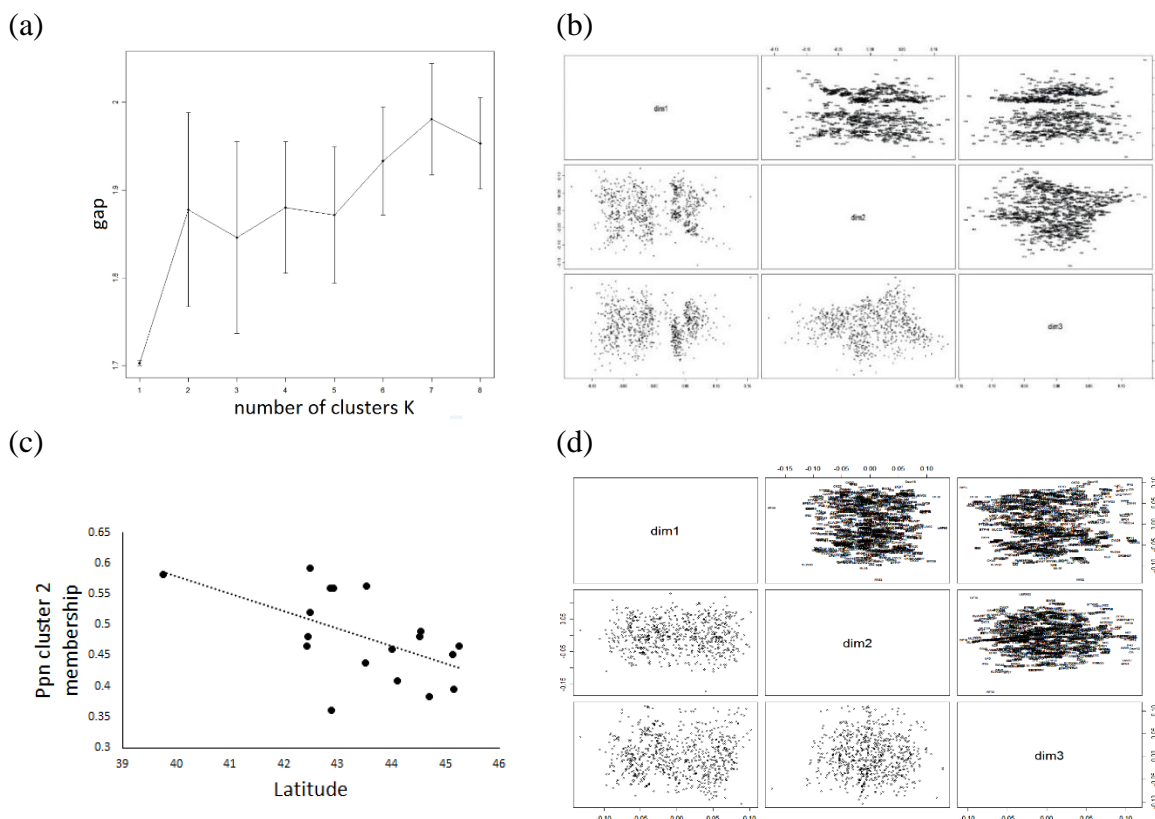
Figure 2.8 AWclust results for *Mytilus galloprovincialis* using the 4 of 6 dataset: (a) the gap statistic showing the optimal number of clusters, (b) 2-D MDS cluster plot of the allele sharing distance, (c) Linear regression analysis of proportion of cluster 2 membership as a function of latitude, $R^2 = 0.308$, p = 0.017, and using 5 neutral loci: (d) 2-D MDS cluster plot of the allele sharing distance.

Contrasting with the AWclust results, the DAPC scatterplot showed only one cluster of individuals and no differentiation between populations for all datasets used (Figure 2.9).
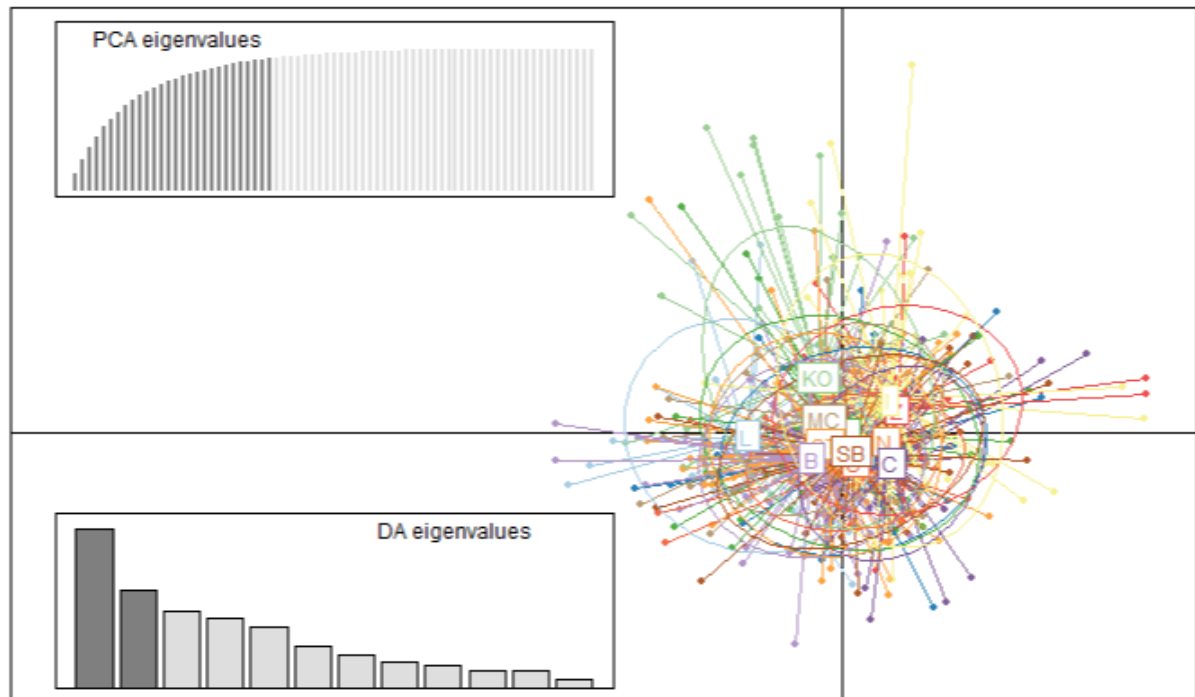
Figure 2.9 DAPC scatterplot of mussel populations. Coloured ellipses represent populations and coloured dots represents individuals. DA and PCA eigenvalue plots show the number of discriminant functions and principal components used in the computations.

Relationship between genetic and geographic distance

Correlations between genetic distance and coastal distance using the 6 of 6 dataset were very low but significant for Mantel tests when log $(1 + CD)$ was used for the distance metric and close to significant without log transformation (Table 2.12 and Figure 2.10).

Table 2.12. Mantel test results with p-value obtained from 9999 permutations based on the 6 of 6 dataset of 577 individuals. Distance matrices used were based on coastal distance (CD) and its logarithm, log (1 + CD) for geographic distance and F_{ST} and linearised F_{ST}, $LinF_{ST} = F_{ST}/ (1 - F_{ST})$ for genetic distance.

Dataset	F_{ST} v CD	$LinF_{ST}$ v CD	F_{ST} v log(1+CD)	$LinF_{ST}$ v log(1+CD)
6 loci	$R^2 = 0.035$	$R^2 = 0.035$	$R^2 = 0.020$	$R^2 = 0.021$
	$p = 0.086$	$p = 0.069$	$p = 0.046$	$p = 0.040$
Neutral	$R^2 = 0.004$	$R^2 = 0.004$	$R^2 = 0.001$	$R^2 = 0.001$
loci	$p = 0.331$	$p = 0.335$	$p = 0.336$	$p = 0.324$

(a)   (b)

Figure 2.10 Scatterplots of Mantel tests for IBD using the 6 of 6 dataset of 577 individuals, the linearised genetic distance, $LinF_{ST}$ and (a) coastal distance, CD ($R^2 = 0.035$, p = 0.069) and (b) logarithm of coastal distance, log(1 + CD) ($R^2 = 0.021$, p = 0.040).

Significant positive autocorrelation was observed at the smallest distance class for all analyses covering the entire study range and using 6 loci, with the observed r falling outside the permuted 95% confidence belt of the null hypothesis and with the bootstrapped 95% confidence interval for r not straddling zero (Figure 2.11a and Supplementary Table 3i, iv - vi).

The overall correlograms were also significant at the 0.01 level (employed here, rather than 0.05, as suggested by Banks and Peakall, 2012) with p-values ranging from 0.001 to 0.007 for all analyses on six loci across the entire study range. Separate analyses of the five neutral loci and the MGE7 locus alone showed that the patterns of spatial autocorrelation were driven by the MGE7 locus (Figure 2.11b & c and Supplementary Table 3ii & iii). None of the statistical tests were significant for the neutral loci and the overall correlogram p-value was 0.105. For the MGE7 locus alone, the tests were significant for the two smallest distance classes and the p-value for the overall correlogram was 0.0003 compared to the corresponding 6-locus p-value of 0.001.

The largest distance class over all analyses that gave significant positive spatial structure was [0 − 35], which included distances between populations of 10, 12 and 20 km (Supplementary Table 3iv). Analyses of small distance classes showed significant spatial structure in the distance classes [0] and (0 − 6] but not for the third (6 − 30] (Supplementary Table 3v). Overall, these analyses indicate that the positive spatial structure observed ceases to be statistically significant at some distance between 12 − 30 km.
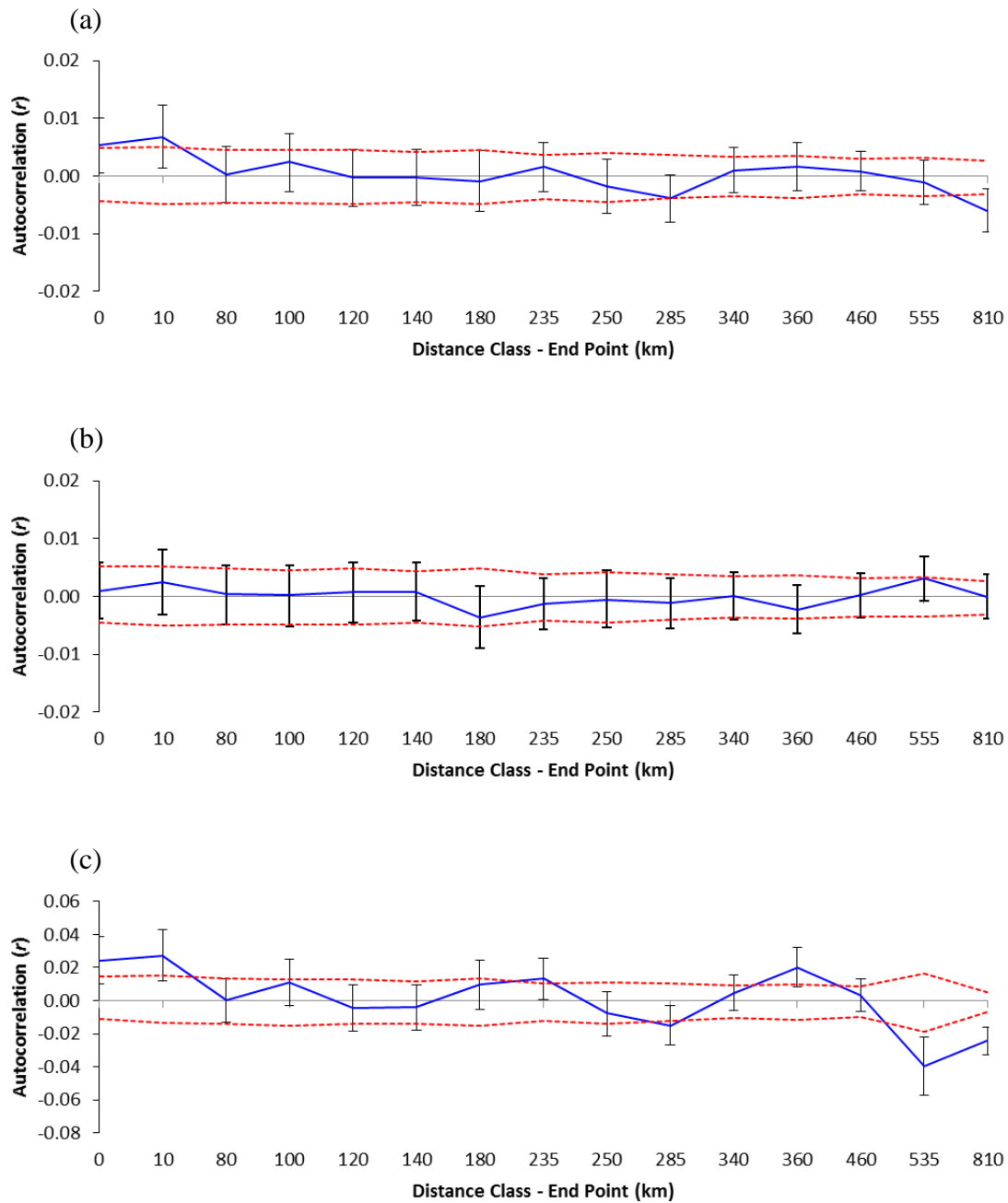
Figure 2.11 Correlogram plots of spatial autocorrelation (*r*) as a function of geographic distance for the 6 of 6 dataset of 577 individuals with (a) all loci, (b) 5 neutral loci only and (c) MGE7 only. The area between the red dashed lines represents the 95% confidence intervals for the null hypothesis of no spatial autocorrelation (9999 permutations) and the error bars represent the bootstrapped (10,000 trials) 95% confidence interval for each observed *r*.

Discussion

In this chapter, patterns of genetic structure and diversity of 18 populations of *Mytilus galloprovincialis* along the eastern coast of the Adriatic Sea were examined. Hypotheses of regional differentiation due to potential barriers to gene flow due to the Northern, Middle and Southern gyres of the Adriatic Sea, and of isolation by distance (IBD) along the 800 km eastern coastline were explicitly tested. The main findings were that (i) five neutral markers revealed little evidence of genetic structure, with populations generally well-connected genetically and being close to panmictic, and (ii) one outlier locus, MGE7, which may be under selection or is linked to another loci that is under selection, provides strong evidence of significant structure along the study range.

Over all analyses performed, when only confirmed neutral loci were considered, there was very little evidence of genetic structure and significant results were not consistent between tests. For example, only the north-central pairwise population comparison was significant by Hierarchical AMOVA while only the north-south pairwise comparison was significant using Exact G-tests. When the MGE7 locus was included, there was significant structure in a range of diverse tests (See Table 2.13 for a summary of the main test results for this study and a comparison of results when the MGE7 locus was included and when only neutral loci were included).

Table 2.13 Comparison of main test results with and without the MGE7 locus. N, northern region, C, central region, S, southern region.

Test	Results with MGE7 included	Results when MGE7 excluded
Hierarchical AMOVA	Overall regional structure $F_{RT} = 0.003$, $p \leq 0.001$ Structure at population level $F_{ST} = 0.010$, $p \leq 0.001$ Strong N-S and C-S structure ($p < 0.0001$) N-C less differentiated ($p = 0.012$)	No regional structure $F_{RT} = 0.001$, $p = 0.085$ Weak structure at population level $F_{ST} = 0.002$, $p = 0.024$ No N-S or C-S structure, some N-C differentiation $F_{ST} = 0.002$, $p = 0.013$
Exact G-tests of differentiation	Overall structure ($p = 2.2 \times 10^{-6}$) N-S and C-S comparisons significant	No overall structure ($p = 0.056$) but N-S differentiation ($p = 0.008$)
Population AMOVA	64 significant pairwise population comparisons out of 153 tests	One significant pairwise population comparison out of 153 tests.
Structure	For K=2, cluster membership is correlated with latitude ($p = 0.002$)	For K=2, cluster membership is not correlated with latitude ($p = 0.178$)
AWclust	2 – 3 clusters observed on MSD plot and for K=2, cluster membership is correlated with latitude ($p = 0.017$)	Only one cluster and no correlation with latitude for K=2 ($p = 0.276$)
Mantel tests	F_{ST} v log (1 + CD) r = 0.143, p = 0.046	F_{ST} v log (1 + CD) r = 0.032, p = 0.336
Spatial autocorrelation	Significant spatial autocorrelation for 0 and 10 km distant classes ($p = 0.001$)	No significance ($p = 0.105$)

There was a strong correlation between the frequency of one of the MGE7 alleles, size 243 bp, and latitude. Given this strong correlation, certain results would be expected from analyses based on datasets including the MGE7 locus. Tests for regional differentiation

between the northern and southern regions would be predicted to be larger than differentiation between the northern and central or central and southern regions. Positive results from tests for the presence of IBD would also be predicted since populations close together on the cline would be expected to have more similar frequencies in the MGE7 allele and therefore be genetically more similar than populations that are further away. These predictions were all confirmed by hierarchical AMOVA, Mantel tests and tests for spatial autocorrelation. Moreover, the significant correlation between cluster membership, for K = 2, and latitude found in the STRUCTURE and AWclust analyses also reflected the allele correlation. These results illustrate the lack of specificity of tests intended to detect certain models of genetic structure since all are a direct consequence of the spatial distribution of MGE7 alleles, not the underlying model.

Considering the hypothesis of genetic differentiation according to the three gyres in the Adriatic Sea, when only neutral loci were used, the hierarchical AMOVA did not overall support regional differentiation (F_{RT} = 0.001, p = 0.085) but there was some support for genetic differentiation between the Northern and Central regions (F_{RT} = 0.002, p = 0.013). The Exact G-tests also found no overall support for the hypothesis for genetic differentiation between the regions (p = 0.056) but the Northern and Southern regions were significantly different p = 0.008. The Bayesian analysis STRUCTURE and non-parametric programs AWclust and DAPC identified only one cluster when the five neutral loci were used. Analyses using the Bayesian spatial method, Geneland (Guillot et al. 2005), and spatial analyses using the programs Barrier vs. 2.2 (Manni et al. 2004), SAMOVA (Dupanloup et al. 2002) and the Monmonier algorithm as implemented in AIS (Miller 2005) were also uninformative (methods and results reported in the Supplementary Material). Overall, the hypothesis of regional genetic structure based on neutral population processes can be rejected. Since there was no strong evidence for regional structure, tests for IBD could be carried out across the whole study range without concern that

regional structure could bias results. When only neutral loci were included in tests, there was no support for IBD, with none of the Mantel tests or tests for spatial autocorrelation statistically significant. The evidence for Chaotic Genetic Patchiness based on neutral loci was scanty and unconvincing compared to empirical examples of CGP in Eldon et al. (2016) and may be related to sampling noise rather than real genetic structure (Waples 1998). For the hierarchical AMOVA using neutral loci, the regional differentiation index, F_{RT}, was not significant while the population differentiation index, F_{ST}, which includes variation between regions and populations, was significant, suggesting that there may be processes leading to genetic differentiation that occur over small inter-population scale, rather than large inter-regional scale. Also, two populations, Zadar (Z) and Mali Lošinj (ML) had unusually high average pairwise F_{ST} values compared to other populations (both 0.008 compared to $0.001 - 0.004$ for all other populations) although only one of 153 pairwise comparisons was significant.

These results confirm that there is little genetic structure among populations along the eastern coast of the Adriatic Sea, but it does not follow that populations are demographically well-connected because low levels of migration, m, can lead to high levels of genetic connectivity when effective population sizes are large (Gagnaire et al. 2015). Weak genetic structure can arise from very different scenarios with, for example, similar F_{ST} values expected for small populations with high demographic connectivity ($N_e = 1000$, $m = 0.1$) and for large populations which are demographically independent ($N_e = 10,000,000$, $m = 0.00001$). Estimates of effective population size were very large for almost all populations investigated which together with low F_{ST} values means that it is difficult to assess the level of migration between populations (Palumbi 2003).

There are several reasons to suspect that blue mussel populations in the Adriatic Sea may not be as well connected demographically as the evidence of panmixia suggests. While the original model of open seas and high demographic connectivity hypothesized that

connectivity was mediated through long-distance larval dispersal, it is now thought that successful long distance larval transport and settlement is a rare event because of the complicated interplay between the timing of life history events and oceanographic conditions that favour long-distance dispersal (Palumbi 2003, Treml et al. 2012, Selkoe et al. 2016). Long-distance dispersal requires, not only that larvae are transported by major offshore currents, but also that larvae are able to escape their natal origin, reach these currents and then leave them to arrive at suitable habitat precisely at the stage of larval development favourable to settlement. The retention of a majority of mussel larvae within an area of 20 – 40 km has been found in several studies in different geographic regions and using different methodologies. Gilg & Hilbish (2003) compared hydrographic modelling with genotyping of newly settled spat in areas around a *Mytilus* hybrid zone in Cornwall and Devon in the United Kingdom; McQuaid & Phillips (2000) monitored the spread of a small population of invasive *M. galloprovincialis*, established in Port Elizabeth, South Africa, over four years; Becker et al. (2007) used elemental fingerprinting to determine the source of settled juveniles of *M. galloprovincialis* and *M. californianus* in California; and Stuckas et al. (2017) used hydrodynamic modelling and multilocus genotyping in a hybrid zone in the Baltic Sea. In addition, biophysical modelling of larval dispersal in the Mediterranean and Adriatic Seas has predicted that the Croatian coast is a larval sink with particularly high rates of larval retention in the northern section due to the combination of complex shoreline, extended continental shelves and relatively weak currents (Dubois et al. 2016, Bray et al. 2017). Finally, while there were only low levels of differentiation discovered using microsatellites along the Croatian coast (Štambuk et al. 2013), there was significant genetic differentiation between a population from Zadar on the central coast of Croatia and one in Ravenna, Italy, on the northwest coast of the Adriatic Sea in studies based on microsatellites and mitochondrial DNA sequence (Giantsis et al. 2014a & b). This

differentiation agrees well with the prediction of larval retention along the Croatian coast found using biophysical modelling by Dubois et al. (2016) and Bray et al. (2017).

Intra-population genetic diversity

Average allelic richness, A_R, and the inbreeding coefficient, F_{IS}, were fairly similar across all populations, with the high values for F_{IS} (0.336 – 0.460) reflecting the heterozygote deficiencies found at 5 loci over all populations. Significant heterozygote deficiency and high values for F_{IS} were also found in two studies of blue mussels, one along the Croatian coast (Štambuk et al. 2013, F_{IS}: 0.195 – 0.319) and one covering eastern and central regions of the Mediterranean Sea (Giantsis et al. 2014a, F_{IS}: 0.284 – 0.416). The different ranges of F_{IS} values in each study may be a result of the different microsatellite loci used since, in this study, the average F_{IS} for each locus over all populations varied widely. The seven populations of farmed mussels had similar levels of diversity as the wild populations, as was found in farmed/wild blue mussel populations in Greece (Giantsis et al. 2012, Giantsis et al. 2014a & b), indicating that aquaculture practices along the eastern coast of the Adriatic and in the eastern Mediterranean do not lead to a loss in genetic variation as can occur in exploited populations and that can reduce the rate of recovery from over-exploitation (Allendorf et al. 2008).

Shellfish culture in Croatia, Bosnia-Herzegovina, Montenegro and Albania is generally small-scale following traditional methods where spat is collected from the sea on ropes or rafts rather than artificially in hatcheries. Production is low and quite variable from year to year, for example: 15 tonnes in Bosnia and Herzegovina in 2004, 400 tonnes in Croatia in 2011 and 3,500 tonnes in Croatia in 2007 compared to > 17,000 tonnes in Greece and > 60,000 tonnes in Italy, (Kovačić et al. 2017, Mandić et al. 2017, FAO 2005-2019, FAO 2018). Although mussel farms often use locally collected spat, the practice of transferring spat from natural spawning grounds to locations with conditions favourable to growth is common (Giantsis et al

2014b, Kovačić et al. 2017). In Greece, spat collected from Chalastra in the Thermaikos Gulf, has been translocated to many Greek mussel farms including some in the Ionian Sea and this practice is thought to have contributed to the low levels of genetic differentiation found among Greek populations (Giantsis et al 2014a). Within the Adriatic Sea, there is also likely to be unintentional translocation by shipping, of larvae via ballast water, and of juveniles and adults via byssal thread attachment to hulls. This form of anthropogenic transport has been well-documented (Apte et al. 2000, Gardner et al. 2016). Marine traffic is heavy in the Adriatic Sea with many cargo ships, tankers and cruise liners connecting the Adriatic Sea with the Mediterranean Sea and beyond, and other vessels such as ferries, fishing vessels and recreational boats making smaller voyages across the Adriatic Sea, crisscrossing between islands and local harbours.

The importance of marker type

Microsatellites are one of the most useful genetic markers used in population genetic studies since they are highly variable and have a recent temporal scale of reference compared to other commonly used genetic markers (Waits & Storfer 2016). They have been used for many marine species (White et al. 2010, Amaral et al. 2012, Wei et al. 2013a, Schiavina et al. 2014, Miller et al. 2018) including the *Mytilus* species complex (Diz & Presa 2008, 2009, Giantsis et al. 2014a, Larraín et al. 2015, Araneda et al. 2016) but surprisingly for such an ecologically and economically important species group, only a limited number of microsatellite loci primers have been reported so far. There has also been limited success in cross-amplification between *Mytilus* species and even between geographical regions within species (Table 2.14 and Supplementary Table 4). Nguyen et al. (2011) found that of 38 loci published in six papers only 6 amplified robustly and consistently and were polymorphic. Similarly, Giantsis et al. (2014a) tested 22 loci from five papers and selected only 10 based on their clear,

reproducible banding. This may in part be explained by high rates of substitution found in bivalves generally, which when occurring in the primer region can lead to null alleles (Harrang et al. 2013). The poor performance of microsatellites markers developed from European mussels in the southern Hemisphere mussels (Nguyen et al. 2011) is not surprising given the distinct lineages found between hemispheres (Gérard et al. 2008, Gardner et al. 2016). It is unexpected, however, that in a study on *Mytilus galloprovincialis* (Giantsis et al. 2014a), markers developed from the same species and same lineage had higher rates of null alleles than markers developed from Baltic Sea *M. trossulus* and Irish Sea *M. edulis* (Presa et al. 2002; Varela et al. 2007; Gardeström et al. 2008; Lallias et al. 2009). Five of the seven markers developed by Presa et al. (2002), presumably from northern Spanish mussels, performed well on *M. galloprovincialis* from the western Mediterranean (Diz and Presa, 2008) but only two were selected for use in a study covering the Ligurian, Adriatic, Ionian and Aegean Seas (Giantsis et al. 2014a), and only two for the Croatian coast (Štambuk et al. 2013).

In this study, of the 8 microsatellite loci initially tested, 4 were derived from *M. galloprovincialis* EST-derived sequence, which have been used successfully in studies of both northern and southern lineages (Westfall 2011, Štambuk et al. 2013, Giantsis et al. 2014a), 2 from *Mytilus trossulus* from the Baltic Sea (Gardeström et al. 2008) and 2 from, presumably Spanish mussels from the northeast Atlantic coast (Presa et al. 2002). The microsatellites developed by Presa et al. (2002) worked well for *M. galloprovincialis* from the western Mediterranean Sea (Diz & Presa 2008), so it is surprising that two of these markers, Mgμ1 and Mgμ5, did not work well in this study, with one discarded completely and the other able to be scored for only 44.1% of individuals, while loci developed for *M. trossulus* were satisfactory.

Table 2.14 Characteristics of published microsatellite loci developed for the *Mytilus* species complex. Het def, number of loci with heterozygote deficiency / total number loci. *Mg, Me, Mt, Mytilus galloprovincialis, edulis* and *trossulus* respectively. TNP, test not performed.

