The Development of Genetic Resources to Futureproof Shellfish Aquaculture against Climate Change

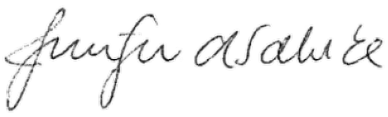
Jennifer Catherine Nascimento Schulze

PhD Thesis

The Development of Genetic Resources to Futureproof Shellfish Aquaculture against Climate Change

Submitted by Jennifer Catherine Nascimento Schulze to the University of Exeter as a thesis for the degree of Doctor of Philosophy in Biological Sciences in January 2023.

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.



Main supervisor: Dr Robert P Ellis

Secondary supervisors:

Dr Tim P Bean, Prof Dr Eduarda Santos, Dr Ceri Lewis, Dr Matthew Sanders

Acknowledgments

In the first place, I would like to thank my main supervisor, Rob Ellis, for believing I was the right student to work on this project. Thank you for not giving up on me and making the whole thing happen - I am eternally grateful I could work on a topic I am truly passionate about. I am equally thankful for all the advice and support provided during the past years, especially for your enthusiasm about science. Most importantly, thank you for the freedom you have given me these past years, allowing me to be curious and explore my own ideas. My heartfelt thanks also goes to my second supervisor, Tim Bean. Thank you for always providing valuable input, for the numerous opportunities to grow as a scientist and travel around the world, and, for helping me see the bigger picture and keep on track and reach my goals. Above all, thank you for always being there to support me when I needed advice. The trust you both had in me was truly motivating and I had the best of times working with you during these past years.

Thank you Eduarda Santos for the great supervision, for the excitement to discuss science and for teaching me the value of being precise and well planned for an experiment. Above all, I am thankful for the emotional support given throughout these years. I also thank Matthew Sanders, my supervisor at Cefas, for all the help provided during my time in Weymouth and for always keeping in touch throughout the years, making sure the project was progressing well. It was great to work with someone truly enthusiastic about bivalve physiology.

Thanks to my supervisor Bonnie Fraser and the team in Exeter, Josie Paris, Jim Withing as well as Carolina Penaloza for introducing me to the world of bioinformatics. I owe you what I know today and will be forever grateful.

Thanks to Rod Wilson, for evaluating the progress of my project during these past years, for the good advice given, for the inspiring talks and for inviting me to be part of such a great team.

I also thanks the National Environmental Research Council (NERC) and Cefas for providing the funding that has enabled me to undertake this PhD. A special thanks for the NERC team, Sara and Emmi, who have given the greatest support and all their kindness to the PhD students, especially during the coronavirus pandemic times.

Thank you for our collaborators at GEOMAR, Frank Melzner and Jahan Vajedsamiei. I feel extremely lucky to continue working with, and learning from, you since 2015 and hope we will develop many other great mussel studies together in the future.

I also would like to say thanks to everyone who has supported me during the course of my PhD. Particularly, thanks to my Hatherly colleagues for making this journey fun and light and for the friendship we have built. Kirsty McLeod, Joanne Morten, Jessica Rudd, Charlie Ellis and Dan Osmond: it was a great pleasure to share conversations, coffees, laughs and climbs with you. Charlie and Louisa, thanks for the lovely mussel sampling adventures.

I special thanks to Armin, who during this past year was the most understanding and supportive person I could have asked for. Thanks for always cheering me up after a long day in front of the computer, for listening to me and for discussing science with me at any time. The whole journey was much more enjoyable with you by my side.

Finally, I thank you my parents for the unconditional support, for always believing in me and for going above and beyond to assure I could follow my dreams. In particular, thank you mum, for the friendship and dedication, especially during these past years and for the beautiful paintings you have done, which are now part of the thesis.

“Andei por andar, andei

E todo caminho deu no mar

Andei pelo mar, andei

Nas águas de Dona Janaína

A onda do mar leva,

A onda do mar traz

Quem vem pra beira da praia, meu bem

Não volta nunca mais

Quem vem pra beira do mar, ai

Nunca mais quer voltar”

Dorival Caymmi

Abstract

The global population is expected to reach 9.7 billion people by 2050. Therefore, sustainable intensification of the food system to support the increasing population under a changing climate is vital. Aquaculture is the fastest expanding food sector globally, but projections suggest this industry will be severely impacted by anthropogenically induced climate change in the coming decades. Genomic selection has been highlighted by the Food and Agriculture Organisation (FAO) as an avenue to ensure the aquaculture industry will thrive under future climatic conditions, fast-tracking breeding of specialised lines. Following this line of reasoning, the work in this thesis aimed to contribute to the advances in genomic breeding of marine molluscs, a highly sustainable source of animal protein. In the first chapter, I reviewed the existing literature on shellfish aquaculture, exploring how the current production systems and aquaculture practices may impact the genetic diversity of the produced stock and, consequently, the success of any proposed genomic breeding in these taxa, alongside identifying the main gaps in our current knowledge in this field. Subsequently, each experimental chapter of this thesis then addressed a separate gap in our current understanding. In chapter II, I developed a multi-species 60 K SNP-array applicable for genotyping four mussel species relevant for aquaculture: *M. edulis*, *M. galloprovincialis*, *M. chilensis* and *M. trossulus*. Following low-coverage whole genome sequencing of 138 mussels, ~60K SNPs were implemented in the platform, from which 23,252 are applicable for genotyping *M. edulis* individuals, 22,165 for *M. chilensis*, 20,504 SNPs *M. galloprovincialis* and 20,149 for *M. trossulus*. This tool therefore represents a major advance in the current technological capability available in this species complex, and will allow researchers to explore the genetic diversity of mussels using a dense number of markers (on the scale of thousands), whilst producing comparable data among studies. It is also applicable for breeding purposes and may facilitate future implementation of genomic selection in these taxa. Subsequently, in chapter III, I applied this SNP-array to explore the genomic structure of blue mussels in South West England; an important area for mussel aquaculture and a region that has been key for our understanding of hybrid zone dynamics, between the species *M. galloprovincialis* and *M. edulis,* historically. Results from this

chapter agree with the previous description of the hybrid zone, with *M. galloprovincialis* genotypes dominating the north coast of Cornwall whilst *M. edulis* is the prevailing species in the southern coast of SW England. Furthermore, the transition zone between the two species was located at Lizard Point. These results validate the utility of the SNP-array for determining speciation across a wide region, and update our knowledge in genotype distribution in this key hybrid zone. In the thesis' fourth chapter, I explored how selection towards environmental resilience impacts the phenotypes and/or genotypes of blue mussels. For this, I compared the genetic diversity and performance of juvenile mussels exposed to a thermal stress event to those of individuals which did not faced it. Results from this chapter suggest that thermal resilience may impact the performance of mussels, as stressed individuals presented a lower dry tissue weight (g) than naïve ones. In addition, a shift towards *M. edulis* genotypes occurred in the selected cohort. With climate change expected to increase the average sea surface temperature of the ocean, results presented in this thesis shows that such changes may impact the genetic background of mussel spat in the Baltic Sea and possibly. Such results may also impact aquaculture production of this species. In the final chapter, I discuss how the research questions explored in this thesis advance our knowledge in marine bivalve genomics and how they may change the field by supporting the development and application of genomic approaches for breeding these taxa.

Key words: sustainable aquaculture, genomic breeding, SNP-array, hybridisation, blue mussels, climate change

Contents

Figure Contents

Table Contents

List of Abbreviations

AF Allele frequency
AXAS Axiom Analysis Software

Bp basepair

CO_2 Carbonic dioxide
CR Call rate

DAPC Discriminant analysis of Principal components
DNA Deoxyribonucleic Acid
DQC DishQC

FAO Food and Agriculture Organisation
FSW Filtered sea water

GAMM Generalised additive mixed model
GS Genomic Selection
GWAS Genome wide association study
GxE Genotype by Environment

H-index Hybrid index

IPCC Intergovernmental Panel on Climate Change

Kb Kilobase

L Litres
LD Linkage disequilibrium

MAF Minor allele frequency
mg milligram
mL millilitre

NGS Next generation sequencing
NS Non-selected
N_e Effective population size
N_b Effective number of breeders during a single breeding event

PCA **P**rincipal component analysis
PC Principal component

QC Quality control

QTL Quantitative trait loci

RNA Ribonucleic acid

S Selected
SE Standard error
SH Shell height
SL Shell length
SNP Single Nucleotide Polymorphism
SST Sea surface temperature
SW South West
SW (Chapter IV)Shell dry weight

TW Tissue weight
UK United Kingdom 1
$\mathbf{V_k}$ Variance in reproductive success

WE Water exchange
WGS Whole genome sequencing

µl microliter

1. Chapter I – General Introduction: Optimising Hatchery Practices for Genetic Improvement of Marine Bivalves

This chapter has been adapted from the version published by the Scientific Journal 'Reviews in Aquaculture' as:

Nascimento-Schulze, J.C., Bean, T.P., Houston, R.D., Santos, E.M., Sanders, M.B., Lewis, C. and Ellis, R.P., 2021. **Optimizing hatchery practices for genetic improvement of marine bivalves. Reviews in Aquaculture**, *13*(4), pp.2289-2304

In this chapter, I was responsible for researching and reviewing articles and for leading the writing of the review paper.

The future of bivalve aquaculture relies on artificial propagation

With the global human population projected to exceed 9 billion by 2050, food production must increase by at least 59% to meet projected demand (Valin *et al.* 2014). Feeding this growing population, whilst maintaining biodiversity and good environmental stewardship, is one of the major global challenges of the 21st century. This issue is exacerbated further by the need to ensure that the future intensification of food production is sustainable, especially in the face of climate change (IPCC 2018; United Nations 2015).

Aquaculture is the fastest growing food production sector globally, expanding on average 6.4% per annum since 2001 (Subasinghe 2017). Nearly half of the current global finfish and shellfish production derives from aquaculture (FAO 2019a), with this sector expected to underpin most future growth in seafood production (Kobayashi *et al.* 2015). Currently, mollusc farming accounts for approximately 21 % of world aquaculture production (Subasinghe 2017).

Important scientific advances in bivalve husbandry practices (i.e. optimisation of diet, fertilisation protocols and larval rearing) occurring within the last century (e.g. Carriker 1956; Galstoff 1938; Loosanoff and Davis 1963) led to the establishment of the first commercial bivalve mollusc hatcheries (Mann 1983) resulting in the global expansion of shellfish aquaculture. The ability to control environmental conditions in indoor facilities enables broodstock conditioning and spat production almost year-round. Most importantly, the development of a constant and reliable source of spat benefits the expansion of the bivalve aquaculture sector, facilitating the predictability of production and enabling the possibility of selective breeding.

Hatcheries are expected to play a key role in the continued expansion of bivalve aquaculture. The potential of hatchery production is highlighted in China, which accounts for 80% of global production of Pacific oysters (*Crassostrea gigas*) (Yang *et al.*, 2014), and the sector now relies almost entirely on hatchery sourced spat (Li *et al*., 2011). Nonetheless, demand for hatchery-produced spat

is often low in areas where natural (wild capture) spat is available and abundant. A similar situation occurs in France, which is responsible for 82% of Pacific oyster production in Europe (92,000 tonnes in 2018 (Eurostat 2020)) where over 60% of spat is captured from wild sources (Richez 2012). This contributes to a slow shift from a natural to hatchery production model (Adamson *et al*., 2017). This same production template is also true for mussels; currently, industries for two of the main farmed species, the blue mussel (genus *Mytilus*) in Europe (Kamermans *et al.* 2013) and the green-lipped mussel (*Perna canaliculus*) in New Zealand (Symonds *et al.* 2019) still rely primarily on natural spat. This process is an inexpensive but unreliable practice, which is vulnerable to habitat disturbances and restricts the development of cultivation technologies such as selective breeding.

The Food and Agriculture Organisation (FAO) has recently proposed a number of key developments which will assist the aquaculture industry in addressing several long-term sustainability challenges (FAO 2016). One priority area highlighted by the FAO is the use of stock management and selective breeding to produce lines with greater reliability and productivity in a wide range of environments (FAO 2019b). To date, encouraging responses to selection have been observed in aquatic species: the average gain in body weight per generation is 8.7% in shrimps, 10.3% in oysters and between 9% and 17.9% among finfish species (see review by Gjedrem and Rye 2018). Although recent estimates show that in the 10 main farmed aquatic species 75% of production benefits from some form of selection (Houston *et al.* 2020), only a small percentage of global aquaculture production (less than 10% in 2012) utilises genetically improved stock (Gjedrem *et al.*, 2012).

Despite an increasing availability of genomic resources for bivalves, the mechanisms underlying domestication (i.e. adaptation to a farmed environment), and genotype-environment interactions (*G*x*E*) ongoing in cultured bivalve species remain poorly studied. The degree to which these processes influence the response to selection in these taxa, and consequently the potential to genetically improve organisms, represent two key knowledge gaps with respect to bivalve selective breeding (Figure 1.1 and 1.2). Accordingly, the potential for losses in genetic diversity during production is exacerbated and likely hinders the efficiency of existing hatchery management and

selective breeding programs, jeopardizing sustainable growth of this sector. There is a fundamental need to clarify the impacts of hatchery-management practices on the genetic and phenotypic constitution of cohorts, and the resulting long-term implications for bivalve production.

In the following sections of this introductory chapter, I explore the mechanisms by which production practices and life-history characteristics can influence shellfish aquaculture production (Figure 1.1). First, I describe the status of selection in bivalve aquaculture globally and the different methods employed for production. Further, I discuss how management practices impact the genetic variability and quality of spat during hatchery-propagation, potentially benefiting or hindering the optimisation of selective breeding approaches for these taxa. Last, I argue why and how a greater control of hatchery-propagation processes can contribute to the sustainable intensification of the bivalve aquaculture industry. There are considerably fewer studies investigating the consequences of domestication selection in bivalves in comparison to other aquatic species. Therefore, in order to infer the possible consequences of artificial propagation in these organisms, I compare the selection pressures acting in hatcheries with those acting in the wild, when applicable. By identifying and summarising the main gaps in knowledge in the field, whilst discussing their possible causes and the consequences they might have to shellfish aquaculture production, I set the stage to the following chapters of this thesis (Chapters II,III and IV), in which I address some of these open questions. More specifically, in the following chapters of the thesis, I i) develop a genomic tool which can be applied to the investigation of population structure in wild and commercial populations of Blue Mussels, *Mytilus* spp, a commercially relevant marine bivalve mollusc species (Chapter II); ii) apply this genotyping platform tool to explore genetic structure in wild populations of Blue Mussels in the Southern West coast of England (Chapter III); iii) investigate the consequences of pre-selection to stressful environmental conditions on the performance and the genetic composition of marine bivalve mollusc early life-stages, exploring how such selection might benefit breeding towards environmental resilience (Chapter IV). Each of these chapters will therefore help, in turn, to elucidate questions that

remain surrounding the contribution of selective breeding practices to the sustainable optimisation of global shellfish aquaculture.

By identifying the main gaps in knowledge and advancing our understanding of these topics, I expect to contribute towards an increased efficiency and accuracy of selection in these taxa, as well as inspire future research which may contribute to increasing efficiency and accuracy of selection in marine bivalves.

Selective breeding in bivalve aquaculture: current status and opportunities

Successful breeding programs have been established for bivalves worldwide and include those applying mass and family selection approaches (Table 1.1) (Hollenbeck and Johnston 2018). In mass selection, individuals are typically selected according to their performance in comparison to the population's mean for a specific trait (e.g. growth) without fully accounting for family structure. This strategy can be effective but runs the risk of inbreeding depression and is only suitable for a focus on one or two traits. Alternatively, family selection is based on pedigree information, and individuals from the top performing families are chosen to form the breeding populations, allowing for effective maintenance of genetic diversity. Family-based selection has been applied in a commercial *M. galloprovinciallis* breeding program. Here, the use of 77 full-sib families resulted in a heritably of 0.35 (SE =0.09) for total weight and 0.23 (SE =0.08) for meat yield as a ratio between meat weight and total weight, after 2 generations; both of which are commercially relevant traits (Nguyen *et al.*, 2014).

Selection has also successfully improved traits such as growth rate (de Melo *et al.*, 2016; Hershberger *et al.*, 1984), disease resistance (Dégremont *et al*., 2015; Dove *et al.*, 2013; Naciri-Graven *et al.*, 1998) and resilience towards environmental perturbation (Parker *et al.*, 2015). Despite these success stories, bivalve aquaculture production still relies greatly on wild type strains (Hollenbeck & Johnston 2018), that may not be adapted to the farming environment (Yáñez *et al.*, 2015). Therefore, significant potential for genomic improvement exists, providing the opportunity to maximise productivity for bivalve aquaculture species worldwide.

The recent development and increasing affordability of high throughput sequencing technologies has facilitated the incorporation of genomic tools in breeding programs of aquatic species (Zenger *et al.* 2019). This has enabled a step forward from family selection, particularly for traits which are difficult or impossible to measure directly on selection candidates, such as disease resistance. For such traits, family selection would only allow for family level breeding values, thereby missing out on within-family genetic variation. Genomic tools allow breeders to access and utilise the within-family component of genetic variation. This can be achieved in two main ways. Firstly, mapping of quantitative trait loci (QTL) allows the identification of genetic markers significantly associated with a specific trait within the species of interest. Selection for traits with large effect QTLs can be improved by applying marker-assisted selection (Zenger *et al.* 2019). Secondly, genomic selection can be applied for selection of polygenic traits (Meuwissen *et al.*, 2016). Such approach can cover a large number of loci across the genome and provides enough information to capture all causative variants for a given trait, as loci are expected to be in linkage disequilibrium with one or more common markers (Meuwissen *et al.*, 2001). Besides, genomic selection captures the within-family variance as markers shared between individuals can be identified, increasing the accuracy of the estimated breeding values and response to selection (see Zenger *et al.*, 2019 and references therein). Additionally, it enables retrospective parental assignment, which allows multiple families to be grown in mixed tanks and reduces the generation of confounding genetic and environmental effects. Genomic selection can also be designed to fit different levels of ploidy (Ashraf *et al.* 2016; Endelman *et al.* 2018) and can be a valuable asset to guarantee a high precision in breeding programs utilising broodstock with higher levels of ploidy.

	Species	Group	Location	Type of selection	Program type	Founded	References
Mediterranean mussel	*Mytilus galloprovincialis*	Mussel	Australia	Family	Industrial	2008	(Nguyen & Ingram 2012)
Greenlip mussel	*Perna canaliculus*	Mussel	New Zealand	Family	Industrial	1999	(Camara & Symonds 2014)
Pacific oyster	*Crassostrea gigas*	Oyster	USA	Family	Industrial	1996	(de Melo *et al.* 2016; Langdon *et al.* 2003)
Pacific oyster	*Crassostrea gigas*	Oyster	Australia	Family; mass	Industrial	1997	(Kube *et al.*, 2011; Ward *et al.*, 2005)
Pacific oyster	*Crassostrea gigas*	Oyster	New Zealand	Family	Industrial	1999	(Camara & Symonds 2014)
Pacific oyster	*Crassostrea gigas*	Oyster	France	Mass	Experimental	2009	(Dégremont *et al.*, 2015)
Pacific oyster	*Crassostrea gigas*	Oyster	China	Mass	Experimental	2007	(Li *et al.*, 2011; Zhong *et al.*, 2016)
Sydney rock oyster	*Saccostrea glomerata*	Oyster	Australia	Family; mass	Industrial	1990	(Dove *et al.*, 2013; Nell *et al.*, 1996; 1999)
Bay Scallop	*Argopecten irradians*	Scallop	China	Mass	Unknown	2001	(Zheng *et al.*, 2004; 2006)

Table 1-1: Large-scale breeding programs for cultured marine bivalve species (adapted from Hollenbeck & Johnston, 2018).

Having a set of tools which link high-resolution genetics with phenotypes is a main requirement for genomic selection. To date, genomes have been assembled for several of the main cultured species (Hollenbeck & Johnston 2018). In addition, the development of DNA markers including microsatellites (Li *et al.*, 2003; Wang *et al.*, 2016) and single nucleotide polymorphisms (SNPs) (Fleury *et al.*, 2009; Nguyen *et al.*, 2014; Sauvage *et al.*, 2007; Vu *et al.*, 2021; Wang *et al.*, 2015) as well as the identification of genomic regions associated to traits of economic importance through

QTL mapping (Guo *et al.*, 2012; Jiao *et al.*, 2014; Sauvage *et al.*, 2010) and genome wide association studies (GWAS) (Gutierrez et al. 2018; Meng et al. 2019), create a genomic toolbox which provides a backbone for future research. Ultimately, this information promotes the development of genomic based selection techniques and the fine-tuning of breeding programs.