Reference Year	No. loci	Origin and testing of microsatellites	Het def	Loci
Presa et al. 2002	7	Derived from *Mg* and tested on *Mg* from NW Spain, *Me* from SE Ireland and *Mt* from S Finland	2/7	Mgµ1 - Mgµ7
Varela et al. 2007	2	Derived from and tested on *Mg* from NW Spain.	TNP	Mg181, 192* and 220.
Yu & Li 2007	8	From *Mg* EST sequences from GenBank. Tested on *Mg* from Haiyang, China	5/8	MGE1 – MGE8
Gardeström et al. 2008	6	Derived from Baltic *Mt* Tested on Baltic *Mt* and Skagerrak *Me*. Only MT203 and MT282 amplified consistently for *Me*.	All loci	MT63, 203, 279, 282, 288 & 289
Lallias et al. 2009	10	Derived from and tested on Menai Strait, Irish Sea *Me*.	6/10	*Med* 362, 367, 379, 397, 722, 733, 737, 740, 744, 747
Nguyen et al. 2011	4	Derived from *Mytilus* EST sequences from GenBank	TNP	My029, 033, 048, 650

*Monomorphic in Varela et al. (2007), polymorphic in Nguyen et al. (2011)

Hardy-Weinberg equilibrium and marine molluscs

Heterozygote deficiencies and non-conformation to Hardy-Weinberg Equilibrium (HWE) have been well-documented in marine bivalves and have been found in both allozyme and microsatellite studies (Zouros & Foltz 1984, Hedgecock et al. 2004, Wei et al. 2013a, Daniels & Litvaitis 2017), and in particular for blue mussels in the Mediterranean (Giantsis et al. 2014a, Štambuk et al. 2013, Diz & Presa 2008). Reasons proposed include technical problems such as

allelic drop out and null alleles, the Wahlund effect, selection, aneuploidy (Teixeira de Sousa et al. 2012), and peripheral dislodgement (Daniels & Litvaitis 2017). Many genetic analyses depend on conformity to HWE conditions in their assumptions and where these analyses have been performed in this study, non-parametric equivalents have been used to confirm results, where possible.

Neutral markers and markers under selection

Of the planned eight microsatellite panel for this study one, Mgμ5, was discarded because of poor amplification, another, Mgμ1, was able to be scored for only 44.1% of individuals and therefore only used in some analyses and a third, MGE7, one of four EST-derived markers, was flagged an outlier and potentially under selection or close to a gene under selection. Subsequent genetic analyses strongly support this finding with significant results based on datasets including the MGE7 locus being reversed or markedly weakened when only neutral loci were used (see Table 2.13).This left only 5 neutral loci for analyses, which is lower than recommended for genetic studies of population structure, especially when there are low levels of genetic differentiation (Selkoe & Toonen 2006). In general, a larger number of markers would increase the statistical power to detect low genetic differentiation although, when effective populations sizes are high or migration rates are high (or both) simply increasing the number of neutral markers may still not be able to resolve the level of demographic connectivity (Gagnaire et al. 2015).

Outlier loci such as MGE7 have had an increasing role in uncovering genetic structure in marine systems where neutral loci are uninformative (Selkoe et al. 2016). For the greenshell mussel, *Perna canaliculus*, out of ten microsatellite loci assayed, one locus was identified as an outlier and revealed a strong signal for a genetic break that occurs across multiple taxa in

central New Zealand, but the nine neutral loci revealed only very weak evidence (Wei et al. 2013a). Similarly, one locus out of 16 presumed neutral microsatellites was a significant outlier for the roundnose grenadier, *Coryphaenoides rupestris*, in the North Atlantic. While the neutral loci detected a weak pattern of IBD and a possible genetic break, the outlier identified strong differentiation between populations at different depths which was related to the frequency of one of the outlier alleles (White et al. 2010). Because neutral markers are by definition not under selection, powerful mathematical models and analyses of population structure that rely on Hardy-Weinberg equilibrium can be used to identify genetic differentiation and diversity and to inform management and conservation effort. In the marine setting, however, these markers often fail to detect genetic structure while it is the genes that are under selection, which have functional roles in shaping genetic structure that do identify genetic structure and therefore matter to ecology, and to efforts to manage and conserve the genetic diversity of populations (White et al. 2010). Outlier loci are not, however, necessarily connected to adaptive processes and can arise from a number of evolutionary processes. In particular, they have been found at high frequencies in long one-dimensional landscapes such as coastlines which affect population processes in a manner that inflates the variance of F_{ST} (Bierne et al. 2013). Although the classical inferences and techniques employed for neutral loci cannot be used on adaptive loci, their potential to discern genetic differentiation and identify locally adapted populations can be of great use to fisheries management and conservation (Gagnaire et al. 2015). Associations between genetic markers and environmental variables have been taken as support for the involvement of selection in maintaining the pattern of genetic variation (Vasemägi 2006). The next step in describing the distribution pattern of the outlier locus MGE7 is to search for associations of MGE7 alleles with environmental variables, and this is the focus for Chapter 3.

Chapter 3: Seascape genetics of *Mytilus galloprovincialis*: linking genetic and environmental variation

Introduction

One of the aims of seascape genetics is to understand how environmental features influence the evolutionary processes that lead to patterns of spatial genetic variation and differentiation in marine organisms (Selkoe et al. 2016). The molecular markers used to estimate genetic variation can however be either selectively neutral, not having any effect on fitness, or adaptive, with direct effects on fitness in different environments (Holderegger et al. 2006). Both types of genetic marker are affected by mutation, gene flow and genetic drift, but only adaptive markers are affected by selection. This means that the information garnered from the different types of genetic marker describe different aspects of microevolution. In seascape genetics, the environmental features that are important to genetic variation will differ between neutral and selective loci according to how each environmental feature affects neutral or selective population genetic processes (Gaggiotti et al. 2009). In Chapter 2, five neutral loci found only low levels of genetic structure, while the MGE7 locus revealed distinct patterns of genetic variation, and may be under selection or linked to a gene under selection.

Seascape genetics

Seascape genetics seeks to describe and quantify the effects of environmental features on the spatial and temporal genetic patterns of marine organisms (Riginos & Liggins, 2013). In seeking biologically relevant environmental factors that may be associated with different genetic patterns, it can be useful to view these factors at various spatial and temporal scales.

For example, in the Adriatic Sea, environmental variables can be viewed at the whole-sea scale where broad east-west and north-south trends exist, at an intermediate level where environmental gradients may be associated with specific seascape features such as estuaries, lagoons, and currents, or at the small scale spanning the intertidal range (Hamer et al. 2008, Hamer et al. 2010, Lipizer et al. 2014). Temporal variation can be viewed at geologic time scales, decadal scales (e.g. El Nino/La Nina), as the yearly cycle through the seasons or at the level of daily or tidal variation, as well as from the perspective of anthropogenic changes such as global warming (Treml et al. 2012, Riginos & Liggins, 2013, Galli et al. 2017).

Which environmental variables are most important in contributing to the population genetic structure of marine organisms, depends on how these spatial and temporal scales interact with biological processes affecting survival, processes such as reproduction, larval transport, recruitment and growth (Cowen & Sponaugle 2009). In the Mediterranean blue mussel, *Mytilus galloprovincialis*, the timing of gametogenesis is heavily dependent on temperature and food supply and, except for a quiescent period during summer, spawning can occur throughout the year, with peak spawning starting as early as late winter and lasting through spring (Da Ros et al. 1985, Gosling 2003). Temperature, salinity and food availability are also important factors affecting larval growth rates and survival. Mediterranean blue mussels are planktotrophic (a feeding larval stage dispersing through the plankton) and therefore depend on the availability of food (phytoplankton and other organic material such as detritus) in the waters through which they disperse. Larvae at early developmental stages are generally less tolerant of temperature and salinity variation than are shelled larvae. There is also a genetic component to growth and survival in different salinity regimes with larvae from high salinity environments having no growth at 19 psu, while larvae from low salinity environments grow even at 14 psu. (Gosling 2003).

Many marine organisms have pelagic (pre-metamorphosis larval) and benthic (post-metamorphosis juvenile and adult) stages which affect how these organisms experience their environment (Liggins et al. 2013, Pascual et al. 2017). During the pelagic stage larvae move relative to static benthic features but may be transported within distinct packets of water, so environmental variables related to seawater may be relatively stable. With settlement comes a change in frame of reference: now attached to the benthos, instead of moving with the sea, juveniles experience the sea moving about them bringing greater fluctuations in environmental conditions (Riginos & Liggins 2013). The transition between pelagic and benthic stages may therefore be a time of particular vulnerability and if extreme conditions are a regular feature of the settlement location, selective mortality of juveniles can lead to genetic structuring in the adult population. Regular annual selective mortality is seen in *Mytilus edulis* in Long Island Sound (New York state, Atlantic coast of the USA), where the strong selection directed against juveniles with the *Lap^{94}* allele in response to low salinity is reflected in the genetic makeup of adult populations. Mussels are osmoconformers maintaining the osmotic pressure of their internal environment at the same levels as the external marine environment. One mechanism for achieving this is via the leucine aminopeptidase enzyme, encoded by the *Lap* gene, which hydrolyses oligopeptides to their component amino acids in order to increase the internal cell osmotic pressure in response to external increases in salinity. Mussels carrying a *Lap^{94}* allele are at a selective disadvantage in low salinity environments and follow a sharp cline with allele frequencies decreasing with salinity (Hilbish et al. 1982, Hilbish & Koehn 1985).

Biological responses to environmental changes

Inhabitants of the intertidal zone endure a remarkably variable environment with regular changes associated with annual, daily and tidal cycles, and unpredictable changes, related to climatic and sea conditions such as storms and heavy rains, which can occur very

quickly over a matter of hours to minutes. Mussels at the benthic life stage, recent settlers and adults, must respond quickly and do so at a number of levels of levels of biological organisation. Two environmental variables known to elicit robust responses at many levels are maximum sea surface temperature and minimum salinity (Lockwood et al. 2015). In mussels, behavioural responses include valve closure to prevent desiccation at high temperatures and to preserve salinity within the mantle cavity at low salinities. Valve closure is accompanied by a switch in physiology and metabolism to allow the mussel to survive in this protected insulated state. With increasing temperature, a sudden drop in heart rate signals a drop in metabolic rate, which is thought to coincide with valve closure (Braby & Somero 2006a), and a decrease in activity of the glycolytic enzyme pyruvate kinase indicates a switch from aerobic to anaerobic metabolism (Anestis et al. 2007).

Responses to environmental stressors such as high temperature and low salinity also occur at the transcriptional and post-translational level. Post-translational responses such as phosphorylation, which change the activity and function of existing cellular proteins, can occur in minutes while transcriptional responses such as the upregulation of heat shock proteins may take hours (Evans & Somero 2010, Lockwood et al. 2015). In *Mytilus galloprovincialis*, P38 MAPK phosphorylation and the activation of stress-related MAPK targetting cascades occur in response to low salinity, initiating many cellular responses to stress (Hamer et al. 2008). An almost universal transcriptional response to environmental stress, present in bacteria, plants and animals, is the heat shock response (HSR) (Chen et al. 2018). Many environmental stressors cause protein damage by altering the structural stability and function of proteins within the cell. Heat shock proteins (HSPs) are a large family of proteins, which are constitutively expressed and prevent the accumulation of damaged proteins by refolding structurally deformed proteins or targetting them for degradation. The heat shock response leads to the upregulation of HSPs and their cognates in response to elevated temperatures and

other protein-damaging pressures such hyposaline stress which can also lead to oxidative damage (Hamer et al. 2008). A related mechanism implicated in protection against environmental stress is the ubiquitin pathway in which the small molecule ubiquitin tags proteins for degradation via the cytosolic proteasome complex. Like HSPs, ubiquitin is constitutively expressed, and has a housekeeping role in ensuring constant turnover of many cellular proteins (Lodish et al. 2000). Ubiquitin also targets damaged proteins and higher levels of ubiquitin-tagged proteins have been found in the gills of mussels collected in the summer than those collected in the winter, which would be expected to have lower levels of heat-related protein damage (Hofmann & Somero 1995).

At the evolutionary level are adaptive responses which may occur in the genetic code of any gene, or associated regulatory sequence, involved at any level of the response. There are many instances in the coastal marine environment where environmental gradients coincide with changes in allele frequency at a locus implicated in adaptation to varying environmental factors. In the common killifish, *Fundulus heteroclitus*, the allele frequency of the heart-type lactate dehydrogenase, LDH-B$_4$, follows a clinal temperature gradient along the eastern coast of North America. Two different alleles are almost fixed at either end of the latitudinal extremes of the habitat range. The two allozymes have different enzyme kinetics and activity at different temperatures resulting in different phenotypes and survivorship, particularly in embryos, when exposed to high temperatures (Crawford & Power 1989). In the intertidal copepod, *Tigriopus californicus*, heat shock gene HSPB1 varies between two populations, in both the promoter and coding regions of the gene. The HSPB1 protein variant from a heat tolerant population is better at protecting protein structure at high temperatures than the northern allozyme, while the variant from the cooler population has a mutation in the promoter region leading to lower levels of transcription of heat shock proteins. (Tangwancharoen et al. 2018).

The Adriatic Sea environment

As described above, salinity and temperature are known to affect genetic structuring in marine populations through adaptive and selective processes. There are large spatial and temporal variations in both sea temperature and salinity of surface waters in the Adriatic Sea which can be regarded as a dilution basin for the Mediterranean Sea (Artegiani et al. 1997, Lipizer at al. 2014). Among rivers with significant discharge rates into the Mediterranean Sea, six flow into the Adriatic Sea. The Po River contributes 11% of the freshwater inflow for the whole Mediterranean Sea while three others, the Neretva, Vjosa and Buna-Drini rivers, enter the sea on the eastern coast of the Adriatic (PERSEUS – UNEP/MAP Report 2015). The Vjosa and Buna-Drini rivers are two of several Albanian rivers whose influence on surface salinity can be clearly seen in climatological maps of surface salinity in spring (e.g., Fig 5.6 in Lipizer et al. 2014).

It is surprising perhaps that, with so many freshwater inputs, the Adriatic Sea is nonetheless very saline with seasonal average salinity for the surface water of each of the three basins varying between 37.9 – 38.2, 37.7 – 38.3 and 35.5 – 37.4 psu for the southern, middle and northern basins, respectively. Evaporation of water from the Adriatic Sea itself and the inflow of high salinity waters from the northern Ionian Sea (> 38.6 psu) are the main processes that maintain this high salinity. Coastal waters are, however, generally less saline than open waters, independent of individual freshwater inputs (Artegiani et al. 1997).

The area of this study along the eastern Adriatic coast is characterised by a limestone karst landscape with accompanying erosion, which has led to a highly indented coast with island chains oriented in parallel to the coastline (UNEP/MAP-RAC/SPA 2015). This erosion has also created an extensive and complex underground network of flow with many freshwater wells and springs, both on land and at the sea bottom. Several of the sampling sites employed in the present study that are in semi-enclosed or elongated bays are particularly influenced by

local springs and rivers: STW in the Bay of Kaštela by Jadro Spring 4 km from the coast, CF, CW, IF and IW in Kotor Bay by the Ljuta and Škurda wells (Joksimović et al.2017). Thus, mussels in these areas, as well as those in the Lim estuary, LKF, can experience wide variations in salinity. Salinity can change very quickly in times of heavy rain and mussel tolerance to salinity may be as tied to the magnitude and speed of changes as to the absolute maximum and minimum values reached.

High surface water temperatures in the summer when reproduction is quiescent would have effects only on established populations, while low temperatures in winter and spring may have large effects on the timing of spawning, larval survival and the duration of larval development during the pelagic stage (Gosling 2003). Within one location, the temperature experienced by mussels can vary enormously depending on position in the intertidal range, time of aerial exposure, and orientation to the sun. Worldwide there is an increasing trend in the occurrence of marine heat waves and temperature-related mass mortality events (Oliver et al. 2018). In the Mediterranean Sea, more frequent and severe heat waves are expected and species inhabiting shallow waters are expected to be severely impacted. It is thought that conditions in the Mediterranean Sea are already close to the physiological limit of mussels (Galli et al. 2017). In the northern Adriatic Sea, recent mass mortality events related to elevated temperatures have resulted in changes in benthic assemblages with a shift from slow- to fast-growing organisms (Di Camillo and Cerrano 2015).

Mussels are filter-feeders removing suspended particles such as bacteria, phytoplankton and detritus from the water via their gills. The concentration of phytoplankton can be estimated through chlorophyll-a concentration, either by direct measurement of water samples or indirectly by satellite-based colour imaging (Böhm et al. 2003, Đurović et al. 2018). Although phytoplankton is considered to be the primary food source for bivalves, its abundance varies throughout the year and by location dependent in part on the availability of nutrients (Gosling

2003). In the Adriatic Sea, nutrients come from river input, particularly the Po River (Italy), and upwelling of nutrient rich deeper waters. In general, chlorophyll-a levels are higher through the winter and spring and lower during the summer though this trend is highly dependent on location and river outflow. High chlorophyll-a concentrations from input from the Po River and several Albanian rivers are clearly visible year-round on satellite images (Figures 3a & b in Böhm et al. 2003).

Food supply is considered the most important factor affecting growth in bivalves and sites favourable to aquaculture are often areas of food abundance. The eastern coast of the Adriatic Sea is generally oligotrophic, but there are areas that naturally tend to higher nutrient concentrations, and many of these have been developed as sites for aquaculture (Frankic 2003). In this thesis, 6 of the 7 farmed populations (LKF in the Lim estuary, NF and BF in Mali Ston Bay, CF and IF in Kotor Bay and SBF in Butrint Lagoon), were in sites that could be classified as mesotrophic or eutrophic, at least for part of the year (Kolitari et al. 2013, Kršinić et al. 2016, Kovačić et al. 2017, Đurović et al. 2018). Food abundance affects not only mussel growth rate but other biological processes relevant to survival and reproduction, such as the timing of gametogenesis and spawning, and the survival of the planktotrophic larvae.

Seascape genetics is a young discipline with analytical methods still in development. GLM has been widely utilised in seascape genetics, including for bivalves, to find associations between environmental variation and genetic structuring (Wei et al 2013b, Selkoe et al. 2016, Silva & Gardner 2016). Generalised Linear Modelling (GLM) is used to describe the relationship between a dependent variable, such as genetic variation, and a number of predictor, or independent, variables. It is an extension of the classical linear model with a variety of distribution models possible and where the response variable is transformed, using a link function (Neuhaus & McCulloch 2011). Isolation-by-environment (IBE), a model comparable to isolation-by-distance (IBD), but using environmental dissimilarity instead of geographic

distance, predicts that genetic distances increase with environmental dissimilarity (Nanninga et al. 2014). As with IBD, IBE can be tested for through Mantel tests, but because geographic distance and environmental dissimilarity are often highly correlated, disentangling the two can be challenging. Partial Mantel tests allow genetic distances to be compared with either geographic distance or environmental dissimilarity by controlling for the effects of the other (Castellano & Balletto, 2002). A different approach is taken by the program Geste, a hierarchical Bayesian method which estimates population-specific F_{ST}-values and then uses GLM to relate these F_{ST}-values to the environmental variables (Foll & Gaggiotti 2006).

In Chapter 2, almost all genetic structure found amongst 18 populations of *Mytilus galloprovincialis* depended on one outlier locus, MGE7 (see Table 2.13 for a summary). This pattern of strong genetic differentiation in one outlier locus against weak differentiation across multiple neutral loci suggests that MGE7 may be close to a gene under selection. Moreover, a strong correlation with latitude was found for the most common allele of this locus, MGE7[243], $R^2 = 0.476$, p = 0.002. Latitude can be considered to be a proxy for environmental variables such as various measures of temperature (e.g., minimum, maximum, range, etc.) and, in the Adriatic Sea, salinity, with which temperature covaries. In this chapter, I aim to investigate the role of environmental variables in explaining the variation in allelic frequency at the MGE7 locus. Selection is a possible explanation for single locus genetic clines which covary with environmental factors but such clines can also arise by chance drift and historic processes (Vasemägi 2006, Schmidt et al. 2008). Identifying which particular environmental variables are most tightly associated with the allele frequency differences is an important step in establishing whether or not selective processes may be involved in the formation of a genetic cline, and better understanding how environmental variation may affect population genetic variation and gene flow (connectivity).

Materials and Methods

In chapter 2, a strong correlation was found between the most common allele of the MGE7 locus, MGE7[243], and latitude, so a first step was to perform correlation tests between all six of the MGE7 alleles and latitude. A generalised linear model using multiple linear regression was then used to analyse the independent relationships between a range of genetic indices, and a number of environmental and geospatial predictor variables. To understand how environmental factors are associated with genetic variation, an array of genetic indices, each quantifying a different aspect of genetic structure, was used. F_{ST}-statistics, which express genetic differentiation between populations, were modified to express a population-specific metric by calculating the mean linearised F_{ST} statistic for each population (from 17 pairwise LinF_{ST} values for all 18 populations) following the method of Wei et al. (2013b). Analyses were performed firstly using the five neutral loci only, to uncover environmental factors that may be involved in neutral population processes such as genetic drift, gene flow and mutation. The next analyses used LinF_{ST} based on six loci, including the locus MGE7 and then LinF_{ST} based only on the MGE7 locus. Finally, to focus on the allele which was strongly correlated with latitude, MGE7[243], the frequency of this allele, f(MGE7[243]) and the frequency of the homozygous genotype for this allele, f(MGE7[243/243]) were used as the genetic response variable. Each measure quantifies a different aspect of genetic structure. For example, while the frequency of allele MGE7[243] is a measure of one allele from one locus, which was strongly correlated with latitude, LinF_{ST} using only the MGE7 locus included information from all MGE7 alleles and LinF_{ST} using six loci included information from all alleles and all loci.

Environmental Data Collection

To capture the response of a genetic variable to changes in the environment the predictor variables should be biologically relevant to both neutral and adaptive processes that may contribute to population genetic variation. For example, an environmental factor relevant to neutral population processes might impede or promote gene flow, while a factor relevant to adaptive processes might confer an advantage in a particular environment. The availability of environmental data varied widely between sites, because there is no standardised environmental data collection within Croatia, or amongst countries with coastline along the Adriatic Sea. For the farmed and paired wild populations (Limski Channel (LKF), Mali Ston Bay (NF, BF, BW), Kotor Bay (CF, CW, IF, IW) and Butrint Lagoon (SBF)) and populations in large harbours (Zadar (Z), Vranjic (STW) and Kotor (IF, IW)) data from published articles was available and more plentiful than for the other collection sites (CK, KLW, OSOR, ML, MLC in the Gulf of Kvarner, U close to the Ugljan-Pašman bridge and STF, Marina Bay). For these latter areas the online resources at oceanbrowser.net and the Croatian sea bathing water quality website baltazar.izor.hr, together with the models presented in Lipizer et al. (2014) and satellite data presented in Böhm et al. (2004), were used and also allowed comparison amongst all sites.
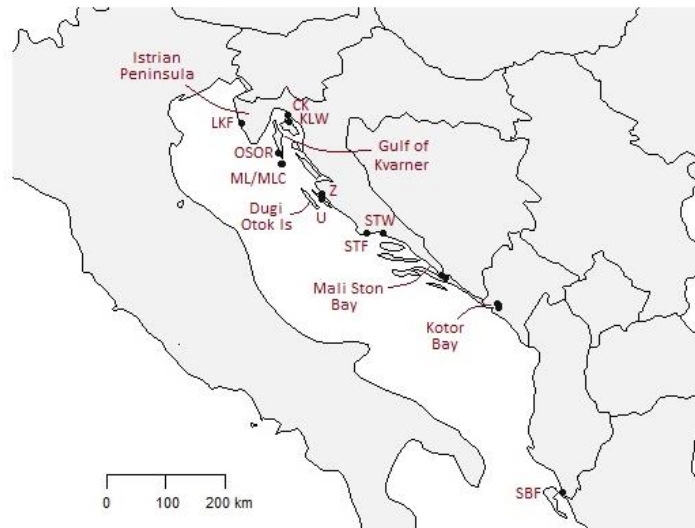
Figure 3.1 Map of the Adriatic Sea showing the location of sampling sites: LKF Limski Channel, farmed; CK Črišnjeva Krk bridge; KLW Klimno, Soline Bay; OSOR Cres- Lošinj bridge, Osor; ML Mali Lošinj Harbour; MLC Mali Lošinj Čikat; Z Gaženica Harbour, Zadar; U Ugljan-Pašman bridge; STF Poljica, farmed; STW Vranjic, Split; Mali Ston Bay: NF Neum, farmed; BF/BW Bistrina Cove, farmed & wild; Kotor Bay: CF/CW farmed & wild, IF/IW farmed & wild, SBF Butrint Lagoon, farmed.

Maximum and minimum measures of three environmental variables, sea surface temperature (SST), salinity (SAL) and Chlorophyll-a concentration (CHL-a), were gathered from published articles and publicly available online databases. Other variables, such as dissolved oxygen and heavy metal concentrations, or indices relating to environmental quality which have relevance to adaptation to hypoxic and oxidative stress, were available only at some sites (Joksimović et al. 2017, Moisiu et al. 2016). The maximum and minimum values were chosen for analyses since they are more important to adaptive processes than a mean or median value. The absolute maximum and minimum values endured for many organisms may not however be as relevant to survival and fitness as the duration of extreme events. Blue mussels respond to high temperatures and low salinities by closing their valves, allowing survival for a few days in a protected (anaerobic) state (Braby & Somero 2006a and references therein). The

data available was not, however, detailed enough to allow use of this sort of maximum or minimum occurring over several days.

Large volumes of data for SST and CHL-a are available via satellite remote sensing, but as explained below, the interpretation of satellite data from coastal areas is challenging (Böhm et al. 2003, Đurović et al. 2018). Standard processing of satellite data uses a series of masks and flags to eliminate data of poorer quality. For instance, automatic cloud detection tends to exclude coastal areas and Böhm et al. (2003) found that some coastal areas were always flagged for cloud cover. For the chlorophyll data, masks for shallow and Case 2 waters, which are optically turbid and generally coastal, were not applied, giving a more complete data set but with overall lower quality. This was appropriate for the Böhm et al. (2003) study which was focussed on large spatial scale circulation patterns but is problematic here since landscape and seascape approaches also focus on fine spatial and temporal scales and local conditions (Manel et al. 2003). Satellite-derived data were not available within $1 - 2$ km of the coast for both SST and CHL-a, and was not available for the Kotor Bay and Butrint lagoon populations (CF, CW, IF, IW and SBF) due to their inland positions relative to the main contour of the coastline. This resolution is not sufficient to detect freshwater influences from the many springs and rivers common along the eastern Adriatic coast especially for the sampling sites in bays and lagoons many kilometres from the open sea. It was therefore decided to collate locally collected data from published papers and online resources with the aim of gathering a coherent and consistent dataset providing between-site comparisons.

Maximum and minimum readings were collected for each variable, although the definition of maximum and minimum varied between resources. For the satellite data from Böhm et al. (2003), the lowest and highest readings were taken from the monthly median (SST) or monthly mean (CHL-a) maps compiled from daily measurements over two years. The data from the DIVA-model (Lipizer et al. 2014) is seasonal over a large number of years so the

maximum and minimum readings are averages over summer and winter. The Croatian sea bathing water quality website gave approximately fortnightly SST and SAL data for the months May to September and of all data sources gave the most geographically complete dataset for the Croatian section of the coast with over 900 monitored sites. This allowed greater confidence in the maximum SST and both minimum and maximum SAL readings with, for example, 4 measuring stations within Klimno Bay (KLW) and 3 in Čikat Bay (MLC). From the published papers the environmental variables were collected between 1998 and 2015 and varied from weekly readings over one year (Drakulović et al. 2012) to two readings taken in one year (Klöppel et al. 2011 and Kovačić et al. 2017).

Overall, there was good agreement for trends and inter-site comparisons between all sources, particularly for SST and SAL. The CHL-a data was not so complete, with more detailed information available for the 10 southernmost populations and only rather sparse data for the 8 northern populations, which relied on satellite-derived data from Böhm et al. (2003) and the website Oceanbrowser.net (see Table 3.1 for the sources of environmental data at each sampling site).

Table 3.1. Sources of data for environmental variables, sea surface temperature (SST), salinity (SAL) and chlorophyll-a (CHL-a), and details of the data collection method.

Source	Environmental variables	Method of data collection	Collection details	Sampling Site
Hamer et al. (2010)	SST, SAL	Direct measurement	3-monthly, 1998-2007	LKF, CK, Z, STW
Klöppel et al. (2011)	SST, SAL	Direct measurement	May and September 2005	LKF
Kovačić et al. (2017)	SST, SAL	Direct measurement	April and July 2010	LKF
Kanduč et al. (2018)	SST, SAL	Direct measurement	Feb, Aug 2010, April, Nov 2011	LKF
Ninčević-Gladan et al. (2015)	SST, SAL, CHL-a	Direct measurement	Monthly, 2001-2007	STW, BF, BW
Krsinic et al. (2016)	SST, SAL, CHL-a	Direct measurement	Weekly or fortnightly 1/2002 – 1/2003	NF, BF, BW
Mladineo et al. (2009)	SST, SAL	Direct measurement	Seasonally 2005	NF, BF, BW
Hafner et al. (2018a)	SST, SAL	Direct measurement	Mar, May, Jul 2010, monthly 2011	NF
Hafner et al. (2018b)	SST, SAL, CHL-a	Direct measurement	Monthly, 2011	NF
Gavrilovic et al. (2011)	SST, SAL	Direct measurement	2-monthly 2009	NF, BF, BW
Mandić et al. 2017	SST, SAL, CHL-a	Direct measurement	Monthly	CF, CW, IF, IW
Krivokapić et al. (2011) Drakulović et al. (2012)	SST, SAL, CHL-a	Direct measurement	Weekly March 2008 – Feb 2009	IF, IW
Đurović et al. (2018)	CHL-a	Direct measurement	14 times from 15/3/2015 to 28/12/2015	CF, CW, IF, IW
Topi et al. (2013)	SST, SAL	Direct measurement	Seasonally May 2010 – Jan 2011.	SBF
Kolitari et al. (2013)	SST, SAL, CHL-a	Direct measurement	Seasonally Oct 2010 – Jun 2011	SBF
Moisiu et al. (2016)	SST, SAL	Direct measurement	Jul-04	SBF
Oceanbrowser.net OR SeaDataNet Climatologies[1]	SST, SAL / CHL-a	Standardised collection of ocean datasets from over 700 institutions	https://www.seadatanet.org/About-us Autumn 2011, Winter 2006 Pixels – 10 km^2	All populations except CF/CW, IF/IW & SBF
Sea quality in the Republic of Croatia[2]	SST, SAL	Direct measurement	Fortnightly from mid-May to mid-September 2013-2017.	All populations except CF/CW, IF/IW & SBF
Böhm et al. (2003)	SST	Daily satellite data (NOAA-14 satellite)	Monthly median 1999 – 2000, Pixels – 1 km^2	All populations except CF/CW, IF/IW & SBF
	CHL-a	Daily satellite data (SeaWiFS ocean colour)	Monthly averaged 1999 – 2000, Pixels – 1 km^2	
Lipizer et al. (2014)	SST, SAL	DIVA modelling based on data held at NODC[3] collected (1911-2009)	Seasonal figures from Data-Interpolating Variational Analysis	All populations

Notes: [1] http://sdn.oceanbrowser.net/web-vis/ and http://gher-diva.phys.ulg accessed 28/10/2018.

[2] http://baltazar.izor.hr/plazepub/kakvoca_detalji10 accessed 2/10/2018

[3] Italian National Oceanographic Data Centre

Nine environmental variables were used in analyses, the minimum and maximum values of sea surface temperature, salinity and chlorophyll-a concentrations (minSST, maxSST, minSAL, maxSAL, minCHL-a and maxCHL-a), together with delta values (deltaSST, deltaSAL and deltaCHL-a) which were simply the difference between the maximum and minimum readings for each variable. Although these variables are simple derivations from the minimum and maximum measures they were thought to be biologically relevant, especially deltaSAL, because changes in salinity are known to occur very rapidly. Three geospatial variables were used, latitude (Lat), longitude (Long) and the total coastal distance (TotalCD), the total of the coastal distances (CD) between any one population and all others. (As in Chapter 2, CD is the shortest distance by sea between each population pair).

Five different genetic indices, described above, were used in the analyses: (i) mean $\text{Lin}F_{ST}$ based on the five neutral loci, (ii) mean $\text{Lin}F_{ST}$ based on six loci, including MGE7, (iii) mean $\text{Lin}F_{ST}$ based only on MGE7, (iv) the frequency of allele MGE7[243], f(MGE7[243]) and (v) the frequency of the homozygous genotype f(MGE7[243/243]). $\text{Lin}F_{ST}$ was as calculated for the analyses in Chapter 2 (by GenAlEx 6.503 (Peakall & Smouse 2012) using the complete dataset of 577 individuals for which data for each locus was available. Another genetic index was trialled, which calculated an overall F_{ST} for each population by pooling all other populations into one and finding the F_{ST} between this population and the pooled populations, but results are not presented here because all models included many variables (overfitting) and were therefore not informative.

Correlation of variables and GLM analysis

A fundamental problem of modelling a dependent variable on two highly correlated independent variables is that a strong signal cannot be attributed to either of the correlated

variables. When a single correlated variable is removed, the signal is explained by the remaining correlated variable and as such should be interpreted as a signal for both variables. The correlation between the twelve environmental and geospatial variables was calculated and visualised using a principal component analysis (PCA) using the software package Statistica Version 7.1 (StatSoft, Tulsa, Oklahoma, USA). Variables were discarded to keep $|R| < 0.9$ for all comparisons using the methodology proposed by Anderson et al. (2008). As described in the results section, when the environmental variables, deltaSST, deltaSAL and deltaCHL-a and the geospatial variable, Long, were removed, leaving a set of 8 independent variables, only two comparisons remained with $|R|$ greater or close to 0.9: maxSST vs minSAL (-0.9615) and minSAL vs maxCHL-a (-0.891). These three variables are relevant to many biological processes that could affect genetic structure, particularly minSAL and maxSST, which have demonstrated roles in genetic selection related to environmental clines (see the Introduction for examples and the Discussion for further text).

Therefore, to keep these biologically relevant but highly correlated variables in the analyses and to obtain the clearest possible picture of the relationship between genetic structure and the independent predictor variables, analyses were performed using eight different subsets of the 8 independent variables (Table 3.2). Each set included a different combination of the three independent variables which were highly correlated (maxSST, minSAL and maxCHL-a). Set No. 1 included all 8 independent variables, Sets No. 2-7 excluded either one or two of the highly correlated variables, while Set No. 8 excluded all three so that the importance of the three highly correlated variables could be evaluated. This sort of hierarchical structure for testing allows for a more nuanced understanding of the role of environmental variables in explaining genetic structure.

Table 3.2 The eight variable sets of environmental and geospatial factors used in GLM analysis, chosen to minimise the number of pairwise correlations with |R|>0.9 and to distinguish which, of the three correlated variables, maxSST, minSAL and maxCHL-a, explains most genetic variation.

Set No.	No. Variables	minSST	maxSST	minSAL	maxSAL	minCHL-a	maxCHL-a	Lat	TotalCD	Excluded variables	Correlated variables
1	8	✓	✓	✓	✓	✓	✓	✓	✓	none	maxSST/minSAL, minSAL/maxCHL-a
2	7	✓		✓	✓	✓	✓	✓	✓	maxSST	minSAL/maxCHL-a
3	7	✓	✓		✓	✓	✓	✓	✓	minSAL	
4	7	✓	✓	✓	✓	✓		✓	✓	maxCHL-a	maxSST/minSAL
5	6	✓			✓	✓	✓	✓	✓	maxSST & minSAL	
6	6	✓		✓	✓	✓		✓	✓	maxSST & maxCHL-a	
7	6	✓	✓		✓	✓		✓	✓	minSAL & maxCHL-a	
8	5	✓			✓	✓		✓	✓	maxSST, minSAL & maxCHL-a	

To find which environmental and geospatial variables were most important in explaining genetic variation between populations, generalized linear models, using the GLZ multiple regression subroutine, were calculated using Statistica Version 7.1 (StatSoft, Tulsa, Oklahoma, USA) with a normal distribution and a Log link function. The best subsets method was used and models ranked according to the Akaike information criterion (AIC), which is a maximum likelihood estimate with a penalty added for each additional parameter included in the model to avoid a bias against parsimony (Sieberts & Schadt 2007). For each genetic index and each set of independent variables, the top-ranked model (lowest AIC value) and a set of best-fitting models were used to evaluate which environmental and geospatial variables were most important in explaining genetic differentiation. The number of models selected for the set of best-fitting models was chosen by inspecting the plot of AIC values against the number of models and finding a point of inflexion in the plot. This is a point at which successive models add only a small and relatively constant amount to the AIC, and are therefore viewed as not being particularly informative. The importance of each independent variable in explaining genetic structure was evaluated by calculating the percentage of best-fit models containing that independent variable. An example of the AIC plot and method of selecting the set of best-fitting models is shown in Supplementary Figure 5 and Supplementary Table 5.

Initial examination of the environmental variables at each sampling site indicated that populations were grouped according to the range of environmental conditions experienced. To formally test this, multivariate routines in Primer v.6 (Clarke & Gorley 2006) were used. The data was transformed by taking the fourth root prior to the creation of a resemblance matrix. An hierarchical cluster analysis (CLUSTER) was performed and a non-metric multidimensional scaling (MDS) plot produced to visualise the grouping. The environmental

variables based on sea surface temperature (maxSST, minSST, deltaSST) and salinity (maxSAL, minSAL, deltaSAL) were used since for one population, KLW, maxCHL-a and minCHL-a had the same value which always led to KLW clustering by itself.

Correlation of allele frequencies with geospatial and environmental variables

An important assumption in correlation testing is that all observations are independent. When testing correlations between allele frequencies and environmental variables this assumption should be carefully considered, since when there are patterns of IBD, geographically close populations are likely to be genetically similar and share environmental similarities, which means that these populations cannot be regarded as independent observations for the correlation test (Coop et al. 2010, McDonald 2014). This is the situation here since strong patterns of spatial autocorrelation driven by the MGE7 locus occurred at the smallest spatial scale for analyses (see Chapter 2, Figure 2.11) with significant results for distances up to 12 km. Populations within 12 km of each other that had similar environmental variables were grouped together as one observation to test the strength of the association between allele frequency and environmental variables. Although the environmental variables were very similar in each group, they were not always the same, so in these cases a weighted average value for the environmental variable was calculated using the number of samples in each population as the weight.

There were moderately high levels of missing data (18% - 29%) from the MGE7 locus for 6 of 18 populations (Z, U, CF, CW, IF, IW) using the 4 of 6 dataset (Table 2.3). Four of these populations (CF, CW, IF and IW) are from Kotor Bay and were among the five populations with lowest minSAL readings (9.2 – 13 psu). Although, the fifth of the low minSAL populations (SBF from Butrint Lagoon) had only one individual with missing data

(2%), the possible association between high frequencies of allele MGE7^{243} and missing data was investigated to ascertain whether the correlation between MGE7^{243} frequency and minSAL could be artefactual. The occurrence of missing data was assessed over all loci and all populations, and then correlations between MGE7^{243} frequency and minSAL were repeated with adjustments to compensate for the potential association of missing data and the MGE7^{243} allele. Correlation analyses were performed, firstly, excluding all populations with more than 10% of MGE7 data missing (both with and without the clustering of populations to ensure independent observations) and then, by assuming that all missing MGE7 data from all populations was not MGE7^{243} allele, thus reducing the initial correlation found (also with and without independent observation clustering).

As a first step in understanding the spatial pattern of change in the frequency of the MGE7^{243} allele, correlations of the genotypic frequencies of the homozygote MGE7$^{243/243}$ and the heterozygote MGE$^{243/X}$, where X = any other allele, were also tested. Because the frequencies of alleles MGE^{237}, MGE7^{240} and MGE7^{252} in the whole dataset were low at 0.009, 0.060 and 0.009, correlations with environmental factors may not be apparent. Tests for possible linkage of these low-frequency alleles with a putative gene under selection together with MGE7^{243} were also performed.

Results

Allelic variation with latitude

Four alleles, MGE7[240], MGE7[243], MGE7[246] and MGE7[249], had a frequency > 0.05 across all populations. The most notable trend in allelic frequency with latitude was the negative correlation of MGE7[243] ($R^2 = 0.476$), which was also the most common allele ($f = 0.538$ over all populations). This was balanced by the positive allelic trends for MGE7[246] and MGE7[249], although the correlations were weaker for these two alleles ($R^2 = 0.061$ and 0.230, respectively). The gradient and correlation for the least frequent of these alleles, MGE7[249], were of greater magnitude than for MGE[246] (Figure 3.2).
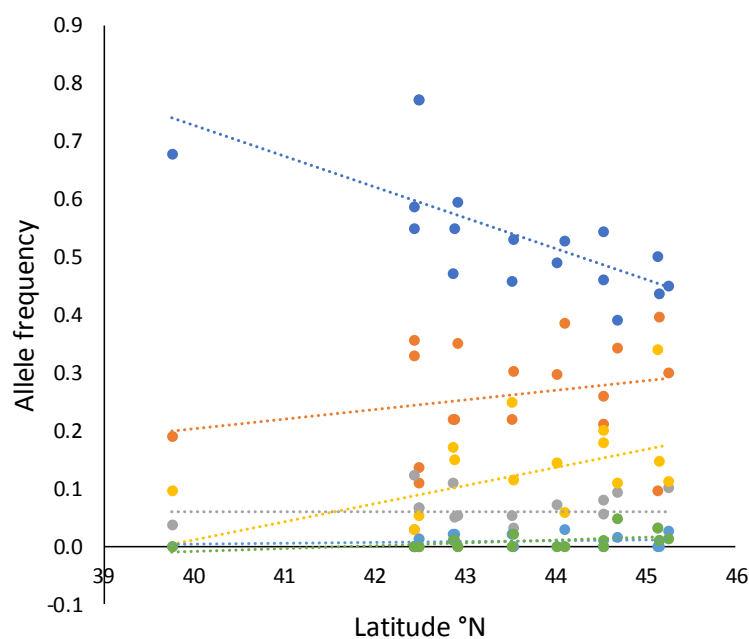
	Allele	f allele	R^2	gradient	p-value
	237	0.008	0.030	0.001	NS
	240	0.060	0.000	-0.003	NS
	243	0.538	0.476	-0.054	0.002
	246	0.263	0.061	0.017	NS
	249	0.124	0.230	0.031	0.044
	252	0.008	0.237	0.005	0.040

Figure 3.2 Correlation of MGE7 allelic frequencies with Latitude, showing frequency of each allele over all populations, f allele, the correlation coefficient R^2, gradient and p-value.

Correlation of variables

Eight of 66 pairwise correlations for 9 environmental and 3 geospatial variables had a correlation coefficient, $|R| > 0.9$, with a further 8 comparisons very close to 0.9. When the environmental variables, deltaSST, deltaSAL and deltaCHL-a, and the geospatial variable, Long, were removed, only 2 comparisons remained with $|R|$ greater or close to 0.9: maxSST v. minSAL (-0.961) and minSAL v maxCHL-a (-0.891). PCA plots of all 12 variables and of 8 variables after removal of 4 highly correlated variables are shown in Figure 3.3 and the correlations between variables shown in Table 3.3.

(a)   (b)
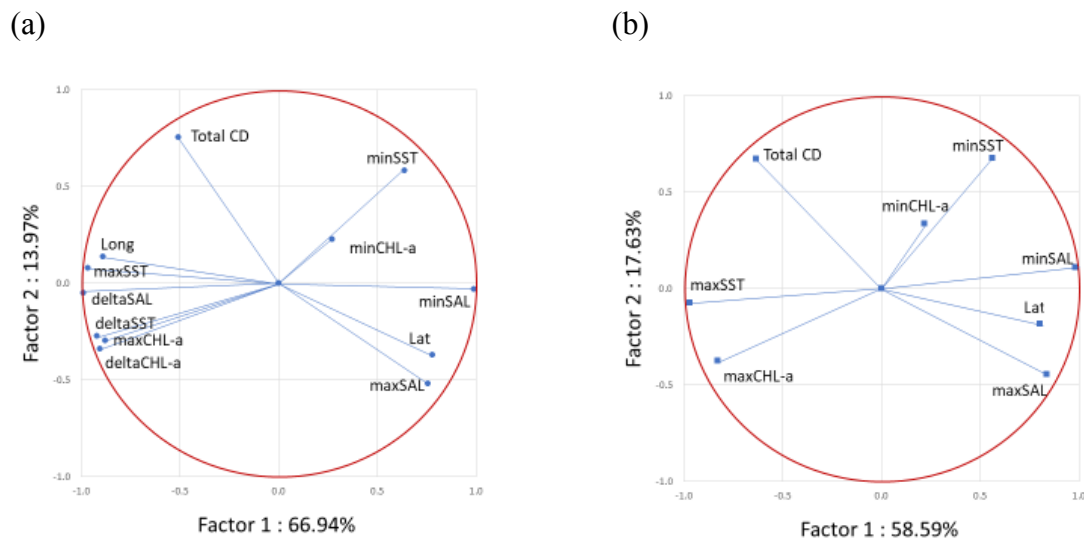
Figure 3.3 Principal component analysis (PCA) of (a) 12 environmental and geospatial variables and (b) 8 variables after removal of 4 highly correlated variables.

Table 3.3 Correlation coefficients, R, between eight environmental and two geospatial variables. After the removal of 4 highly correlated variables, |R| is still greater or close to 0.9 for two comparisons, maxSST vs. minSAL and minSAL vs. maxCHL-a, shown in bold red.

	minSST	maxSST	minSAL	maxSAL	minCHL-a	maxCHL-a	Lat	TotalCD
minSST	1.00							
maxSST	-0.59	1.00						
minSAL	0.57	**-0.96**	1.00					
maxSAL	0.15	-0.74	0.78	1.00				
minCHL-a	0.22	-0.12	0.24	0.03	1.00			
maxCHL-a	-0.65	0.85	**-0.89**	-0.59	-0.04	1.00		
Lat	0.24	-0.76	0.76	0.62	0.36	-0.46	1.00	
TotalCD	-0.06	0.54	-0.50	-0.81	0.02	0.17	-0.59	1.00

GLM analyses

Results of GLM analyses using the genetic index mean $LinF_{ST}$ based only on 5 neutral loci fell into two distinct groups. The first group, comprising variable sets 1 and 2, appeared to be overfitted (i.e., included all variables in the top-ranked model except minSST and maxSAL) and had low p-values (0.0002 and 0.0003), while the other group, comprising variable sets 3 – 8 included only one variable, minSST in the top-ranked model with high p-values (0.177 for all). MinSST was excluded from models in the first group, which included most of the other variables, while it was the only variable included in the 6 models in the second group. This is in contrast with the other genetic indices where models across the variable sets were consistent and had similar p-values (see Table 3.4 and Supplementary Table 6). These inconsistent results mean that no strong association between neutral genetic structure and environmental variables exists, which may reflect the low levels of genetic differentiation present.

Full results of the GLM analyses are set out in the Supplementary Section with the top-ranked model (ranked by AIC) for each analysis (8 sets of environmental and geospatial variables × 5 genetic indices) detailed in Supplementary Table 6 and the set of best-fitting

models for each analysis in Supplementary Table 7. (Details of the eight variable sets are shown in Table 3.2).

Focussing now on the four genetic indices which do include information from the MGE7 locus, results for the two LinF_{ST} genetic indices, one based on 6 loci and one based only on MGE7, were similar, as were results for the two frequency genetic indices, f(MGE7^{243}) and f(MGE7$^{243/243}$). Summary results for LinF_{ST} based on 6 loci and allele frequency f(MGE7^{243}) are presented in Table 3.4 and assessed below.

For LinF_{ST}, the top-ranked model for variable sets 1 - 7 comprised two variables, one of which was Lat. The second variable was one of the three highly correlated variables maxSST, minSAL or maxCHL-a. The top-ranked model for variable set 8 (which excluded all three of the highly correlated variables) included Lat, minSST and maxSAL. The p-value for this model was 0.344 compared to p-values between 0.071 – 0.171 for the 7 other models, indicating that the three highly correlated variables were important for the overall fit of the model. Of the three highly correlated variables, minSAL was selected for the top-ranked model if it was included in the variable set (sets 1, 2, 4 & 6), otherwise either maxSST (sets 3 & 7) or maxCHL-a (set 5) were included. Lat had a significant all effects p-value for 7/8 variable sets while minSAL had a significant all effects p-value when neither maxSST nor maxCHL-a were available (set 6). None of the other variables had a significant all effects p-value. Therefore, although the three highly correlated variables were important for the fit of the models, Lat was the most important variable in explaining genetic variation that was based on LinF_{ST} over 6 loci.

Table 3.4 Summary results from GLM analysis showing the top-ranked model for two of the five genetic indices: Frequency of allele MGE7^{243} and LinF_{ST} for the eight variable sets. The variable sets have different combinations of the three highly correlated variables, maxSST, minSAL and maxCHL-a. For each variable set, the table shows which of these variables is excluded, the top-ranked model with p-value and the all effects p-value for the intercept. An asterisk in a variable column indicates that the all effects p-value for that variable is significant.

Variables included in the top-ranked model

Variable Set	Variables excluded from the Variable set	minSST	maxSST	minSAL	maxSAL	minCHL-a	maxCHL-a	Lat	TotalCD	Model p-value	All effects Intercept p-value
Frequency of allele MGE7^{243}											
1	-			✓						<0.0001	0.601
2	maxSST			✓						<0.0001	0.498
3	minSAL						✓	✓		0.0001	0.428
4	maxCHLa			✓						<0.0001	0.478
5	maxSST minSAL						✓	✓		0.0001	0.535
6	maxSST maxCHLa			✓*						<0.0001	0.400
7	minSAL maxCHLa		✓							<0.0001	0.867
8	maxSST minSAL maxCHLa				✓			✓	✓	0.0014	0.002
LinF_{ST} over 6 loci											
1	-			✓				✓*		0.071	0.020
2	maxSST			✓				✓*		0.071	0.012
3	minSAL		✓					✓*		0.098	0.032
4	maxCHLa			✓				✓*		0.071	0.014
5	maxSST minSAL						✓	✓*		0.171	0.029
6	maxSST maxCHLa			✓*				✓*		0.071	0.010
7	minSAL maxCHLa		✓					✓*		0.099	0.075
8	maxSST minSAL maxCHLa	✓			✓			✓		0.344	0.076

In contrast, when genetic differences at the allelic level (f(MGE7^{243})) were considered, the three highly correlated variables were more important than Lat. Lat appeared in only 3 of 8 variable sets, each of which did not include minSAL (sets 3, 5 and 8). When minSAL was in the variable set (sets 1, 2, 4 & 6), it was the sole variable selected for the top-ranked model. It had a significant all effects p-value only in set 6 which did not include either maxSST or maxCHL-a, illustrating that these correlated variables could compensate for one another in the

models (as expected of strongly correlated variables). When minSAL was not in the variable set, maxSST (set 7) or maxCHL-a and Lat (sets 3 & 5) were selected in the top-ranked model. The top-ranked model for variable set 8 (which excluded all three of the highly correlated variables) included Lat, maxSAL and TotalCD. For this variable set, the p-value for the top-ranked model was 0.0014 compared to p-values ≤ 0.0001 for variable sets $1 - 7$, again showing the relative importance of maxSST, minSAL and maxCHL-a in explaining allelic frequency. The all effects p-value for the intercept for variable set 8 was 0.002 compared to $0.400 - 0.867$ for variable sets $1 - 7$ suggesting that a significant proportion of genetic variation could not be explained by the environmental and geospatial variables available.

In summary, as the measure for genetic variation changed from neutral loci (LinF_{ST} based on neutral loci), to measures including the MGE7 locus (LinF_{ST} based on 6 loci and MGE7 alone) and finally measures focussed on allele and genotype frequencies (MGE7^{243} and MGE7$^{243/243}$), the three highly correlated environmental variables, maxSST, minSAL and maxCHL-a became more important in GLM modelling.

Correlation of allelic frequencies with geospatial and environmental variables

The GLM analyses identified 3 important but correlated environmental variables that explain differences in the frequencies of the MGE7^{243} allele and the MGE7$^{243/243}$ homozygotes. The correlations between allele frequency for each MGE7 allele and the geospatial variable Lat and the three environmental variables identified by GLM analysis are shown in Table 3.5. The environmental variable with the strongest relationship with MGE7^{243} was minSAL, followed by maxSST, which were both more strongly correlated than latitude with MGE7^{243} frequency, while maxCHL-a had a weaker correlation than latitude. The second most frequent allele, MGE7^{246}, was not correlated with any of the variables, but allele MGE7^{249}, which had

an overall frequency of 0.124, had a much stronger correlation with all three environmental variables than latitude.

Table 3.5 Correlation of main MGE7 alleles (frequency > 0.05) and Latitude, minSAL, maxSST and maxCHL-a.

Allele	frequency		Latitude	minSAL	maxSST	maxCHL-a
240	0.06	R^2	0	0.013	0.005	0.005
		gradient	-0.003	-0.0003	0.001	0.001
		p-value	NS	NS	NS	NS
243	0.538	R^2	0.476	0.635	0.563	0.429
		gradient	-0.054	-0.0077	0.037	0.028
		p-value	0.002	<0.0001	<0.001	0.003
246	0.263	R^2	0.061	0.053	0.063	0.008
		gradient	0.017	0.0019	-0.011	-0.003
		p-value	NS	NS	NS	NS
249	0.124	R^2	0.23	0.432	0.327	0.399
		gradient	0.031	0.0054	-0.024	-0.023
		p-value	0.044	0.003	0.013	0.005

Populations can be grouped according to abiotic environmental factors

The hierarchical cluster analysis performed in PRIMER clustered the 18 populations into 5 groups (Figure 3.4). Of all the populations, the three island populations in the Gulf of Kvarner, OSOR, ML and MLC, are closest to the open sea and an open sea environment with low fluctuations in salinity values and moderate maxSST values.
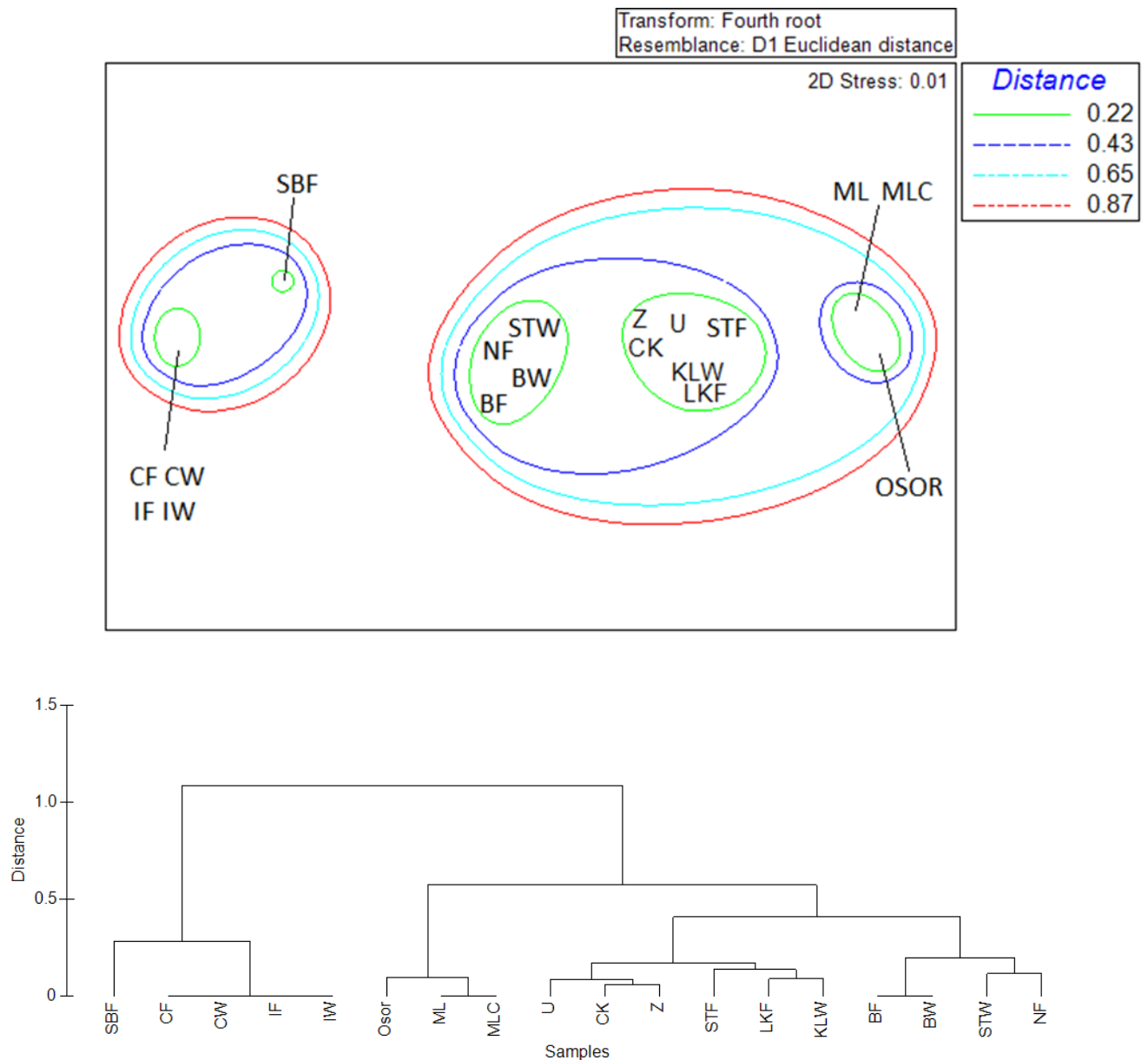
Figure 3.4 MDS plot (above) and hierarchical clustering (below) of populations grouped by maxSST, minSST, deltaSST, maxSAL, minSAL and deltaSAL.