Marine bivalves share complex genomic and life-history features, including high levels of nuclear genetic diversity, high heterozygosity, and elevated numbers of deleterious mutations and null alleles (Plough and Hedgecock, 2011; Bierne *et al.*, 1998; Gerdol *et al.*, 2019; Hollenbeck and Johnston, 2018 and references therein). In addition, reproductive attributes (broadcast spawning, high fecundity, high early mortality rates), and a high variance in reproductive success (V_k) among individuals (Hedgecock & Pudovkin 2011), are commonly described in these taxa. Variance in reproductive success can result in low effective population sizes (N_e) and low numbers of effective breeders (N_b) relative to census size, termed 'sweepstake reproduction', which has been observed in both wild and hatchery-propagated stock (Boudry *et al.*, 2002; Hedgecock, 1994; Hedgecock & Sly, 1990; Plough & Hedgecock, 2011). Heterozygous deficiencies relative to Hardy-Weinberg equilibrium, and segregation distortion of markers described in paired crosses, are also commonly reported in bivalves (Launey and Hedgecock, 2001; Peñaloza *et al.*, 2014).

These properties of the bivalve genome, together with specific life-history characteristics of these organisms may influence the efficiency and applicability of genomic resources in breeding programs. Therefore, efforts to elucidate the role these features play on the selection process are vital to enhance production in this sector. Selection must focus on traits that enhance larval performance and productivity, whilst simultaneously selecting for traits which are relevant in later development. For example, selection for growth and survival during the larval stage can reduce production time and benefits the industry by reducing production costs. However, selection in later stages must also be applied as the fastest growing larvae may not be the fastest growing adults. Additionally, *G*x*E* interactions may also play a role and selection must be therefore focused on specific regions and/or

environmental conditions. Thus, another key priority is to understand the genetic basis of these traits, as well as their genetic and developmental correlations.

Genomic resources have the potential to revolutionise aquaculture, contributing to the rapid expansion and optimisation of marine bivalve production. Nonetheless, socioeconomic factors also play a key role in the implementation of new technologies in existing production systems, and may slow down the pace of genomic breeding in aquaculture, especially in developing countries (Kumar *et al.*, 2018). To date, industrial applications of genomic selection in aquatic species are limited, and largely restricted to finfish species (Zenger *et al.* 2019).

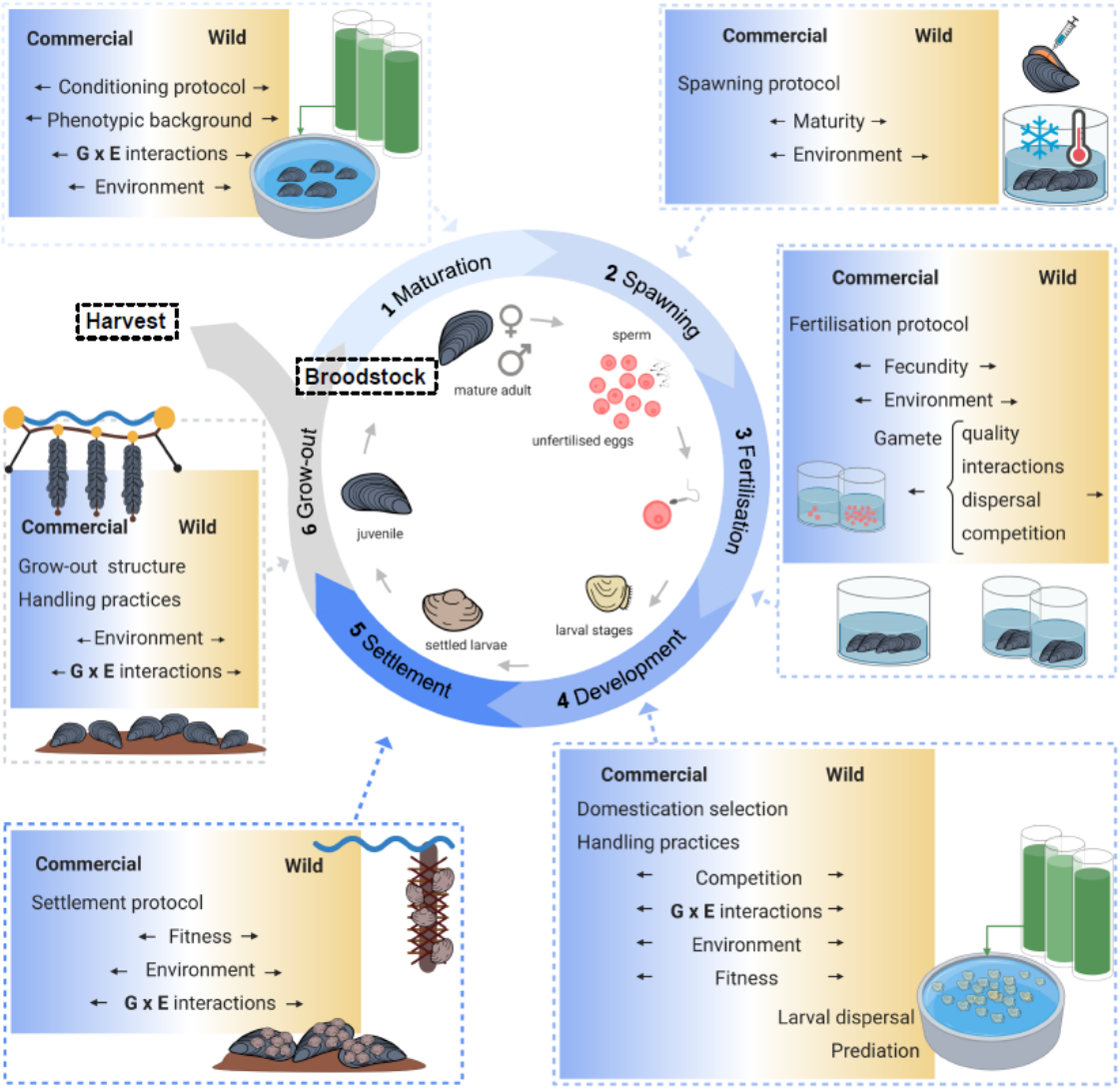


Figure 1-1: Concept diagram highlighting the selection pressures acting upon natural and artificially propagated bivalve stocks, across different life-cycle stages and corresponding production steps. Selection pressures (coloured boxes) acting upon wild populations (indicated with yellow background), hatchery populations (indicated with blue background) or both (indicated with flanking arrows), as generated by the artificial or

natural environment. Selection pressures correspond to each production step discussed within the review (steps 1 - 5; maturation, spawning, fertilization, development, settlement and grow-out; outer circle), as well as the corresponding life-cycle stage (inner circle). Figure created with BioRender.com.

Broodstock conditioning and its implications on genetic variability

Contrary to natural ecosystems, hatcheries offer a largely uniform environment to cultivate broodstock, reducing sources of stress caused by sub-optimal or fluctuating conditions. In these artificial systems, environmental conditions can be manipulated to trigger gametogenesis in broodstock throughout the year, extending the period through which mature breeders are available (Goulletquer 2004). Overall, the process of induced gametogenesis, known as conditioning, aims to maximise the fecundity of progenitors, synchronising spawning, whilst maintaining the high quality of gametes and larval viability (Lannan *et al.*, 1980; Utting & Millican, 1997). For aquaculture purposes, broodstock are either collected in their natural environment or taken from previous generations of hatchery stock and are held in flow-through systems (Goulletquer 2004). During the conditioning process, quality and availability of food resources have a direct effect on adult fecundity levels and reproductive output (see Utting and Millican, 1997 and references therein), with lipid and proteins obtained from food accumulated during oogenesis (Li, *et al.*, 2000). A significant correlation between biochemical content of oocytes and early developmental success (Boulais *et al.*, 2015; Corporeau *et al.*, 2012; Massapina *et al.*, 1999), highlights the vital role that conditioning can play in production, and consequently, in the genetic makeup of cohorts (Figure 1.1).

To date, standard conditioning protocols have been established for the main cultured bivalve species (Helm 2004). However, a large (up to 2-fold) variation in length of conditioning period is reported among strategies adopted by different hatcheries, and a quality check of broodstock gonad development is not consistently undertaken among hatcheries (Reynaga-Franco *et al.*, 2020) .Without equal opportunity for success in breeding, N_e/N ratio is lowered. In addition, an unsynchronized response of broodstock to conditioning may reduce the potential number of breeding pairs, promote discrepancies of both V_k among individuals and performance among families (Boudry *et al.* 2002), with inbreeding levels within a breeding program consequently increasing. Such issues rapidly nullify

predictive ability of selective breeding methods and impose a challenge for the implementation of genomic selection in these taxa.

Genomic and phenotypic consequences of hatchery propagation

In hatcheries, spawning of broodstock can be triggered either by non-lethal techniques (thermal cycling, intermittent exposure to air and/or introduction of potassium chloride, hydrogen peroxide, steroids or neurotransmitters in the mantle cavity or adductor muscle) or by stripping (scarifying) the gonads of individuals (Helm *et al.*, 2004). The adoption of gonad-stripping or chemically induced spawning protocols can help to standardize the time of gamete release, reducing the deterioration of gametes. However, such approaches do not discriminate between mature and immature gametes present in the gonad. The lack of control of gamete quality during artificial spawning may lead to a high variability in developmental rate within a batch (Tanyaros & Tarangkoon 2016). In fact, for some species such as *M. edulis*, gonad stripping is a non-viable approach which impairs production (Kamermans *et al.* 2013). Moreover, the required sacrifice of pedigreed broodstock individuals (where identified) may render this approach unfavourable for selective breeding.

Owing to its practicality, mass spawning (combining gametes from multiple females with an aliquot of pooled male gametes) is a common procedure for artificial fertilisation (Helm 2004; SPAT 2012). This approach does not control parental contribution and can result in reduced numbers of effective parents in the program. Moreover, as best performing individuals may be excluded from crosses, mass selection can limit the accuracy of the breeding program. However, fecundity levels observed in bivalves are high and fertilisation is commonly successful, and a sufficient number of offspring is often achieved. Inbreeding load, as well as impaired development, can be concealed by management practices (e.g. culling) where the low performing individuals are eliminated from a batch by size-selection (Taris *et al.* 2006). As genomic and marker assisted selection endeavour to capture

favourable genetic variation, it is vital to identify and control for the possible impacts of spawning and fertilisation protocols on the genetic variability and performance of cohorts (Figure 1.2).

To overcome issues with parental contribution, pairs can be individually crossed. Paired crossing is less commonly adopted in hatcheries as it is a more laborious approach, requiring the control of fertilisation rates of individual crosses and investment in personnel and equipment. This method demands additional physical space to separate mating pairs and subsequent offspring during larval development. Furthermore, the rearing of juveniles in family-specific tanks presents an issue with confounding of genetic and common environmental effects, which would require multiple replicate tanks per family to resolve. Subsequently mixing families and growing them together in a common environment can mitigate against this issue. However, the gamete density used in artificial crosses is substantially higher than in nature. Empirical evidence demonstrates that mass spawning increases V_k among males and pair crossing individuals increases the variance in reproductive success among females (Hornick & Plough 2019). Handling practices may additionally contribute to increase variation in family sizes, which often goes undetected. Long term, such practices can bottleneck genetic variability in artificially propagated stock and dramatically reduces the ability to predict success for selective breeding. Therefore, altered genetic diversity of hatchery-propagated stock is an inevitable consequence of the chosen fertilisation approach (Figure 1.1) (Hornick & Plough 2019).

Phenomena occurring at the gamete level may also play a role in determining parental contribution in crosses and act as an early selective pressure. For example, the distance which sperm must travel to reach oocytes, gamete phenotype (biochemical composition, sperm motility and behaviour, oocyte size and age) and gamete interactions are factors that influence the success of fertilisation (Boulais, *et al.*, 2015,2017; Levitan 2006; Suquet *et al.*, 2010). Genetic compatibility can influence fertilisation success, favouring crosses between less related individuals (Lymbery *et al.*, 2017). In sea urchins, low sperm densities favoured crosses between common genotypes which match at the gamete binding locus (oocyte-sperm compatibility locus) (Levitan and Ferrell, 2006). Sperm-saturation, in turn, promoted reproductive success of individuals with less frequent genotypes. These

findings highlight the putative role of gamete density in sperm choice behaviour. Factors such as affinity between crosses (Kekäläinen & Evans 2017; Lymbery et al. 2017) and sperm longevity (Crean *et al*., 2012) have been linked to increased postzygotic fitness (Oliver & Evans 2014). However, the extent to which interactions at the gamete-level, as well as gamete phenotype, influences fertilisation success in external fertilisation is not yet fully understood (Breed & Moore 2016).

The precise determination of oocyte-sperm ratios and controlled-crossing approaches may benefit fertilisation success and enhance the contribution of individual broodstock (Song *et al*., 2009). Gamete density used in artificial crosses is substantially higher than in nature, increasing competition among individuals, and acts on N_b. Oocyte mechanisms acting against polyspermy are not 100% effective, thus, increased competition can lower the rates of fertilisation success among crosses. Commonly, substantial variation in gamete phenotype and fertilisation rates are observed among and within individuals (Breed & Moore 2016). During hatchery propagation, gamete quality is assessed via crude visual observations of sperm motility and concentration, as well as shape (roundness), size and colouration of oocytes. Individuals classified with high quality gametes are selected for fertilisation, whilst those not meeting the quality criteria are excluded from crosses. Correlations between gamete phenotypes (e.g. oocyte biochemical composition) and larval viability in artificially bred bivalves and other invertebrates have been previously described (Boulais *et al.*, 2015; Crean *et al.*, 2012; Massapina *et al.*, 1999). Nonetheless, the extent to which gamete traits and gamete-level interactions influence V_k and genetic variability of offspring is not fully understood. If such implications can be carried over throughout the individuals' life, the expression of key genotypes may be modulated by pre fertilisation selection. However, further investigation is required to clarify how physiological and molecular mechanisms underlie gamete phenotype and affinity of crosses during external fertilisation (Figure 1.2). Such knowledge can benefit the development of mating systems that maximise fertilisation and homogenise V_k among breeders.

Hatchery-propagated larvae are reared in a controlled environment, avoiding the risks imposed by oceanic drift and predation. In this environment, water quality parameters are maintained

at, or close to, conditions considered optimal for the survival of the species being cultured. This optimised environment enables the levels of production to be improved, maximising larval growth and settlement rates of the produced species. In the long term, domestication contributes to enhance performance under these artificial rearing conditions. However, domestication selection can lower environmental resilience when exposed to natural conditions. A lower fitness of individuals in the wild has been observed in fish species which are currently in transition to a domesticated status (Araki *et al.*, 2008). Moreover, genomic footprints of domestication in fish species can vary greatly between populations from independent origins selected for the same trait (López *et al.* 2019), as a result of the specific characteristics of a rearing environment (Vandeputte *et al.* 2009). Recent findings indicate that selectively bred *C. virginica* larvae were less able to tolerate starvation compared to wild cohorts, experiencing significantly higher mortality rates (McFarland *et al.*, 2020). However, the genomic mechanisms underlying domestication selection of marine bivalve species remains poorly investigated in comparison to finfish species. Optimisation of selective breeding of these taxa will require these factors to be better understood and controlled for.

Culling, or size selection, is commonly practiced in hatcheries throughout larval development. Selection for similar growth rates under culture conditions generally improves overall spat production and reduces variation in development within a cohort (Taris *et al.* 2006), but may potentially mask the signs of inbreeding depression (Taris *et al.*, 2007). Therefore, such a practice may benefit early production stages. However, the effect of size selection on a stocks' genetic variation are not clear (Figure 1.2). Culling may result in accidental removal of individuals that may reach market size quickest in later development, individuals with alternative traits of interest (e.g. disease resistance) or traits that are relevant during later stages of production (e.g. robustness), directly impacting a breeding scheme. A practical example is seen in *Mercenaria mercenaria* larvae, where initially small individuals present in culture tanks are be capable of surpassing the size of individuals that were initially larger, at later stages of development (Gionet *et al.*, 2010). In addition, this process can reduce genetic variability of offspring (Taris *et al.* 2006) acting as a genetic bottleneck in hatcheries. Losses of entire

cohorts could result from sudden shifts in conditions when *G*x*E* interactions when animals selected as optimal under a hatchery production environment perform poorly in a subsequent grow-out environment.

Developmental plasticity (input during early development persisting in adult phenotype) can modify the performance of individuals in their later life. For example, exposure of quagga mussel larvae (*Dreissena bugensis)* to a range of temperatures has been correlated with the development of different shell morphotypes in adults (Peyer *et al.*, 2010). If early exposure to stressors can imprint performance of organisms in later life, alternative culling strategies (e.g. application of a salinity or temperature shock during early development) would enable selection for robustness to future environmental conditions. Shellfish restoration programmes relying on hatchery seed supply could benefit from such strategy, as produced individuals would be more likely to withstand fluctuating environments and/or exacerbated conditions resulting from climate change. Accordingly, the hatchery environment and management practices may themselves help or hinder spat development, potentially affecting the performance of individuals at grow-out sites (Reynaga-Franco *et al.*, 2019).

Currently, research into the implications of hatchery practices on the genetic characteristics of bivalves is restricted to a few studies (Boudry *et al*., 2002; Hornick & Plough, 2019; Lallias, Boudry, Lapègue *et al.*, 2010; McFarland *et al.*, 2020; Taris *et al.*, 2007, 2006). Unravelling the genomic basis of environmental resilience will allow the potential of selection towards robustness, or generalist phenotypes, and its association with other commercially relevant QTL to be determined (Vu *et al.* 2021). Additionally, the development of physiological indices of larval performance, and their association with the individual genotype, can contribute to improve selection in these taxa (Pan *et al.*, 2016). Further studies clarifying the correlation between performance of hatchery bred larvae and individual performance during grow-out will contribute to the optimisation of production throughout the entire life-cycle of these taxa. Furthermore, such information can contribute to the development of breeding strategies in marine bivalves.

Hatchery bred spat, which have reached the settlement stage, are often induced to settle. This practice not only facilitates efficient husbandry but avoids any adverse consequences (e.g. depleted energy reserves) of spending too long in the pediveliger stage. Uniformity in settlement time can be achieved by manipulating environmental stimuli such as temperature shocks, or via the addition of fine shell particles or other material to induce settlement in tanks (Helm 2004). Alternatively, settlement of larvae can be chemically induced by exposure to neurotransmitters (Grant *et al.*, 2013; Joyce & Vogeler, 2018; Sánchez-Lazo & Martínez-Pita, 2012). Further investigation is needed to elucidate the role such approaches play as a selective pressure in the hatchery environment and whether these can be used to select or induce favourable characteristics (Figure 1.2).

Settlement and metamorphosis are critical moments in the life cycle of bivalves. Substantial mortalities occur during these stages in both natural populations and artificially propagated stock (Hunt & Scheibling 1997; Plough 2016; Plough & Hedgecock 2011), with survival at the post-settlement stage reaching only 2.8 % of the original population in some cases (Plough 2016). Genotype-dependent mortality linked to deleterious recessive mutations can occur immediately before or during metamorphosis (Plough 2016; Plough & Hedgecock 2011). Insights on genotype-dependent mortality during settlement have opened the opportunity to investigating the applicability of QTLs to select for uniformity of settlement timing (Plough 2016). Settlement rates lower than 5% have been previously reported in other studies including those on *Mytilus* spp (Knöbel et al. 2021). Nonetheless, breeding facilities provide adequate substrate for settlement asides using different chemical substances to induce larval settlement (Silveira et al. 2011; Tanyaros 2011). In oyster hatcheries settlement success normally ranges between 20 % to over 50%, depending on the density and employed settlement approach (Helm 2004; Silveira et al. 2011; Tanyaros & Chuseingjaw 2016). In contrast, mortality in the period immediately post-settlement is lower, with no indication of being genotype-dependent (Dégremont *et al*., 2007; Plough, 2016).