The five populations CK, KLW, Z, U and STF are in environments with low freshwater influence, consistent with the slightly lower salinity observed close to the coast compared to the open sea in the Adriatic Sea (Artegiani et al. 1997). (U is also an island population but has maxSST and minSAL in accord with the CK, KLW, Z and STF and is sheltered from the open sea by another long island, Dugi Otok). Although LKF was also clustered in this group with

maxSST and minSAL fitting in the range (see Table 3.6), LKF is in the estuary of the Lim River with a salinity gradient ranging from 12.2-35.0 psu (Hamer et al. 2010, Kanduč 2018).

Table 3.6 Populations grouped according to site description and environmental variables maxSST and minSAL.

Sites	maxSST (°C)	minSAL (psu)	Description
OSOR, ML, MLC	26.0	36.7	Island close to open sea environment
CK, KLW, Z, U, STF	25.6 - 28.0	32.2 – 35.4	Coastal levels of freshwater influence
LKF	27.2	34.6	Estuarine
STW, NF, BF, BW	28.3 - 28.5	28.4 – 30.1	Bays with moderate freshwater input
CF, CW, IF, IW, SBF	31.0 - 31.4	9.2 - 13.0	Enclosed bay or lagoon with high freshwater input and/or low water exchange with the sea

The 4 sites STW, NF, BF and BW all have moderate freshwater influence and high maxSST. STW is situated in the semi-enclosed Bay of Kaštela which has freshwater discharging from the Jadro Springs 4 km from the coast. NF, BF and BW are all in the important mussel farming area of Mali Ston Bay into which the Neretva River discharges, although it is unknown how much the Neretva waters, which enter the bay 15 km north of NF, influence salinity at NF, BF and BW compared to more local freshwater inputs.

The 5 populations CF, CW, IF, IW and SBF are sited in enclosed bays or lagoons and experience the lowest salinity conditions of all sampling sites with high freshwater input and/or low water exchange with the sea. Kotor Bay, the site of four populations CF, CW, IF and IW, is one of two inner bays of the large and highly indented Boka Kotorska Bay in Montenegro. The two inner bays are connected to the sea via narrow straits and generally have low levels of water exchange with the sea (Bellafiore et al. 2011). There are numerous sources of freshwater which strongly influence the salinity, temperature and nutrient levels in the Bay. Exchange of

water between Butrint Lagoon, site of population SBF, and the open sea, is severely restricted by Vivari Channel, which is 2.5 km long and 6 m deep. Water flows from the sea into the lagoon only during very high tides and storms (Moisiu et al. 2016). Table 3.6 summarises the groupings with details of environmental variables and corresponding habitat features (with SBF grouped together with CF, CW, IF and IW and LKF separated from CK, KLW, Z, U and STF because of its situation in an estuary).

To formally test the correlation between allelic frequency and environmental variables, populations within 12 km of each other were grouped together to reduce lack of spatial independence, as described in the Methods section. The populations grouped together were: CK and KLW, ML and MLC, Z and U, BF and BW, CF, CW, IF and IW, giving 11 independent observations. For allele MGE7[243], correlations with minSAL and maxSST, but not maxCHL-a, were significant (Table 3.7). None of the correlations was significant for allele MGE7[249], but this may be because of stochastic variation because of the much lower overall frequency of this allele. For MGE7[243], the correlation coefficients for the 11 independent observations were higher than for the 18 populations separately. This may be because, although the environmental variables in Kotor (CF, CW, IF and IW) were uniform, the allele frequency was not as close as with other groupings (0.770, 0.770, 0.549 and 0.586 respectively, see Supplementary Table 2). This analysis was therefore repeated with CF and CW in one group and IF and IW in another, with 12 independent observations. The relationships were still significant with $R^2 = 0.651$, p = 0.0015 for minSAL, and $R^2 = 0.691$, p = 0.0008 for maxSST.

Table 3.7. The correlation coefficient of MGE7 alleles 243 and 249 with minSAL, maxSST and maxCHL-a calculated with populations within 12 km of each other in similar environments grouped together to ensure independence of observations.

Allele	frequency		minSAL	maxSST	maxCHL-a
243	0.538	R^2	0.804	0.738	0.331
		p-value	<0.001	<0.001	NS
249	0.124	R^2	0.289	0.204	0.238
		p-value	NS	NS	NS

In relation to the possible association between missing data and the MGE7[243] allele, examination of missing data across all populations and all loci showed that the MT282 locus had a similar trend of missing data from Kotor Bay, between 14– 31%. The average percentage of missing values over all loci (except Mgµ1) was between 8 – 13% in Kotor Bay, 9% and 10% from the 2 populations NF and Z, and between 2% - 6% for the 12 other populations which included the other population with low minSAL, SBF (3%). When correlations between allele frequency and minSAL were performed after dropping all populations with more than 10% missing MGE7 data (i.e. the 4 populations from Kotor and Z and NF), the correlation coefficient, R^2, increased from 0.635, p < 0.001 (18 populations) to 0.645, p = 0.002, (12 populations). This correlation was, however, dependent on the remaining low salinity population, SBF, and when this population was also removed, the correlation decreased, $R^2 = 0.230$, p = 0.051. When the independent observation clustering was used as detailed above, $R^2 = 0.771$, p = 0.002 (9 populations) and, without SBF, $R^2 = 0.732$, p = 0.039 (8 populations). For a worst case scenario when all missing data was assumed not to be MGE7[243], $R^2 = 0.299$, p = 0.019, and, when the independent clustering was also carried out, $R^2 = 0.382$, p = 0.032 (12 populations). Thus, even when compensating for missing MGE7 data in several ways, the correlation between MGE7[243] with minSAL was still significant, except for the one case when the population SBF was removed (p = 0.051).

To explore how environmental variables affect the genetic structure of populations, allele and genotype frequencies were examined in relation to minSAL since it had the strongest correlation of all the independent variables with the frequency of MGE7[243] and was also the most important variable in the GLM analyses at explaining differences in genetic variation.

Allele/Genotype	R^2	gradient	p-value
243 allele	0.635	-0.008	<0.001
243/243 homozygote	0.644	-0.008	<0.001
243/X heterozygote	0.002	0.000	NS
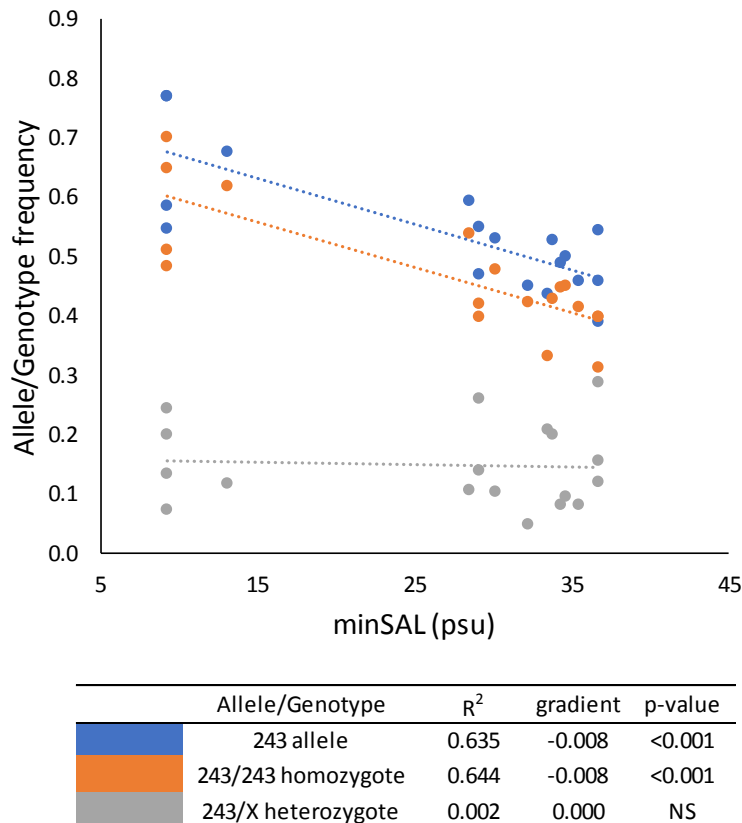
Figure 3.5. Correlation of MGE7[243] allele, homozygote and heterozygote frequency with minSAL.

The correlation of the homozygous MGE7[243/243] genotype with minSAL was similar to that of the MGE7[243] allele with minSAL, but all heterozygote combinations MGE7[243/X] (where X is any other allele) had a correlation of almost zero with a flat gradient (that is, MGE7[243/X] was independent of minSAL - Figure 3.5). Of 112 individuals with a heterozygous MGE7[243/X] genotype, only 7 individuals were MGE7[243/249], but since the MGE7[249] allele has an opposite correlation to MGE7[243], the heterozygote regression with minSAL was repeated excluding this

genotype. The results were very similar with $R^2 = 0.010$ and m = -0.0006. Correlations tested on the grouped populations to avoid dependent observations were also similar with the correlation between minSAL and the frequency of MGE7[243] homozygotes being very similar to the allelic correlation and the heterozygote frequency being almost flat. In contrast, the other allele with a strong correlation with minSAL, MGE7[249], did not have dissimilar allele, homozygote and heterozygote correlations ($R^2 = 0.432$, 0.374, and 0.338, p-value = 0.003, 0.007 and 0.011, respectively).

Although there is a strong association between the frequency of MGE7[243] allele and the environmental factor minSAL, this relationship need not be unique among MGE7 alleles. Working on the hypothesis that the MGE7 is close to a gene that is under selection and that MGE7[243] is linked to a gene variant which is favourable in low salinity conditions, one of the low-frequency MGE7 alleles may also be linked to the favourable gene variant, but stochastic variation at low frequencies may have hidden the trend (Figure 3.2) Since the frequencies of alleles MGE[237], MGE7[240] and MGE7[252] in the whole dataset were low at 0.009, 0.060 and 0.009, respectively, linkage to the gene variant under selection may not be apparent unless their contribution is added to the contribution from MGE7[243] in which case the correlation with minSAL would be expected to remain the same or be strengthened. If these alleles are not linked to the gene variant, then the correlation with minSAL would be expected to weaken. There were only 12 MGE[237] and 12 MGE7[252] alleles in the dataset, but there were 90 MGE[240] alleles (just under 6% of all alleles), so the hypothesis that MGE[240] is also linked to the variant of the gene under selection that is favourable to low salinity conditions was tested by counting genotypes MGE7[240/240], MGE7[240/243] and MGE7[243/243] as homozygotes, and each genotype that includes just one of these alleles as heterozygotes (Table 3.8). For comparison, the effect of testing the same hypothesis for MGE7[237], with only 12 alleles, is also shown.

Table 3.8 Testing the hypotheses that allele MGE7^{240} and allele MGE7^{237} are also linked to the favourable gene variant: correlation tests for frequency of allele, homozygote and heterozygote against minSAL, where, for MGE7^{240}, for example, genotypes MGE7$^{240/240}$, MGE7$^{240/243}$ and MGE7$^{243/243}$ are counted as homozygotes and MGE7$^{240/X}$ and MGE7$^{243/X}$ are counted as heterozygotes.

		243 alone	243 and 240	243 and 237
Allele	R^2	0.635	0.681	0.619
	p-value	<0.0001	<0.0001	0.0001
	gradient	-0.0077	-0.0081	-0.0075
Homozygote	R^2	0.644	0.668	0.629
	p-value	<0.0001	<0.0001	<0.0001
	gradient	-0.0076	-0.0082	-0.0075
Heterozygote	R^2	0.002	0.001	0.014
	p-value	NS	NS	NS
	gradient	-0.0003	0.0002	0.0007

Testing whether MGE^{237}, which had only 12 alleles in the dataset, is also linked to the putative gene under selection, led to a reduction in the correlation coefficient, the magnitude of the gradient and an increase in the p-value for both the allele and homozygote frequencies while testing whether MGE7^{240}, which had 90 alleles in the dataset, is also linked to the putative gene under selection led to an increase in the correlation coefficient and magnitude for the gradient and a decrease in the p-value (Table 3.8).

GLM analyses were repeated using the sum of the frequencies of the MGE7$^{240/240}$, MGE7$^{240/243}$ and MGE7$^{243/243}$ genotypes as the genetic index. Results were rather similar to those based on MGE7^{243} allele and homozygote frequencies, with minSAL included in the top-ranked model for every environmental variable set containing it. The all effects p-value for this genetic index was significant for 3 of the 4 environmental variables sets that included minSAL, compared to only 1 of 4 for the MGE7^{243} allele frequency, indicating that the other independent variables were not able to compensate for minSAL in explaining this genetic index.

Table 3.9 Correlation of main genotypes (frequency > 0.05) with minSAL.

Genotype	frequency	gradient	R^2	p-value
243/243	0.464	-0.0076	0.644	<0.0001
243/246	0.105	0.0001	<0.001	NS
246/246	0.174	0.0004	0.002	NS
246/249	0.057	0.0028	0.379	0.007
249/249	0.083	0.0033	0.374	0.007

The genotype correlations (Table 3.9) followed the pattern found for MGE7 alleles (Table 3.5) with the MGE7$^{243/243}$ homozygote having the strongest correlation with minSAL ($R^2 = 0.644$, p < 0.0001), followed by the MGE7$^{246/249}$ heterozygote and the MGE7$^{249/249}$ homozygote which were nearly identical with $R^2 = 0.379$ and 0.374, respectively, p = 0.007 for both. The MGE7$^{246/246}$ homozygote and the MGE7$^{243/246}$ heterozygote were not correlated with minSAL. Surprisingly, the MGE7$^{243/249}$ heterozygote occurs at a lower frequency (f < 0.01) than the MGE7$^{246/249}$ heterozygote even though the MGE7^{243} allele occurs at a higher frequency (f = 0.538) than the MGE7^{246} allele (f = 0.263).

MGE7 was one of the five loci out of 7 that showed significant heterozygote deficiencies across all populations. In this background of heterozygote deficiency, it is difficult to test whether any individual heterozygous genotypes occurred at lower or higher than expected rates. For the MGE7 locus over all populations, the ratio of observed to expected heterozygotes was 0.40. The ratio of observed to expected heterozygote frequency for the four most abundant alleles (f > 0.05) for total numbers and averaged over all populations are shown in Table 3.10. The MGE7$^{243/249}$ genotype had the lowest ratio at 0.070 reflecting its low frequency (f < 0.01), and this was followed by the MGE7$^{240/249}$ at 0.269. Other ratios ranged between 0.370 – 0.875.

Table 3.10. Ratio of observed to expected frequency for the six most common genotypes.

MGE7 genotype	Ratio of observed to expected frequency across all populations	Average ratio of observed to expected per population
240/243	0.496	0.659
240/246	0.380	0.322
240/249	0.269	0.119
243/246	0.370	0.400
243/249	0.070	0.119
246/249	0.875	0.855

Discussion

The two main findings of the population genetic analyses of Chapter 2 were that (i) five neutral markers revealed little evidence of genetic structure, and (ii) one outlier locus, MGE7, provided strong evidence of significant structure along the study range of ~ 800 km. The strong genetic structure associated with the MGE7 locus in a background of very weak differentiation based on neutral loci suggests that MGE7 is close to a gene that is under selection, or is under selection itself, since demographic and population processes should affect neutral loci similarly (Riginos et al. 2002). In this chapter, the association between genetic variation and environmental factors was examined, using information from both neutral loci and the MGE7 locus. A strong association was found between the frequency of allele MGE7[243] and three highly correlated environmental variables, maxSST, minSAL and maxCHL-a. When MGE7 genotypes were examined, this strong association extended to homozygotes of the MGE7[243] allele, but not to its heterozygotes. This chapter represents the first steps in understanding the factors that may be involved in driving the distribution pattern of MGE7 alleles.

The strength of the correlation between the MGE7[243] allele and the environmental variables maxSST, minSAL and maxCHL-a is suggestive of a role for selection acting directly or indirectly on the MGE7 locus (Schmidt et al. 2008). The MGE7 locus may be a neutral hitchhiker locus situated close to a gene under selection or it may itself be part of the coding region of a gene under selection (Gagnaire et al. 2015). This would not be surprising, since the locus was derived from EST sequences from GenBank (Yu & Li 2007); it was however unexpected since this locus has previously been used successfully as a neutral marker (Westfall 2011, Štambuk et al. 2013). Microsatellite repeat DNA is well-known through its use as a genetic marker and also through its role in several neurological diseases, such as Huntington's

disease and some cancers (Vieira et al. 2016, Bagshaw 2017). They do however have less well-known functions in normal cell processes which are only beginning to be explored. In humans, polymorphic microsatellites located within gene promoter regions have been implicated in the modulation of gene expression via several different mechanisms. (Sawaya et al. 2013, Vieira et al. 2016, Bagshaw 2017). Less is known, however, about microsatellite function within coding regions, but they are overwhelming tri- or hexa-nucleotide repeats due to selective pressure against microsatellites that would change the reading frame (the repeat motif for MGE7 is CAA) (Tóth et al. 2000).

The associations between genetic variation and environmental and geospatial variables identified by GLM analyses became stronger as the genetic index used honed in on the distribution of the MGE7[243] allele. The genetic index based on neutral loci gave only inconsistent results over the eight variable sets, while measures that included information about all six MGE7 alleles were strongly associated with the geospatial variable Lat and, to a lesser extent, the three highly correlated environmental variables maxSST, minSAL and maxCHL-a. Genetic indices based solely on the frequency of the allele MGE7[243] or the homozygous genotype MGE7[243/243] were strongly associated with the three environmental variables, with little contribution from Lat in the models.

The correlation of the MGE7[243] allele frequency and latitude uncovered in Chapter 2, may be coincidental since by chance the northernmost populations in Croatia (shown in the first three lines of Table 3.6) all had high minSAL (32.2 − 36.7 psu) and moderate maxSST (25.6 − 28.0 °C) while the five southernmost populations, in Kotor Bay in Montenegro and Butrint Lagoon in Albania, had low minSAL (9.2 − 13.0 psu) and high maxSST (31.0 − 31.4 °C) values. There are, however, areas of mussel habitat on the Croatian coast with minSAL in the 10 − 13 psu range (Hamer et al. 2010; Kanduč et al. 2018). In particular, the LKF population

with minSAL of 34.6 psu was collected from a mid-point of the Lim estuary, which has minSAL of 12.2 psu further inland. Four seasonal readings for this area of Lim estuary ranged between 12.2 – 22.0 psu. There may also be sections of the Montenegrin-Albanian coast away from freshwater influence that have higher minSAL and more moderate maxSST values than the four Montenegrin sampling sites in Kotor Bay and the Albanian sampling site in Butrint Lagoon, which are highly influenced by freshwater.

Variations in allele frequencies that coincide with changes in an environmental variable point towards selection, but may also arise from chance effects and historic events (Schmidt et al. 2008). The distribution patterns of the $MGE7^{243}$ allele, the homozygous $MGE7^{243/243}$ genotype, and all heterozygous genotypes involving the $MGE7^{243}$ allele are, however, distinctive. The differential results with respect to homozygotes and heterozygotes suggest a selective pressure since this pattern is much less likely to arise by chance alone. The allelic and homozygote frequencies per population ranged between 0.39 – 0.77 and 0.31 – 0.70 respectively, and the correlation with minimum salinity was strongly negative for both ($R^2 =$ 0.635 and 0.644, respectively). The heterozygous genotype frequencies ranged between 0.05 – 0.29 and were not correlated with minimum salinity ($R^2 = 0.002$) (See Table 3.8). The correlations and gradients of alleles $MGE7^{243}$, $MGE7^{246}$ and $MGE7^{249}$ with minSAL suggest there are two or three different alleles under differential selection. Examination of the correlation of genotypes with minSAL support this contention, with the homozygotes of the three most common alleles having different correlations with minSAL (negative for $MGE7^{243/243}$, positive for $MGE7^{249/249}$ and no correlation for $MGE7^{246/246}$). The heterozygote correlations are informative, with the $MGE7^{243/246}$ heterozygote being uncorrelated with minSAL, while the $MGE7^{246/249}$ heterozygote has a similar correlation to the homozygote $MGE7^{249/249}$ and the $MGE7^{243/249}$ heterozygote occurring at a lower than expected frequency ($f < 0.01$, correlation not performed). This presents a complex picture of relationships between

the different genotypes and the environmental variable minSAL and interactions between alleles which suggest that there may be three different alleles under selection.

The GLM and individual correlation tests revealed that minSAL explained genetic differences between populations better than maxSST and maxCHL-a. Results from Mantel and partial Mantel tests, and analyses using the Bayesian program Geste agreed with this ordering (see the Supplementary Material for methods and results). However, the high correlations between the three environmental variables together with the quality of the data set itself, which was gathered from a large number of sources using different methodologies (direct measurement, satellite imagery and modelling), means that conclusions about which is the most important variable to the mechanisms of the putative selective advantage cannot be made. As described in the results section, the variables also seem to be descriptors of certain coastal environments which are characterised by different levels of environmental variation (for example, the open sea populations enjoy the mildest environment with the lowest maxSST and the highest minSAL, while the populations in the enclosed waters of Kotor Bay endure the most extreme conditions with highest maxSST and lowest minSAL). Thus, if selection is the driving force behind the variation in MGE7 allele frequencies, the mechanism may be related to a general stress response, such as pathways that deal with oxidative stress and the ability to endure anoxic conditions with valves closed.

Strong temperature and salinity-related selection has been observed in *Mytilus galloprovincialis* congeners, particularly in relation to the distribution of mussels in the *Mytilus edulis* species complex around hybrid zones. For example, in central California, *Mytilus trossulus*, invasive species *Mytilus galloprovincialis* and their hybrids coexist in distributions clearly linked to temperature and salinity variation, with *M. trossulus* more abundant than *M. galloprovincialis* at low salinity sites and at locations with warmer summer and cooler winter

temperatures (Braby & Somero 2006b). In this California study, mean temperature and mean salinity were strongly correlated ($R^2 = 0.59$) and the authors did not draw conclusions on the basis of their distribution data as to which of the environmental variables was most important in defining the distribution patterns. From previous work comparing the responses of the two congeners to thermal challenge and which indicated that *M. galloprovincialis* is more tolerant of high temperature, the authors suggested that low salinity may be the determining factor. In the previous work, basal levels (for mussels held at 13°C) of HSP70 isoforms and ubiquitin-tagged proteins are higher in *M. trossulus* than *M. galloprovincialis* suggesting that *M. trossulus* proteins are less thermally stable. In a different approach, the activity of proteins after electrophoretic separation and then heat treatment of showed no difference in the frequencies of heat sensitive alleles between cold-adapted *M. edulis* and warm-adapted *M. galloprovincialis* in SW England (Gardner and Skibinski 1990).

A functional link between a single allele and environmental selection has been established for *Mytilus edulis* in Long Island Sound, New York State. Regular annual selective mortality occurs with strong selection directed against juveniles with the *Lap⁹⁴* allele in response to low salinity, and this selection is reflected in the genetic makeup of adult populations. Mussels are osmoconformers, maintaining the osmotic pressure of their internal environment at the same levels as the external marine environment. One mechanism for achieving this is via the leucine aminopeptidase enzyme, encoded by a *Lap* gene, which hydrolyses oligopeptides to their component amino acids to increase the internal cell osmotic pressure in response to external increases in salinity. The *Lap⁹⁴* allele has greater catalytic efficiency which comes at the expense of greater losses of amines to the environment (Hilbish et al. 1982). Mussels carrying a *Lap⁹⁴* allele are at a selective disadvantage in low salinity environments and follow a sharp decrease in allele frequency from 0.55 in the ocean to 0.12 within Long Island Sound, a distance of 30 km (Hilbish & Koehn 1985). *Lap⁹⁴* is highly

dominant with biochemical analyses not able to distinguish between homozygotes and heterozygotes (Hilbish & Koehn 1985).

This pattern of selection was also observed in Wellington Harbour, New Zealand, where Lap^3 allele frequencies were lower in larger mussels at sites experiencing large variations in salinity while the allele frequency remained the same in small and large mussels (Southern hemisphere lineage *M. galloprovincialis*) at sites of low salinity variation (Gardner & Palmer 1998). In the Adriatic Sea, the MGE7[243] allele has the opposite pattern to the Long Island model: the heterozygote shows no changes in frequency over the cline while the homozygote is more frequent in low salinity environments. The MGE7[249] allele does, however, match the Lap^{94} profile, with the heterozygote and homozygote genotypes having the same trend across salinity changes and an association with high salinity sites. The frequency of the MGE7[249] allele is, however, lower than Lap^{94} ranging from 0 – 0.34 and 0.12 over all populations, compared to 0.12 – 0.5 for Lap^{94}.

There is some evidence of different tolerances to thermal stress in different *M. galloprovincialis* populations in the Mediterranean Sea. During the severe Mediterranean marine heatwave of 2003 there were mass mortality events across many invertebrate species (Galli et al. 2017), including total mortality of mussels in a small partially enclosed bay in north eastern Spain where temperatures reached 28°C for 2 weeks (Ramón et al. 2007). In contrast, an experiment testing responses to thermal stress of mussels from the Gulf of Thermaikos in Greece reported 30% mortality of mussels exposed to 28°C water temperatures for 15 days (Anestis et al. 2007). At 30°C there was 80% mortality after 15 days. If only small differences in the transcriptional response to thermal stress result in different tolerances across congeners (refer to Lockwood et al. 2015 – text below), then differences in thermal tolerances between different populations of the same species are also likely to be due to small differences in response and could potentially be related to just one gene. Intriguingly, the allele frequency of

the MGE7[243] allele in two populations in Kotor, CF and CW, which experienced maxSST over 30°C was 0.70.

At the level of transcriptional and post-translational responses to thermal and osmotic shock, several studies reviewed by Lockwood et al. (2015) showed that while overall transcriptional and proteomic responses to stresses are similar between *M. trossulus* and *M. galloprovincialis*, there are a few important differences that are implicated in individual species tolerances. For example, *M. trossulus* induces proteins involved in the ubiquitin pathway (protein degradation) more strongly, while *M. galloprovincialis* induced HSP24 (protein refolding). The authors speculate that since relatively few differences in the transcriptional response result in a +3°C differential in thermal tolerance between the species then possible future adaptation to increased global temperatures may be substantial, especially given the large effective population sizes of *Mytilus* species and the large extent of genetic diversity that they may contain. This speculation may be over-simplistic since it relies only on the transcriptional response of an organism to stress and ignores other levels of adaptation such as the thermal stability of proteins. In both species, there are differences between the responses to osmotic stress versus thermal stress. The response to hypo-osmotic stress at the transcriptional level is relatively muted in both species while post-translational modifications such as phosphorylation are more marked and are likely to occur more quickly than changes in transcription. This may reflect the need to respond more quickly to changes in salinity than in changes to temperature which may take time. There are differences between the responses of the congeners which lead to higher low salinity tolerance in *M. trossulus*.

Of great interest, given the correlation between temperature and salinity in this and other studies (Braby & Somero 2006b, Hamer et al. 2010, Nanninga et al. 2014) was that of 46 genes found to have changes in expression in response to both temperature and salinity

challenge, 45 of them responded in the opposite direction for the two stressors (Lockwood et al. 2015). For example, expression of ion-transport related genes is repressed in low salinity conditions and induced at high temperatures. The one gene that was induced in the same direction for both stressors encodes a thioredoxin reductase enzyme, involved in oxidative stress. Oxidative stress commonly occurs concurrently with environmental stresses, including hyposaline and thermal stress. Additional stresses such as ocean acidification and pollution, add synergistically to overall levels of oxidative stress (Michaelidis et al. 2014) and any adaptive mechanism that enhances the ability of mussels to respond to oxidative stress is expected to be favoured in the environments described above which experience extremes in both sea surface temperature and salinity (Hamer et al. 2008, Michaelidis et al. 2014).

Concluding remarks

MGE7 may be a candidate gene that matters to ecology, or linked to such a gene (Schmidt et al. 2008). The relationship between MGE7[243] allele frequency and environmental variables at a large scale has been shown (down the east coast) and suggests the possible influence of selection. This association between an allele frequency and environmental clines would be more compelling if parallel clines could be demonstrated at other spatial scales or in another location. Next steps could include examining allele frequencies in an estuary such as Lim, in northern Croatia, or any other smaller scale section of coast with a wide range of environmental readings, such as the Boka Kotorska Bay in Montenegro. If allele shifts are replicated in other environmental clines, then the locus can be considered a strong candidate locus for selection (Schmidt et al. 2008).

Chapter 4: General Discussion

The overall aim of this thesis was to investigate the genetic structure and population connectivity of the Mediterranean blue mussel, *Mytilus galloprovincialis*, in part of its native range using neutral microsatellite markers. Population genetic methods were used to test hypotheses of genetic differentiation and connectivity derived from patterns of circulation and particle transport in the Adriatic Sea (Liubartseva et al. 2016, Bray et al. 2017). In a seascape genetics approach, statistical methods were used to identify associations between genetic and environmental data (Manel et al. 2003). The discovery of an outlier locus, MGE7, which might itself be under selection or be linked to a gene under selection, presented the opportunity to detect genetic differentiation potentially based on selective processes. This is the first time that an association between MGE7 allele frequencies and environmental variables has been reported.

The ecological and evolutionary processes that influence population structure and diversity in marine organisms can be difficult to elucidate because gene flow is often mediated by a planktonic larval stage that cannot be measured directly (Waples 1998, Cowen & Sponaugle 2009). Indirect methods of inferring connectivity through population genetic approaches using neutral codominant genetic markers, such as allozymes, microsatellites and SNPs, have had variable success (Palumbi 2003, Gagnaire et al. 2015). In some cases, these methods have been able to estimate genetic connectivity and detect barriers to gene flow (Diz & Presa 2008, Schiavina et al. 2014, Pascual et al. 2017) but in others, genetic differentiation has not been detected, even in widely dispersed or seemingly isolated populations (Waples 1998, Addison et al. 2008, Allendorf et al. 2008). Increasingly, loci under selection or neutral

loci linked to genes under selection (hitchhiker loci) have been employed as well as neutral markers to detect population structure (Gaggiotti et al. 2009, White et al. 2010, Wei et al. 2013a). One of the advantages of using markers under selection for marine organisms with large effective populations is that the ability of selection to counteract the homogenising effect of gene flow and to thereby produce population structure is enhanced with larger populations. For neutral population processes, the ability of genetic drift to counteract gene flow decreases with effective population size making detection of genetic structure more difficult (Gagnaire et al. 2015).

Neutral markers can reveal genetic differentiation between populations that is informative regarding neutral evolutionary processes such as gene flow and genetic drift (Liggins et al. 2013). Thus, currents, gyres and eddies might affect neutral population structure by promoting or inhibiting gene flow, while favourable environments which can support large breeding populations might reduce genetic drift by lowering the likelihood of chance events altering allelic frequencies. In contrast, markers under selection may reveal differentiation between populations based on underlying environmental differences linked to the mechanism of selection (Gaggiotti et al 2009, White et al. 2010).

The qualitative difference in the type of information revealed through the neutral and selected markers has been shown in a number of studies. For example, in a study on the roundnose grenadier (*Coryphaenoides rupestris*), neutral loci allowed a barrier to gene flow coinciding with the subpolar front to be identified, while a locus under selection revealed genetic structuring related to depth (White et al. 2010). For the Atlantic herring, *Clupea harengus*, genetic differences due to feeding migrations were closely associated with variation in neutral loci, while salinity was associated with an outlier locus (Gaggiotti et al. 2009). Neutral loci showed only low levels of genetic differentiation in the greenshell mussel, *Perna*

canaliculus, while one outlier locus revealed a north-south genetic discontinuity which has subsequently been shown to be present in many marine taxa (Wei et al. 2013a). In this thesis, there was little evidence of genetic differentiation among 18 populations sampled along the entire eastern Adriatic Sea coast using only neutral loci, but a strong association was found between environmental variables maxSST, minSAL and maxCHL-a and the outlier locus MGE7.

The correlation of MGE7 alleles with environmental variables on a large scale (the east coast of the Adriatic Sea) suggests there may be a selective basis to the distribution of MGE7 alleles. While the seascape genetics approach used here can detect associations between genetic structure and biologically relevant environmental variables, it cannot detect a causal relationship (Liggins et al 2013). Before performing experiments capable of demonstrating a mechanistic link between the genetic and environmental correlations, a number of tests using the same seascape genetics approach can be employed to ascertain the generality of the association over multiple locations and scales. For intertidal organisms, Schmidt et al. (2008) made predictions of how candidate genes under selection would behave at multiple scales and parallel environments in the North Atlantic. If selection is a driver of the genetic variation observed, then a clinal association observed at one spatial scale may also be observed at other spatial scales (latitudinal, estuarine, intertidal) and in multiple locations (east and west coasts). These ideas could be extended to the Adriatic Sea where, for example, variations in salinity exist at the basin-scale (Lipizer et al. 2014), within bays (e.g. Kotor Bay, Mandić et al. 2017) and at the estuarine scale (e.g. Lim estuary, Hamer et al. 2010, Kovačić et al. 2017). Future experiments can establish whether the association of MGE7 extends to the smaller scale, such as Kotor Bay in Montenegro, or Lim estuary in northern Croatia. Similarly, if a clinal

relationship was observed along the parallel western Adriatic coast, this would be strong circumstantial evidence of selection, especially if observed at widely-spaced estuaries.

Variations in allele frequencies observed in different sized mussels (juvenile vs. adults) and mussels sampled at different times of year may also be suggestive of selection and may point to the time when selection is occurring. For example, selection related to the *Lap* locus occurs regularly at the juvenile stage and populations in low salinity environments show corresponding size-dependent differences in allele frequencies (Hilbish & Koehn 1985, Gardner & Palmer 1998). In the Adriatic Sea, the timing of major spawning, dispersal and settlement events coincide quite well with annual patterns of low salinity that typically occur in spring and early summer (da Ros et al. 1985, Hamer et al. 2010, Krivokapić et al 2011). It is plausible that, if the selective pressure linked to MGE7 is related to minSAL, that selection occurs at early life stages every generation. If selection is related to maxSST, selection may not occur uniformly across years, but rather coincide with years of marine heatwaves when mass mortality is seen across many marine taxa (Di Camillo et al. 2015 and Galli et al. 2017). Selection could also occur differentially at the same site depending on the location in the intertidal zone and exposure to the sun (Gardner & Skibinski 1990). Judicious sampling at different times of the year and over several years of different sized mussels may enable the timing of selection to be established which could also shed light on which environmental variable is most important to selection.

Future experiments that could be used to explore possible mechanistic links between MGE7 allele or genotype frequency and environmental variables include common garden experiments, where mussels of known genotype are exposed to certain environmental conditions (Liggins et al. 2013). These experiments could be used to pick apart the complex picture of varying correlations between the different MGE7 genotypes and environmental

variables uncovered in Chapter 3 (Tables 3.9 and 3.10). Because the MGE7 allele and genotype frequencies were correlated with environmental variables that seem to be descriptors of particular coastal environments, then, if selection is driving these patterns, they could be related to a general stress response rather than to a single environmental variable. This suggests experiments which target particular responses to stress, such as MAPK signaling (Hamer et al. 2008, Evans & Somero 2010), valve closure (Braby & Somero 2006a, Anestis et al. 2007) or responses to protein damage such as the heat shock response (Lockwood et al. 2015, Chen et al. 2018). There are also a large number of polymorphic allozymes of known function which could be checked for association with MGE7 which was derived from EST-sequence (Yu & Li 2007, Schmidt et al. 2008). The protein sequence of these allozymes could be examined for the existence of the MGE7 microsatellite (depending on the reading frame, a CAA (or TTG) repeat motif can be translated into several different amino acids). Although it is unlikely that any of these allozymes correspond to the MGE7 locus, any association between a particular allozyme and the MGE7 locus could indicate that the functions of the allozyme and the marker are similar, that they are involved in the same regulatory network or that they are polygenic (quantitative trait loci for the same phenotypic trait). Allozyme assays could therefore be performed to check for correspondence with MGE7 alleles and to compare distribution patterns along the environmental cline.

Although whole-genome sequencing has been conducted on *Mytilus galloprovincialis*, only a limited nuclear DNA genome sequence is so far available, with 39 of 248 core eukaryotic genes completely sequenced (Murgarella et al. 2016). A BLAST search of the 534 bp MGE7 sequence did not result in any significant hits. Given that the genome of the Pacific oyster, *Crassostrea gigas*, has been sequenced, complete sequencing of the *Mytilus galloprovincialis* genome should be a priority (Zhang et al. 2012). In the meantime, a technique such as primer walking (Shyamala & Ames 1989) could be employed to provide additional sequence

information around the existing MGE7 EST-sequence which may allow identification or likely function of a putative linked gene.

The function of MGE7 could be explored through gene expression profiling and proteomics. Differences in the level of transcripts have been observed for *M. galloprovincialis* and *M. trossulus* subjected to thermal and hypo-osmotic stress (Lockwood et al. 2015). This method could be extended to conspecific individuals with different genotypes. If the putative gene under selection is involved in the regulation of transcription at any level, differences in gene expression may be observed in individuals with different genotypes. Microsatellites have known functions in gene regulation such as transcription factor binding, enhancer functions and spacing between regulatory elements, however, because MGE7 is EST-derived, a regulatory role may be less likely (Sawaya et al. 2013, Vieira 2016, Bagshaw 2017). Comparative proteomic approaches have also identified differential expression of proteins involved in stress responses in *M. galloprovincialis* and *M. trossulus* (Tomanek 2012) and again, this approach could be extended to different *M. galloprovincialis* genotypes and different experimental stressors (hyposaline stress, heat stress, oxidative stress) to probe differential protein expression.

The small number of loci available in this study is a limitation and does raise the question of whether the locus MGE7 can really be regarded as an outlier. With the loci available in this study there is a clear line between results including MGE7 which detected genetic structure and results excluding MGE7 that did not find any structure (summarised in Table 2.13). If, however, ten more microsatellite loci were added to the analyses, this line might become clearer still or more blurred; the outlier status of MGE7 might be strengthened or weakened. Each microsatellite locus has its own distinct evolutionary history so there is natural variation in the ability of each locus to detect neutral evolutionary processes (Selkoe & Toonen 2006). The correlation between latitude, environmental variables and MGE7 allele frequencies

has been considered from the perspective of possible selective processes. But how likely are these patterns from the perspective of neutral processes? If MGE7 was a neutral marker and the pattern of allelic distribution was a result of regional differentiation then would the observed correlations of MGE7^{243} with latitude, maxSST and minSAL have been observed? Clearly, there is scope for further research here to test the spatial and temporal pattern of allelic frequency differences, and for the process(es) responsible for it, to be better understood.

In Chapter 2 the regional structure detected in the AMOVA analyses when MGE7 was included was explained by the correlation with latitude: that is, if the correlation between MGE7^{243} and latitude is given, then certain results would be expected from AMOVA analyses based on datasets including the MGE7 locus. The differentiation detected was exactly as predicted, with greater differentiation found between the northern and southern regions than between the northern and central or central and southern regions. Here, we are considering the reverse situation: if the regional differentiation detected by a "neutral" MGE7 locus and mediated by the three Adriatic Sea gyres is given, then we would predict a correlation between latitude and MGE7 allele frequencies, which would, in turn, lead to a correlation between MGE7 and the environmental variables, since by chance, there is a strong correlation between latitude and environmental variables. This second correlation (between MGE7 and environmental variables) would, however, be expected to be weaker, which is not the case. The correlation between f(MGE7^{243}) and minSAL ($R^2 = 0.635$, $p < 0.001$) is stronger than the correlation between f(MGE7^{243}) and latitude ($R^2 = 0.476$, $p = 0.002$). A more cogent argument against the hypothesis of neutrality at the MGE7 locus is that it cannot explain the striking difference in the correlation between minSAL and MGE7$^{243/243}$ homozygotes and MGE7$^{243/X}$ heterozygotes shown in Figure 3.5.

Turning back to a scenario where the differences in MGE7 allele frequencies are associated with the environmental variables, as discussed in Chapter 3, the correlation between MGE7 alleles and latitude was described as being contingent on the correlation between MGE7 and the environmental variables. It is, however, possible that it is the correlation with the environmental variables, and not latitude, that is coincidental. A possible explanation of a pattern of MGE7 distribution related to latitude not environmental variables, is that a favourable new variant of a gene linked to allele MGE7[243] (and possibly also MGE7[240]) has arisen outside the Adriatic Sea and is now being propagated northward through the Adriatic Sea. In this scenario, the gene under selection would be linked to MGE7, not MGE7 itself, since MGE7[243] already exists at a considerable frequency in the northern populations. Because of the recessive nature of the advantageous allele (only homozygotes show the correlation) the change in allele frequency is initially slow: most early instances of the rare new variant are in heterozygotes but once allele frequencies increase to a level where homozygotes start appearing more often the frequency of the allele increases quickly (Andrews 2010). In this scenario, because the MGE7 locus is a hitchhiker, a northward progression of the frequency shift would be observed, together with a gradual erosion of the association between the new gene variant and MGE7 due to recombination events so that the hitchhiking effect softens as the favourable mutation travels away from its origin (Gagnaire et al. 2015). This is an unlikely scenario because it is a transient phenomenon which would only be observed while the favourable gene variant is being propagated northward and while the MGE7[243] allele is still associated by linkage with the favourable variant. This scenario could be tested by sampling over evenly spaced populations over several years (Andrews 2010).

Based on only neutral loci, populations along the eastern coast of the Adriatic Sea were close to panmictic. Genetic differentiation has been found between a population in Ravenna,

on the northwestern Adriatic coast, and Zadar on the eastern coast using microsatellites (not MGE7) and mitochondrial sequence (Giantsis et al. 2014a & b). Although these studies covered 13 and 8 populations, respectively, from the central and eastern Mediterranean Sea, only two populations from the Adriatic Sea were covered, comprising 36 individuals each in the microsatellite study. The original hypothesis of regional genetic differentiation defined by the three Adriatic Sea gyres could be replaced by an hypothesis of genetic differentiation between the northwestern Adriatic Sea and the Croatian coast mediated by the strong southward flow of the Western Adriatic Current, which originates just south of the Po River delta and north of Ravenna (Lipizer et al. 2014), and high levels of larval retention predicted for the Croatian coast from models of larval dispersal (Dubois et al. 2016, Bray et al. 2017).

The association between MGE7 alleles and the environmental variables maxSST, minSAL and maxCHL-a may have important relevance for *M. galloprovincialis* with regards to climate change and aquaculture management. Climate change is characterised not only by an overall warming trend and ongoing ocean acidification, but also by increased frequencies of extreme weather events such as marine heatwaves and extreme precipitation leading to low salinity conditions (Lejeusne et al. 2010, Di Camillo & Cerrano 2015, Galli et al. 2017). The changing climate has altered the geographic range of many species and is likely to be implicated in the success of *M. galloprovincialis* as an invasive species (Tomanek 2012). It has supplanted native *M. trossulus* in southern California possibly due to its higher tolerance to warm temperatures (Schneider & Helmuth 2007, Tomanek 2012). Climate models predict that the Mediterranean Sea will be one of the regions hardest hit by warming trends and conditions are already close to the upper limit of tolerance for blue mussels with several instances of mass mortality associated with marine heatwaves in Spain, Greece and in the Adriatic Sea (Lejeusne et al. 2010, Michaelidis et al. 2014, Galli et al. 2017, personal communication J Gardner). In

the Mediterranean Sea moreover, shifts in species' distributional ranges are limited by the northern boundaries of the sea.

High temperatures are only one example of the environmental stress confronting marine species in the Mediterranean Sea. Stressors such as ocean acidification may also be driven by climate change, but other stressors are more generally anthropogenic, such as increasing levels of pollution, eutrophication and habitat destruction (Lejeusne et al 2010). There is a synergistic effect when two or more stressors coincide with the range of thermal tolerance narrowing. Organisms exposed to heavy metals have increased sensitivity to thermal stress and heat stressed organisms are more vulnerable to disease (Anestis et al. 2010, Michaelidis et al. 2014). Environmental stresses increase energy demands generally in order to maintain cellular homeostasis, repair damaged proteins and detoxify reactive oxygen species (ROS). There is also a trade-off between energy metabolism which generates ROS and expending energy to detoxify ROS which can reach a tipping point under environmental stress (Michaelidis et al. 2014). In addition, increased energy demand due to thermal stress can often not be met through intake of food, since food availability is at its lowest during the summer throughout the Mediterranean Sea and also because thermal stress impairs the ability of mussels to ingest food (Anestis et al. 2007, Krivokapić et al 2011, Michaelidis et al. 2014).

The findings of this study have important implications for management of aquaculture and for conservation in relation to climate change. Firstly, the possible relationship of MGE7 and selection related to environmental stress underlines the importance of natural genetic variation in blue mussels as climate challenges intensify. Although relatively few changes in transcriptional regulation between *M. galloprovincialis* and *M. trossulus* resulted in substantial increases in thermal tolerances between the congeners (Lockwood et al. 2015), the severe impact of heat stress on highly evolved and conserved systems such as energy metabolism and

detoxification argue that there is little room for further evolutionary adaptation to thermal stress (Tomanek 2012, Michaelidis et al. 2014).

Secondly, the existence of genetic differentiation between different environments argues for care in aquaculture where the transfer of spat from natural spawning grounds to aquaculture sites with favourable conditions for growth is widely practiced (Giantsis et al. 2014a, Kovačić et al. 2017). Aquaculture sites are often in bays and lagoons which can experience greater variation in temperature and salinity than the open sea (Galli et al. 2017). The variation in MGE7[243] alleles with minSAL and maxSST suggests that a large proportion of juveniles collected from areas of low environmental variability may not have the genetic background to survive and grow in areas of high environmental variability. Therefore, since the traditional methods of mussel farming employed along the eastern Adriatic coast are extremely labour intensive the matching of environmental conditions between the natal origin and the aquaculture site, as far as possible, is recommended (Mandić et al. 2017).

Thirdly, recovery of mussel populations from marine heatwaves and mass mortality events will depend on the ability of larvae to disperse from surviving populations to affected areas. The difficulty of estimating larval dispersal and gene flow using classical population genetic methods, especially when effective populations are large, limits the ability of conservationists and aquaculture management to predict which areas are vulnerable to long-term depopulation (Palumbi 2003, Gagnaire et al. 2015). The level of migration necessary to maintain genetic connectivity and which may occur from relatively rare long-distance dispersal events, is very much lower than the migration needed to repopulate an area after mass mortality. As discussed in Chapter 2, the retention of the majority of mussel larvae within an area of 20 − 40 km has been found in a number of studies on *Mytilus* species using different methodologies and based in different geographic regions (McQuaid & Phillips 2000, Gilg & Hilbish 2003, Becker et al. 2007, Stuckas et al. 2017). This suggests that recovery from mass mortality events

over large areas will take time. Development of species-specific biophysical modelling which incorporates fine scale hydrodynamic modelling, larval behaviour and biological processes during spawning and recruitment is a priority.

Blue mussels are a common intertidal species along the mostly rocky eastern coast of the Adriatic Sea (Hamer et al. 2010, UNEP/MAP-RAC/SPA 2015). They are important to ecosystem functioning, as ecosystem engineers and as filter-feeders, and important economically in aquaculture (Fly et al. 2015, Kovačić et al. 2017). The predicted increase in frequency and severity of marine heatwaves and other climate change effects such as heavy rains with sudden drops in salinity pose a serious risk to low-motility organisms such as mussels (Harley et al. 2006, Galli et al 2017). Loss of large proportions of filter feeding animals will impact ecosystem functioning and water quality and increase the risk of eutrophication and further mass mortality events (Di Camillo & Cerrano 2015). Conservation and aquaculture practices should therefore be directed at maintaining genetic diversity, and monitoring the quality of the environment to reduce the synergistic effects of pollution, eutrophication and thermal stress (Harley et al. 2006, Lejeusne et al. 2010).

References

Addison JA, Ort BS, Mesa KA, Pogson GH (2008) Range-wide genetic homogeneity in the California sea mussel (*Mytilus californianus*): a comparison of allozymes, nuclear DNA markers, and mitochondrial DNA sequences. *Molecular Ecology*, **17**, 4222-4232.

Allendorf FW, England PR, Luikart G, Ritchie PA, Ryman N (2008) Genetic effects of harvest on wild animal populations. *Trends in Ecology and Evolution*, **23**(6), 327-337.

Amaral AR, Beheregaray LB, Bilgmann K, Boutoy D, Freitas L, Robertson KM, Sequeira M, Stockin KA, Coelho MM, Möller LM (2012) Seascape Genetics of a Globally Distributed, Highly Mobile Marine Mammal: The Short-Beaked Common Dolphin (Genus *Delphinus*). *PLoS one*, **7**(2), e31482.

Anderson MJ, Gorley RN, Clarke KR (2008) Permanova+ for PRIMER: Guide to Software and Statistical Methods. PRIMER-e, Plymouth, UK.

Andrews CA (2010) Natural Selection, Genetic Drift, and Gene Flow Do Not Act in Isolation in Natural Populations. *Nature Education Knowledge*, **3**(10), 5.

Anestis A, Lazou A, Pörtner HO, Michaelidis B (2007) Behavioural, metabolic, and molecular stress responses of marine bivalve *Mytilus galloprovincialis* during long-term acclimation at increasing ambient temperature. *American journal of physiology, regulatory, integrative and comparative physiology,* **293**, R911-R921.

Anestis A, Pörtner HO, Karagiannis D, Angelidis P, Staikou A, Michaelidis B (2010) Response of *Mytilus galloprovincialis* (L.) to increasing seawater temperature and to marteliosis: Metabolic and physiological parameters. *Comparative Biochemistry and Physiology, Part A,* **156**, 57-66.

Apte S, Holland BS, Godwin LS, Gardner JPA (2000) Jumping ship: a stepping stone event mediating transfer of a non-indigenous species via a potentially unsuitable environment. *Biological Invasions*, **2**, 75-79.

Araneda C, Noriega X, Jilberto F, Ruíz J, Jara C, Lafarga F, Larraín MA (2016) Development of genome-wide microsatellite resources in a commercially important mussel species (*Mytilus chilensis*), *Animal Genetics*, **47**(3), 387-388.

Archambault P, McKindsey CW, Bourget E (1999) Large-scale Shoreline Configuration Influences Phytoplankton Concentration and Mussel Growth. *Estuarine, Coastal and Shelf Science*, **49**, 193-208.

Arribas LP, Donnarumma L, Paloma MG, Scrosati RA (2014) Intertidal mussels as ecosystem engineers: their associated invertebrate biodiversity under contrasting wave exposures. *Marine Biodiversity*, **44**, 203-211.

Artegiani A, Bregant D, Paschini E, Pinardi N, Raicich F, Russo A (1997) The Adriatic Sea General Circulation. Part I: Air-Sea Interactions and Water Mass Structure. *Journal of Physical Oceanography*, **27**, 1492-1514.

Bagshaw ATM (2017) Functional Mechanisms of Microsatellite DNA in Eukaryotic Genomes. *Genome Biology and Evolution*, **9**(9), 2428-2443.

Banks SC, Peakall R (2012) Genetic spatial autocorrelation can readily detect sex-biased dispersal. *Molecular Ecology*, **21**, 2092-2105

Beaumont AR, Turner G, Wood AR, Skibinski DOF (2004) Hybridisations between *Mytilus edulis* and *Mytilus galloprovincialis* and performance of pure species and hybrid veliger larvae at different temperatures. *Journal of Experimental Marine Biology and Ecology*, **302**(2), 177-188.

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society B*, **263**, 1619-1626.

Becker BJ, Levin LA, Fodrie FJ, McMillan PA (2007) Complex larval connectivity patterns among marine invertebrate populations. *Proceedings of the National Academy of the United States of America*, **104**(9), 3267-3272.

Bellafiore D, Guarnieri A, Grilli F, Penna P, Bortoluzzi G, Giglio F, Pinardi N (2011) Study of the hydrodynamical processes in the Boka Kotorska Bay with a finite element model. *Dynamics of Atmospheres and Oceans*, **52**, 298-321.

Bierne N, Roze D, Welch JJ (2013) Pervasive selection or is it…? Why are F_{ST} outliers sometimes so frequent? *Molecular Ecology*, **22**, 2061-2064.

Böhm E, Banzon V, D'Acunzo E, D'Ortenzo F, Santoleri R (2003) Adriatic Sea surface temperature and ocean colour variability during the MFSPP. *Annales Geophysicae*, **21**, 137-149.

Braby CE, Somero GN (2006a) Following the heart: temperature and salinity effects on heart rate in native and invasive species of blue mussels (genus *Mytilus*). *Journal of Experimental Biology, **209**,* 2554-2566.

Braby CE, Somero GN (2006b) Ecological gradients and relative abundance of native (*Mytilus trossulus*) and invasive (*Mytilus galloprovincialis*) blue mussels in the California hybrid zone. *Marine Biology*, **148**, 1249-1262.

Branch GM, Steffani CN (2004) Can we predict the effects of alien species? A case-history of the invasion of South Africa by *Mytilus galloprovincialis* (Lamarck). *Journal of Experimental Marine Biology and Ecology*. **300**, 189-215.

Bray L, Kassis D, Hall-Spencer JM (2017) Assessing larval connectivity for marine spatial planning in the Atlantic. *Marine Environmental Research*, **125**, 73-81.

Carson HS, López-Duarte PC, Rasmussen L, Wang D, Levin LA (2010) Reproductive Timing Alters Population Connectivity in Marine Metapopulations. *Current Biology*, **20**(21), 1926-1931.

Castellano S, Balletto E (2002) Is the partial Mantel test inadequate? *Evolution*, **56**(9), 1871-1873.

Chambers GK, MacAvoy ES (2000) Microsatellites: consensus and controversy. *Comparative biochemistry and physiology. Part B, Biochemistry & molecular biology*, **126**, 455-476.

Chen B, Feder ME, Kang L (2018) Evolution of heat-shock protein expression underlying adaptive responses to environmental stress. *Molecular Ecology*, **27**, 3040-3054.

Clarke KR, Gorley RN (2006) PRIMER v6: User Manual/Tutorial. PRIMER-E, Plymouth.

Coop G, Witonsky D, Di Rienzo A, Pritchard JK (2010) Using Environmental Correlations to Identify Loci Underlying Local Adaptation. *Genetics*, **185**(4), 1411-1423.

Cowen RK, Sponaugle S (2009) Larval dispersal and marine population connectivity. *Annual Review of Marine Science*, **1**, 443-466.

Crawford DL, Power DA (1989) Molecular basis of evolutionary adaptation at the lactate dehydrogenase-B locus in the fish *Fundulus heteroclitus*. *Proceedings of the National Academy of the United States of America*, **86**, 9365-9369.