Genotype by Environment responses to the grow-out environment

All spat, both wild and hatchery propagated, are exposed to environmental variability experienced within the coastal and estuarine zones in which grow-out occurs, and are thus susceptible to this daily and seasonal variability (Figure 1.1). To thrive in such demanding environments, individuals must either be genetically adapted to extreme conditions, or possess highly plastic physiological responses which allow them to regulate internal mechanisms.

Accordingly, Pacific oysters have demonstrated the ability to regulate genes involved in stress response pathways when facing abiotic stress conditions, including elevated temperature and air exposure (Zhang *et al.* 2012). These findings suggest that a high level of plasticity is a strategy which has allowed these sessile organisms to successfully colonise stressful environments. The expansion of gene families that function as part of the organism's response against biotic and abiotic stress, as well as immune response, suggest that this group has adapted to a sessile life in fluctuating environments (i.e. intertidal coastal and estuarine waters) (Wang, *et al.*, 2012). A better understanding of plasticity mechanisms in bivalves can contribute to the development of culturing conditions which improve performance in desirable traits. The selection processes experienced during early development in hatcheries could contribute to direct effects on the performance of cohorts, as well as increasing the likelihood of stochastic *G*x*E* interactions.

*G*x*E* is an important factor dictating performance of aquaculture species (see review by Sae-Lim *et al.*, 2016 and references therein). Where animals from a breeding programme are reared in different environments, it can result in a re-ranking of families or genotypes. This can negatively impact genetic gain and the effectiveness of a breeding programme. These effects have been observed in previous studies investigating the variation in *C. gigas* performance among families across grow-out sites (Evans and Langdon, 2006b; Langdon *et al.*, 2003). In other cases, selected genotypes outperform in certain conditions, but become poor performers when exposed to a different set of conditions – i.e. re-ranking of genotypes (Dégremont *et al.*, 2005; Evans and Langdon, 2006b; Langdon *et al.*, 2003; Wang *et al.*, 2013). In marine bivalves, between-family variance described for traits such as growth,

survival and environmental resilience (Dégremont *et al*., 2005; Dégremont *et al*., 2015; Scanes *et al*., 2020) indicates the genetic basis of traits associated with performance (Vu *et al.* 2021). The re-ranking of genotypes, in turn, highlights the intrinsic effect of *G*x*E* on overall performance of a family or cohort. There is also genetic variation in how well animals perform across diverse environmental conditions, and this robustness of genotypes to diverse conditions can be analysed using reaction norms, and potentially incorporated into breeding goals to help tackle the impact of *G*x*E* (Hill & Mulder 2010).

Epigenetic mechanisms (e.g. DNA methylation, histone modifications, non-coding RNAs) are a relevant component of *G*x*E* interactions, through exposure mediated *G*x*E*. These mechanisms can modify a phenotype without changing the DNA sequence and can have long-lasting effects (Jablonka & Lamb 2002). In the last decades, new technologies have facilitated the study of epigenetics, providing insights into the contribution of the epigenome to the expressed phenotypes in response to the environment. Among the wide scale of techniques available to study epigenetic regulation, DNA methylation has received the most attention in marine bivalves. In *C. gigas*, DNA methylation patterns have been associated with gene function (Gavery & Roberts 2010) and have been linked to gene regulation (Olson & Roberts 2014; Riviere *et al.*, 2013). Environmental heterogeneity has been associated with divergent DNA methylation patterns among *C. virginica* populations (Johnson & Kelly 2020). In *Mytilus galloprovincialis* and the New Zealand pygmy mussel *Xenostrobus secures*, methylation patterns of invasive populations differ from populations in their native range (Ardura *et al.*, 2018). Whilst epigenetics can contribute to rapid and transient plasticity in response to stress and environment in marine bivalves, future studies combining genomic and epigenomic information are needed to elucidate the processes underlying *G*x*E* interactions and phenotype expression in these taxa.

Recent evidence also underlines the adaptive nature of phenotypic plasticity in traits involved in environmental resilience (Li *et al.*, 2018). Domestication, or the reduction of environmental variation in early life stages, is unlikely to select for plasticity and may lead to epigenetic profiles that are less suited to the farm environment. However, the impact of artificial-breeding and early life

hatchery condition on the epigenome of marine bivalve species remains unresolved. Selective pressures acting during hatchery-propagation most likely favour domestication rather than adaptation towards variable natural environments. Therefore, the potential of hatchery-propagated stock to cope with environmental stress may be reduced during breeding and hatchery processes. Indeed, the epigenome of artificially bred Atlantic salmon differs greatly from wild populations, and the reduced fitness of hatchery-propagated stock in comparison to wild populations is likely a consequence of such variation (Le Luyer *et al.* 2017). Here, we emphasize that domestication selection in early life stages could lower the mean performance of cohorts through *G*x*E* interactions, once spat is deployed in the grow-out systems. Signs of lower tolerance to environmental stress in *C. virignica* have been linked to domestication selection (McFarland *et al.* 2020). However, further studies are required to disentangle the implication of domestication selection, if any, on the performance of hatchery produced stock.

To understand the role of *G*x*E* interactions in the expression of phenotypes, the performance of different lines needs to be tested in a range of environments. Strong *G*x*E* interactions could be countered by the creation of specific breeding programs targeting specific grow-out environments (Dégremont *et al.* 2007). However, it would first be wise to understand if any of the potential hatchery stressors or selection events described herein are contributors to *G*x*E* events, and if they can be mitigated through alteration in early life selection. As a crude hypothesis, growth in *C. gigas*, a trait which has been under selection pressure in the hatchery, seems to be highly dependent on the environment, whereas other traits such as survival in the presence of disease seem dependent on the family (Dégremont *et al.*, 2005; Evans and Langdon, 2006b). For the latter, expression of phenotypes can be maintained within families across a range of environments by epigenetic mechanisms (Gavery & Roberts 2017; Uren Webster *et al.* 2018).

Other omic techniques, such as proteomics and metabolomics, provide a direct measurement of expressed phenotypes and are therefore valuable tools to explore the genotype-phenotype link and evaluate performance (Laudicella *et al.* 2020). Further studies investigating the relation between

specific environmental conditions utilising a holistic omic approach may allow to understand and control for *G*x*E* in bivalve breeding programmes and are critical to improve aquaculture (Figure 1.2).

Selective breeding under a changing climate

Shifts in sea surface salinity, temperature and ocean chemistry (e.g. ocean acidification), alterations in precipitation patterns as well as stronger and more frequent heat waves, are some of the main consequences of climate change to the marine environment predicted for the coming decades (IPCC 2018).

As ectothermic calcifying organisms, marine bivalves are particularly vulnerable to climate change. Shell dissolution and decreased shell growth caused by ocean acidification has been described in marine bivalves (Melzner *et al.* 2011). Higher sea surface temperatures, especially in summer months, may challenge species with lower thermal tolerance (Steeves *et al.*, 2018). Fluctuating sea surface salinities may have deleterious implications for shell growth (Riisgård *et al.*, 2012) whilst the interactions of this factor with increased temperature or hypercapnia (elevated CO_2) can increase mortality (Rybovich *et al.*, 2016) and reduce hardness and resistance of shells (Dickinson *et al.* 2012). Phytoplankton communities are likely to be impacted by climate change (Käse & Geuer 2018), and temporal shifts in species abundance and composition may impact the nutrient uptake in marine bivalves, limiting physiological and biological processes. Climate change may also contribute to lowering the immune response of bivalves (Mackenzie *et al.*, 2014), and modify host-pathogen interactions, increasing sensitivity towards diseases (Asplund *et al.* 2014).

The grow-out phase of bivalve aquaculture takes place in the natural environment. Therefore, the implications of climate change are not restricted to wild populations. Strong changes in local environmental conditions may limit production and force the relocation of grow-out sites to suitable areas. Environmental changes and increased disease outbreaks might lead to severe mortality and considerable economic losses in this industry and restrictions in spat commercialisation may be needed to avoid the further spread of diseases. Thinner and weaker shells will facilitate their rupture

during transportation. Hatchery propagation may also be impaired by climate change to a certain degree, as those rely in natural sea water supply. Therefore, there is an imminent need for research to develop bivalve strains robust to climate change and resilient towards diseases.

Epigenetic processes can contribute to rapid adaptation towards environmental stressors generated by climate change. In *S. glomerata*, short-term exposure to elevated CO_2 concentration not only increases resilience of exposed individuals, but can also be passed through generations (Parker *et al.* 2015). Such resilience has been associated with a change in regulation of genes associated with stress related functions (Goncalves *et al.* 2016). Accordingly. empirical evidence demonstrates that low pH stress (pH 7.4) can modify the methylation patterns of *Crassostrea hongkongenssis* pediveliger larvae (Lim *et al.* 2021). Genomic processes, in turn, are involved in long term adaptation to environmental changes.

Identifying the mechanisms acting behind *G*x*E* interactions which increase performance of a species under climate change-associated stressors is an important step to characterise the genomic and epigenomic profile of robust genotypes. Accordingly, *G*x*E* can be exploited in breeding programmes to increase environmental resilience and the application of genomic selection can fast-track the development of such lines (Mulder 2016). The growing body of high quality assembled genomes facilitate the precise identification of genomic regions linked to traits responsible for environmental resilience. The application of genomic selection, or gene editing approaches, can then facilitate the development of robust lines, or lines able to withstand sub-optimal environmental conditions relevant to a certain grow-out region (e.g. elevated temperature, low pH). As hatcheries allow for the control of genetic stocks, indoor propagation is undoubtedly an essential asset to guarantee the development of lines able to thrive under future predicted environmental scenarios.

Summary and future perspectives

A gap in knowledge remains on how domestication selection and husbandry practices can constrain genetic variability of hatchery-propagated stock during early life stages in marine bivalves

(Figure 1.2). Despite the negative implications of inbreeding load on performance, the control of reproductive output, differential performance of genotypes and genetic variability of stock remains relatively low in bivalve production. Such lack of control may hinder spat performance and consequently aquaculture production. Future advances in bivalve production and selective breeding require an understanding and optimization of hatchery production processes in order to maximise genetic gain.

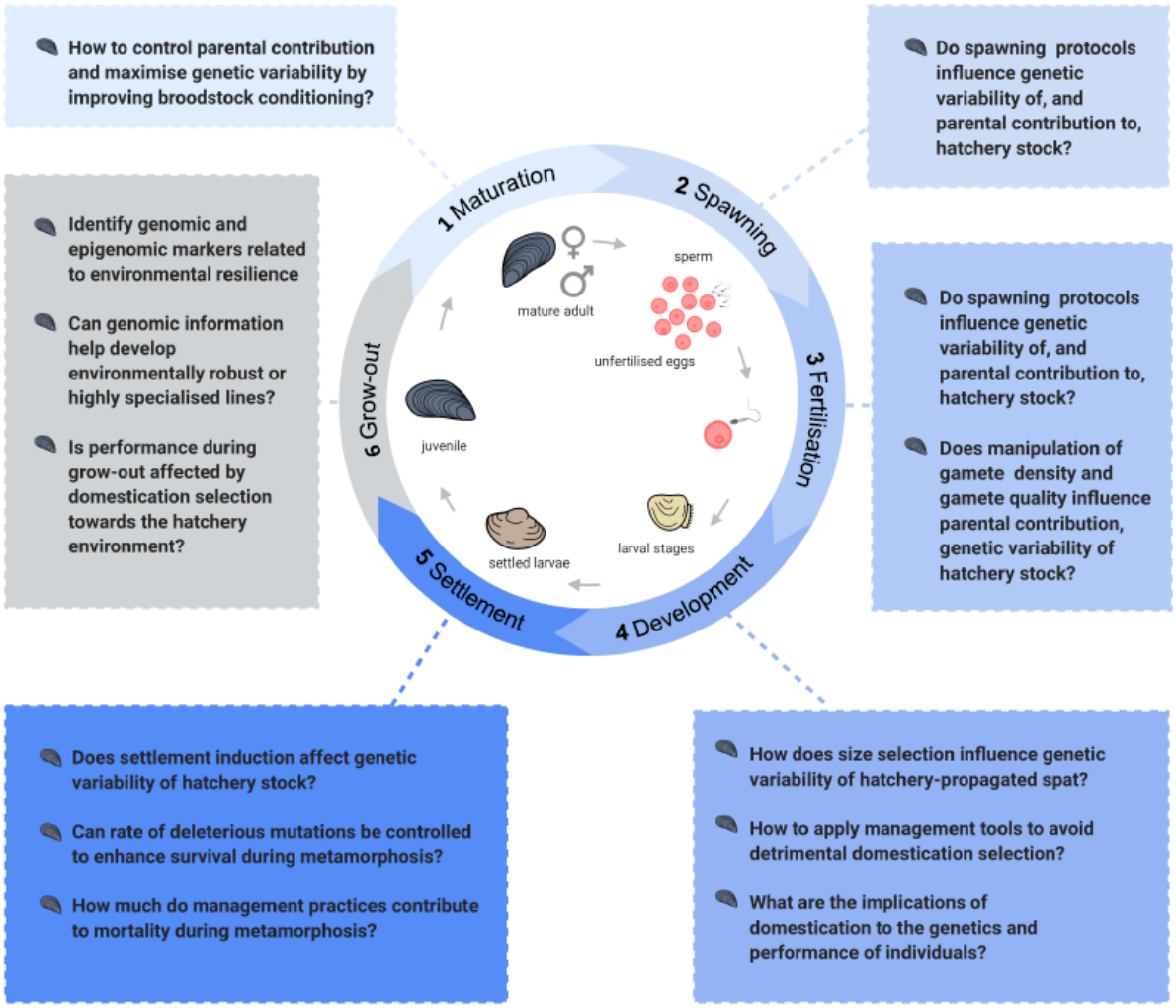


Figure 1-2: Critical knowledge gaps currently preventing the widespread implementation of genomic breeding approaches in bivalve aquaculture. Main knowledge gaps as identified in this review and considered as key priorities for future research (coloured boxes). Knowledge gaps are linked to each production step discussed in this review (steps 1 – 5; maturation, spawning, fertilisation, development, settlement and grow out; outer circle) as well as the corresponding life-cycle stage (inner circle). Figure created with BioRender.com.

Selection pressures acting in hatcheries differ from those acting on the populations in their natural environment. In contrast to wild populations that face fluctuating environmental conditions, farmed stocks are produced under relatively stable, benign, conditions, but exposed to other stresses such as elevated densities and handling practices. Organisms reared in hatcheries, naïve to the wild, might lack resilience to environmental variation due to domestication selection and/or epigenetic mechanisms. In addition, negative *G*x*E* interactions can be detrimental for production and can be compounded by artificial bottlenecking or epigenetic alterations caused by the hatchery environment. However, it is still not fully understood if selection for phenotypes that enhance hatchery production contribute to adult performance during the grow-out phase. Therefore, whilst performance of larval stages must remain as an important component of breeding programmes and hatchery production, it is key to consider traits related to the challenges the spat will face during the grow-out phase.

Genomic resources will contribute with the understanding of evolutionary and adaptive processes, as well as those which are linked to domestication (Yáñez *et al.* 2015). Elucidating the (epi)genomic mechanisms which underpin the expressed phenotypes will allow the divergence of selection from classic commercial traits towards broad environmental resilience (either outperforming or generalist genotypes). Integrating robustness as a founding criterion for selection can potentially contribute to increase grow-out productivity, especially in light of climate change.

Genomic selection can favour the development of genetically-improved lines for multiple traits, facilitate the management of genetic variability (D'Ambrosio *et al.* 2019) and potentially reduce environmental sensitivity accounting for *G*x*E* (Mulder 2016). Most importantly, the implementation of such selection approaches is key to the sustainable optimisation of bivalve aquaculture production, particularly in the light of climate change. It is crucial to focus resources on developing environmentally robust lines. However, progress of marker assisted and genomic selection in bivalve aquaculture will require a greater control of hatchery practices to allow sources of unaccounted genetic variation to be minimised, and genetic gain to be maximised.

The considerations highlighted in this section are particularly relevant to the oyster industry, as hatchery systems for most of the main consumed oyster species are well advanced. For other commonly consumed bivalve species, such as mussels, especially *Mytilus* spp, few hatchery protocols have been developed (Saurel et al. 2022). Examples of successful development and implementation of hatchery systems for Green-lipped mussels (*Perna canaliculus*) and blue mussels (*M. galloprovincialis*) are found in New Zealand and Australia, respectively.

Project Aims and Chapter-specific Objectives

The previous section of this thesis has highlighted the unquestionable need to futureproof marine bivalve mollusc aquaculture towards environmental change, and that such process could be fast-tracked by coupling the implementation of hatcheries to the development of genomic based breeding programs. For species which hatchery technologies are yet to be developed, maturing effective production protocols and adequate systems becomes therefore crucial. Accordingly, all bivalve species can benefit from the development of genomic tools, as those can contribute only to optimise production, but also to the development of conservation programmes. Such resources also facilitate the unravelling of evolutionary processes ongoing in these taxa, contributing to our understanding of speciation and adaptation.

These advances however cannot be reached without a basic understanding on the existing genetic diversity present in these species. In order to breed environmental resilience into key commercial bivalve stocks, understanding how environmental stress can influence performance in major economic traits of these species, and highlighting how these changes are correlated to the genetic background of individuals is critical to further develop this sector. Therefore, the main aim of this thesis is to address the major knowledge gaps which at present hamper the optimisation of the shellfish aquaculture industry, by providing empirical evidence of how genomic and phenotypic information of wild and cultured marine bivalve stock can overcome many of the barriers to successful selective breeding. Each chapter will contribute to this main aim from a different perspective:

Chapter II describes the creation and validation process of a genomic tool, a 60K SNP array for the Blue Mussel species complex *M. edulis, M. galloprovinciallis, M. trossulus* and the southern hemisphere species *M. chilensis*. This multipurpose tool can be applied for genetic makeup characterisation of wild and aquaculture derived marine bivalve populations. Such a tool can contribute to generate consistent genotyping data over a genome-wide scale, and benefit our understanding of ongoing processes such as selection, adaptation, as well as the development of GWAS studies for traits of interest.

Chapter III outlines the application of the Blue Mussel 60K SNP array genotyping platform to assess the genetic diversity of wild marine bivalve populations (*Mytilus* spp) in the South West (SW) coast of England. This hybrid zone, composed by two species: *M. edulis* and *M. galloprovincialis*, is a highly productive area relevant for the mussel aquaculture industry in the UK. The genetic background of populations distributed along the SW hybrid zone has been previously described using fragment length polymorphisms analysis (Hilbish *et al.* 2002). Contrary to using a genome-wide diversity scan approach to analyse diversity, using multiple SNPs, this technology analyses the difference in length at a specific locus of the genome among individuals. Therefore, a second specific aim of this chapter was to compare the genotyping efficiency of the Blue Mussel SNP array to other genotyping technologies available to these taxa. Subsequently, the relationship between the genotype distribution and habitat choice was inferred by comparing low tide vs high tide populations and estuary vs rocky tidal populations. Such analysis can bring valuable insights regarding management of mussel farms, influencing practices such as site choice for seed collection and grow-out culture.

Chapter IV assesses the implications of early exposure to environmental stress at a genomic and phenotypic level. More specifically, this chapter describes an experiment investigating the implications of short-term exposure to elevated seawater temperatures on the genetic makeup of

Baltic juvenile stock, using the 60K SNParray developed for the Blue Mussels species complex. Responses at a phenotypic level were investigated in the same experiment by comparing the performance of pre-exposed (warm selected) vs non-exposed (non-selected) mussel stock followed by 25 days cultured under a range of temperature conditions.

Chapter V Presents the general discussion and main conclusions of the thesis.

References

Adamson E, Syvret M, Woolmer A (2017). Shellfish Seed Supply for Aquaculture in the UK Report on Views Collected from the Industry in 2017. London

Araki H, Berejikian BA, Ford MJ, Blouin MS (2008). SYNTHESIS: Fitness of hatchery-reared salmonids in the wild. *Evolutionary Applications*

Ardura A, Clusa L, Zaiko A, Garcia-Vazquez E, Miralles L (2018). Stress related epigenetic changes may explain opportunistic success in biological invasions in Antipode mussels. *Scientific Reports* **8**: 10793.

Ashraf BH, Byrne S, Fé D, Czaban A, Asp T, Pedersen MG, et al. (2016). Estimating genomic heritabilities at the level of family-pool samples of perennial ryegrass using genotyping-by-sequencing. *Theoretical and Applied Genetics* **129**: 45–52.