Da Ros L, Bressan M, Marin MG (1985) Reproductive cycle of the mussel (*Mytilus galloprovincialis* Lmk) in Venice Lagoon (North Adriatic). *Italian Journal of Zoology*, **52**(3-4), 223-229.

Daniels EQ, Litvaitis MK (2017) Does peripheral dislodgement contribute to heterozygote deficiencies in blue mussels? *Invertebrate Biology*, **136**(3), 301-308.

Di Camillo CG, Cerrano C (2015) Mass mortality events in the NW Adriatic Sea: Phase shift from slow- to fast- growing organisms. *PLoS ONE*, **10**(5), e0126689.

Diz AP, Presa P (2008) Regional patterns of microsatellite variation in *Mytilus galloprovincialis* from the Iberian Peninsula. *Marine Biology*, **154**, 227-286.

Diz AP, Presa P (2009) The genetic diversity pattern of *Mytilus galloprovincialis* in Galician Rías (NW Iberian estuaries). *Aquaculture*, **287**, 278-285.

Do C, Waples RS, Peel D, Macbeth GM, Tillett BJ, Ovenden JR (2014) NeEstimator V2: re-implementation of software for the estimation of contemporary effective population size (N_e) from genetic data. *Molecular Ecology Resources*, **14**(1), 209-214.

Drakulović D, Pestorić B, Cvijan M, Krivokapić S, Vuksanović N (2012) Distribution of phytoplankton community in Kotor Bay (south-eastern Adriatic Sea). *Central European Journal of Biology*. **7**(3), 470-486.

Dubois M, Rossi V, Ser-Giacomi E, Arnaud-Haond S, López C, Hernández-García E (2016) Linking basin-scale connectivity, oceanography and population dynamics for the

conservation and management of marine ecosystems. *Global Ecology and Biogeography*, **25**, 503-515.

Dupanloup I, Schneider S, Excoffier L (2002) A simulated annealing approach to define the genetic structure of populations. *Molecular Ecology*, **11**, 2571-2581.

Đurović B, Đurović I, Joksimović A, Crnojević V, Đukanović S, Pestorić B (2018) Monitoring the eutrophication using Landsat 8 in the Boka Kotorska Bay. *Acta Adriatica*, **59**(1), 17-34.

Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetic Resources*, **4**, 359-361.

Eldon B, Riquet F, Yearsley J, Jollivet D, Broquet T (2016) Current hypotheses to explain chaos under the sea. *Current Zoology*, **62**(6), 551-566.

Evanno G. Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611-2620.

Evans TG, Somero GN (2010) Phosphorylation Events Catalyzed by Major Cell Signalling Proteins Differ in Response to Thermal and Osmotic Stress among Native (*Mytilus californianus* and *Mytilus trossulus*) and Invasive (*Mytilus galloprovincialis*) Species of Mussels. *Physiological and Biochemical Zoology*, **83**(6), 984–996

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

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479-91.

FAO (2005-2019) National Aquaculture Sector Overview Fact Sheets: Albania, text by Cobani M, updated 6 March 2015; Bosnia and Herzegovina, text by Hamzic A, updated 1 February 2005; Croatia, text by Pecigoš B, updated 10 March 2017; Montenegro, text by Krasic M, Mandic M, updated 27 March 2017. In *FAO Fisheries and Aquaculture Department* [online]. Rome. [Cited 20 January 2019].

FAO (2018) FAO Yearbook of Fishery and Aquaculture Statistics 2016. Aquaculture Production.

Fly EK, Hilbish TJ, Wethey DS, Rognstad RL (2015) Physiology and biogeography: The response of European mussels (*Mytilus* spp.) to climate change. *American Malacological Bulletin*, **33**(1), 136-149.

Foll M, Gaggiotti O (2006) Identifying the Environmental Factors That Determine the Genetic Structure of Populations. *Genetics*, **174**, 875-891.

Frankic A (2003) Integrated Coastal Management & Sustainable Aquaculture Development in the Adriatic Sea, Republic of Croatia. Downloaded from http://ccrm.vims.edu/publications/pubs/Adriaticaquaculture.pdf on 27-02-2019.

Gaggiotti OE, Bekkevold D, Jørgensen HBH, Foll M, Carvalho GR, Andre C, Ruzzante DE (2009) Disentangling the effects of evolutionary, demographic, and environmental factors influencing genetic structure of natural populations: Atlantic herring as a case study. *Evolution*. **63**(11), 2939-2951.

Gagnaire PA, Broquet T, Aurelle D, Viard F, Souissi A, Bonhomme F, Arnaud-Haond S, Bierne N (2015) Using neutral, selected, and hitchhiker loci to assess connectivity of marine populations in the genomic era. *Evolutionary Applications*, **8**, 769-786.

Galli G, Solidoro C, Lovato T (2017) Marine Heat Waves Hazard 3D Maps and the Risk for Low Motility Organisms in a Warming Mediterranean Sea. *Frontiers in Marine Science*. **4**, 136.

Gao X, Starmer JD (2008) AWclust: point-and-click software for non-parametric population structure analysis. *BMC Bioinformatics*, **9**, 77.

Gardeström J, Pereyra RT, André C (2008) Characterization of six microsatellite loci in the Baltic blue mussel *Mytilus trossulus* and cross-species amplification in North Sea *Mytilus edulis*. *Conservation Genetics*, **9**, 1003-1005.

Gardner JPA, Palmer NL (1998) Size-dependent, spatial and temporal genetic variation at a leucine aminopeptidase (LAP) locus among blue mussel (*Mytilus* galloprovincialis) populations along a salinity gradient. *Marine Biology*, **132**, 275-281.

Gardner JPA, Skibinski DOF (1988) Historical and size-dependent genetic variation in hybrid mussel populations. *Heredity*, **61**, 93-105.

Gardner JPA, Skibinski DOF (1990a) Genotype-dependent fecundity and temporal variation of spawning in hybrid mussel (*Mytilus*) populations. *Marine Biology*. **105**, 153-162.

Gardner JPA, Skibinski DOF (1990b) Thermostability differences of allozyme loci in *Mytilus edulis*, *Mytilus galloprovincialis* and hybrid mussels. *Marine Ecology Progress Series*, **64**, 99-105.

Gardner JPA, Skibinski DOF (1991) Biological and physical factors influencing genotype-dependent mortality in hybrid mussel populations. *Marine Ecology Progress Series*, **71**, 235-243.

Gardner JPA, Zbawicka M, Westfall KM, Wenne R (2016) Invasive blue mussels threaten regional scale genetic diversity in mainland and remote offshore locations: the need for baseline data and enhanced protection in the Southern Ocean. *Global Change Biology*, **22**, 3182-3195.

Gavrilovic A, Jug-Dujakovic J, Marinovic Bonacic A, Conides A, Bonacic K, Ljubicic A, Van Gorder S (2011) The influence of environmental parameters on the growth and meat quality of the Mediterranean mussel *Mytilus galloprovincialis* (Mollusca: Bivalvia).

Aquaculture, Aquarium, Conservation and Legislation International Journal of the Bioflux Society, **4**(5), 573-583.

Gérard K, Roby C, Bierne N, Borsa P, Féral JP, Chenuil A (2015) Does natural selection explain the fine scale genetic structure at the nuclear exon *Glu-5'* in blue mussels from Kerguelen? *Ecology and Evolution*, **5**, 1456-1473.

Giantsis IA, Kravva N, Apostolidis AP (2012) Genetic characterization and evaluation of anthropogenic impacts on genetic patterns in cultured and wild populations of mussels (*Mytilus galloprovincialis*) from Greece. *Genetics and Molecular Research*, **11**, 3814-3823.

Giantsis IA, Mucci N, Randi E. Abatzopoulos TJ, Apostolidis AP (2014a) Microsatellite variation of mussels (*Mytilus galloprovincialis*) in central and eastern Mediterranean: genetic panmixia in the Aegean and the Ionian Seas. *Journal of the Marine Biological Association of the United Kingdom*, **94**(4), 797-809.

Giantsis IA, Abatzopoulos TJ, Angelidis P, Apostolidis AP (2014b) Mitochondrial Control Region Variability in *Mytilus galloprovincialis* Populations from the Central-Eastern Mediterranean Sea. *International Journal of Molecular Sciences*, **15**(7), 11614-11625.

Gilg MR, Hilbish TJ (2003) The geography of marine larval dispersal: coupling genetics with fine-scale physical oceanography. *Ecology*, **84**(11), 2989-2998.

Global Invasive Species Database (2019) Species profile: *Mytilus galloprovincialis*. Downloaded from http://www.iucngisd.org/gisd/speciesname/Mytilus+galloprovincialis on 27-02-2019.

Gosling EM (2003) *Bivalve Molluscs: Biology, Ecology and Culture.* Fishing News Books, a division of Blackwell Publishing, Oxford, UK.

Griffiths CL, Hockey PAR, Van Erkom Schurink C, Le Roux PJ (1992) Marine invasive aliens on South African shores: Implications for community structure and trophic functioning. *South African Journal of Marine Science*, **12**, 713-722.

Gruber MAM, Hoffmann BD, Ritchie PA, Lester PJ (2013) The conundrum of the yellow crazy ant (*Anoplolepis gracilipes*) reproductive mode: no evidence for dependent lineage genetic caste determination. *Insectes Sociaux*, **60**, 135–145.

Guillot G, Estoup A, Mortier F, Cosson JF (2005) A spatial statistical model for landscape genetics. *Genetics*, **170**, 1261-1280.

Guillot G, Leblois R, Coulon A, Frantz AC (2009) Statistical methods in spatial genetics. *Molecular Ecology*. **18**, 4734-4756.

Guillot G, Renaud S, Ledevin R, Michaux J, Claude J (2012) A unifying model for the analysis of phenotypic, genetic, and geographic data. *Systematic Biology*, **61**, 897-911.

Hafner D, Jasprica N, Car A (2018a) Taxonomic Survey of Benthic Diatoms in Neum Bay, Southeastern Adriatic. *Natura Croatica*, **27**(1), 1-26.

Hafner D, Car A, Jasprica N, Kapetanović T, Dupčić Radić I (2018b) Relationship between marine epilithic diatoms and environmental variables in oligotrophic bay, NE Mediterranean. *Mediterranean Marine Science*, **19**(2), 223-239.

Hamer B, Jakšić Ž, Pavičić-Hamer D, Perić L, Medaković D, Ivanković D, Pavičić J, Zilberberg C, Schröder HC, Müller WEG, Smodlaka N, Batel R (2008) Effect of hypoosmotic stress by low salinity acclimation of Mediterranean mussels *Mytilus galloprovincialis* on biological parameters used for pollution assessment. *Aquatic Toxicology*, **89**, 137-151.

Hamer B, Medaković D, Pavičić-Hamer D, Jakšić Ž, Štifanić M, Nerlović V, Travizi1 A, Precali R, Kanduč T (2010) Estimation of freshwater influx along the eastern Adriatic coast as a possible source of stress for marine organisms. *Acta Adriatica*, **51**(2), 181-194.

Hamer B, Korlević M, Durmiši E, Nerlović V, Bierne N (2012) Nuclear marker Me 15/16 analyses of *Mytilus galloprovincialis* populations along the eastern Adriatic coast. *Cahiers de Biologie Marine*, **53**, 35-44.

Harrang E, Lapègue S, Morga B, Bierne N (2013) A High Load of Non-neutral Amino-Acid Polymorphisms Explains High Protein Diversity Despite Moderate Effective Population Size in a Marine Bivalve With Sweepstakes Reproduction. *G3: Genes, Genomes, Genetics*, **3**(2), 333-341.

Harley CDG, Hughes AR, Hultgren KM, Miner BG, Sorte CJB, Thomber CS, Rodriguez LF, Tomanek L, Williams SL (2006) The impacts of climate change in coastal marine systems. *Ecology Letters*, **9**, 228-241.

Hedgecock D (1994) Does variance in reproductive success limit effective population sizes of marine organisms? In *Genetics and evolution of aquatic organisms* (ed Beaumont A), pp 122-134. Chapman and Hall, London.

Hedgecock D, Li G, Hubert S, Bucklin K, Ribes V (2004) Widespread null alleles and poor cross-species amplification of microsatellite DNA loci cloned from the Pacific oyster, *Crassostrea gigas*. *Journal of Shellfish Research*, **23**(2), 379-385.

Hedgecock D, Barber PH, Edmands S (2007) Genetic Approaches to Measuring Connectivity. *Oceanography*, **20**(3), 70-79.

Hedgecock D, Pudovkin AI (2011) Sweepstakes Reproductive Success in highly fecund marine fish and shellfish: a review and commentary. *Bulletin of Marine Science*, **87**(4), 971-1002.

Hilbish TJ, Deaton LE, Koehn RK (1982) Effect of an allozyme polymorphism on regulation of cell volume. *Nature*. **298**, 688-689.

Hilbish TJ, Koehn RK (1985) Dominance in Physiological Phenotypes and Fitness at an Enzyme Locus. *Science*, **229**, 52-54.

Hilbish TJ, Mullinax A, Dolven SI, Meyer A, Koehn RK, Rawson PD (2000) Origin of the antitropical distribution pattern in marine mussels (*Mytilus* spp.): routes and timing of transequatorial migration. *Marine Biology,* **136**, 69-77.

Hill WG (1981) Estimation of effective population size from data on linkage disequilibrium. *Genetical Research Cambridge*, **38**, 209-216.

Hofmann GE Somero GN (1995) Evidence for protein damage at environmental temperatures: seasonal changes in levels of ubiquitin conjugates and HSP70 in the intertidal mussel *Mytilus trossulus. The Journal of Experimental Biology*, **198**, 1509-1518.

Holderegger R, Kamm U, Gugerli F (2006) Adaptive vs. neutral genetic diversity: implications for landscape genetics *Landscape Ecology*, **21**, 797-807.

Holderegger R, Wagner HH (2006) A brief guide to Landscape Genetics. *Landscape Ecology*, **21**, 793-796.

International Hydrographic Organization (1953) Limits of Oceans and Seas (Special Publication No 28). p17 [online: https://www.iho.int/iho_pubs/standard/S-23/S-23_Ed3_1953_EN.pdf]. Montecarlo. [Cited 22 January 2019].

Jenkins DG, Carey M, Czerniewska J, Fletcher J, Hether T, Jones A, Knight S, Knox J, Long T, Mannino M, McGuire M, Riffle A, Segelsky S, Shappell L, Sterner A, Strickler T, Tursi R (2010) A meta-analysis of isolation by distance: relic or reference standard for landscape genetics? *Ecography*, **33**: 315-320.

Jensen JL, Bohonak AJ, Kelley ST (2005) Isolation by distance web service. *BMC Genetics*, **6**(13).

Johnson MS, Black R (1982) Chaotic genetic patchiness in an intertidal limpet, *Siphonaria* sp. *Marine Biology*, **70**, 157-164.

Joksimović D, Castelli A, Mitrić M, Martinović R, Perošević A, Nikolić M, Stanković S (2017) Metal Pollution and Ecotoxicology of the Boka Kotorska Bay. In: *The Boka Kotorska Bay Environment* (eds Joksimović A, Djurović M, Semenov AV, Zonn IS, Kostianoy AG), pp. 395-409. Springer, Cham.

Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**(11), 1403-1405.

Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, **11**, 94.

Jombart T, Ahmed I (2011) adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*, **27**(21), 3070-3071.

Kalinowski ST (2005) HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. *Molecular Ecology Notes*, **5**, 187-189.

Kanduč T, Šlejkovec Z, Falnogaa I, Mori N, Budič B, Kovačić I, Pavičić-Hamer D, Hamer B (2018) Environmental status of the NE Adriatic Sea, Istria, Croatia: Insights from

mussel Mytilus galloprovincialis condition indices, stable isotopes and metal(loid)s. *Marine Pollution Bulletin*, **126**, 525-534.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics Methods and Protocols.* **28**, 1647-1649.

Kijewski T, Zbawicka M, Strand J, Kautsky H, Kotta J, Rätsep M, Wenne R (2019) Random forest assessment of correlation between environmental factors and genetic differentiation of populations: Case of marine mussels *Mytilus. Oceanologia.* **61**, 131-142.

Klöppel A, Messal C, Pfannkuchen M, Matschullat J, Zucht W, Hamer B, Brümmer F (2011) Abiotic Sponge Ecology Conditions, Limski Kanal and Northern Adriatic Sea, Croatia. *Open Journal of Marine Science*, **1**, 18-30.

Kolitari J, Gjyli L, Mukli L, Gjyli S, Vukaj J (2013) Distribution of Chlorophyll a in Lagoon of Butrint waters comparing with environment factors (Albania). *Albanian Journal of Agricultural Sciences*, **12**(1), 87-93.

Kovačić I, Pavičić-Hamer D, Kanduč T, Hamer B (2017) Adaptation of cultured mussel *Mytilus galloprovincialis* Lamarck, 1819 from the northern Adriatic Sea to nearby aquaculture sites and translocation. *Acta Adriatica*, **58**(2), 285-296.

Krivokapić S, Pestorić B, Bosak S, Kušpilić G, Wexels Riser C (2011) Trophic State of Boka Kotorska Bay (South-eastern Adriatic Sea). *Fresenius Environmental Bulletin*, **20**(8), 1960-1969.

Kršinić F, Čalić M, Carić M (2016) The population structure of planktonic protists and small metazoans in Mali Ston Bay (Adriatic Sea) - implications for determination of trophic state and shellfish culturing potential. *Acta Adriatica*, **57**(1), 17-38.

Lallias D, Stockdale R, Boudry P, Lapègue S, Beaumont AR (2009) Characterization of ten microsatellite loci in the blue mussel *Mytilus edulis*. *Journal of Shellfish Research*, **28**(3), 547-551.

Larraín MA, Díaz NF, Lamas C, Uribe C, Jilberto F, Araneda C (2015) Heterologous microsatellite-based genetic diversity in blue mussel (*Mytilus chilensis*) and differentiation among localities in southern Chile. *Latin American Journal of Aquatic Research.* **43**(5), 998-1010.

Launey S, Ledu C, Boudry P, Bonhomme F, Naciri-Graven Y (2002) Geographic Structure in the European Flat Oyster (*Ostrea edulis* L.) as Revealed by Microsatellite Polymorphism. *Journal of Heredity*, **93**(5), 331-338.

Leblois R, Estoup A, Rousset F (2003) Influence of Mutational and Sampling Factors on the Estimation of Demographic Parameters in a "Continuous" Population Under Isolation by Distance. *Molecular Biology and Evolution*, **20**, 491-502.

Lejeusne C, Chevaldonné P, Pergent-Martini C, Boudouresque CF, Pérez T (2010) Climate change effects on a miniature ocean: the highly diverse, highly impacted Mediterranean Sea. *Trends in Ecology and Evolution*, **25**(4), 250-260.

Liggins L, Treml EA, Riginos C (2013) Taking the Plunge: An Introduction to Undertaking Seascape Genetic Studies and using Biophysical Models. *Geography Compass*, **7**, 173-196.

Lipizer M, Partescano E, Rabitti A, Giorgetti A, Crise A (2014) Qualified temperature, salinity and dissolved oxygen climatologies in a changing Adriatic Sea. *Ocean Science*, **10**, 771-797.

Liubartseva S, Coppini G, Lecci R, Creti S (2016) Regional approach to modelling the transport of floating plastic debris in the Adriatic Sea. *Marine Pollution Bulletin*, **103**, 115-127.

Lockwood BL, Somero GN (2011) Invasive and native blue mussels (genus *Mytilus*) on the California coast: The role of physiology in a biological invasion. *Journal of Experimental Marine Biology and Ecology*, **400**, 167-174.

Lockwood BL, Connor KM, Gracey AY (2015) The environmentally tuned transcriptomes of *Mytilus* mussels. *The Journal of Experimental Biology*, **218**, 1822-1833.

Lodish H, Berk A, Zipursky SL, Matsudaira P, Baltimore D, Darnell J (2000) *Molecular Cell Biology (4th Ed)*. WH Freeman, New York.

McDonald JH, Seed R, Koehn RK (1991) Allozymes and morphometric characters of three species of *Mytilus* in the Northern and Southern hemispheres. *Marine Biology*, **111**, 323-333.

McDonald JH (2014) *Handbook of Biological Statistics (3rd Ed)*. Sparky House, Baltimore.

McQuaid CD, Phillips TE (2000) Limited wind-driven dispersal of intertidal mussel larvae: *in situ* evidence from plankton and the spread of the invasive species *Mytilus galloprovincialis* in South Africa. *Marine Ecology Progress Series*, **201**, 211-220.

Mandić M, Ikica Z, Gvozdenović S (2017) Mariculture in the Boka Kotorska Bay: Tradition, Current State and Perspective. In: *The Boka Kotorska Bay Environment* (eds Joksimović A, Djurović M, Semenov AV, Zonn IS, Kostianoy AG), pp. 395-409. Springer, Cham.

Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. *TRENDS in Ecology and Evolution*, **18**(4), 189-197.

Manni F, Guérard E, Heyer E (2004) Geographic patterns of (genetic, morphologic, linguistic) variation: how barriers can be detected by "Monmonier's algorithm". *Human Biology*, **76**(2): 173-190.

Meirmans PG (2012) The trouble with isolation by distance. *Molecular Ecology*, **21**, 2839–2846.

Michaelidis B, Pörtner H-O, Sokolova I, Tomanek L (2014) Advances in Predicting the Impacts of Global Warming on the Mussels *Mytilus galloprovincialis* in the Mediterranean Sea. In: *The Mediterranean Sea: Its history and present challenges* (eds Goffredo S, Dubinsky Z), pp319-339. Springer, Dordrecht.

Miller MP (2005) Alleles in Space (AIS): Computer software for the joint analysis of interindividual spatial and genetic information. *Journal of Heredity*, **96**: 722-724.

Miller K, Baird HP, van Oosterom J, Mondon J, King CK (2018) Complex genetic structure revealed in the circum-Antarctic broadcast spawning sea urchin *Sterechinus neumayeri*. *Marine Ecology Progress Series*, **601**, 153-166.

Mladineo I, Trumbić Ž, Jozić S, Šegvić T (2009) First Report of Cryptosporidium sp. (Coccidia, Apicomplexa) Oocysts in the Black Mussel (*Mytilus galloprovincialis*) reared in the Mali Ston Bay, Adriatic Sea, *Journal of Shellfish Research*, **28**(3), 541-543.

Moisiu L, Panagiotopoulos IP, Durmishi Ç, Kapsimalis V, Anagnostou C (2016) The anoxic Butrint Lagoon, SW Albania. *Environmental Earth Sciences*, **75**, 1443.

Murgarella M, Puiu D, Novoa B, Figueras A, Posada D, Canchaya C (2016) A First Insight into the Genome of the Filter-Feeder Mussel *Mytilus galloprovincialis. PLoS one*, **11**(3), e0151561.

Nanninga GB, Saenz-Agudelo P, Manica A (2014) Environmental gradients predict the genetic population structure of a coral reef fish in the Red Sea. *Molecular Ecology*, **23**, 591-602.

Neuhaus J, McCulloch C (2011) Generalized linear models. *Wiley Interdisciplinary Reviews: Computational Statistics*, **3**(5), 407-413.

Nguyen TTT, Hayes BJ, Guthridge K, Ab Rahim ES, Ingram BA (2011) Use of a microsatellite-based pedigree in estimation of heritabilities for economic traits in Australian blue mussel, *Mytilus galloprovincialis*. *Journal of Animal Breeding and Genetics*. **128**, 482-490.

Ninčević-Gladan Z, Bužančić M, Kušpilić G, Grbec B, Matijević S, Skejić S, Marasović I, Morović M (2015) The response of phytoplankton community to anthropogenic pressure gradient in the coastal waters of the eastern Adriatic Sea. *Ecological Indicators*, **56**, 106-115.

Oceanbrowser.net and SeaDataNet http://sdn.oceanbrowser.net/web-vis/ and http://gher-diva.phys.ulg.ac.be/emodnet/ accessed 28/10/2018.

Oliver ECJ, Donat MG, Burrows MT, Moore PJ, Smale DA, Alexander LV, Benthuysen JA, Feng M, Gupta AS, Hobday AJ, Holbrook NJ, Perkins-Kirkpatrick SE, Scannell HA, Straub SC, Wernberg T (2018) Longer and more frequent marine heatwaves over the past century. *Nature Communications*. **9**, 1324.

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) Micro-Checker: Software for Identifying and Correcting Genotyping Errors in Microsatellite Data. *Molecular Ecology Notes*, **4**, 535–538.

Palumbi SR (2003) Population Genetics, Demographic Connectivity, and the Design of Marine Reserves. *Ecological Applications*, **13**(1) Supplement, S146-S158.

Pascual M, Rives B, Schunter C, MacPherson (2017) Impact of life history traits on gene flow: A multispecies systematic review across oceanographic barriers in the Mediterranean Sea. *PLoS one*, **12**(5), e0176419.

Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic structure for teaching and research – an update. *Bioinformatics*, **28**, 2537-2539.

Perez MF, Franco FF, Bombonato JR, Bonatelli IAS, Khan G, Romeiro-Brito M, Fegies AC, Ribeiro P, Silva GAR, Moraes EM (2018) Assessing population structure in the face of isolation by distance: Are we neglecting the problem? *Diversity and Distributions*, **24**, 1883-1889.

PERSEUS – UNEP/MAP Report, (2015). *Atlas of Riverine Inputs to the Mediterranean Sea.*

Presa P, Pérez M, Diz AP (2002) Polymorphic microsatellite markers for blue mussels (*Mytilus* spp.). *Conservation Genetics*, **3**, 441-443.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945-959.

Puebla O, Bermingham E, McMillan WO (2012) On the spatial scale of dispersal in coral reef fishes. *Molecular Ecology*, **21**, 5675-88.

Quesada H, Zapata C, Alvarez G (1995) A multilocus allozyme discontinuity in the mussel *Mytilus galloprovincialis*: the interaction of ecological and life-history factors. *Marine Ecology Progress Series*, **116**, 99-115.

R Core Team (2014) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. [http://www.R-project.org/]

Ramón M, Fernández M, Galimany E (2007) Development of mussel (*Mytilus galloprovincialis*) seed from two different origins in a semi-enclosed Mediterranean Bay (N.E. Spain). *Aquaculture*, **264**, 148-159.

Rawson PD, Harper FM (2009) Colonization of the northwest Atlantic by the blue mussel, *Mytilus trossulus* postdates the last glacial maximum. *Marine Biology*, **156**, 1857-1868.

Riginos C, Cunningham CW (2005) Local adaptation and species segregation in two mussel (*Mytilus edulis* × *Mytilus trossulus*) hybrid zones. *Molecular Ecology*, **14**, 381-400.

Riginos C, Liggins L (2013) Seascape Genetics: Populations, Individuals, and Genes Marooned and Adrift. *Geography Compass*, **7**(3), 197-216.

Riginos C, Sukhdeo K, Cunningham CW (2002) Evidence for selection at Multiple Allozyme Loci across a Mussel Hybrid Zone. *Molecular Biology and Evolution*, **19**(3), 347-351.

Riquet F, Comtet T, Broquet T, Viard F (2017) Unexpected collective larval dispersal but little support for sweepstakes reproductive success in the highly dispersive brooding mollusc *Crepidula fornicata. Molecular Ecology*, **26**, 5467-5483.

Rioux Paquette S (2011) PopGenKit: useful function for (batch) file conversion and data resampling in microsatellite datasets. R package.

Rose CG, Paynter KT, Hare MP (2006) Isolation by distance in the eastern oyster, *Crassostrea virginica*, in Chesapeake Bay. *Journal of Heredity*, **97**(2), 158-170.

Rousset F (1997) Genetic Differentiation and Estimation of Gene Flow from *F*-Statistics Under Isolation by Distance. *Genetics*, **145**, 1219-1228.

Rousset F (2008) GENEPOP '007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources*, **8**(1), 103-106.

Roux C, Fraïsse C, Castric V, Vekemans X, Pogson GH, Bierne N (2014) Can we continue to neglect genomic variation in introgression rates when inferring the history of speciation? A case study in a *Mytilus* hybrid zone. *Journal of Evolutionary Biology*, **27**, 1662-1675.