Asplund ME, Baden SP, Russ S, Ellis RP, Gong N, Hernroth BE (2014). Ocean acidification and host–pathogen interactions: blue mussels, *Mytilus edulis*, encountering Vibrio tubiashii. *Environmental Microbiology* **16**: 1029–39.

Bierne N, Launey S, Naciri-Graven Y, Bonhomme F (1998). Early effect of inbreeding as revealed by microsatellite analyses on *Ostrea edulis* larvae. *Genetics* **148**: 1893–1906.

Boudry P, Collet B, Cornette F, Hervouet V, Bonhomme F (2002). High variance in reproductive success of the Pacific oyster (*Crassostrea gigas,* Thunberg) revealed by microsatellite-based parentage analysis of multifactorial crosses. *Aquaculture* **204**: 283–96.

Boulais M, Charlotte C, Arnaud H, Ismaël B, Claudie Q, Virgile Q, et al. (2015). Assessment of oocyte and trochophore quality in Pacific oyster, *Crassostrea gigas*. *Aquaculture*, **437:**201-207.

Boulais M, Chenevert KJ, Demey AT, Darrow ES, Robison MR, Roberts JP, et al. (2017). Oyster reproduction is compromised by acidification experienced seasonally in coastal regions. *Scientific Reports* **7**: 1–9.

Breed MD, Moore J (2016). Chapter 11 - Mating Systems. , eds. MD Breed. *Animal Behaviour* pp. 357–93. San Diego: Academic Press.

Camara MD, Symonds JE (2014). Genetic improvement of New Zealand aquaculture species: Programmes, progress and prospects. *Journal of Marine and Freshwater Research*, **48(3)**: 466-491.

Carriker MR (1956). Biology and propagation of young hard clams, *Mercenaria mercenaria*. *Journal of the Elisha Mitchell Scientific Society* **72**: 57–60.

Corporeau C, Vanderplancke G, Boulais M, Suquet M, Quéré C, Boudry P, et al. (2012). Proteomic identification of quality factors for oocytes in the Pacific oyster *Crassostrea gigas*. *Journal of Proteomics* **75**: 5554–63.

Crean AJ, Dwyer JM, Marshall DJ (2012). Fertilization is not a new beginning: the relationship between sperm longevity and offspring performance. *PloS one* **7**: e49167–e49167.

D'Ambrosio J, Phocas F, Haffray P, Bestin A, Brard-Fudulea S, Poncet C, et al. (2019). Genome-wide estimates of genetic diversity, inbreeding and effective size of experimental and commercial rainbow trout lines undergoing selective breeding. *Genetics Selection Evolution* **51**: 1-5.

de Melo C, Durland E, Langdon C (2016). Improvements in desirable traits of the Pacific oyster, *Crassostrea gigas*, as a result of five generations of selection on the West Coast, USA. *Aquaculture* **460**: 105-115.

Dégremont L, Bédier E, Soletchnik P, Ropert M, Huvet A, Moal J, et al. (2005). Relative importance of

family, site, and field placement timing on survival, growth, and yield of hatchery-produced Pacific oyster spat (*Crassostrea gigas*). *Aquaculture* **249**: 213-229.

Dégremont L, Ernande B, Bédier E, Boudry P (2007). Summer mortality of hatchery-produced Pacific oyster spat (*Crassostrea gigas*). I. Estimation of genetic parameters for survival and growth. *Aquaculture,* **262(1)**:41-53.

Dégremont L, Garcia C, Allen SK (2015a). Genetic improvement for disease resistance in oysters: A review. *Journal of Invertebrate Pathology* **131**: 226–41.

Dégremont L, Nourry M, Maurouard E (2015b). Mass selection for survival and resistance to OsHV-1 infection in *Crassostrea gigas* spat in field conditions: response to selection after four generations. *Aquaculture* **446**: 111–21.

Dickinson GH, Ivanina A V, Matoo OB, Pörtner HO, Lannig G, Bock C, et al. (2012). Interactive effects of salinity and elevated CO2 levels on juvenile eastern oysters, *Crassostrea virginica*. *Journal of Experimental Biology* **215**: 29–43.

Dove MC, Nell JA, Mcorrie S, Wayne AO (2013a). Assessment of Qx and Winter Mortality Disease Resistance of Mass Selected Sydney Rock Oysters, *Saccostrea glomerata* (Gould, 1850), in the Hawkesbury River and Merimbula Lake, NSW Australia. *Journal of Shellfish Research* **32**: 681.

Dove MC, Nell JA, O'Connor WA (2013b). Evaluation of the progeny of the fourth-generation Sydney rock oyster *Saccostrea glomerata* (Gould, 1850) breeding lines for resistance to QX disease (*Marteilia sydneyi*) and winter mortality (*Bonamia roughleyi*). *Aquaculture Research* **44**: 1791–1800.

Emlet RB, Hoegh-Guldberg O (1997). Effects of Egg Size On Postlarval Performance: Experimental Evidence From a Sea Urchin. *Evolution* **51**: 141–52.

Endelman JB, Carley CAS, Bethke PC, Coombs JJ, Clough ME, da Silva WL, et al. (2018). Genetic Variance Partitioning and Genome-Wide Prediction with Allele Dosage Information in Autotetraploid Potato. *Genetics* **209**: 77–87.

Eurostat (2020). *Fisheries statistics* https://ec.europa.eu/eurostat/web/fisheries/data/database

Evans S, Langdon C (2006). Effects of genotype × environment interactions on the selection of broadly adapted Pacific oysters (*Crassostrea gigas*). *Aquaculture* **261(2)**: 522-534.

FAO (2016). The state of world fisheries and aquaculture 2016. Contributing to food security and nutrition for all.

FAO (2019a). FAO yearbook. Fishery and Aquaculture Statistics 2017/FAO annuaire. Statistiques des pêches et de l'aquaculture 2017/ FAO anuario. Estadísticas de pesca y acuicultura 2017. Rome

FAO (2019b). The State of the World's Aquatic Genetic Resources for Food and Agriculture. Commission on Genetic Resources for Food and Agriculture. 2019.

Fleury E, Huvet A, Lelong C, de Lorgeril J, Boulo V, Gueguen Y, et al. (2009). Generation and analysis of a 29,745 unique Expressed Sequence Tags from the Pacific oyster (*Crassostrea gigas*) assembled into a publicly accessible database: the GigasDatabase. *BMC genomics* **10**: 1-15.

Gallager SM, Mann R (1986). Growth and survival of larvae of Mercenaria mercenaria (L.) and *Crassostrea virginica* (Gmelin) relative to broodstock conditioning and lipid content of eggs. *Aquaculture* **56**: 105–21.

Galstoff PS (1938). Physiology of Reproduction of *Ostrea virginica*. II . Stimulation of Spawning in the Female Oyster. **75**: 286–307.

Gavery MR, Roberts SB (2010). DNA methylation patterns provide insight into epigenetic regulation in the Pacific oyster (*Crassostrea gigas*). *BMC Genomics* **11**: 1-9.

Gavery MR, Roberts SB (2017). Epigenetic considerations in aquaculture. *PeerJ* **5**: e4147–e4147.

Gerdol M, Moreira R, Cruz F, Gómez-Garrido J, Vlasova A, Rosani U, et al. (2019). Massive gene presence/absence variation in the mussel genome as an adaptive strategy: first evidence of a pan-genome in Metazoa. *bioRxiv,*781377.

Gionet C, Mayrand E, Landry T (2010). Elimination of animals with best growth potential as a possible effect of the culling of *Mercenaria mercenaria notata* (L.) larvae in hatchery procedure. *Aquaculture International,* **18**: 801–12.

Gjedrem T, Robinson N, Rye M (2012). The importance of selective breeding in aquaculture to meet future demands for animal protein: A review. *Aquaculture* **350**–**353**: 117–29.

Gjedrem T, Rye M (2018). Selection response in fish and shellfish: a review. *Reviews in Aquaculture*, **10(1)**: 168-179.

Goncalves P, Anderson K, Thompson EL, Melwani A, Parker LM, Ross PM, et al. (2016). Rapid transcriptional acclimation following transgenerational exposure of oysters to ocean acidification. *Molecular ecology* **25**: 4836–49.

Goulletquer P (2004). Cultured Aquatic Species Information Programme. *Crassostrea gigas*.

Grant MN, Meritt DW, Kimmel DG (2013). Chemical induction of settlement behavior in larvae of the eastern oyster *Crassostrea virginica* (Gmelin). *Aquaculture* **402**–**403**: 84–91.

Guo X, Li Q, Wang QZ, Kong LF (2012). Genetic Mapping and QTL Analysis of Growth-Related Traits in the Pacific Oyster. *Marine Biotechnology,* **14:** 218-226.

Gutierrez AP, Hooper C, Bean T, Stenton CA, Sanders MB, Paley RK, et al. (2018). A Genome-Wide Association Study for Host Resistance to Ostreid Herpesvirus in Pacific Oysters (*Crassostrea gigas*). *G3: Genes|Genomes|Genetics* **8**: 1273 LP – 1280.

Hedgecock D (1994). Does variance in reproductive success limit effective population sizes of marine organisms? *Genetics and Evolution of Aquatic Organisms,* **122:** 122-134.

Hedgecock D, Pudovkin AI (2011). Sweepstakes Reproductive Success in Highly Fecund Marine Fish and Shellfish : A Review and Commentary. *Bulletin of Marine Science,* **87**: 971–1002.

Hedgecock D, Sly F (1990). Genetic drift and effective population sizes of hatchery-propagated stocks of the Pacific oyster, *Crassostrea gigas*. *Aquaculture,* **88(1)**: 21-38.

Helm MM (2004). *Hatchery Culture of Bivalves: A Practical Manual*, Vol. 639.2 F3 v Rome: Food and agriculture organization of the united nations.

Hershberger WK, Perdue JA, Beattie JH (1984). Genetic selection and systematic breeding in Pacific oyster culture. *Aquaculture* **39**: 237–45.

Hilbish T, Carson E, Plante J, Weaver L, Gilg M (2002). Distribution of *Mytilus edulis, M. galloprovincialis*, and their hybrids in open-coast populations of mussels in southwestern England. *Marine Biology* **140**: 137–42.

Hill WG, Mulder HANA (2010). Genetic analysis of environmental variation. *Genetics Research* **92**: 381–95.

Hollenbeck CM, Johnston IA (2018). Genomic tools and selective breeding in molluscs. *Frontiers in Genetics* **9**: 1–15.

Hornick KM, Plough L V. (2019). Tracking genetic diversity in a large-scale oyster restoration program: effects of hatchery propagation and initial characterization of diversity on restored vs. wild reefs. *Heredity,* **123(2)**: 92-105.

Houston RD, Bean TP, Macqueen DJ, Gundappa MK, Jin YH, Jenkins TL, et al. (2020). Harnessing

genomics to fast-track genetic improvement in aquaculture. *Nature Reviews Genetics,* **21(7)**: 389-409.

Hunt HL, Scheibling RE (1997). Role of early post-settlement mortality in recruitment of benthic marine invertebrates. *Marine Ecology Progress Series* **155**: 269-301.

IPCC (2018). IPCC Special Report 1.5 - Summary for Policymakers. In *Global Warming of 1.5°C. An IPCC Special Report on the Impacts of Global Warming of 1.5°C above Pre-Industrial Levels and Related Global Greenhouse Gas Emission Pathways, in the Context of Strengthening the Global Response to the Threat of Climate Change, sustainable development, and efforts to eradicate poverty. Sustainable Development, and Efforts to Eradicate Poverty.* Geneva, Switzerland: World Meteorological Organization, 32.

Jablonka E, Lamb MJ (2002). The changing concept of epigenetics. *Annals of the New York Academy of Sciences* **981**: 82–96.

Jiao W, Fu X, Dou J, Li H, Su H, Mao J, et al. (2014). High-resolution linkage and quantitative trait locus mapping aided by genome survey sequencing: building up an integrative genomic framework for a bivalve mollusc. *DNA research : an international journal for rapid publication of reports on genes and genomes* **21**: 85–101.

Johnson KM, Kelly MW (2020). Population epigenetic divergence exceeds genetic divergence in the Eastern oyster *Crassostrea virginica* in the Northern Gulf of Mexico. *Evolutionary Applications* **13**: 945–59.

Joyce A, Vogeler S (2018). Molluscan bivalve settlement and metamorphosis : Neuroendocrine inducers and morphogenetic responses. *Aquatculture,* **487**: 64–82.

Kamermans P, Galley T, Boudry P, Fuentes J, McCombie H, Batista FM, et al. (2013). *Blue Mussel Hatchery Technology in Europe*, Vol. 2009 Woodhead Publishing Limited.

Käse L, Geuer JK (2018). Phytoplankton Responses to Marine Climate Change – An Introduction BT - YOUMARES 8 – Oceans Across Boundaries: Learning from each other. pp. 55–71. Cham: Springer International Publishing.

Kekäläinen J, Evans JP (2017). Female-induced remote regulation of sperm physiology may provide opportunities for gamete-level mate choice. *Evolution,* **71**: 238–48.

Knöbel L, Nascimento-Schulze JC, Sanders T, Zeus D, Hiebenthal C, Barboza FR, et al. (2021). Salinity Driven Selection and Local Adaptation in Baltic Sea Mytilid Mussels. *Frontiers in Marine Science*, **8:** 692078.

Kobayashi M, Msangi S, Batka M, Vannuccini S, Dey MM, Anderson JL (2015). Fish to 2030: The Role and Opportunity for Aquaculture. *Aquaculture Economics and Management,* **19:**282-300.

Kube P, Cunningham M, Dominik S, Parkinson S, Henshall J, Finn B, et al. (2011). Enhancement of the Pacific Oyster Selective Breeding Program. *Hobart: FRDC and Seafood CRC. 177p.(Project no. 2006/227)*

Kumar G, Engle C, Tucker C (2018). Factors Driving Aquaculture Technology Adoption. *Journal of The World Aquaculture Society* **49**: 447–76.

Lallias D, Boudry P, Lapègue S, King JW, Beaumont AR (2010). Strategies for the retention of high genetic variability in European flat oyster (*Ostrea edulis*) restoration programmes. *Conservation Genetics,* **11:** 1899-1910.

Langdon C, Evans F, Jacobson D, Blouin M (2003). Yields of cultured Pacific oysters Crassostrea gigas Thunberg improved after one generation of selection. *Aquaculture,* **220(1-4):** 227-244.

Lannan JE, Robinson A, Breese WP (1980). Broodstock management of *Crassostrea gigas*. II.

Broodstock conditioning to maximize larval survival. *Aquaculture,* **21(4):** 337-345.

Laudicella VA, Whitfield PD, Carboni S, Doherty MK, Hughes AD (2020). Application of lipidomics in bivalve aquaculture, a review. *Reviews in Aquaculture* **12**: 678–702.

Launey S, Hedgecock D (2001). High Genetic Load in the Pacific Oyster *Crassostrea gigas*. *Genetics* **159**: 255 LP – 265.

Le Luyer J, Laporte M, Beacham TD, Kaukinen KH, Withler RE, Leong JS, et al. (2017). Parallel epigenetic modifications induced by hatchery rearing in a Pacific salmon. *Proceedings of the National Academy of Sciences* **114**: 12964 LP – 12969.

Levitan DR (1996). Predicting Optimal and Unique Egg Sizes in Free-Spawning Marine Invertebrates. *The American Naturalist* **148**: 174–88.

Levitan DR (2006). The Relationship between Egg Size and Fertilization Success in Broadcast-Spawning Marine Invertebrates. *Integrative and Comparative Biology* **46**: 298–311.

Levitan DR, Ferrell DL (2006). Selection on Gamete Recognition Proteins Depends on Sex, Density, and Genotype Frequency. *Science* **312**: 267 LP – 269.

Li G, Hubert S, Bucklin K, Ribes V, Hedgecock D (2003). Characterization of 79 microsatellite DNA markers in the Pacific oyster *Crassostrea gigas*. *Molecular Ecology Notes* **3**: 228–32.

Li Q, Osada M, Mori K (2000). Seasonal biochemical variations in Pacific oyster gonadal tissue during sexual maturation. *Fisheries Science,* **66(3):** 502-508.

Li Q, Wang Q, Liu S, Kong L (2011). Selection response and realized heritability for growth in three stocks of the Pacific oyster *Crassostrea gigas*. *Fisheries Science* **77**: 643–48.

Lim Y-K, Cheung K, Dang X, Roberts SB, Wang X, Thiyagarajan V (2021). DNA methylation changes in response to ocean acidification at the time of larval metamorphosis in the edible oyster, *Crassostrea hongkongensis*. *Marine Environmental Research* **163**: 105217.

Loosanoff VL, Davis HC (1952). Temperature Requirements for Maturation of Gonads of Northern Oysters. *Biological Bulletin,* **103(1)**: 80-96.

Loosanoff VL, Davis HC (1963). Rearing of Bivalve Mollusks. In Vol. 1, ed. FSBT-A in MB Russell, pp. 1–136. Academic Press.

López ME, Benestan L, Moore J-S, Perrier C, Gilbey J, Di Genova A, et al. (2019). Comparing genomic signatures of domestication in two Atlantic salmon (*Salmo salar* L.) populations with different geographical origins. *Evolutionary Applications* **12**: 137–56.

Lymbery RA, Kennington WJ, Evans JP (2017). Egg chemoattractants moderate intraspecific sperm competition. *Evolution Letters* **1**: 317–27.

Mackenzie CL, Lynch SA, Culloty SC, Malham SK (2014). Future Oceanic Warming and Acidification Alter Immune Response and Disease Status in a Commercial Shellfish Species, *Mytilus edulis* L. *PLOS ONE* **9**: e99712.

Mann R (1983). Bivalve mollusc hatcheries: A critical appraisal of their development and a review of their potential value in enhancing the fisheries of developing nations. *Mems Asoc. Latinoam. Acuicult.* **5**: 97–105.

Mao Y, Lin F, Fang J, Fang J, Li J, Du M (2019). Bivalve Production in China. In *Goods and Services of Marine Bivalves*, eds. AC Smaal, JG Ferreira, J Grant, JK Petersen, Ø Strand, pp. 51–72. Cham: Springer International Publishing.

Massapina C, Joaquim S, Matias D, Devauchelle N (1999). Oocyte and embryo quality in *Crassostrea gigas* (Portuguese strain) during a spawning period in Algarve, South Portugal. *Aquatic Living*

Resource, **12(5):** 327-333.

McFarland K, Plough L V, Nguyen M, Hare MP (2020). Are bivalves susceptible to domestication selection? Using starvation tolerance to test for potential trait changes in eastern oyster larvae. *bioRxiv* 2020.02.26.966176.

Melzner F, Stange P, Trübenbach K, Thomsen J, Casties I, Panknin U, et al. (2011). Food Supply and Seawater pCO2 Impact Calcification and Internal Shell Dissolution in the Blue Mussel *Mytilus edulis*. *PLOS ONE* **6**: e24223.

Meng J, Song K, Li C, Liu S, Shi R, Li B, et al. (2019). Genome-wide association analysis of nutrient traits in the oyster *Crassostrea gigas*: genetic effect and interaction network. *BMC Genomics* **20**: 625.

Meuwissen T, Hayes B, Goddard M (2016). Genomic selection: A paradigm shift in animal breeding. *Animal Frontiers* **6**: 6–14.

Meuwissen THE, Hayes BJ, Goddard ME (2001). Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics* **157**: 1819 LP – 1829.

Mulder HA (2016). Genomic selection improves response to selection in resilience by exploiting genotype by environment interactions. *Frontiers in genetics* **7**: 178.

Naciri-Graven Y, Martin A-G, Baud J-P, Renault T, Gérard A (1998). Selecting the flat oyster *Ostrea edulis* (L.) for survival when infected with the parasite Bonamia ostreae. *Journal of Experimental Marine Biology and Ecology* **224**: 91–107.

Nell JA, Sheridan AK, Smith IR (1996). Progress in a Sydney rock oyster, *Saccostrea commercialis* (Iredale and Roughley), breeding program. *Aquaculture* **144**: 295–302.

Nell JA, Smith IR, Sheridan AK (1999). Third generation evaluation of Sydney rock oyster *Saccostrea commercialis* (Iredale and Roughley) breeding lines. *Aquaculture* **170**: 195–203.

Nelson TC (Hristian) (1938). The feeding mechanism of the oyster. I. On the pallium and the branchial chambers of *Ostrea virginica, O. Edulis and O. Angulata*, with comparisons with other species of the genus. *Journal of Morphology* **63**: 1–61.