Sawaya S, Bagshaw A, Buschiazzo E, Kumar P, Chowdhury S, Black MA, Gemmell N (2013) Microsatellite Tandem Repeats are Abundant in Human Promoters and Are Associated with Regulatory Elements. *PLoS one*, **8**(2), e54710.

Schiavina M, Marino IAM, Zane L, Melià P (2014) Matching oceanography and genetics at the basin scale. Seascape connectivity of the Mediterranean shore crab in the Adriatic Sea. *Molecular Ecology*, **23**, 5496-5507.

Schmidt PS, Serrão EA, Pearson GA, Riginos C, Rawson PD, Hilbish TJ, Brawley SH, Trussell GC, Carrington E, Wethey DS, Grahame JW, Bonhomme F, Rand DM (2008) Ecological Genetics In The North Atlantic: Environmental Gradients And Adaptation At Specific Loci. *Ecology*, **89**(11), S91–S107.

Schneider KR, Helmuth B (2007) Spatial variability in habitat temperature may drive patterns of selection between an invasive and native mussel species. *Marine Ecology Progress Series*, **339**, 157-167.

Sea quality in the Republic of Croatia http://baltazar.izor.hr/plazepub/kakvoca_detalji10 accessed 2/10/2018.

Selkoe KA, Toonen RJ (2006) Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology Letters*, **9**, 615-629.

Selkoe KA, Scribner KT. Galindo HM (2016) Waterscape Genetics: Applications of landscape genetics to rivers, lakes, and seas. In: *Landscape Genetics* (eds Balkenhol N, Cushman SA, Storfer AT, Waits LP), pp. 220-246. Wiley Blackwell, West Sussex.

Shyamala V, Ames GFL (1989) Genome walking by single-specific-primer polymerase chain reaction: SSP-PCR. *Gene*, **1**, 1-8.

Sieberts SK, Schadt EE (2007) Inferring Causal Associations between Genes and Diseases via the Mapping of Expression Quantitative Trait Loci. In: *Handbook of Statistical Genetics 3rd Edition* (eds Balding DJ, Bishop M, Cannings C), pp. 296-326. Wiley, West Sussex.

Silva CNS, Gardner JPA (2016) Identifying environmental factors associated with the genetic structure of the New Zealand scallop: linking seascape genetics and ecophysiological tolerance. *ICES Journal of marine Science*, **73**(7), 1925-1934.

Štambuk A, Šrut M, Šatović Z, Tkalec M, Klobučar GIV (2013) Gene flow vs. pollution pressure: Genetic diversity of *Mytilus galloprovincialis* in eastern Adriatic. *Aquatic Toxicology*, **136-137**, 22-31.

Stuckas H, Knöbel L, Schade H, Breusing C, Hinrichsen H. Bartel M, Langguth K, Melzner F (2017) Combining hydrodynamic modelling with genetics: can passive larval drift shape the genetic structure of Baltic *Mytilus* populations? *Molecular Ecology*, **26**, 2765-2782.

Tangwancharoen S, Moy GW, Burton RS (2018) Multiple Modes of Adaptation: Regulatory and Structural Evolution in a Small Heat Shock Protein Gene. *Molecular Biology and Evolution*. **35**(9), 2110-2119.

Teixeira de Sousa J, Joaquim S, Matias D, Ben-Hamadou R, Leitão A (2012) Evidence of non-random chromosome loss in bivalves: differential chromosomal susceptibility in aneuploid metaphases of *Crassostrea angulata* (Ostreidae) and *Ruditapes decussatus* (Veneridae). *Aquaculture* 344-349, 239-241.

Teske PR, Sandoval-Castillo J, van Sebille E, Waters J, Beheregaray LB (2015) On-shelf larval retention limits population connectivity in a broadcast spawner. *Marine Ecology Progress Series*, **532**, 1-12.

Tomanek L (2012) Environmental Proteomics of the Mussel *Mytilus*: Implications for Tolerance to Stress and Change in Limits of Biogeographical Ranges in Response to Climate Change. *Integrative and Comparative Biology*, **52**(5), 648-664.

Topi T, Bani A, Sulçe S (2013) Physico chemical characteristics and heavy metal contents of water from Butrinti lagoon, Albania. *Albanian Journal of Agricultural Sciences*, **12**(2), 321-326.

Tóth G, Gáspári Z, Jurka J (2000) Microsatellites in different eukaryotic genomes: Surveys and analysis. *Genome Research*, **10**(7), 967-981.

Treml EA, Roberts JJ, Chao Y, Halpin PN, Possingham HP, Riginos C (2012) Reproductive Output and Duration of the Pelagic Larval Stage Determine Seascape-Wide Connectivity of Marine Populations. *Integrative and Comparative Biology*, **52**(4), 525-537.

Tudor M, Janeković I (2016) Modelling origin and transport fate of waste materials on the southeastern Adriatic coast (Croatia). *Geofizika*, **33**, 53-77.

UNEP/MAP-RAC/SPA (2015) Adriatic Sea: Description of the ecology and identification of the areas that may deserve to be protected. By Cerrano C. Edited by Cebrian D and Requena S, RAC/SPA, Tunis.

Varela MA, González-Tizón A, Mariñas L, Martínez-Lage A (2007) Genetic Divergence Detected by ISSR Markers and Characterization of Microsatellite Regions in *Mytilus* Mussels. *Biochemical Genetics*. **45**, 565-578.

Vasemägi A (2006) The Adaptive Hypothesis of Clinal Variation Revisited: Single-Locus Clines as a Result of Spatially Restricted Gene Flow. *Genetics*, **173**, 2411-2414.

Verhoeven KJF, Simonsen KL, McIntyre LM (2005) Implementing false discovery rate control: increasing your power. *OIKOS*, **3**, 643-647.

Vieira MLC, Santini L, Diniz AL, Munhoz CF (2016) Microsatellite markers: What they mean and why they are so useful. *Genetics and Molecular Biology*, **39**(3), 3112-328.

Villacorta-Rath C, Souza C, Murphy NP, Green BS, Gardner C, Strugnell JM (2017) Temporal genetic patterns of diversity and structure evidence chaotic genetic patchiness in a spiny lobster. *Molecular Ecology*, **27**, 54-65.

Waits LP, Storfer A (2016) Basics of Population Genetics: Quantifying neutral and adaptive genetic variation for landscape genetic studies. In: *Landscape Genetics* (eds Balkenhol N, Cushman SA, Storfer AT, Waits LP), pp. 35-57. Wiley Blackwell, West Sussex.

Waples RS (1998) Separating the wheat from the chaff: patterns of genetic differentiation in high gene flow species. *Journal of Heredity*, **89**, 438-450.

Waples RS (2006) A bias correction for estimates of effective population size based on linkage disequilibrium at unlinked loci. *Conservation Genetics*, **7**(2), 167-184.

Waples RS, Do C (2010) Linkage disequilibrium estimates of contemporary N_e using highly variable genetic markers: A largely untapped resource for applied conservation and evolution. *Evolutionary Applications*, **3**(3), 244-262.

Wei K, Wood AR, Gardner JPA (2013a) Population genetic variation in the New Zealand greenshell mussel: locus-dependent conflicting signals of weak structure and high gene flow balanced against pronounced structure and high self-recruitment. *Marine Biology*, **160**, 931–949.

Wei K, Wood AR, Gardner JPA (2013b) Seascape genetics of the New Zealand greenshell mussel: sea surface temperature explains macrogeographic scale genetic variation. *Marine Ecology Progress Series*, **477**, 107–121.

Westfall KM, Gardner JPA (2010) Genetic diversity of Southern hemisphere blue mussels (Bivalvia: Mytilidae) and the identification of non-indigenous taxa. *Biological Journal of the Linnean Society*. **101**, 898-909.

Westfall KM (2011) Molecular ecology and systematics of blue mussels (genus *Mytilus*) (Mytilidae; Bivalvia; Mollusca) in the Southern hemisphere. PhD thesis, Victoria University of Wellington.

White TA, Stamford J, Hoelzel AR (2010) Local selection and population structure in a deep-sea fish, the roundnose grenadier (*Coryphaenoides rupestris*). *Molecular Ecology*, **19**, 216-226.

Wilson-Sanders SE (2011) Invertebrate Models for Biomedical Research, Testing, and Education. *ILAR Journal*, **52**(2), 126-152.

Wright, S (1943) Isolation by Distance. *Genetics*. **28**: 114–138.

Yu H, Li Q (2007) Development of EST-SSRs in the Mediterranean mussel, *Mytilus galloprovincialis*. *Molecular Ecology Notes*, **7**, 1308-1310.

Zhang G, Fang X, Guo X, Li L, Luo R, Xu F, Yang P, Zhang L, Wang X, Qi H et al. (2012) The oyster genome reveals stress adaptation and complexity of shell formation. *Nature*, **490**(7418), 49-54.

Zonn IS, Kostianoy AG (2017) The Adriatic Sea. In: *The Boka Kotorska Bay Environment* (eds Joksimović A, Djurović M, Semenov AV, Zonn IS, Kostianoy AG), pp. 19-41. Springer, Cham.

Zouros E, Foltz DW (1984) Possible Explanations of Heterozygote Deficiency in Bivalve Molluscs. *Malacologia*, **25**(2), 583-591.

Zulliger DE, Tanner S, Ruch M, Ribi G (2009) Genetic structure of the high dispersal Atlanto-Mediterranean sea star *Astropecten aranciacus* revealed by mitochondrial DNA sequences and microsatellite loci. *Marine Biology*, **156**, 597-610.

Supplementary Material

Table of contents

140

Chapter 2 Supplementary Figures

Supplementary Figure 1 Rarefaction curves to assess whether sample sizes were sufficiently large to characterise the allelic variation of each population.

Supplementary Figure 2 Structure results for *Mytilus galloprovincialis* using 6 loci: Structure Harvester results using (a) the Admixture model with a maximum at K = 5 and local maximum at K = 2, and (b) the No admixture model with maximum at K = 2; bar plots of structure outputs using the Admixture model for (c) K = 5 and (d) K = 2 and (e) K = 2 with individuals sorted by proportion cluster membership within population.

Chapter 2 Supplementary Tables

Supplementary Table 1. Number of alleles per locus per population and total for the whole study range.

Population	MGE3	MT282	MGE4	MGE5	MGE7	MT203	Mgµ1
LKF	7	8	6	7	5	7	11
CK	6	10	7	10	6	13	12
KLW	6	8	6	9	5	13	17
OSOR	5	7	5	8	6	13	14
ML	4	11	5	8	5	11	11
MLC	7	8	7	9	4	15	10
Z	6	7	5	8	4	14	13
U	6	9	6	6	4	13	12
STF	9	10	6	9	5	13	17
STW	6	9	7	8	5	13	11
NF	5	10	6	5	3	12	18
BF	6	8	5	11	6	12	18
BW	7	8	5	7	6	13	16
CF	6	7	7	6	4	14	16
CW	7	7	6	9	5	11	14
IF	6	10	7	8	3	13	10
IW	6	8	7	9	4	12	15
SBF	6	8	6	10	4	12	19
Total	13	16	8	15	6	24	48

Supplementary Table 2 Allele frequencies for seven *Mytilus galloprovincialis* microsatellite loci over 18 populations.

Locus Allele	LKF	CK	KLW	Osor	ML	MLC	Z	U	STF	STW	NF	BF	BW	CF	CW	IF	IW	SBF
MGE3																		
302	-	-	-	-	-	-	-	-	0.011	-	-	-	-	-	-	-	-	-
365	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
371	0.016	-	-	-	-	-	0.010	0.010	-	-	-	-	-	-	-	-	-	-
374	0.032	0.023	0.031	0.015	-	0.010	0.020	-	0.021	0.011	-	0.021	0.030	0.041	0.031	0.010	-	0.023
377	0.677	0.616	0.729	0.706	0.844	0.710	0.770	0.730	0.617	0.702	0.745	0.638	0.670	0.735	0.673	0.720	0.670	0.640
380	0.065	0.070	0.073	0.059	0.083	0.090	0.080	0.100	0.138	0.064	0.102	0.096	0.070	0.071	0.071	0.050	0.057	0.151
383	0.016	0.047	0.021	-	-	0.020	-	0.010	0.011	0.011	0.031	0.011	0.010	0.010	0.020	-	-	0.035
386	-	-	-	-	-	-	-	0.020	0.011	-	-	-	-	-	-	0.010	-	-
452	-	-	-	-	-	-	-	-	0.011	-	-	-	-	-	-	-	-	-
470	-	-	-	-	-	0.010	-	-	0.011	-	-	-	0.010	-	-	0.011	-	-
473	0.048	0.070	0.031	0.074	0.021	0.030	0.030	0.040	0.074	0.064	0.061	0.096	0.080	0.092	0.092	0.110	0.125	0.047
476	0.145	0.174	0.115	0.147	0.052	0.130	0.090	0.100	0.106	0.149	0.061	0.138	0.130	0.051	0.092	0.100	0.125	0.105
479	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.011	-
MGE4																		
156	0.017	0.012	0.022	-	0.010	0.010	-	-	0.010	0.022	0.011	0.035	0.021	-	0.011	0.033	-	0.024
165	0.138	0.151	0.244	0.182	0.115	0.240	0.220	0.125	0.167	0.167	0.098	0.174	0.213	0.143	0.160	0.185	0.125	0.202
168	0.034	0.070	-	0.030	0.042	0.073	0.070	0.034	-	-	0.065	-	-	0.041	0.043	0.043	0.138	0.060
171	0.534	0.570	0.533	0.621	0.656	0.448	0.410	0.557	0.594	0.600	0.630	0.535	0.500	0.612	0.649	0.511	0.600	0.524
174	0.103	0.058	0.078	0.030	0.021	0.094	0.030	0.080	0.094	0.033	0.087	0.035	0.074	0.010	0.032	0.076	0.038	0.036
177	0.172	0.128	0.100	0.136	0.167	0.125	0.270	0.193	0.125	0.156	0.109	0.221	0.191	0.173	0.106	0.141	0.063	0.155
180	-	0.012	0.022	-	-	-	-	-	0.010	0.011	-	-	-	0.010	-	0.011	0.025	-
186	-	-	-	-	-	0.010	-	0.011	-	0.011	-	-	-	0.010	-	-	0.013	-

Supplementary Table 2 continued.

Locus Allele	LKF	CK	KLW	Osor	ML	MLC	Z	U	STF	STW	NF	BF	BW	CF	CW	IF	IW	SBF
MGE5																		
208	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
214	-	-	-	-	-	-	-	-	0.011	-	0.011	0.010	-	-	-	0.010	-	-
217	-	0.023	0.021	0.015	0.030	0.030	0.032	-	-	-	-	-	-	-	-	0.010	0.012	0.012
220	-	0.023	0.031	0.015	0.010	-	-	-	0.011	0.053	-	0.020	0.010	0.012	0.031	0.010	0.023	0.024
223	0.113	0.023	0.052	0.015	0.030	0.020	0.021	0.041	0.021	0.032	0.011	0.051	0.063	0.048	0.042	0.082	0.035	0.083
226	0.016	0.035	0.031	0.044	0.020	0.030	0.032	-	0.021	0.011	-	0.010	0.052	-	0.010	0.041	0.035	0.012
229	0.016	-	0.021	-	-	-	0.032	-	-	-	-	0.010	-	-	0.010	-	-	0.012
232	-	0.012	-	-	-	0.010	-	-	0.011	0.011	-	0.020	-	0.071	-	-	0.012	0.012
235	0.016	0.012	0.073	0.029	0.030	0.060	0.053	0.031	0.043	0.074	-	0.020	0.031	0.071	0.031	0.041	0.012	0.036
238	0.145	0.233	0.188	0.279	0.270	0.230	0.149	0.163	0.202	0.181	0.170	0.133	0.188	0.190	0.229	0.173	0.151	0.250
241	0.661	0.512	0.552	0.529	0.530	0.580	0.596	0.663	0.606	0.564	0.755	0.592	0.594	0.571	0.563	0.561	0.616	0.524
244	-	0.116	0.031	0.074	0.080	0.030	0.085	0.092	0.074	0.074	0.053	0.102	0.063	0.107	0.073	0.082	0.105	0.036
247	0.032	-	-	-	-	0.010	-	-	-	-	-	-	-	-	-	-	-	-
250	-	-	-	-	-	-	-	0.010	-	-	-	-	-	-	0.010	-	-	-
253	-	0.012	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
MGE7																		
237	-	0.025	-	0.016	-	-	0.029	-	0.021	-	-	0.020	0.020	-	0.014	-	0.029	-
240	0.032	0.100	0.010	0.094	0.056	0.080	-	0.071	0.052	0.031	0.054	0.110	0.050	0.068	0.068	0.122	-	0.036
243	0.500	0.450	0.438	0.391	0.544	0.460	0.529	0.490	0.458	0.531	0.595	0.470	0.550	0.770	0.770	0.549	0.586	0.679
246	0.097	0.300	0.396	0.344	0.211	0.260	0.386	0.296	0.219	0.302	0.351	0.220	0.220	0.108	0.135	0.329	0.357	0.190
249	0.339	0.113	0.146	0.109	0.178	0.200	0.057	0.143	0.250	0.115	-	0.170	0.150	0.054	0.014	-	0.029	0.095
252	0.032	0.013	0.010	0.047	0.011	-	-	-	-	0.021	-	0.010	0.010	-	-	-	-	-

Supplementary Table 2 continued.

Locus Allele	LKF	CK	KLW	Osor	ML	MLC	Z	U	STF	STW	NF	BF	BW	CF	CW	IF	IW	SBF
MT282																		
314	-	-	-	-	-	-	-	-	0.022	-	-	-	-	-	-	-	-	-
316	-	0.029	-	-	0.013	-	-	-	-	-	-	-	-	-	-	-	0.027	-
326	-	-	-	-	-	-	-	0.024	-	-	-	-	-	-	-	-	-	-
330	0.017	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
332	0.103	0.015	0.174	0.143	0.066	0.033	0.056	-	0.044	0.047	0.044	-	0.054	-	0.013	0.083	0.054	0.012
334	0.466	0.515	0.488	0.357	0.408	0.489	0.403	0.298	0.422	0.442	0.444	0.534	0.413	0.588	0.474	0.452	0.473	0.442
336	0.138	0.147	0.163	0.161	0.171	0.174	0.167	0.262	0.244	0.186	0.178	0.148	0.207	0.103	0.269	0.190	0.149	0.186
338	-	0.044	0.012	0.107	0.053	0.054	-	0.155	0.056	0.081	0.022	0.102	-	0.059	0.090	0.119	-	0.070
340	0.190	0.132	0.116	0.161	0.171	0.141	0.208	0.143	0.122	0.070	0.144	0.148	0.152	0.118	0.051	0.071	0.135	0.140
342	0.034	0.044	0.012	-	0.026	0.033	0.069	0.048	0.033	0.047	0.022	0.034	0.087	0.059	0.051	0.012	0.041	0.058
344	0.034	0.029	0.023	-	0.026	-	0.069	0.048	0.011	0.070	0.067	0.011	0.022	0.059	0.051	0.012	0.054	0.081
346	-	0.029	-	-	-	0.043	-	-	0.011	0.023	0.011	-	0.022	-	-	-	-	-
348	-	-	-	0.054	0.013	0.033	-	0.012	0.033	-	0.022	0.011	0.043	0.015	-	0.024	-	0.012
350	0.017	0.015	0.012	-	0.039	-	0.028	-	-	-	0.044	0.011	-	-	-	0.024	0.068	0.012
354	-	-	-	0.018	0.013	-	-	-	-	0.035	-	0.011	-	-	-	-	-	-
356	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.012	-	-

Supplementary Table 2 continued.

Locus Allele	LKF	CK	KLW	Osor	ML	MLC	Z	U	STF	STW	NF	BF	BW	CF	CW	IF	IW	SBF
MT203																		
167	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.021	-	-	-
169	-	-	-	-	-	-	-	-	-	-	0.010	-	-	-	-	-	-	-
171	-	0.024	-	-	0.031	0.020	-	-	0.021	-	-	-	0.020	-	-	-	-	0.024
173	-	-	0.021	0.015	0.031	0.020	0.061	-	0.021	-	-	-	-	-	-	-	-	-
175	0.448	0.500	0.426	0.353	0.296	0.340	0.327	0.468	0.340	0.389	0.500	0.418	0.367	0.391	0.468	0.449	0.341	0.393
177	0.138	0.179	0.319	0.235	0.286	0.210	0.204	0.117	0.202	0.211	0.167	0.224	0.235	0.228	0.160	0.102	0.122	0.286
179	-	0.024	0.011	0.015	0.041	0.020	0.010	0.011	0.032	0.011	-	-	0.010	0.011	-	-	0.024	0.012
181	0.103	0.071	0.043	0.015	0.092	0.120	0.102	0.128	0.181	0.189	0.146	0.112	0.153	0.098	0.160	0.102	0.122	0.048
183	-	-	0.011	0.044	0.020	0.050	0.010	0.011	0.032	0.022	0.021	0.031	-	0.011	0.011	0.061	0.098	0.036
185	-	0.036	0.021	0.059	-	0.050	0.020	0.032	0.043	0.011	0.010	0.041	0.031	0.065	0.053	0.051	0.037	0.036
187	0.034	0.036	0.043	0.088	0.092	0.050	0.071	0.053	0.053	0.022	0.021	0.020	0.020	0.033	0.032	0.051	0.073	0.036
189	0.103	0.036	0.011	0.044	0.061	0.030	0.031	0.064	0.021	0.011	0.042	0.061	-	0.022	0.021	0.031	0.049	0.024
191	-	0.012	0.032	0.044	0.041	0.020	0.020	-	0.032	-	0.021	-	0.031	0.022	-	0.010	-	-
193	0.103	0.036	0.032	0.044	-	0.070	0.102	0.074	0.011	0.067	0.021	0.051	0.061	0.033	0.021	0.041	0.073	0.083
195	-	0.024	-	0.015	-	0.010	-	0.011	0.011	0.033	-	-	0.020	0.011	-	0.020	-	-
197	-	0.012	-	-	-	0.010	0.020	-	-	-	-	0.010	-	0.011	-	0.010	0.012	-
199	0.069	0.012	0.021	-	-	-	-	0.011	-	-	-	-	-	-	-	0.010	0.012	0.012
201	-	-	-	-	-	-	-	0.011	-	0.011	-	0.010	-	-	0.011	-	-	-
203	-	-	-	-	-	-	0.010	0.011	-	-	-	-	0.010	0.022	-	-	-	-
205	-	-	-	-	-	0.010	-	-	-	-	-	-	-	-	-	-	-	-
207	-	-	-	0.029	-	-	-	-	-	-	0.021	0.010	0.010	-	-	-	-	-
209	-	-	-	-	0.010	-	0.010	-	-	-	-	-	-	-	-	-	0.012	-
237	-	-	0.011	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
239	-	-	-	-	-	-	-	-	-	0.011	-	-	-	-	-	-	-	-

Supplementary Table 2 continued.

Locus Allele	LKF	CK	KLW	Osor	ML	MLC	Z	U	STF	STW	NF	BF	BW	CF	CW	IF	IW	SBF
Mgu1																		
124	-	-	-	-	-	-	-	-	0.018	-	-	-	-	0.026	-	-	0.038	-
128	-	-	-	-	-	-	-	-	-	-	0.028	-	-	-	0.025	-	0.038	-
130	-	0.094	-	-	0.026	0.026	0.059	-	0.018	-	0.083	-	-	-	-	-	-	0.020
134	-	-	-	-	-	-	0.029	-	-	-	-	-	-	-	-	-	-	-
138	-	-	-	-	-	-	0.029	0.083	-	-	-	-	-	0.026	0.075	-	0.038	-
144	0.038	-	-	-	-	-	0.029	-	-	-	0.028	0.033	-	-	-	-	-	-
146	-	-	-	0.022	-	-	-	-	-	-	-	-	-	-	0.100	0.038	0.038	-
148	-	0.063	-	0.043	0.053	-	-	0.036	0.036	-	0.028	0.017	0.034	0.079	0.025	0.067	-	0.080
150	-	-	-	0.022	-	-	-	0.071	-	-	-	-	0.069	0.026	0.100	0.200	0.077	0.040
152	-	-	0.016	-	0.053	-	0.059	0.071	0.196	0.205	0.194	0.033	0.052	0.026	0.100	-	0.077	-
154	0.038	-	0.016	0.022	-	-	-	-	-	-	-	0.033	-	-	-	0.033	-	-
156	0.192	0.063	0.047	0.196	0.211	0.053	0.059	0.071	0.179	0.114	-	0.067	0.017	0.079	0.100	-	0.077	0.040
158	-	0.156	0.156	-	0.053	0.316	0.147	0.250	0.214	0.091	0.056	0.083	0.138	0.079	0.125	0.167	0.077	0.100
160	0.346	0.313	0.406	0.130	0.289	0.263	0.265	0.179	0.179	0.318	0.056	0.183	0.293	0.105	0.225	0.233	0.231	0.080
162	0.038	0.063	0.031	0.109	-	-	0.059	0.179	0.071	0.045	0.056	-	-	0.158	-	0.033	-	0.020
164	0.038	-	0.016	-	-	-	-	-	-	0.045	-	0.033	0.034	-	-	-	-	-
166	0.038	-	0.016	-	0.026	0.026	-	-	-	-	-	-	-	-	-	-	-	-
168	-	0.031	0.016	-	-	-	-	-	-	-	-	-	0.017	-	0.025	-	-	-
170	-	0.031	-	-	-	0.026	-	-	-	-	0.056	0.033	-	0.026	0.050	0.033	-	0.040
172	-	0.031	-	0.043	0.053	-	0.059	0.107	-	0.045	0.083	0.100	0.086	0.053	0.050	-	0.038	-
174	-	-	0.047	0.217	-	0.026	0.088	0.107	0.054	-	0.028	0.050	0.052	0.026	-	-	0.038	0.080
176	-	-	0.016	0.043	-	-	-	-	-	-	0.028	-	-	-	-	0.067	0.038	-
178	-	0.031	-	-	-	0.105	-	-	0.018	-	0.056	0.033	0.017	0.079	-	-	0.077	0.020
180	-	-	0.031	0.043	-	-	-	0.036	-	-	0.028	-	-	-	-	-	-	0.100

Supplementary Table 2 continued.

Locus	Allele	LKF	CK	KLW	Osor	ML	MLC	Z	U	STF	STW	NF	BF	BW	CF	CW	IF	IW	SBF
Mgu1 cont'd	182	-	-	0.016	-	-	-	-	-	-	-	-	-	-	-	0.050	-	-	0.060
	184	0.077	0.031	0.063	0.022	0.132	-	0.059	-	0.036	0.023	0.083	0.033	0.034	-	0.050	-	-	0.100
	186	0.077	0.063	0.016	0.022	0.079	-	-	-	0.036	0.045	0.056	0.100	0.086	0.158	0.050	0.133	0.115	0.080
	188	0.077	0.063	-	0.065	-	-	-	-	-	0.023	0.028	0.017	0.017	-	0.025	-	-	0.060
	190	-	-	-	-	-	-	-	-	-	-	-	0.033	-	-	-	-	-	-
	192	-	-	0.031	-	-	-	-	-	0.018	-	-	-	-	-	-	-	-	-
	194	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.038	-
	198	-	-	-	-	-	-	0.059	-	-	-	-	0.017	-	-	-	-	-	0.020
	200	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	202	-	-	0.031	-	-	-	-	-	0.018	-	-	-	0.017	-	0.025	-	0.038	-
	204	-	-	0.047	-	-	-	-	-	0.018	-	-	-	-	-	-	-	-	-
	206	0.038	-	-	-	-	0.053	0.029	0.071	0.036	-	-	0.100	0.034	-	-	-	-	0.020
	208	-	-	-	-	-	-	-	0.036	-	-	-	-	-	-	-	-	-	-
	212	-	-	0.016	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	214	-	-	0.016	-	-	-	-	-	0.018	-	-	-	-	0.026	-	-	-	-
	216	-	-	-	-	-	-	-	0.036	0.036	-	-	-	-	0.026	-	-	-	-
	218	-	-	-	-	-	-	-	0.036	-	-	-	-	-	-	-	-	-	-
	224	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.038	-
	226	-	-	-	-	-	-	-	0.036	-	-	-	-	-	-	-	-	-	-
	232	-	-	-	-	-	-	-	-	-	-	0.028	-	-	-	-	-	-	-
	234	-	-	-	-	-	-	-	-	0.018	-	-	-	-	-	-	-	-	0.020
	236	-	-	-	-	0.026	0.053	-	-	-	-	-	-	-	-	-	-	-	-
	242	-	-	-	-	0.026	-	-	-	-	-	-	-	-	-	-	-	-	-
	262	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.020

Supplementary Table 3. Results of spatial autocorrelation analyses of *M. galloprovincialis*. Note: overall correlograms are considered significant when p < 0.01 as suggested by Banks and Peakall, 2012.