Nguyen TTT, Hayes BJ, Ingram BA (2014). Genetic parameters and response to selection in blue mussel (*Mytilus galloprovincialis*) using a SNP-based pedigree Potential broodstock candidates. *Aquaculture* **420**–**421**: 295–301.

Nguyen TTT, Ingram BA (2012). Progress on selective breeding program for blue mussel in Victoria. Deakin University.

Oliver M, Evans JP (2014). Chemically moderated gamete preferences predict offspring fitness in a broadcast spawning invertebrate. *Proceedings. Biological sciences* **281**: 20140148.

Olson CE, Roberts SB (2014). Genome-wide profiling of DNA methylation and gene expression in *Crassostrea gigas* male gametes. *Frontiers in physiology*, **5:** 224.

Pan TCF, Applebaum SL, Lentz BA, Manahan DT (2016). Predicting phenotypic variation in growth and metabolism of marine invertebrate larvae. *Journal of Experimental Marine Biology and Ecology* **483**: 64–73.

Parker LM, O'Connor WA, Raftos DA, Pörtner HO, Ross PM (2015). Persistence of positive carryover effects in the oyster, *Saccostrea glomerata*, following transgenerational exposure to ocean acidification. *PLoS ONE,* **10(7)**: e0132276.

Peñaloza C, Bishop S, Toro J, Houston R (2006). RAD Sequencing reveals genome-wide heterozygote deficiency in pair crosses of the Chilean mussel *Mytilus* spp. *10th World Congr. Genet. Appl. to Livest. Prod.*

Peyer SM, Hermanson JC, Lee CE (2010). Developmental plasticity of shell morphology of quagga mussels from shallow and deep-water habitats of the Great Lakes. *Journal of Experimental Biology* **213**: 2602–9.

Plough L V (2016). Fine-scale temporal analysis of genotype-dependent mortality at settlement in the Pacific oyster *Crassostrea gigas*. *bioRxiv* 84616.

Plough L V, Hedgecock D (2011). Quantitative trait locus analysis of stage-specific inbreeding depression in the Pacific oyster. *Genetics,* **189**: 1473–86.

Reynaga-Franco F-J, Grijalva-Chon J-M, *Crassostrea gigas* Castro-Longoria R, Barraza-Guardado R-H, Arreola-Lizárraga J-A, Chávez-Villalba J (2019). Biological performance of *Crassostrea gigas* stocks produced at different hatcheries and cultivated under same environmental conditions. *Aquaculture Research,* **50**: 621–33.

Reynaga-Franco F de J, Grijalva-Chon J-M, Castro-Longoria R, Barraza-Guardado R-H, Arreola-Lizárraga J-A, Chávez-Villalba J (2020). Designing a protocol to evaluate *Crassostrea gigas* spat production in hatcheries: Identification of critical aspects. *Aquacultural Engineering,* **89**: 102055.

Richez F (2012). Report on the impact of recent *Crassostrea gigas* mortality in France and its consequences to oyster farming in Northern Ireland. Aquaculture Initiative, Department of Agriculture and Rural Development (NI) and European Fisheries Fund. Department of Agriculture and Rural Development (Northern Ireland), Belfast, UK.

Riisgård HU, Bøttiger L, Pleissner D (2012). Effect of Salinity on Growth of Mussels , *Mytilus edulis*, with Special Reference to Great Belt (Denmark). *Open Journal of Marine Science* **2**: 167–76.

Riviere G, Wu G-C, Fellous A, Goux D, Sourdaine P, Favrel P (2013). DNA Methylation Is Crucial for the Early Development in the Oyster *C. gigas*. *Marine Biotechnology* **15**: 739–53.

Rybovich M, La Peyre MK, Hall SG, La Peyre JF (2016). Increased temperatures combined with lowered salinities differentially impact oyster size class growth and mortality. *Journal of Shellfish Research* **35**: 101–13.

Sae-Lim P, Gjerde B, Nielsen HM, Mulder H, Kause A (2016). A review of genotype-by-environment interaction and micro-environmental sensitivity in aquaculture species. *Reviews in Aquaculture*, **8(4)**: 369-393.

Sánchez-Lazo C, Martínez-Pita I (2012). Induction of settlement in larvae of the mussel *Mytilus galloprovincialis* using neuroactive compounds. *Aquaculture* **344**–**349**: 210–15.

Saurel C, Ng C, Barreau P, Connellan I, Hannon C, Hughes A, et al. (2022). Hatchery protocols for production of blue mussel seeds.

Sauvage C, Bierne N, Lapègue S, Boudry P (2007). Single Nucleotide polymorphisms and their relationship to codon usage bias in the Pacific oyster *Crassostrea gigas. Genes,* **406**: 13–22.

Sauvage C, Boudry P, De Koning D-J, Haley CS, Heurtebise S, Lapègue S (2010). QTL for resistance to summer mortality and OsHV-1 load in the Pacific oyster (*Crassostrea gigas*). *Animal Genetics* **41**: 390–99.

Scanes E, Parker LM, O'Connor WA, Dove MC, Ross PM (2020). Heatwaves alter survival of the Sydney rock oyster, *Saccostrea glomerata*. *Marine Pollution Bulletin* **158**: 111389.

Silveira RC, Silva FC, Gomes CHM, Ferreira JF, Melo CMR (2011). Larval settlement and spat recovery rates of the oyster *Crassostrea brasiliana* (Lamarck, 1819) using different systems to induce metamorphosis. *Brazilian Journal of Biology* **71**: 557–62.

SPAT (2012). Reference Manual for SPAT Oyster Gardeners. SPAT—Southold Project in Aquaculture Training. Cornell Cooperative Extension of Suffolk Marine Program, 75.

Steeves LE, Filgueira R, Guyondet T, Chassé J, Comeau L (2018). Past, Present, and Future: Performance of Two Bivalve Species Under Changing Environmental Conditions. *Frontiers in Marine Science*, **5**: 184.

Subasinghe R (2017). *World Aquaculture 2015 : A Brief Overview*, Vol. 1140.

Suquet M, Labbe C, Brizard R, Donval A, Le Coz JR, Quere C, et al. (2010). Changes in motility, ATP content, morphology and fertilisation capacity during the movement phase of tetraploid Pacific oyster (*Crassostrea gigas*) sperm. *Theriogenology* **74**: 111–17.

Symonds JE, Clarke SM, King N, Walker SP, Blanchard B, Sutherland D, et al. (2019). Developing Successful Breeding Programs for New Zealand Aquaculture : A Perspective on Progress and Future Genomic Opportunities. *Frontiers in Genetics* **10**: 1–7.

Tanyaros S (2011). The effect of substrate conditioning on larval settlement and spat growth of the cupped oyster, *Crassostrea belcheri* (Sowerby), in a Hatchery. *Agriculture and Natural Resources* **45**: 629–35.

Tanyaros S, Chuseingjaw S (2016). A partial substitution of microalgae with single cell detritus produced from seaweed (*Porphyra haitanensis*) for the nursery culture of tropical oyster (*Crassostrea belcheri*). *Aquaculture Research* **47**: 2080–88.

Tanyaros S, Tarangkoon W (2016). Variability in larval period, post-setting growth and survival of the oyster *Crassostrea belcheri* produced by gamete stripping method. *Agriculture and Natural Resources* **50**: 295–98.

Taris N, Batista FM, Boudry P (2007). Evidence of response to unintentional selection for faster development and inbreeding depression in *Crassostrea gigas* larvae. *Aquaculture* **272**: 69–79.

Taris N, Sauvage C, Batista FM, Baron S, Ernande B, Haffray P, et al. (2006). Phenotypic and genetic consequences of size selection at the larval stage in the Pacific oyster (*Crassostrea gigas*). *Journal of Experimental Marine Biology and Ecology,* **333(1):** 147-158.

United Nations (2015). Transforming our world: The 2030 agenda for sustainable development. *General Assembley 70 session.* New York, NY, USA.

Uren Webster TM, Rodriguez-Barreto D, Martin SAM, Van Oosterhout C, Orozco-terWengel P, Cable J, et al. (2018). Contrasting effects of acute and chronic stress on the transcriptome, epigenome, and immune response of Atlantic salmon. *Epigenetics* **13(12)**: 1191-1207.

Utting SD, Millican PF (1997). Techniques for the hatchery conditioning of bivalve broodstocks and the subsequent effect on egg quality and larval viability. *Aquaculture,* **155(1-4)**: 45-54.

Valin H, Sands RD, van der Mensbrugghe D, Nelson GC, Ahammad H, Blanc E, et al. (2014). The future of food demand: understanding differences in global economic models. *Agricultural Economics* **45**: 51–67.

Vandeputte M, Dupont-Nivet M, Haffray P, Chavanne H, Cenadelli S, Parati K, et al. (2009). Response to domestication and selection for growth in the European sea bass (*Dicentrarchus labrax*) in separate and mixed tanks. *Aquaculture,* **286(1-2)**: 20-27.

Vu S V, Gondro C, Nguyen NTH, Gilmour AR, Tearle R, Knibb W, et al. (2021). Prediction Accuracies of Genomic Selection for Nine Commercially Important Traits in the Portuguese Oyster (*Crassostrea angulata*) Using DArT-Seq Technology. *Genes,* **12**: 210.

Wang C, Chai X, Wang H, Tang B, Liu B (2013). Growth performance of the clam, *Meretrix meretrix*, breeding-selection populations cultured in different conditions. *Acta Oceanologica Sinica,* **32**: 82–87.

Wang J, Qi H, Li L, Que H, Wang D, Zhang G (2015). Discovery and validation of genic single nucleotide

polymorphisms in the Pacific oyster *Crassostrea gigas*. *Molecular Ecology Resources* **15**: 123–35.

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

Wang X, Li Q, Yu H, Kong L (2016). Genetic variation assessed with microsatellites in mass selection lines of the Pacific oyster (*Crassostrea gigas*) in China. *Journal of Ocean University of China* **15**: 1039–45.

Ward RD, Thompson PA, Appleyard SA, Swan AA, Kube PD (2005). Sustainable genetic improvement of Pacific oysters in Tasmania and South Australia. *Fisheries Research and Development Corporation final report. Canberra, Australia.*

Yáñez JM, Newman S, Houston RD (2015). Genomics in aquaculture to better understand species biology and accelerate genetic progress. *Frontiers in Genetics,* **9**:693.

Yang H, Sturmer LN, Baker S (2014). Molluscan Shellfish Aquaculture and Production. 1–8.

Zenger KR, Khatkar MS, Jones DB, Khalilisamani N, Jerry DR, Raadsma HW (2019). Genomic Selection in Aquaculture: Application, Limitations and Opportunities With Special Reference to Marine Shrimp and Pearl Oysters. *Frontiers in genetics*, **9:**, 693.

Zhang G, Fang X, Guo X, Li L, Luo R, Xu F, et al. (2012). The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* **490**: 49–54.

Zheng H, Zhang G, Liu X, Guo X (2006). Sustained response to selection in an introduced population of the hermaphroditic bay scallop *Argopecten irradians irradians* Lamarck (1819). *Aquaculture* **255**: 579–85.

Zheng H, Zhang G, Liu X, Zhang F, Guo X (2004). Different responses to selection in two stocks of the bay scallop, *Argopecten irradians irradians* Lamarck (1819). *Journal of Experimental Marine Biology and Ecology* **313**: 213–23.

Zhong X, Feng D, Yu H, Kong L, Li Q (2016). Genetic variation and breeding signature in mass selection lines of the Pacific oyster (*Crassostrea gigas*) assessed by SNP markers. *Plos one* **11**: e0150868.

2. Chapter II - SNP Discovery and Genetic Structure in Blue Mussel species Using Low Coverage Sequencing and a Medium Density 60K SNP array

This chapter resulted from a collaborative work of myself,h Alexis Simon, Jim P Whiting, Josie Paris, Carolina Penaloza, Bonnie P Fraser, Nicolas Bierne, Ross D. Houston Tim P. Bean and Robert Ellis. The manuscript has been submitted to the Scientific Journal “Evolutionary applications” in October 2022 and is currently ongoing the review process. I was involved in work carried in the laboratory, for analysing the data and developing the array and writing the manuscript.

Abstract

Blue mussels from the genus *Mytilus* are an abundant component of the benthic community, found in the high latitude habitats. These foundation species are relevant to the aquaculture industry, with over 2 million tonnes produced globally each year. Mussels withstand a wide range of environmental conditions and the species readily hybridise in regions where their distributions overlap. Significant effort has been made to investigate the consequences of environmental stress on mussel physiology, reproductive isolation and local adaptation. Yet our understanding on the genomic mechanisms underlying such processes remains limited. In this study, we developed a multi species medium-density 60K SNP-array including four species of the *Mytilus* genus. SNPs included in the platform were called from 138 mussels from 23 globally distributed mussel populations, sequenced using a whole-genome low coverage approach. The array contains polymorphic SNPs which capture the genetic diversity present in mussel populations thriving across a gradient of environmental conditions (~59K SNPs) and a set of published and validated SNPs informative for species identification and for diagnosis of transmissible cancer (610 SNPs). The array will allow the consistent genotyping of individuals, facilitating the investigation of ecological and evolutionary processes in these taxa. The applications of this array extend to shellfish aquaculture, contributing to the optimisation of this industry via genomic selection of blue mussels, parentage assignment, inbreeding assessment and traceability. Further applications such as genome wide association studies (GWAS) for key production traits and those related to environmental resilience are especially relevant to safeguard aquaculture production under climate change.

Keywords: mussels, SNP, population genomics, aquaculture, SNP chip, *Mytilus*

Background

Blue mussels from the genus *Mytilus* are an abundant component of the benthos, found in high latitude habitats (Gosling 2015). These foundation species can aggregate in high densities, forming extensive beds or reefs, which provide a number of important ecosystem services (e.g. providing spatial structure, undertaking nutrient cycling and forming an important food source) (van der Schatte Olivier *et al.* 2020). Additionally, mussels play an important economic role, as both a fishery and aquaculture species, accounting for approximately 12%, or ~2 million tonnes, of global mollusc production (Subasinghe 2017). Most landings (>90%) derive from aquaculture (Avdelas *et al.* 2021), with farmed bivalves identified as one of the most sustainable sources of animal protein (Hilborn *et al.* 2018). From a nutritional perspective, mussels contain high levels of omega-3 polyunsaturated fatty acids and essential amino acids, which in the human diet have significant health benefits (Carboni *et al.* 2019).

Commercial blue mussel production in Europe relies almost exclusively on collection of naturally-settled spat (i.e. settled juveniles) (Kamermans *et al.* 2013). Several environmental factors (e.g. water temperature, salinity, food availability and local currents) influence the reproductive cycle, in addition to triggering spawning events and determining larval dispersal patterns. There are eight species within the *Mytilus* genus (Gaitán-Espitia *et al.* 2016), which are able to readily hybridise wherever their geographic range overlaps throughout the world: in the west coast of the United Kingdom (Gardner 1996; Hilbish *et al.* 2002; Vendrami *et al.* 2020), the north east Atlantic (Bierne *et al.* 2002, 2003b; Fraïsse *et al.* 2016; Simon *et al.* 2021), north west Atlantic (Koehn *et al.* 1984; Rawson *et al.* 2001; Toro *et al.* 2004), the Baltic sea (Riginos & Cunningham 2005; Stuckas *et al.* 2017; Väinölä & Hvilsom 1991), subarctic and arctic mussels (Mathiesen *et al.* 2017), the north east Pacific (Rawson *et al.* 1999; Saarman & Pogson 2015), south and east Pacific (Larraín *et al.* 2019; Popovic *et al.* 2020) and south west Atlantic (Zbawicka *et al.* 2018). Broad-scale population structure and population dynamics in this species complex is therefore predominantly shaped by interactions between oceanography and the biology of each species. Both pre- and post-settlement selection drives

geographical and ecological segmentation, and contributes to determining species distribution and in shaping hybrid zones (Bierne *et al.* 2002, 2003a; Knöbel *et al.* 2021; Koehn *et al.* 1980). Further research is needed to fully understand the genetic barriers that determine distribution of these taxa, as well as the genomic mechanisms driving such processes across different environments. Moreover, the success of mussel aquaculture is tightly coupled to the environment, across all stages of production, with environmental change also influencing key performance traits including growth, survival and susceptibility to disease (Nascimento-Schulze *et al.* 2021).

Selective breeding has been highlighted as a key tool to facilitate sustainable intensification of bivalve aquaculture, allowing the development of specialised breeding lines resilient to environmental and pathogenic challenges (FAO 2016; Nascimento-Schulze *et al.* 2021; Potts *et al.* 2021). Selection has benefited the production of many cultured aquatic taxa (Gjedrem and Rye 2018), including the most important marine bivalve molluscs such as mussels (*Perna canaliculus*) and oysters (Hollenbeck and Johnston, 2018). Methods of selection vary from mass selection, for example breeding from the fastest growers in a population, through to methods using genetic markers spread across the genome, known as genomic selection (GS). GS is particularly powerful as it can improve accuracy of selective breeding, contributing to highly targeted results, even in the case of polygenic traits, whilst allowing for full control of the genetic relationships of the offspring (Houston *et al.* 2020; Meuwissen *et al.* 2001).

To utilise GS, genome-wide markers are required. Single nucleotide polymorphisms (SNPs) are the most common form of genetic variation and, therefore, the marker of choice for GS. SNPs are present in genes/regulatory regions, and non-coding regions. Whilst these markers may be contributing to the expressed phenotype, using a density of markers can assure that they are in linkage disequilibrium (LD) with causal mutations. Advances in sequencing and computational technologies have lowered costs of SNP discovery and enable the generation of large quantities of sequencing data and high throughput screening of SNPs. Consequently, these genetic markers have been widely applied for the development of further genomic resources.

SNP arrays use a probe-based approach to generate high-quality genotype data whilst requiring less investment in sample preparation. Computational analysis of genotype data generated by arrays is less demanding than techniques such as genotyping-by-sequencing that provide a similar amount of data. Furthermore, SNP arrays enable genotyping in multiple pre-defined loci, guaranteeing reproducibility of analysis (Robledo et al. 2018). Such information contributes to our understanding of genomic processes underpinning research in evolutionary genomics, population genetics, conservation and ecology (Allendorf *et al.* 2010; Andrews *et al.* 2016). Therefore, SNP arrays have valuable applications for the study of both wild and farmed populations and have been successfully applied across multiple taxa (Houston *et al.* 2014; Kranis *et al.* 2013; Michelizzi *et al.* 2011; Stoffel *et al.* 2012). Arrays are currently available for several farmed aquatic species including the Pacific (*Crassostrea gigas*) and European flat oysters (*Ostrea edulis*) (Gutierrez *et al.* 2017; Lapègue *et al.* 2014; Qi *et al.* 2017), Atlantic salmon (*Salmo salar*) (Houston *et al.* 2014), Rainbow trout (*Onchorynchus mykiss*) (Bernard *et al.* 2022; Palti *et al.* 2015) and Nile Tilapia (*Oreochromis niloticus*) (Joshi *et al.* 2018; Peñaloza *et al.* 2020), offering a rapid, accessible and cost-effective approach for medium- and high-density genotyping.

In blue mussels, nuclear and mitochondrial DNA markers have been widely used to investigate evolutionary processes (Quesada *et al.* 1998; Rawson & Hilbish 1995; Zouros *et al.* 1994), species genetics (Koehn 1991; McDonald *et al.* 1991), hybridisation patterns (Bierne *et al.* 2003b; Riginos & Cunningham 2005; Stuckas *et al.* 2017) and selection (Bierne *et al.* 2003a; Fraïsse *et al.* 2016; Knöbel *et al.* 2021; Koehn *et al.* 1980). Low-density SNP-panels have been used to delineate species and genetic lineages within the genus (Simon *et al.* 2021; Wilson *et al.* 2018), to investigate the shell trait-species correlation (Carboni *et al.* 2021) and also for pedigree reconstruction (Nguyen et al. 2014a). Studies have also applied different genotyping methods including low-density panels (Gardner *et al.* 2016; Saarman & Pogson 2015; Wenne *et al.* 2016; Zbawicka *et al.* 2014, 2018), as well as next generation sequencing (NGS), to investigate genetic diversity structure in populations across the globe (e.g. Fraïsse *et al.* 2016; Nguyen *et al.* 2014b; Vendrami *et al.* 2020), which has led to the

development of an 81-SNP Fluidigm genotyping assay panel (Mathiesen *et al.* 2017) and a 212-SNP ancestry informative panel (Simon *et al.* 2021). Finally, the recently assembled genomes of *M. edulis*, *M. galloprovincialis, M. coruscus, M. chilensis and M. californianus* (Corrochano-Fraile *et al.* 2021; Gallardo-Escárate *et al.* 2022; Gerdol *et al.* 2020; Paggeot *et al.* 2022; Simon 2022; Yang *et al.* 2021), will undoubtedly facilitate the development of inclusive and genome-wide tools for these taxa. Low density marker panels are underpowered for many important questions, including GWAS to identify polygenic QTLs, population structure and speciation in these taxa. High density panels can resolve such open questions, consequently fast-tracking genetic improvement in mussels.