Analysis	Dataset	Number of loci	Distance class endpoint (km) Bold* if significant				X-intercept (km)	Overall p-value for correlogram Bold* if significant
(i)	6 of 6	6	**0***	**10***	80	100	118	**0.001***
(ii)	6 of 6	5 neutral	0	10	80	100	147	0.105
(iii)	6 of 6	1- MGE7	**0***	**10***	80	100	114	**0.0003***
(iv)	6 of 6 large distance classes	6	**35***	140	250	340	223	**0.007***
(v)	6 of 6 small distance classes	6	**0***	**6***	30	80	29 and 125	**0.002***
(vi)	6 of 6 #2 small distance class	6	**0***	**12***	80	100	74	**0.004***

Supplementary Table 4. Microsatellite markers developed for the *Mytilus edulis* species complex and studies using them.

Source	Study	Diz & Presa 2008	Diz & Presa 2009	Westfall 2011	Nguyen et al. 2011	Štambuk et al. 2013	Giantsis et al. 2014a	Larraín et al. 2014	This thesis 2018
	Region	Spain & Portugal	N.W. Spain	New Zealand	Australia	Croatia	Central & E. Med.	Chile	E. Adriatic
	No. pops	17	5	13		10	13	6	18
	Loci	6	6	7(8)	10	8	10	5	7(8)
Presa et al. 2002	Mgμ1	✓	✓	✓	✕		✕	✓	✓
	Mgμ2	✓	✓		✕	✓	✕		
	Mgμ3	✓	✓		✓	✓	✓	✓	
	Mgμ5	✓	✓	✕	✕		✕		✕
	Mgμ6	✓ 6a & b	✓ 6a & b		✕		✕		
	Mgμ7				✕		✓		
Yu & Li 2007	MGE1				✕	✓	✓		
	MGE2				✕		✕		
	MGE3			✓	✕	✓			✓
	MGE4			✓	✕				✓
	MGE5			✓	✕	✓	✓		✓
	MGE6				✕		✓		
	MGE7			✓	✓	✓	✕		✓
	MGE8				✕	✓	✓		
Varela et al. 2007	Mg181				✕	✓	✓		
	Mg192*				✓				
	Mg220				✕		✕		
Gardeström et al. 2008	MT203			✓	✕		✓	✓	✓
	MT282			✓	✕		✓	✓	✓
Lallias et al. 2009	Med362				✕		✕		
	Med367				✕		✓		
	Med379				✕				
	Med397				✕				
	Med722				✕				
	Med733				✓				
	Med737				✕		✕	✓	
	Med740				✕		✕		
	Med744				✓		✕		
	Med747				✕				

✓ - the locus was used in the study

✕ the locus was tested but not used in the study.

Geneland Analyses

Like Structure, the Geneland program vs. 4.0.8 (Guillot et al. 2012) uses Bayesian analysis implemented through Markov Chain Monte-Carlo algorithms to group genetic data into clusters, and both programs require conformation to HWE and LD assumptions. Geneland also incorporates geographical information of the geographic location of populations through the use of Voronoi tessellation.

Analyses were performed with 100,000 iterations, with removal of the first 200 iterations and subsequently each 100th iteration saved. Analyses used different combinations of the models available, the correlated and uncorrelated allele frequencies models, the null allele model and the spatial model. The uncorrelated model is recommended as a starting point in analysis, with the correlated model able to detect more subtle differentiation but also being more sensitive to violations of model assumptions such as HWE. Although the geographic location of the samples were known, the uncertainty on coordinates options was selected for some analyses since it allows for samples with the same coordinates to be assigned to different populations. Analyses were performed on the 4 of 6 loci dataset including the locus MGE7 which may be under selection and on the 4 of 5 dataset of neutral loci only.

The Geneland runs on the 4 of 6 dataset using the uncorrelated allele frequency model identified either one or two clusters of populations, with the smaller cluster made up of all the populations from the Bay of Kotor (CF, CW, IF, IW) with Zadar (Z) and Neum (NF) from the Central region (Supplementary Figure 3). The correlated allele frequency model identified between 2 and 10 clusters with no overall pattern observed. Unlike the Structure and AWclust analyses, this grouping of the six populations Z, NF, CF, CW, IF and IW cannot be explained by MGE7-allele frequencies alone because although all these populations have relatively high frequencies of allele 243 there are other populations with allele frequencies in this range,

particularly SBF from Albania which had the third highest frequency of allele 243. Analyses on the 4 of 5 dataset were uninformative.
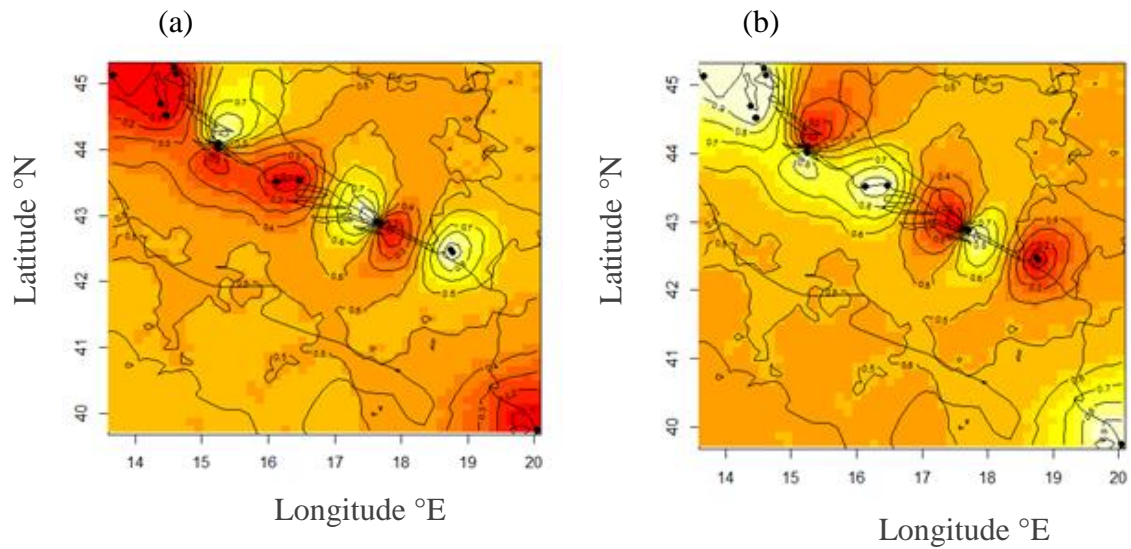

Supplementary Figure 3 Spatial Bayesian analysis using the Geneland program and the uncorrelated allele frequency and spatial models and 6 loci: (a) and (b) assignment of individuals to clusters K = 1 and 2. The palest yellow indicates $p \geq 0.9$ and the deepest orange $p \leq 0.1$ with contour lines at 0.1 increments.

Analyses to identify genetic discontinuities

Three spatial programs, Alleles in Space (AIS, Miller 2005), Barrier (Manni et al. 2004) and SAMOVA (Dupanloup et al. 2002), were used to identify genetic breaks in geographic space. While hierarchical AMOVA assumes clusters a priori, and programs Structure and Geneland assume the number of clusters $K = 1 - N$ a priori, the AIS, Barrier and SAMOVA programs identify genetic discontinuities as the algorithm proceeds. All three programs use Voronoi diagrams and Delaunay triangulation to derive spatial connectivity networks. The AIS and Barrier programs both then apply the Monmonier algorithm to this network, differing in the genetic distance metric used to calculate genetic barriers. The AIS program gives three options computed within the program, raw, residual and "pseudoslope" genetic distance while the Barrier program requires a user-defined genetic distance matrix. Two different genetic distance matrices were used in Barrier, the mean genotypic distance and the AMOVA-derived F_{ST}, both calculated in GenAlEx 6.503 (Peakall & Smouse 2012). Negative F_{ST} values were set to zero. The SAMOVA program uses a "simulated annealing" (SA) algorithm to maximize the F_{CT} as implemented in Arlequin 3.5 for any given number of K groups. Parameters for the SA algorithm were set with S, the number of steps, at 10,000 and A, the cooling factor, at 0.915811. Analyses were performed for $K = 2 - 6$, on the 4 of 7, 4 of 6 and 6 of 6 datasets.

None of these analyses were particularly informative, with inconsistent results across programs and datasets. The Barrier program appeared to reflect the correlation between MGE7 allele frequency and latitude, with several roughly East-West genetic discontinuities identified, which depended on the dataset and genetic distance metric used (Supplementary Figure 4). These geographical divisions depended on the MGE7 locus; when analyses were conducted using neutral loci only, barriers were predominantly placed around single populations.

(a)   (b)
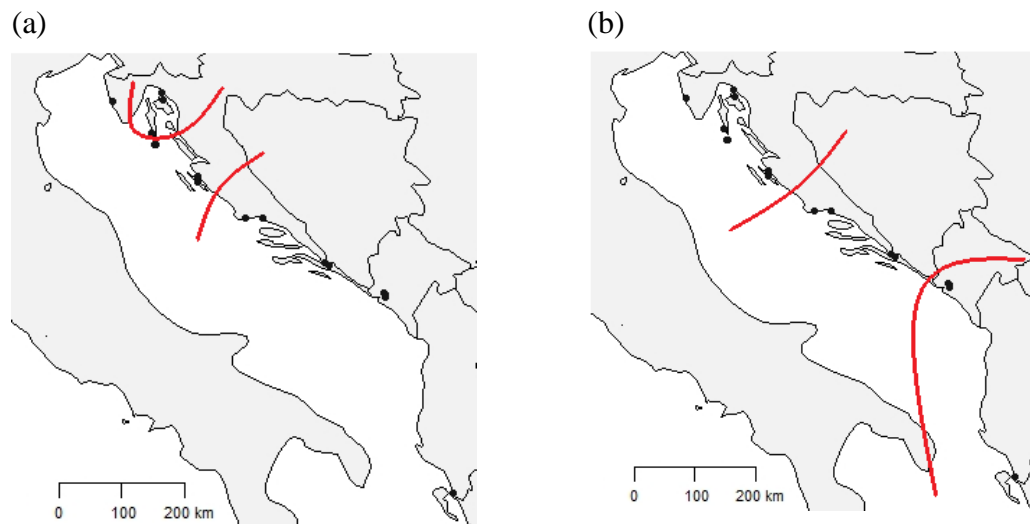
Supplementary Figure 4 Barrier results for *Mytilus galloprovincialis* using F_{ST} as the distance metric for (a) the 4 of 7 dataset and (b) 4 of 6 dataset. Analyses based on datasets without the MGE7 locus were uninformative.

Derivation of the set of best-fitting models
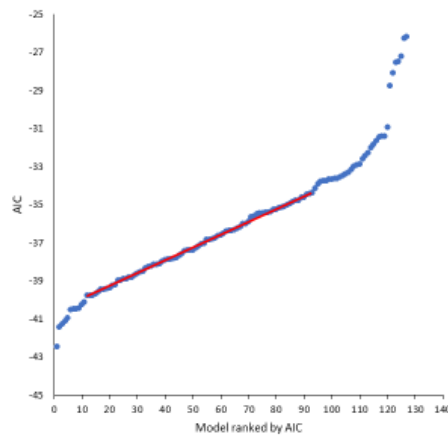
Supplementary Figure 5 AIC plot for variable set 2 (7 variables excluding maxSST) using Frequency of allele MGE7[243] as the genetic measure. The group of best-fitting models reported included all models below the point of inflexion of the AIC plot, here: models 1 – 11. The top-ranked model which includes minSAL alone is reported in Supplementary Table 6 and the group of 11 best fitting models is reported in Supplementary Table 7. 10 of these 11 models included the variable minSAL.

Supplementary Table 5 The 11 best-fitting models for variable set 2 (7 variables excluding maxSST) using Frequency of allele MGE7[243] as the genetic measure, also showing the percentage of models including each environmental variable.

Model Rank	minSST	minSAL	MaxSAL	minCHL-a	MaxChl-a	Lat	Total CD
1	-	✓	-	-	-	-	-
2	-	✓	-	-	-	✓	-
3	-	✓	-	-	✓	-	-
4	✓	✓	-	-	-	-	-
5	-	✓	-	✓	-	-	-
6	-	✓	✓	-	-	-	-
7	✓	✓	✓	-	-	-	-
8	-	✓	-	-	✓	-	✓
9	-	✓	-	-	-	-	✓
10	-	-	-	-	✓	✓	-
11	-	✓	✓	-	✓	-	-
All best-fit models	18.2%	90.9%	27.3%	9.1%	36.4%	18.2%	18.2%

Summary of GLM analyses

Supplementary Table 6 The top-ranked model for each analysis (8 subsets of the environmental and geospatial variables) and 5 genetic measures showing the variables included in the model and the model p-value.

Variable Set	Genetic Measure	Top-ranked model	p
	Mean LinFST based on 5 neutral loci	maxSST minSAL minCHL-a maxCHL-a Lat TotalCD	0.0002
1	Mean LinFST based on 6 loci	minSAL Lat	0.071
8 variables	Mean LinFST based on MGE7 only	minSAL Lat	0.023
	Frequency of MGE7[243] allele	minSAL	<0.0001
	Frequency of MGE7[243] homozygote	minSST minSAL minCHL-a	<0.0001
	Mean LinFST based on 5 neutral loci	minSAL minCHL-a maxCHL-a Lat TotalCD	0.0003
2	Mean LinFST based on 6 loci	minSAL Lat	0.071
7 variables	Mean LinFST based on MGE7 only	minSAL Lat	0.023
(maxSTT excluded)	Frequency of MGE7[243] allele	minSAL	<0.0001
	Frequency of MGE7[243] homozygote	minSAL minSST minCHL-a	<0.0001
	Mean LinFST based on 5 neutral loci	minSST	0.177
3	Mean LinFST based on 6 loci	maxSST Lat	0.098
7 variables	Mean LinFST based on MGE7 only	maxCHL-a	0.026
(minSAL excluded)	Frequency of MGE7[243] allele	maxCHL-a Lat	0.0001
	Frequency of MGE7[243] homozygote	minSST maxSST minCHL-a	<0.0001
	Mean LinFST based on 5 neutral loci	minSST	0.177
4	Mean LinFST based on 6 loci	minSAL Lat	0.071
7 variables	Mean LinFST based on MGE7 only	minSAL Lat	0.023
(maxCHL-a excluded)	Frequency of MGE7[243] allele	minSAL	<0.0001
	Frequency of MGE7[243] homozygote	minSST minSAL minCHL-a	<0.0001
5	Mean LinFST based on 5 neutral loci	minSST	0.177
6 variables	Mean LinFST based on 6 loci	maxCHL-a Lat	0.171
(maxSTT & minSAL	Mean LinFST based on MGE7 only	maxCHL-a	0.026
excluded)	Frequency of MGE7[243] allele	maxCHL-a Lat	0.0001
	Frequency of MGE7[243] homozygote	maxCHL-a Lat	0.0002
6	Mean LinFST based on 5 neutral loci	minSST	0.177
6 variables	Mean LinFST based on 6 loci	minSAL Lat	0.071
(maxSTT & maxCHL-a	Mean LinFST based on MGE7 only	minSAL Lat	0.023
excluded)	Frequency of MGE7[243] allele	minSAL	<0.0001
	Frequency of MGE7[243] homozygote	minSST minSAL minCHL-a	<0.0001
7	Mean LinFST based on 5 neutral loci	minSST	0.177
6 variables	Mean LinFST based on 6 loci	maxSST Lat	0.099
(minSAL & maxCHL-a	Mean LinFST based on MGE7 only	maxSST Lat	0.034
excluded)	Frequency of MGE7[243] allele	maxSST	<0.0001
	Frequency of MGE7[243] homozygote	minSST maxSST minCHL-a	<0.0001
8	Mean LinFST based on 5 neutral loci	minSST	0.177
5 variables	Mean LinFST based on 6 loci	minSST	0.344
(maxSST, minSAL &	Mean LinFST based on MGE7 only	minSST maxSAL Lat	0.039
maxCHL-a excluded)	Frequency of MGE7[243] allele	maxSAL Lat TotalCD	0.00143
	Frequency of MGE7[243] homozygote	maxSAL Lat TotalCD	0.0006

Supplementary Table 7 Results of GLM analyses with heatmap showing the percentage of each variable included in the set of best-fitting models. Eight different sets of environmental and geospatial variables were used to explain genetic variation (Table 3.2). Five different measures of genetic variation are shown: the mean LinF$_{ST}$ for the 5 neutral loci, mean LinF$_{ST}$ for all six loci, mean LinF$_{ST}$ for the MGE7 locus, frequency of the MGE7^{243} allele, frequency of the homozygous genotype MGE7$^{243/243}$. The number of models for which results are displayed depends on the plot of AIC against model number (see Supplementary Figure 5). Significant values according to the all effects test are in italic bold. The p-values for the all effects test on the intercept (null hypothesis is the intercept is at zero) are also shown.

Variable Set	Genetic Measure	# model	min SST	max SST	min SAL	max SAL	min CHL-a	max CHL-a	Lat	Total CD	All effects Intercept p-value
1 8 variables	Mean LinF$_{ST}$ based on 5 neutral loci	38	52.6	52.6	***100***	57.9	***50***	***100***	57.9	***84.2***	0.860
	Mean LinF$_{ST}$ based on 6 loci	22	18.2	36.4	40.9	18.2	9.1	22.7	***81.8***	18.2	0.020
	Mean LinF$_{ST}$ based on MGE7 only	13	30.8	7.7	46.2	30.8	15.4	46.2	84.6	30.8	0.135
	Frequency of MGE7^{243} allele	46	32.6	30.4	84.8	28.3	26.1	32.6	26.1	28.3	0.601
	Frequency of MGE7^{243} homozygote	50	68	30	88	34	46	46	24	34	0.241
2 7 variables (maxSTT excluded)	Mean LinF$_{ST}$ based on 5 neutral loci	18	55.6		***100***	55.6	***50***	***100***	55.6	***88.9***	0.950
	Mean LinF$_{ST}$ based on 6 loci	14	21.4		57.1	21.4	7.1	28.6	***78.6***	21.4	0.012
	Mean LinF$_{ST}$ based on MGE7 only	12	33.3		50	33.3	16.7	50	83.3	***33.3***	0.062
	Frequency of MGE7^{243} allele	11	18.2		90.9	27.3	9.1	36.4	18.2	18.2	0.498
	Frequency of MGE7^{243} homozygote	31	67.7		***100***	38.7	38.7	51.6	25.8	39.7	0.286
3 7 variables (minSAL excluded)	Mean LinF$_{ST}$ based on 5 neutral loci	9	44.4	***11.1***		22.2		11.1	22.2	22.2	0.178
	Mean LinF$_{ST}$ based on 6 loci	24	25	33.3		0	0	0	***75***	25	0.032
	Mean LinF$_{ST}$ based on MGE7 only	15	40	20		26.7	46.7	***73.3***	93.3	26.7	0.140
	Frequency of MGE7^{243} allele	19	15.8	73.7		15.8	36.8	42.1	47.4	21.1	0.428
	Frequency of MGE7^{243} homozygote	21	***61.9***	90.5		23.8	***81.0***	33.3	38.1	23.8	0.122
4 7 variables (maxCHL-a excluded)	Mean LinF$_{ST}$ based on 5 neutral loci	9	***44.4***	11.1	11.1	22.2	0		22.2	22.2	0.007
	Mean LinF$_{ST}$ based on 6 loci	26	26.9	30.8	42.3	26.9	19.2		***76.9***	30.8	0.014
	Mean LinF$_{ST}$ based on MGE7 only	23	39.1	39.1	56.5	39.1	13		***78.3***	21.7	0.253
	Frequency of MGE7^{243} allele	9	22.2	22.2	88.9	22.2	11.1		11.1	11.1	0.478
	Frequency of MGE7^{243} homozygote	24	83.3	***37.5***	79.2	37.5	58.3		25	20.8	0.082
5 6 variables (maxSTT & minSAL excluded)	Mean LinF$_{ST}$ based on 5 neutral loci	8	50			25	0	12.5	25	25	0.707
	Mean LinF$_{ST}$ based on 6 loci	7	28.6			28.6	0	42.9	***57.1***	14.3	0.029
	Mean LinF$_{ST}$ based on MGE7 only	12	41.7			33.3	50	***91.7***	***91.7***	33.3	0.021
	Frequency of MGE7^{243} allele	5	20			20	20	100	100	20	0.535
	Frequency of MGE7^{243} homozygote	6	66.7			16.7	***66.7***	100	66.7	33.3	0.375
6 6 variables (maxSTT & maxCHL-a excluded)	Mean LinF$_{ST}$ based on 5 neutral loci	8	***50***		***12.5***	25	0		25	25	0.014
	Mean LinF$_{ST}$ based on 6 loci	10	30		***70***	30	10		***80***	20	0.010
	Mean LinF$_{ST}$ based on MGE7 only	15	53.3		73.3	53.3	13.3		***73.3***	26.7	0.143
	Frequency of MGE7^{243} allele	19	42.1		***100***	42.1	31.6		31.6	31.6	0.400
	Frequency of MGE7^{243} homozygote	15	***80***		***100***	40	46.7		33.3	26.7	0.108
7 6 variables (minSAL & maxCHL-a excluded)	Mean LinF$_{ST}$ based on 5 neutral loci	8	***50***	12.5		25	0		25	25	0.027
	Mean LinF$_{ST}$ based on 6 loci	2	0	100		0	50		***100***	0	0.075
	Mean LinF$_{ST}$ based on MGE7 only	10	50	70		30	10		70	20	0.497
	Frequency of MGE7^{243} allele	13	18.2	100		18.2	45.5		27.3	27.3	0.867
	Frequency of MGE7^{243} homozygote	14	57.1	***100***		29	***79***		35.7	28.6	0.250
8 5 variables (maxSST, minSAL & maxCHL-a excluded)	Mean LinF$_{ST}$ based on 5 neutral loci		57			29	14		29	29	0.294
	Mean LinF$_{ST}$ based on 6 loci	3	***66.7***			33.3	0		***66.7***	0	0.076
	Mean LinF$_{ST}$ based on MGE7 only	1	100			100	0		100		0.184
	Frequency of MGE7^{243} allele	5	60			60	0		***100***	40	0.002
	Frequency of MGE7^{243} homozygote	3	0			***66.7***	33		***100***	66.7	0.001

Isolation by Environment

In Chapter 2, Mantel tests for Isolation by Distance were inconclusive with some analyses not significant and others weakly significant. Mantel tests for Isolation by Environment (IBE) and Partial Mantel tests to control for geographic distance were carried out using linearised F_{ST}, coastal distance, CD, and the main variables found to explain environmental structure identified by the GLM analyses and deltaSAL. If both genetic distance and environmental distance are correlated to geographic distance, then a Mantel test between genetic and environmental distance will reveal the correlation of each to geographic distance instead of the underlying relationship independent of geographic distance. Partial Mantel tests avoid this problem by testing the correlation between a genetic distance matrix and an environmental difference matrix while controlling for a geographic distance matrix (i.e. keeping the geographic distance statistically constant). The partial Mantel test removes any geographic structure common to both the genetic and the environmental matrices and allows the underlying relationship between genetic and environmental distance to be revealed (Castellano & Balletto, 2002). Mantel and partial Mantel tests were performed using environmental distances based on the environmental variables minSAL, maxSST and deltaSAL. The environmental distance between any two populations was the difference between for the variables for each population. Mantel and partial Mantel tests were performed in IBD Web Service (Jensen et al. 2005) using 1000 randomisations and all other default settings.

Results for the Mantel test LinF_{ST} against log(1+CD) are slightly different here ($R^2 = 0.022$, p = 0.032) from those presented in Chapter 2 ($R^2 = 0.021$, p = 0.040) due to the different methodologies used in GenAlEx and IBD Web Service. The GenAlEx computation

changed negative F_{ST} values to zero while the IBD Web Service uses negative values in the test.

Supplementary Table 8 Mantel and partial Mantel test results for 18 populations of *Mytilus galloprovincialis* between linearised F_{ST} (Lin*F*st) and environmental variables minSAL, deltaSAL, maxSST and geographic distance, log coastal distance.

		R	R²	p
Lin*F*st and log coastal distance				
Correlation		0.149	0.022	0.032
Lin*F*st and minSAL				
Correlation		0.261	0.068	0.014
Partial correlation	Lin*F*st and minSAL controlling for log coastal distance	0.216	0.047	0.056
Partial correlation	Lin*F*st and log coastal distance controlling for minSAL	0.006	<0.001	0.481
Lin*F*st and deltaSAL				
Correlation		0.279	0.078	0.006
Partial correlation	Lin*F*st and deltaSAL controlling for log coastal distance	0.239	0.057	0.040
Partial correlation	Lin*F*st and log coastal distance controlling for deltaSAL	-0.007	<0.001	0.480
Lin*F*st and maxSST				
Correlation		-0.232	0.054	0.030
Partial correlation	Lin*F*st and maxSST controlling for log coastal distance	-0.186	0.035	0.088
Partial correlation	Lin*F*st and log coastal distance controlling for maxSST	0.049	0.002	0.369

Mantel tests between Lin*F*st and minSAL, deltaSAL, maxSST were all significant (p = 0.014, 0.006 and 0.030 respectively). The p-value for the Mantel test between Lin*F*st and log(1+CD), p = 0.032 was higher than those for minSAL and deltaSAL but approximately the same as that for maxSST. When log coastal distance was controlled in a partial Mantel test, the p-values were higher (p = 0.056, 0.040 and 0.088 respectively) with only test with deltaSAL being significant. The partial Mantel tests controlling for the environmental variables were not close to significant, indicating that the strength of the Mantel test between Lin*F*st and log(1+CD) in Chapter 2 relied on the stronger association of Lin*F*st and the environmental variables.

Geste Analyses

The Bayesian method, Geste (Foll & Gaggiotti, 2006), was used to test the effect of environmental and geospatial factors on genetic differentiation. Geste estimates population-specific F_{ST} values using a hierarchical Bayesian approach. The variables considered here were maxSST, deltaSAL and maxCHL-a along with geospatial factors Lat and Long. Two genetic datasets were used: 6 loci made up of 5 neutral loci and MGE7 and the MGE7 locus alone.

For the 6 loci dataset, the highest posterior probability corresponded to the model with the constant term only (p = 0.694). When only the MGE7 locus was used, the constant-term model had much lower highest posterior probability (0.064 when 3 environmental variables were tested and 0.034 when 2 spatial variables were also included). Using MGE7 alone, deltaSAL had the greatest influence on genetic differentiation followed by maxSST (sum of posterior probabilities of all models including the factor, p=0.485 and 0.405 respectively, when 3 environmental and 2 geospatial variables were used).

Supplementary Table 9 Results of Geste Bayesian analysis of effect of environmental and geospatial variables on genetic differentiation – for each variable the sum of posterior probabilities of all models that include that variable are shown. 2 datasets are used, 6 loci including 5 neutral loci and MGE7 and the MGE7 locus alone.

Factors	6 loci 5 neutral & MGE7 3 variables	MGE7 alone 3 variables	MGE7 alone 5 variables
Constant	0.694	0.064	0.034
maxSST	0.108	0.401	0.405
deltaSAL	0.126	0.457	0.485
maxCHL-a	0.141	0.386	0.361
Lat	-	-	0.221
Long	-	-	0.285