Here we describe the generation of a SNP database for the *Mytilus edulis* species complex, including blue mussel species found in the southern hemisphere (i.e. *M. chilensis*). A subset of this database comprising 60K SNPs was then used to develop a medium density SNP-array on the Affymetrix ThermoFisher platform. Polymorphic markers were discovered using low-coverage whole-genome sequencing (lcWGS) data of 23 globally distributed blue-mussel populations, including multiple pure and hybrid genotypes. This tool will facilitate the investigation of population genetic processes such as local adaptation and reproductive isolation in this species complex, and increase the potential of selective breeding in mussel aquaculture.

Material and Methods

Sample collection, DNA extraction and sequencing

Twenty-three *Mytilus* spp populations were sampled from across their global distribution in 2018 (Figure 2.1, Table 2.1), incorporating 3 species in this complex (*M. edulis*, *M. trossulus and M. galloprovincialis)* and their hybrids, together with the southern *M. chilensis*. DNA was extracted from ethanol (96%) preserved adductor muscle tissue using the E.Z.N.A.® mollusc DNA kit (Omega Bio-Tek). Extraction followed the manufacturers protocol except the following adjustment: 30 mg of tissue was first homogenised (FastPrep-24™ 5G, MP Biomedicals™) for 1 minute (30 + 30 secs) at 6 m sec $^{-1}$ in 350 mL of ML1 buffer, and subsequently digested for 1-4 h at 50°C (until no tissue was visible in the vials)

with 25 µl of proteinase K, prior to extraction. DNA was assessed using a NanoDrop One Spectrophotometer (A260/280 and 260/230 ratios) and a Qubit BR assay. Whole-genome sequencing (WGS) library preparation and sequencing of 6 individuals per population (n = 138) was performed by the University of Exeter Sequencing Service. Libraries were prepared with the NEBNext Ultra II FS DNA Sample Preparation Kit (NEB) and Illumina TruSeq adapters following the manufacturer's guidelines. Library quality was initially checked by Tapestation D1000 before samples were pooled, the accuracy of pooling was checked on an Illumina Miseq, before sequencing (150 paired-end) on an Illumina Novaseq 6000 (Illumina, Inc., California) to 5-fold coverage.

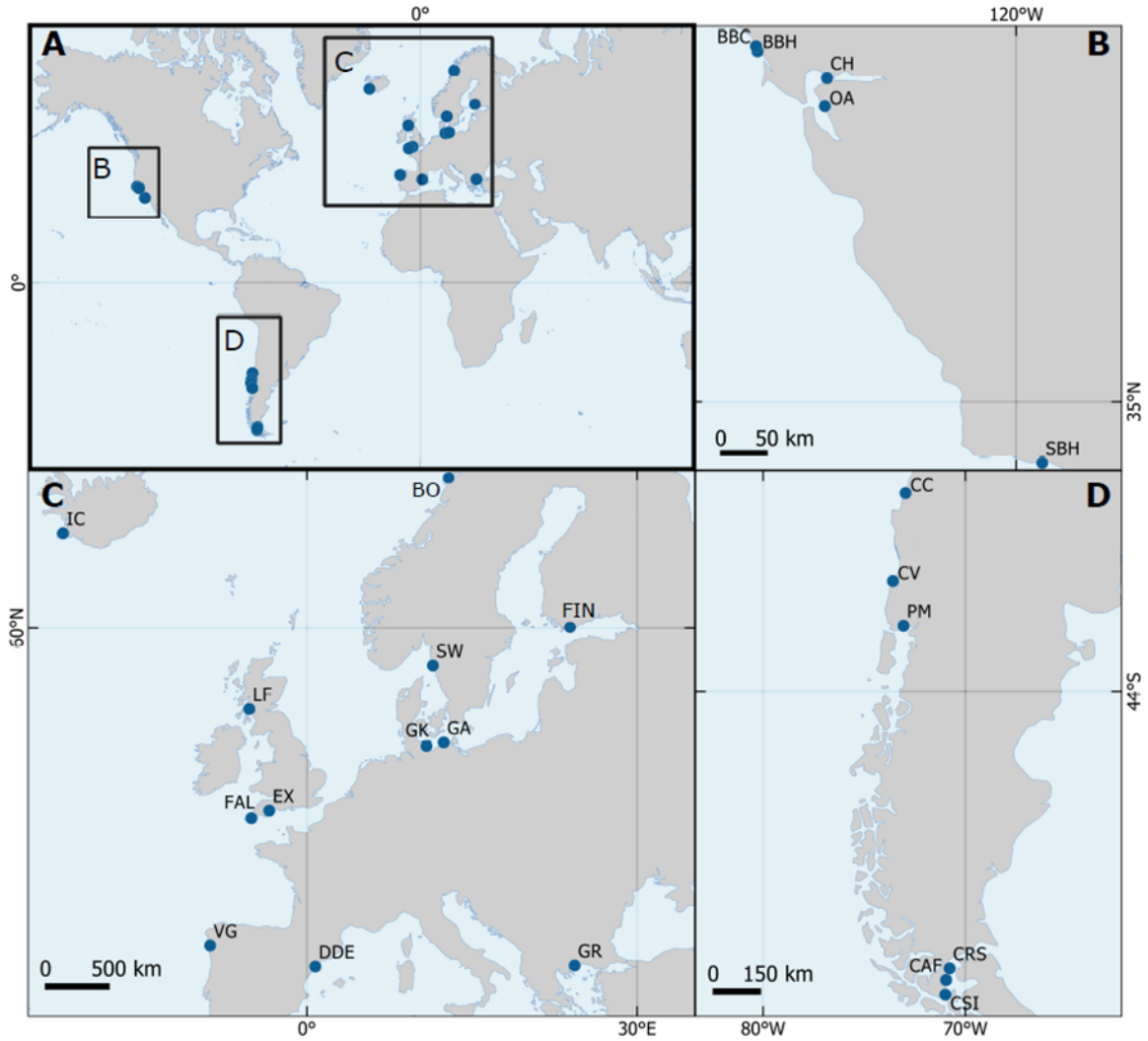


Figure 2-1: Global distribution (A) of the 23 Blue mussel populations used in this study to generate WGS data. Segments (B),(C) and (D) of the figure detail the location of populations in the North East Pacific, North East Atlantic, Baltic and Mediterranean Seas and South East Pacific, respectively.

Table 2-1: Sampling location, geographical coordinates and number of sequenced individuals in each of the populations used to generate WGS sequencing data in this study.

Samples	Sampling location	Geographical coordinates	n°of sequenced individuals
CH	Carquinez Harbour (USA) - North east pacific	38.066; -122.231	6
BBH	Bodega Bay Harbour (USA) - North east pacific	38.313; -123.051	6
BBC	Bodega Bay Coastal (USA) - North east pacific	38.361; -123.070	6
SBH	Santa Barbra Harbour (USA) - North east pacific	34.407; -119.691	6
OA	Oakland (USA) - North east pacific	37.805; -122.257	6
CC	Coliumo, Concepcion (Chile) - South east pacific	-36.537; -72.958	6
CV	Niebla, Valdivia (Chile) - South east pacific	-39.948; -73.402	6
PM	Puerto Montt (Chile) - South east pacific	-41.620; -73.058	6
CSI	San Isidro (Chile) - South east pacific	-53.797; -70.992	6
CRS	Rio Seco (Chile) - South east pacific	-53.024; -70.780	6
CAF	Agua Fresca (Chile) - South east pacific	-53.368; -70.992	6
DDE	Delta de Ebro, Barcelona (Spain) - Mediterranean	40.774; 0.763	6
GR	Eleutherón, Nea Peramos (Greece) - Mediterranean	40.850; 24.321	6
VG	Playa del Vao, Vigo (Spain) - North east Atlantic	42.200; -8.789	6
SW	Kristineberg (Sweden) - North east Atlantic	58.250; 11.447	6
BO	Bodø (Norway) - North east Atlantic	67.200; 14.420	6
GK	Kiel (Germany) - Baltic Sea	54.195; 10.860	6
GA	Ahrenshoop (Germany) - Baltic Sea	54.386; 12.427	6
FIN	Tvärminne (Finland) - Baltic Sea	59.838; 23.208	6
LF	Loch Fyne (Scotland) - North east Atlantic	56.117; -5.240	6
EX	River Exe, Exmouth (England) - North east Atlantic	50.362; -3.232	6
IC	Straumsvik (Iceland) - North east Atlantic	64.020; -22.157	6
FAL	River Fal, Mylor (England) - North east Atlantic	50.152; -5.024	6

Assessment of population structure and introgression in *Mytilus* spp. using WGS data.

To investigate genetic structure and ancestry in the full WGS dataset, an initial filtering was applied in which raw reads were checked for quality and subsequently trimmed and cleaned using fastp v0.19.7 (Chen *et al.* 2018), removing all reads < 100 bp and those with a quality score lower than 20. Clean reads were aligned to the Mediterranean *M. galloprovincialis* genome assembly (NCBI genome accession JAKGDF000000000), using BWA mem v0.7.17 (Li & Durbin 2009). Generated files were manipulated with SAMtools v1.3.1 (Li *et al.* 2009) and BCFtools 1.9 (Li 2011) and duplicate reads marked using MarkDuplicates (Picard v2.6.0) (http://broadinstitute.github.io/picard). Variants were

called using Freebayes v1.3.1 (Garrison & Marth 2012) from ~26K contigs larger than > 10,000 bp and with < 20 x coverage across aligned individuals. Resulting SNPs were filtered with GATK v4.0.5.1 (O'Connor & Van Auwera 2020) using the parameters suggested in the GATK hard filtering germline short variant pipeline (QD < 2.0, FS > 60.0, MQ < 40.0, HaplotypeScore > 13 and MappingQualityRankSum < 13). This resulted in the discovery of over 63 million SNPs. Only biallelic variants were retained, whilst all SNPs with a depth coverage (<4 and >10) were discarded.

Subsequently, variants missing in > 50% of individuals within each population were removed and then all samples were merged, before removing SNPs with a minor allele frequency < 0.01 in all samples using VCFtools v.0.1.15 (Danecek *et al.* 2011). This resulted in a dataset of ~31 Million SNPs. Putatively linked loci using window size of 50 Kb, a step size of 10Kb and an r^2 threshold of 0.1 were pruned from the dataset using Plink v1.9 indep-pairwise (Purcell *et al.* 2007). We first investigated population structure with a principal component analysis (PCA) (Plink v1.9) (Purcell *et al.* 2007) and conducted admixture analysis (admixture v1.3) (Alexander & Lange 2011) among the 23 mussel populations. To determine the best value of *K* clusters (value with lowest cross-validation (CV) error), we ran admixture testing *K* values between 2 and 10. Data missingness was estimated using the vcfR package in R v4.0.2 (R Core Team 2014). To account for the high frequency of missing data in individuals (79% ± 4), we ran the final PCA and admixture analysis only using SNPs that were genotyped across all populations (n = 4,367), which were identified using BCFtools v1.9 flag isec (Li 2011). The results from the population structure and admixture analyses were further applied to determine individual-level species composition used in the development of the SNP array. Species could be inferred as samples were taken from locations in which inhabiting species have been previously determined. Furthermore, results were supported by a visual check of the grouping of samples in the PCA as the array contained SNPs previously published which can safely distinguish between species (Tables 2.1, 2.2).

Initial quality check, alignment, variant calling and filtering for array development

SNP filtering for the SNP array genotyping platform followed the initial filtering steps as described in the section above, but different subsequent steps were applied for filtering: first we removed markers with overall minor allele frequency < 0.01, subsequently we kept only markers with monomorphic flanking regions (35 bp in each direction) and, we excluded those which were present in < 50% of individuals within each of the 23 populations. This resulted in 1,018,259 SNPs being kept for further analysis.

SNP selection for Axiom Blue mussel array

The set of 1,018,259 SNPs with putative monomorphic flanking probes (71-meter bp) was submitted to Affymetrix ThermoFisher for *in silico* evaluation. During the quality control (QC) process, each submitted SNP received a design score (p-convert value) for each 35 bp probe flanking the variant. Based on p-convert value, probes were classified as 'Recommended', 'Neutral', 'Not recommended', or 'Not possible'. For subsequent analysis, only variants that scored as 'Recommended' for both flanking probes were included.

Subsequently, SNPs were filtered within each of the four species group (*M. edulis*, *M. galloprovincialis*, *M. trossulus* and *M. chilensis*) by removing markers with a MAF < 0.02 within species. Samples were visually assigned to each species group (Figure 2.2), by analysing sample clusters in the PCA, and markers were classified as either genotyped in all species, three species, two species or species-exclusive, with those shared on any level among two or more groups being included on the platform (~54K SNPs). To reach the remaining target of 60K SNPs, markers unique for each species complex (10K SNPs) were selected, which will enable the tool to investigate within-species diversity. For this, 2.5K SNPs from each species were randomly filtered using Plink v1.9 flag –thin-count (Purcell *et al.* 2007). In addition, a set of previously published informative SNPs associated with species identification, transmissible cancer, phenotypic sex and population structure in this species complex were provided by collaborators (Table 2.2). The final composition of the panel of SNPs is presented in Table 2.3.

Table 2-2: Information on the previously published set of SNPs added in the platform.

SNP function	Targeted species	nº of SNPs provided	Reference
Species Identification	*M. edulis / M. galloprovincialis / M. trossulus*	12	*(Wilson et al. 2018)*
Phenotypic sex	*M. trossulus / M. edulis*	140	*(Burzyński & Śmietanka 2009; Śmietanka & Burzyński 2017; Śmietanka et al. 2010, 2016)*
Transmissible cancer (set 1)	*M. galloprovincinalis / M. edulis*	301	*Alexis Simon pers com unpublished results*
Transmissible cancer (set 2)	*M. galloprovincinalis / M. edulis*	35	*(Metzger et al. 2016; Vassilenko et al. 2010; Yonemitsu et al. 2019)*
Population structure	*M. galloprovincinalis / M. edulis*	113	*(Hammel et al. 2022; Simon et al. 2021)*
Population structure/fluidigm	*M. trossulus / M. edulis*	113	*(Mathiesen et al. 2017)*
Population structure	*M. edulis / M. chilensis*	96	*(Bach et al. 2019; Gardner et al. 2016; Larraín et al. 2018; Wenne et al. 2016, 2020; Zbawicka et al. 2014, 2018)*

Table 2-3: Final design of the Blue Mussel 60K Array, including information on SNP origin and function.

SNP function	Origin	SNPs (n)
Shared among all species	WGS	54,007
Unique to *M. galloprovinciallis*	WGS	1,339
Unique to *M. edulis*	WGS	1,384
Unique to *M. chilensis*	WGS	1,361
Unique to *M. trossulus*	WGS	1,360
Others	previous published studies (Table 2)	691
Total		**60,142**

SNP array validation

SNP summary statistics

One hundred and twenty-seven samples from the original discovery population were genotyped using the Blue Mussel array platform. For each of the markers in the platform, three intensity genotyping clusters were generated using the Axiom Analysis Suite software (v 5.1.1 Thermo Fisher Scientific) to segregate alleles from a single locus.

The standard protocol for marker quality check (QC) using the AxaS software includes selecting thresholds for the following filters: marker call rate (CR), average sample CR and DishQC (DQC). DQC is a quality control metric that measures the amount of overlap between two homozygous peaks created by a subset of non-polymorphic probes. In this study, we did not identify a subset of non-polymorphic probes shared by all the genotyped subspecies, most likely due to not identifying polymorphisms present in the flaking regions of the markers during initial filtering, rendering DQC inappropriate as a sample QC metric. For this reason, we used marker CR as a direct sample QC for the array validation, and five different runs were performed in the software using the following approach: firstly, combining all discovery population samples (run-ALL) and subsequently using samples classified as *M. galloprovinciallis* (run-GALLO), *M. edulis* (run-EDU); *M. trossulus* (run-TROS), and *M. chilensis* (run-CHIL). These were performed to understand whether a higher proportion of SNPs were called when grouping all species together rather than when analysing each individual species independently. For each of the runs, we analysed the final number of SNPs called with a marker CR equal or above 90% to 95%. Finally, we accounted for polymorphism among markers within and between species.

Assessment of array performance and applications

To assess the array performance, we compared genotype information within the four species generated by two different methods. The first dataset, obtained from the Axiom Analysis Software, consisted of genotype information generated by the 60K Blue mussel SNP array obtained by filtering markers with a CR of ≥ 95%, in the four individual species groups (*M. galloprovincialis, M. edulis, M.*

trossulus and *M. chilensis*). For the second dataset, we extracted genotype information on the same set of markers for the four species groups from the lcWGS data. Markers and their genotypes were extracted from this dataset using VCFtools v0.1.15.

We explored marker frequency and data missingness in both datasets using Plink (v1.9) (Purcell *et al.* 2007). We then compared called SNP genotypes in each of the individual species. For this, in each individual, we excluded SNPs with missing genotypes. This was done in both genotype datasets separately (lcWGS and SNP array). We then assessed the percentage of genotypes that were equivalently called for the SNPs retained in both datasets. Results are presented as (mean ± standard error) per species.

Results

Population structure and introgression of *Mytilus* sp generated from lcWGS data

The structure of the 23 *Mytilus* spp populations used to generate the 60K SNP array was assessed by PCA using the 4,367 SNPs which intersected among all populations (Fig. 2.2 and Supplementary Figure 2.1). The first two eigenvectors accounted for over 43% of the total variance (Supplementary Figure 2.2) with the samples clustering into four primaries, or distinct groups, when analysed visually (Figure 2.2). Using PC1 and PC2 it is possible to identify a grouping of the four putative species clusters. Several individuals were located in between the main species clusters (*M. galloprovincialis, M. trossulus* and *M. edulis*), suggesting the presence of hybrids in the sampled populations. The percentage of missing data across this reduced 4,367 SNP set was 33 % ± 5.

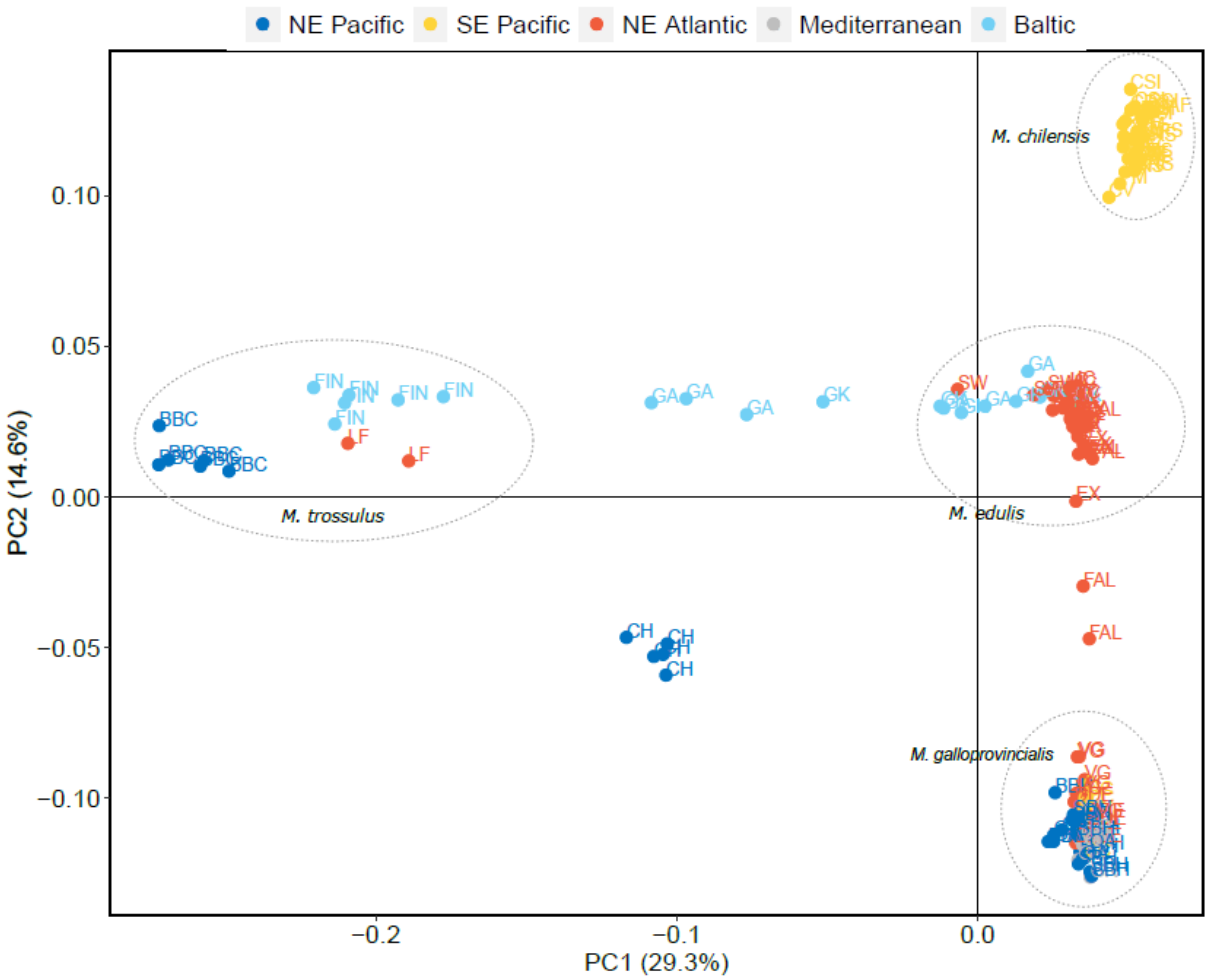


Figure 2-2: Scatter plots of individual variation in PC 1 and 2 scores resulting from PCA applied to the WGS dataset using SNPs intercepting among all blue mussel individuals from the 23 populations. The proportion of overall variation explained by each PC are given in percentages.

Ancestry was explored in blue mussel individuals from each of the 23 populations used for generating lcWGS data in this study with admixture (Figure 2.3). Following evaluation of cross-validation error for each tested *K* (built-in the software), the most likely number of identified clusters in our dataset was *K*=4, corresponding to the four species we expected to uncover in the lcWGS dataset (i.e. *M. edulis*, *M. galloprovincialis*, *M. trossulus* and *M. chilensis*). As we have not included any *M. trossulus* samples originating from 'pure' populations, as previously described in the literature (e.g. Stuckas *et al.* 2017), assignments of clusters to the four specific species in our results are putative, though they agree with previous studies describing the distribution of species (Araneda *et al.* 2016; Michalek *et al.* 2016; Saarman & Pogson 2015; Simon *et al.* 2020; Vendrami *et al.* 2020). Admixture

coefficients (Q) inference of each individual are presented in Figure 2.3, where in each column the proportions of the different colours show the inferred contribution of the four clusters to the genomic composition of the given individual.

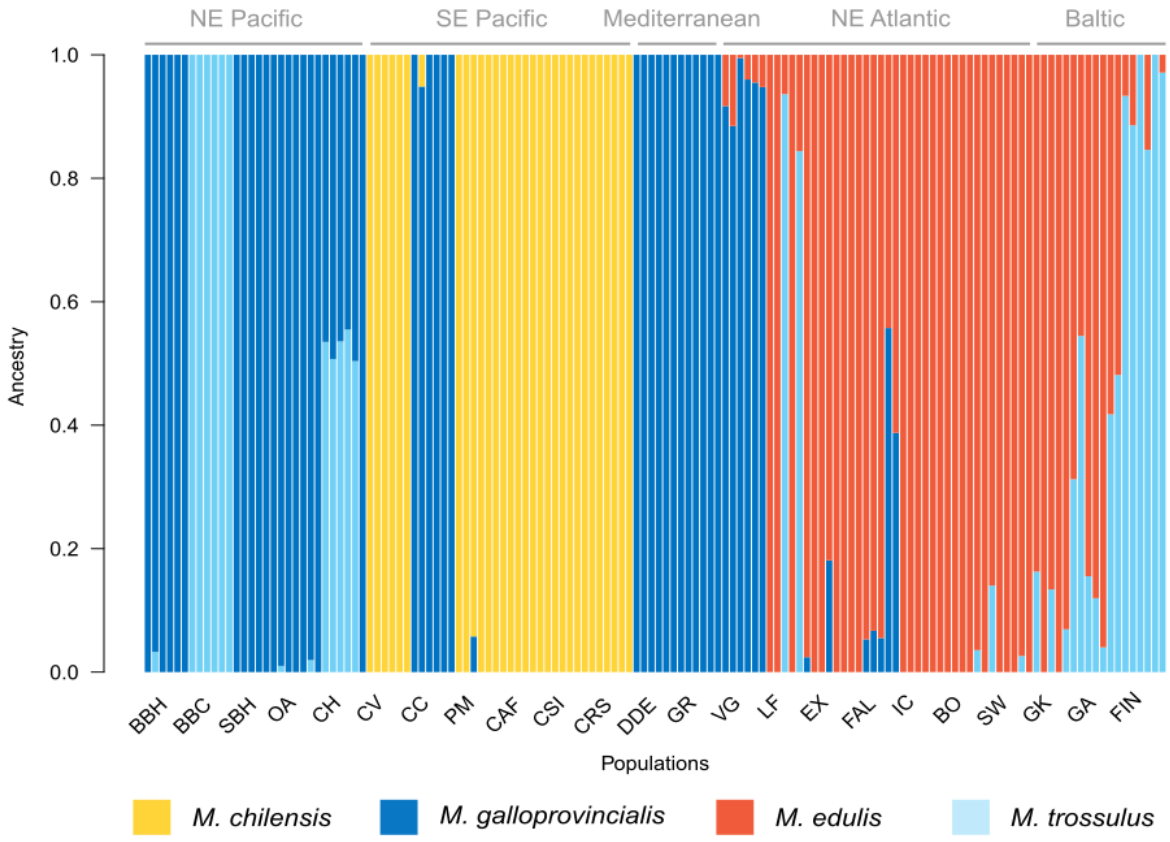


Figure 2-3: Results of genetic admixture analysis and ancestry inference. In each column the proportions of the different colours show the inferred contribution of the four clusters (Q) to the genomic composition of the given individual. Solid bars represent an individual from a single species background.

Clustering roughly recapitulated predicted species distribution with the yellow cluster representing *M. chilensis*, the dark blue cluster *M. galloprovincialis*, the red cluster *M. edulis* and the light blue cluster *M. trossulus* (Figure 2.3). Populations sampled in the south east Pacific were predominantly assigned to *M. chilensis* apart from one population (CC) which was assigned *M. galloprovincialis*. Samples from the north east Pacific were assigned either to *M. galloprovincialis* or *M. trossulus*, with some hybridisation in one of the populations between these two species. Hybrids of all 3 species of the *Mytilus edulis* species complex (*M. edulis*, *M. galloprovincialis* and *M. trossulus*) are present in north Atlantic populations. Baltic populations sampled in this study were assigned to

be admixed between the *M. trossulus and M. edulis* clusters, in a pattern common to introgressed populations.

SNP selection and array development

The alignment of the QC filtered reads against the *M. galloprovincialis* reference genome (genome accession JAKGDF000000000), attainment of bi-allelic SNPs and post-alignment QC filters led to the discovery of ~63 million putative polymorphisms. Following additional filtering criteria of excluding SNPs with MAF < 0.01, those present in less than 50% of genotypes within each population, and those with polymorphic alleles in the 35 bp flanking regions, 1,018,259 SNPs (as 71-mer nucleotide sequences) were assessed through the Affymetrix ThermoFisher *in silico* probe scoring. Markers classified as 'Recommended' on both flanking probes by ThermoFisher probe QC (233,488 SNPs) were used in subsequent analysis. From this set, SNPs with a MAF < 0.02 were filtered from each of the four species groups (Figure 2.4). The final 60,142 SNP array design (Table 2.3) contained: i) 54,007 SNPs shared by either all species (3,268), three species (13,185) or two species (37,554) (Figure 2.5), based on species assignment using admixture analysis; ii) SNPs exclusive for each species (*M. galloprovinciallis* 1,339; *M. edulis* 1,384; *M. trossulus* 1,360 and *M chilensis* 1,361) and iii) 691 SNPs from the 810 previously identified informative SNPs detailed in Table 2.

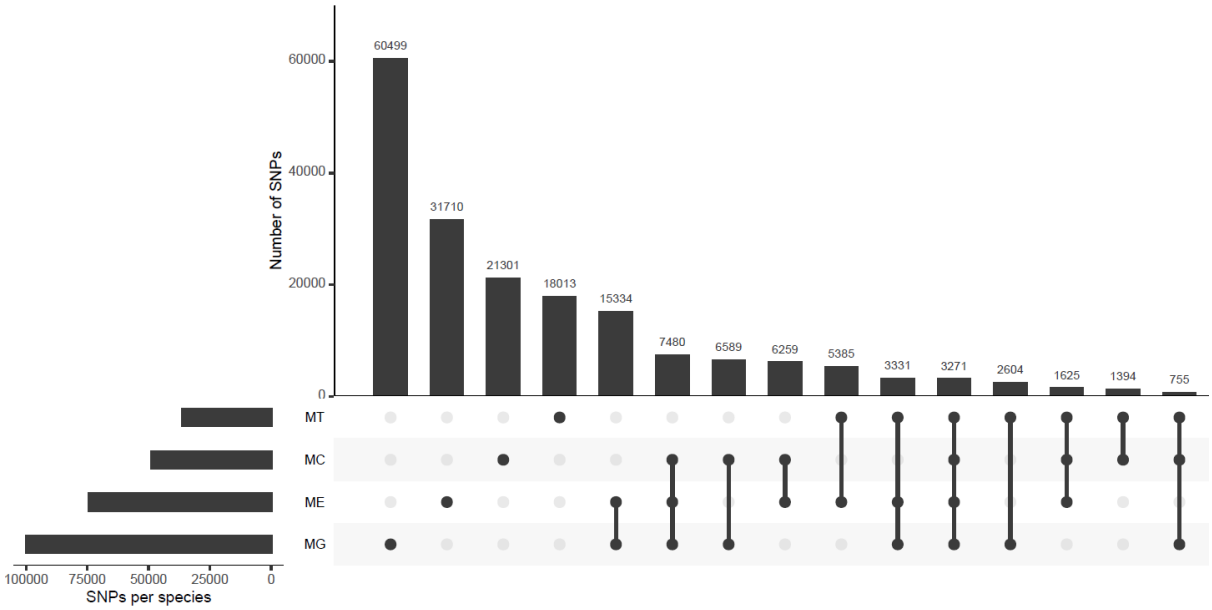


Figure 2-4: UpSet plot of final set of SNPs with MAF ≥ 0.02 for each of the species present in the WGS dataset approved by Affymetrix QC and assigned as 'recommended' on both flanking probes and their intersection among the four different species group: *M. edulis* (ME), *M. galloprovincialis* (MG), *M. trossulus* (MT) and *M. chilensis* (MC).

SNP array validation

SNP summary statistics

In this study, SNPs were retained for genotyping based on marker CR and not sample DQC. In all marker CR filtering scenarios, the number of usable SNPs generated was higher when analysing individual species groups than when analysing all discovery population samples in combination (Table 4). We set the CR of ≥ 95% as a standard for subsequent analysis, allowing genotyping of individuals in approximately 30% of 60K markers in all four individual species. With this CR filter, 23,252 markers were retained for genotyping *M. edulis*, 22,165 for *M. chilenesis*, 20,504 for *M. galloprovincialis* and 20,149 for *M. trossulus*.

We visualised the output of SNPs applicable for genotyping *Mytilus* spp with a CR of ≥ 95% resulting from the analysis of individual species (Table 2.4, Figure 2.5). When combining all groups, 46,420 SNPs were retained as applicable for genotyping, representing 77.2% of SNPs in the array. From this number, 12.1% of the markers were unique for *M. edulis*, 11.4% for both *M. chilensis* and *M.*

trossulus, whilst the lowest number of species exclusive markers was observed for *M. galloprovincialis*, with 9.3%. Approximately 6.92% of the SNPs were shared among all species.

Table 2-4: Number of SNPs applicable for genotyping resulting from multiple call rate (CR) scenarios (from 90% to 96%) using different group of samples: i) all discovery populations samples and ii) samples from the different Blue Mussel species used in the array.

Samples	Sampled individuals (n)	CR (%)	probes (n)
All samples combined	127	90	21,542
		91	20,074
		92	18,576
		93	15,570
		94	13,976
		95	12,440
		96	10,837
M. galloprovincialis **(run-GALLO)**	30	90	33,938
		91	27,509
		92	27,509
		93	27,509
		94	20,504
		95	20,504
		96	20,504
M. edulis **(run-EDU)**	22	90	31,908
		91	23,252
		92	23,252
		93	23,252
		94	23,252
		95	23,252
		96	13,207
M. trossulus **(run-TROS)**	14	90	33,058
		91	33,058
		92	33,058
		93	20,149
		94	20,149
		95	20,149
		96	20,149
M. chilensis **(run-CHIL)**	23	90	30,264
		91	30,264
		92	22,165
		93	22,165
		94	22,165
		95	22,165
		96	13,341

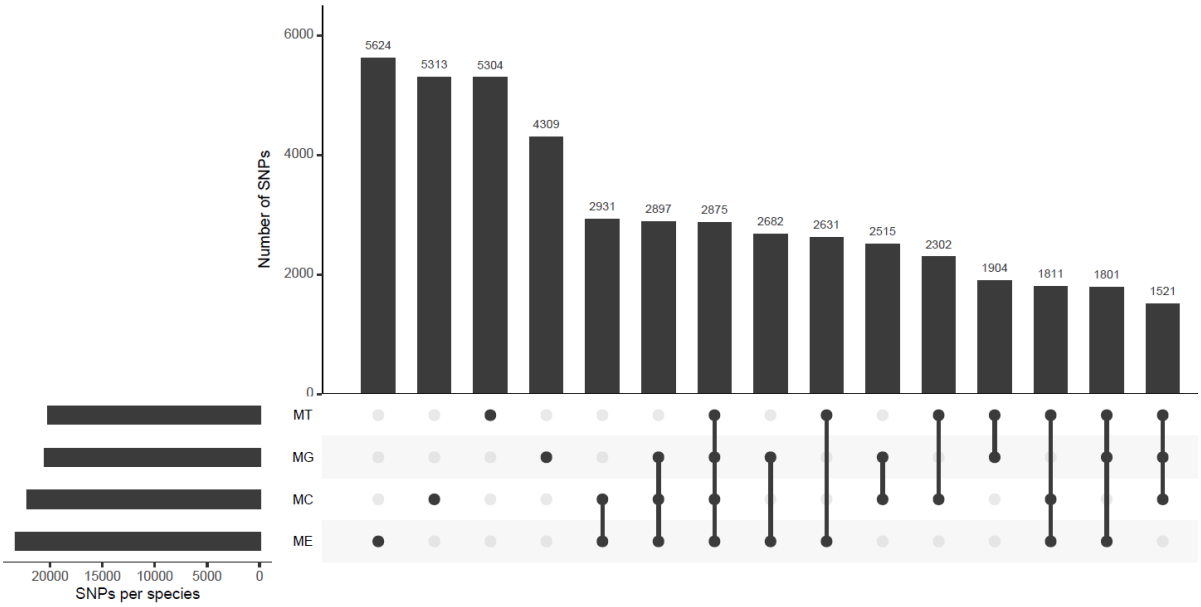


Figure 2-5: UpSet plot representing probes with a CR of ≥ 95% retained from the analysis of individual species, *M. edulis* (ME), *M. galloprovincialis* (MG), *M. trossulus* (MT) and *M. chilensis* (MC), generated by the Axiom Analysis Suite Software and their intersection among the four species group.

The majority of markers passing a CR threshold ≥ 95% were polymorphic within species (Table 2.5). Among the four species, *M. edulis* had the highest number of unique polymorphic markers (15.1%, Figure 2.6) compared to *M. chilensis* (13.5%), *M. galloprovincialis* (11.9%) and *M. trossulus* (10.6%).

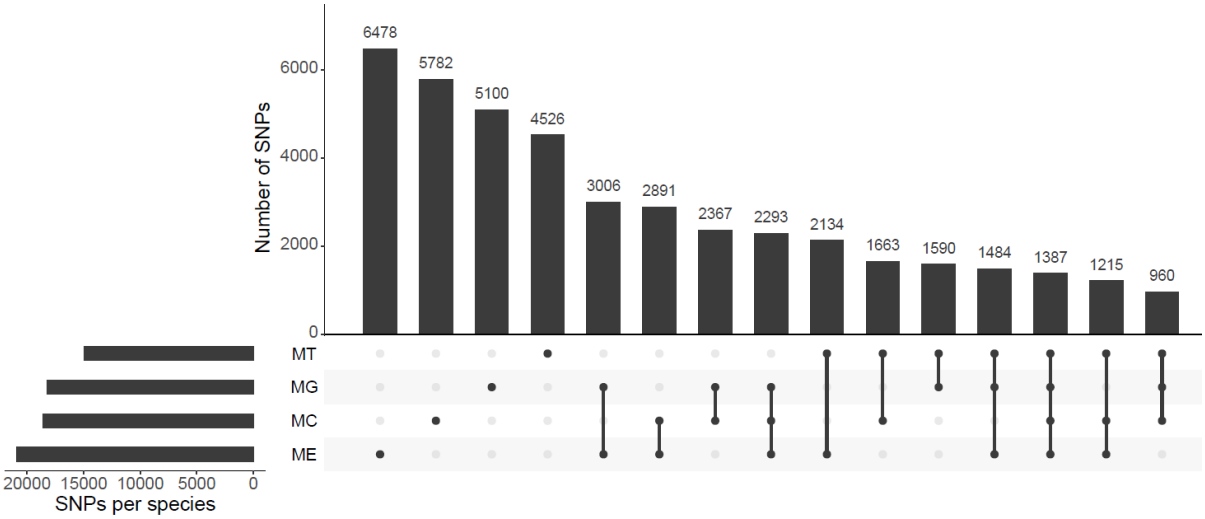


Figure 2-6: UpSet plot representing polymorphic probes with a CR of ≥ 95% retained from the analysis of individual species, *M. edulis* (ME), *M. galloprovincialis* (MG), *M. trossulus* (MT) and *M. chilensis* (MC), generated by the Axiom Analysis Suite Software and their intersection among the four species group.

Individual missing genotypes per individual species in the lcWGS dataset was generally high in the four species group: *M galloprovincialis* (63.11 % ± 0.58), *M. edulis* (64.45 % ± 0.55)), *M. trossulus* (77.24 % ± 0.6) and *M. chilensis* (66 % ± 0.56) in comparison to the array genotyping dataset: *M galloprovincialis* (1 % ± 0.5), *M. edulis* (2 % ± 0.1)), *M. trossulus* (0 ± 0) and *M. chilensis* (2 % ± 0.1), meaning that across the SNPs commonly called using both technologies, a far greater percentage were successfully called using the array.

For *M. galloprovincialis*, an average of 71.02% ± 4.07 of the genotypes were called equivalently using the two technologies, followed by *M. chilensis*, for which an average of 57.5% ± 1.59 of genotypes were consistently called in both genotyping platforms, *M. edulis* (50.1% ± 3.6) and *M. trossulus* (37.58% ± 3).

Table 2-5: Summary of useable SNPs analysing species individually (*M. edulis, M. galloprovincialis, M. trossulus* and *M. chilensis*) applying a call rate over 95% filter.

Samples (DNA)	Individuals (n)	SNPs category	SNPs (n, per category)
M. galloprovincialis	30	Total	20,504
		Monomorphic	2,317
		Polymorphic	18,187
		Polymorphic unique *to M. galloprovincialis*	5,100
M. edulis	22	Total	23,252
		Monomorphic	2,637
		Polymorphic	20,888
		Polymorphic unique *to M. edulis*	6,478
M. trossulus	14	Total	20,149
		Monomorphic	5,190
		Polymorphic	14,959
		Polymorphic unique *to M. trossulus*	4,526
M. chilensis	14	Total	22,165
		Monomorphic	3,607
		Polymorphic	18,558
		Polymorphic unique *to M. chilensis*	5,782

Discussion

In this study, we developed the first high throughput genotyping assay for the four main cultured species in the *Mytilus* genus, a multi-species medium density 60K SNP array. SNPs were selected from 138 individual mussels, originating from 23 wild, widely distributed populations (Figure 2.1, Table 2.1). These successfully called markers were either shared by three species (13,185 SNPs), two species (37,554 SNPs) or all four species (3,268 SNPs), or unique to each of the individual groups: *M. galloprovincialis* (1,343 SNPs), *M. edulis* (1,384 SNPs), *M. trossulus* (1,360 SNPs) and *M. chilenses* (1,361 SNPs). In addition, 691 SNPs representing markers generated by previous studies were included in the genotyping platform (Table 2.2). Following the removal of markers with a CR equal or lower than 95%, the final array design retained 23,252 applicable for genotyping *M. edulis* individuals, 22,165 for *M. chilensis*, 20,504 SNPs *M. galloprovincialis* and 20,149 for *M. trossulus*.

The CR of ≥ 95% threshold selected in this study has been successfully used in GWAS in different taxa to obtain informative results on population structure, admixture and identification of SNPs significant for breeding purposes (Becker *et al.* 2020; Cheng *et al.* 2022; Davenport *et al.* 2020; Liu *et al.* 2019; Wijesena *et al.* 2019). Using the sample CR filter instead of DQC as a genotype data filter has been previously applied in studies of several taxa, including sheep (Davenport *et al.* 2020), chicken (Liu *et al.* 2019) and swine (Cheng *et al.* 2022; Wijesena *et al.* 2019). Here we chose to apply this alternative filtering method as samples genotyped in the 60K blue mussel array consistently received low scores with the DQC parameter (Supplementary Table 2.1) and were removed from the analysis by the software, indicating the existence of polymorphisms present in the DQC 71 bp flanking probes. This is likely a consequence of the very high polymorphism present in mussel genomes (Romiguier *et al.* 2014), coupled with the chosen sampling approach of 23 mussel populations including four species. Whilst low coverage sequencing allowed the discovery of a large number of SNPs within the samples, it also resulted in patchy data. Choosing to sequence individuals with low coverage might have led to a high missingness, which in turn resulted in a high proportion of the

genetic variation being missed during the SNP selection process. This may have also resulted in the removal of accurate data on the presence of SNPs in flanking regions.

In this study, we analysed the performance of SNPs across different sample sets analysed in the AxaS software. A higher number of markers were kept when genotyping samples from each species individually (run-GALLO, run-EDULIS, run-TROS, run-CHIL) than when genotyping all samples combined (run-ALL). Clusters generated by the Axiom Analysis Software classify the genotypes as either homozygous for one of the two alleles in the probe, or heterozygous. Grouping samples of similar genetic background for analysis reduced the number of genotypes being classified ambiguously (i.e. "Other" and "OTV"), due to distortion of clear genotype clustering that sometimes arose with multispecies sample sets. We appreciate it is not always possible to previously infer the genetic background of samples that will be analysed using the array. However, our results highlight that grouping samples of similar genetic background improves the number of high-quality SNPs applicable for genotyping individuals with the array. To increase the number of functional markers applied for genotyping of a sample batch, we propose, when applicable, the employment of an analysis pipeline where samples are first disaggregated by species grouping before subsequently being analysed across the entire marker set. Nonetheless, the comparable population structure inferred from sample genotypes generated by the array (CR filter of ≥ 95%) to that generated by the lcWGS dataset, reinforces the suitability of developing a genotyping platform in this species complex.

Subsequently, we analysed the proportion of polymorphic markers retained within and among the four species groups. A high proportion of the successfully called SNPs on the array are polymorphic within species (*M. edulis* 89.8%, *M. galloprovincialis* 88.7%, *M. chilensis* 83.7% and *M. trossulus* 74.2%). The proportion of unique polymorphic probes per species was considerable (*M. edulis* 15.1%, *M. galloprovincialis* 11.9%, *M. chilensis* 13.5% and *M. trossulus* 10.6%). Polymorphism within each of the mussel species in the array was higher in comparison to other multi-species arrays, such as the ~49.9K European Pine tree multispecies array, in which approximately one quarter of the converted SNPs was polymorphic in each species in the array (Perry *et al.* 2020). Nonetheless, in our

study, only 960 polymorphic markers were shared among all four mussel species. These observations reassure that although markers are shared between either two, three or four mussel species constituting the array, this marker set is appropriate for investigating population structure and species assignment. Blue mussels, as is the case for other marine bivalve mollusc species, are known to have a highly complex and polymorphic genome (Calcino *et al.* 2021; Gerdol *et al.* 2020). Although a high proportion of markers is lost following the QC filtering, a sufficient (~20K) is retained for each of the four species when running QC for batches of samples with similar genetic background. Approximately a quarter of these markers are unique per species. These values reassure that the diversity existing within and among species can be captured with the Blue mussel 60K array. The addition of 691 markers, which have been previously confirmed as species-diagnostic and relevant for sex determination and identification of cancer-positive samples (cited in Table 2), further guarantees the applicability of the genomic tool generated in this study as highly valuable for fine scale investigation of genetic structure in these taxa.

It is very likely that two factors: i) the higher number of *M. galloprovincialis* individual in the founding and validation populations and ii) the alignment to a *M. galloprovincialis* genome assembly, may be contributing to a biased comparison. Besides, the high level of missing data across the lcWGS data suggests that this sequencing method was not the optimal approach for SNP selection for this species complex. For future research, we recommend a higher depth coverage for sequencing individuals. Finally, the percentage of genotypes called equivalently using both approaches was higher for *M. galloprovincialis* than for the other tested species. Such discrepancies might result from variance in the flanking probes, due to divergence in the genomes of these sister species, interfering in the produced signal or in the physical binding of the genetic material.

We assessed population structure and sample ancestry in the lcWGS dataset using only SNPs genotyped in all populations. With a PCA, we could visually distinguish among four species within the genus *Mytilus*, allocating populations as *M. edulis*, *M. trossulus*, *M galloprovincialis,* and *M. chilensis* (or hybrids between species),supported by previous studies of genetic differentiation (Araneda et al.

2016; Fraïsse et al. 2016; Gardner 1994; Hilbish et al. 2002; Koehn 1991; Michalek et al. 2016; Saarman & Pogson 2015; Väinölä & Hvilsom 1991; Vendrami et al. 2020; Zbawicka et al. 2018). Through this analysis we could observe some striking ancestry patterns in the sampled populations: a complete shift of dominant species composition was observed between sites of Bodega Bay mussel populations (BBC compared to BBH, California, USA). These sites are positioned less than 20 km apart from each other but differ in their nature: one being a coastal site (BBC) and the other an enclosed harbour (BBH). Hybrids in Carquinez Harbour (CH, California, USA) were composed of approximately 50:50 genetic contributions from two species, *M. trossulus* and *M. galloprovincialis*. Whilst this pattern is consistent with hybrid genotypes, recent evidence has suggested that 'dock mussels' (i.e. mussels inhabiting port environments) in Europe have a similar admixture pattern between *M. galloprovincialis* and *M. edulis* (Simon *et al.* 2020), and CH populations might be an additional example of such mussels in the north east Pacific. However, further research is required to understand the nature of such admixture patterns in these North West American populations. The Baltic population from Finland (FIN) shows a more advanced introgression pattern between two of the *K* clusters, likely between *M. edulis* and *M. trossulus*. While populations from Kiel and Ahrenshoop (GK, GA) in the contact zone show a mixing of introgressed *M. edulis* and hybrid genotypes, which is in agreement with previous literature (Stuckas *et al.* 2017). Along the northern Pacific coast, *M. galloprovincialis* is present in sheltered waters, contrasting to its preference in north east Atlantic populations where this species predominantly populates the exposed rocky tidal environments along the coast (Bierne *et al.* 2003b; Hilbish *et al.* 2002), with the exception of commercial ports (Simon *et al.* 2020). This observation is potentially consistent with an inverted coupling between local adaption genes and intrinsic species barriers as previously suggested for the inverted genetic-environment relationship observed with *M. edulis* and *M. trossulus* in the eastern and western Atlantic (Bierne *et al.* 2011). Finally, mussel populations in Exmouth, southern England, which have previously been described as pure *M. edulis* (Hilbish *et al.* 2002), are shown to be introgressed with *M. galloprovincialis*, an observation supported by a recent RAD sequencing study (Vendrami *et al.* 2020). This can be explained by a better efficiency to detect

introgression across a semi-permeable barrier to gene flow with an increased number of markers distributed along a larger portion of the genome, as discussed by Fraïsse *et al.* (2016). In conclusion, the species distribution patterns revealed by lcWGS data in this study support previous data for the Baltic Sea (Knöbel *et al.* 2021; Stuckas *et al.* 2017), Southwest England (Hilbish *et al.* 2002; Vendrami *et al.* 2020), Mediterranean (Boukadida *et al.* 2021; Simon *et al.* 2021), USA (Saarman & Pogson 2015) and Chilean (Araneda *et al.* 2016) populations. Such observations provide a valuable insight into species distributions and admixture in the blue mussel species-complex.

Conclusions and future perspectives

We have developed the first medium density multi species blue mussel SNP panel, subsequently validating its performance on 127 individuals from 23 Blue mussel populations. The blue mussel array includes variants present in each of the four main species: *M. edulis*, *M. galloprovincialis*, *M. trossulus* and *M. chilensis.* This is an open access tool which allows genotyping in a consistent marker set distributed across the *Mytilus* genome. The Blue Mussel 60 K array can be applied for breeding purposes (i.e. parentage assignment, inbreeding level assessment and species/product identification and provenance), contributing to the understanding of genetic architecture of traits of interest. Such advances will consequently contribute to the optimisation of blue mussel aquaculture via genomic selection, fast-tracking hatchery production, if those become available in an industrial scale. Equally, this tool can be used to deepen our understanding of population genetic processes in these taxa. Shedding light on t relevant ongoing molecular adaptive responses in mussels, as well as their genetic diversity existing in wild populations, can contribute to the development of conservation guidelines and effective management strategies of both wild and farmed mussel stock.

Data availability

Raw sequence reads from the blue mussel samples used for SNP discovery have been deposited in NCBI's Sequence Read Archive (SRA, https://www.ncbi.nlm.nih.gov/sra). The Blue mussel SNP array is available for from ThermoFisher (E-mail: BioinformaticsServices@thermofisher.com).

Acknowledgments

We would like to first acknowledge the extensive list of colleagues, collaborators and friends who have provided mussel samples for inclusion in this study. This includes Prof. Anne Todgham and Ms. Chessie Cooley-Rieders (UC Davis, California), Dr Mauricio Urbina (Universite de Concepcion, Chile), Dr Sam Dupont (University of Gothenberg, Sweden), Prof. Frank Melzner and Dr. Lara Schmittmann (GEOMAR, Germany), Dr Lydia Kapsenberg (CSIC Institute of Marine Sciences, Barcelona), Dr Stefano Carboni (IMC International Marine Centre, Italy), Dr Hrönn Egilsdóttir and Dr Sindri Gíslason (Marine and Freshwater Institute, Iceland), Dr Angel Diz (University of Vigo, Spain), Dr Manuela Truebano (University of Plymouth, UK), Dr Chrysoula Gubili (Fisheries Research Institute, Greece), Dr Isabel Garcia and Dr Joost Raeymaekers (NORD University, Norway), and Dr Mats Westerbom (University of Helsinki, Finland). We would also like to acknowledge colleagues who provided access and permission to use previously published SNPs for *Mytilus* including Prof. Artur Burzyński, Prof. Roman Wenne, Prof. Małgorzata Zbawicka and Dr Beata Śmietanka (Institute of Oceanology of the Polish Academy of Sciences, Poland), Dr Michael Bekaert (University of Stirling, UK), Prof. Einar Eg Nielsen (DTU Aqua, Denmark) and Dr Michael Metzger (Pacific Northwest Research Institute, USA). This project utilised equipment funded by the Wellcome Trust Multi-User Equipment Grant (award number 218247/Z/19/Z), Institutional Strategic Support Fund (WT097835MF) and Multi User Equipment Award (WT101650MA), as well as equipment funded by the UK Medical Research Council (MRC) Clinical Research Infrastructure Initiative (award number MR/M008924/1) and the BBSRC LOLA award (BB/K003240/1).

The authors acknowledge funding by a NERC GW4+ Doctoral Training Partnership CASE PhD studentship in partnership with CEFAS (NE/L002434/1; DP901X-SCN798; awarded to JCNS), as well as funding from the UK Biotechnology and Biological Sciences Research Council (BBSRC) and the UK Natural Environment Research Council (NERC), via NERC Industrial Innovation Fellowship (NE/R013241/1; awarded to RE) and BBSRC Institute Strategic Programme grants (BBS/E/D/30002275 and BBS/ E/D/1000207).

References

Alexander DH, Lange K (2011). Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics* **12**: 246.

Allendorf FW, Hohenlohe PA, Luikart G (2010). Genomics and the future of conservation genetics. *Nature Reviews Genetics* **11**: 697–709.

Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics* **17**: 81–92.

Araneda C, Larraín MA, Hecht B, Narum S (2016). Adaptive genetic variation distinguishes Chilean blue mussels (*Mytilus* chilensis) from different marine environments. *Ecology and Evolution* **6**: 3632–44.

Avdelas L, Avdic-Mravlje E, Borges Marques AC, Cano S, Capelle JJ, Carvalho N, et al. (2021). The decline of mussel aquaculture in the European Union: causes, economic impacts and opportunities. *Reviews in Aquaculture* **13**: 91–118.

Bach L, Zbawicka M, Strand J, Wenne R (2019). *Mytilus trossulus* in NW Greenland is genetically more similar to North Pacific than NW Atlantic populations of the species. *Marine Biodiversity* **49**: 1053–59.

Becker GM, Davenport KM, Burke JM, Lewis RM, Miller JE, Morgan JLM, et al. (2020). Genome-wide association study to identify genetic loci associated with gastrointestinal nematode resistance in Katahdin sheep. *Animal Genetics* **51**: 330–35.

Bernard M, Dehaullon A, Gao G, Paul K, Lagarde H, Charles M, et al. (2022). Development of a High-Density 665 K SNP Array for Rainbow Trout Genome-Wide Genotyping. Frontiers in Genetics, 13.

Bierne N, Bonhomme F, David P (2003a). Habitat preference and the marine-speciation paradox. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **270**: 1399–1406.

Bierne N, Borsa P, Daguin C, Jollivet D, Viard F, Bonhomme F, et al. (2003b). Introgression patterns in the mosaic hybrid zone between *Mytilus edulis* and *M. galloprovincialis*. *Molecular Ecology* **12**: 447–61.

Bierne N, David P, Langlade A, Bonhomme F (2002). Can habitat specialisation maintain a mosaic hybrid zone in marine bivalves? *Marine Ecology Progress Series* **245**: 157–70.

Bierne N, Welch J, Loire E, Bonhomme F, David P (2011). The coupling hypothesis: why genome scans may fail to map local adaptation genes. *Molecular Ecology* **20**: 2044–72.

Boukadida K, Mlouka R, Clerandeau C, Banni M, Cachot J (2021). Natural distribution of pure and hybrid *Mytilus* sp. along the south Mediterranean and North-east Atlantic coasts and sensitivity of D-larvae stages to temperature increases and metal pollution. *Science of The Total Environment* **756**: 143675.

Burzyński A, Śmietanka B (2009). Is Interlineage Recombination Responsible for Low Divergence of Mitochondrial nad3 Genes in *Mytilus galloprovincialis*? *Molecular Biology and Evolution* **26**: 1441–45.

Calcino AD, Kenny NJ, Gerdol M (2021). Single individual structural variant detection uncovers widespread hemizygosity in molluscs. *Philosophical Transactions of the Royal Society B:*

Biological Sciences **376**: 20200153.

Carboni S, Evans S, Tanner KE, Davie A, Bekaert M, Fitzer SC (2021). Are Shell Strength Phenotypic Traits in Mussels Associated with Species Alone?

Carboni S, Kaur G, Pryce A, Mckee K, Desbois AP, Dick JR, et al. (2019). Mussel Consumption as a "Food First" Approach to Improve Omega-3 Status. *Nutrients* **11**: 1–11.

Chen S, Zhou Y, Chen Y, Gu J (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**: i884–90.

Cheng J, Lim K, Putz AM, Wolc A, Harding JCS, Dyck MK, et al. (2022). Genetic analysis of disease resilience of wean-to-finish pigs under a natural disease challenge model using reaction norms. *Genetics Selection Evolution* **54**: 11.

Corrochano-Fraile A, Davie A, Carboni S, Bekaert M (2021). Evidence of multiple genome duplication events in *Mytilus* evolution. *bioRxiv* 2021.08.17.456601.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. (2011). The variant call format and VCFtools. *Bioinformatics (Oxford, England)* **27**: 2156–58.

Davenport KM, Hiemke C, McKay SD, Thorne JW, Lewis RM, Taylor T, et al. (2020). Genetic structure and admixture in sheep from terminal breeds in the United States. *Animal Genetics* **51**: 284–91.

FAO (2016). The state of world fisheries and aquaculture 2016. Contributing to food security and nutrition for all.

Fraïsse C, Belkhir K, Welch JJ, Bierne N (2016). Local interspecies introgression is the main cause of extreme levels of intraspecific differentiation in mussels. *Molecular Ecology* **25**: 269–86.

Gaitán-Espitia JD, Quintero-Galvis JF, Mesas A, D'Elía G (2016). Mitogenomics of southern hemisphere blue mussels (Bivalvia: Pteriomorphia): Insights into the evolutionary characteristics of the *Mytilus edulis* complex. *Scientific reports* **6**: 1–10.

Gallardo-Escárate C, Valenzuela-Muñoz V, Nuñez-Acuña G, Valenzuela-Miranda D, Tapia F, Yévenes M, et al. (2022). The native mussel *Mytilus chilensis* genome reveals adaptative molecular signatures facing the marine environment. *bioRxiv* 2022.09.06.506863.

Gardner JPA (1994). The *Mytilus edulis* species complex in Southwest England: Multi-locus heterozygosity, background genotype and a fitness correlate. *Biochemical Systematics and Ecology* **22**: 1–11.

Gardner JPA (1996). The *Mytilus edulis* species complex in southwest England: effects of hybridization and introgression upon interlocus associations and morphometric variation. *Marine Biology* **125**: 385–99.

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**: 3182–95.

Garrison E, Marth G (2012). Haplotype-based variant detection from short-read sequencing. arXiv preprint arXiv:1207.3907.

Gerdol M, Moreira R, Cruz F, Gómez-Garrido J, Vlasova A, Rosani U, et al. (2020). Massive gene presence-absence variation shapes an open pan-genome in the Mediterranean mussel. *Genome Biology* **21**: 275.

Gjedrem T, Rye M (2018). Selection response in fish and shellfish: a review. *Reviews in Aquaculture*, **10(1):** 168-179.

Gosling E (2015). *Marine Bilvalve Molluscs* John Wiley and Sons, Ltd. 2nd ed.

Gutierrez AP, Turner F, Gharbi K, Talbot R, Lowe NR, Peñaloza C, et al. (2017). Development of a Medium Density Combined-Species SNP Array for Pacific and European Oysters (*Crassostrea gigas* and *Ostrea edulis*). *G3 Genes|Genomes|Genetics* **7**: 2209–18.

Hammel M, Simon A, Arbiol C, Villalba A, Burioli EA V, Pépin J-F, et al. (2022). Prevalence and polymorphism of a mussel transmissible cancer in Europe. *Molecular Ecology* **31**: 736–51.

Hilbish T, Carson E, Plante J, Weaver L, Gilg M (2002). Distribution of *Mytilus edulis, M. galloprovincialis*, and their hybrids in open-coast populations of mussels in southwestern England. *Marine Biology* **140**: 137–42.

Hilborn R, Banobi J, Hall SJ, Pucylowski T, Walsworth TE (2018). The environmental cost of animal source foods. *Frontiers in Ecology and the Environment*, **16(6)**: 329-335.

Houston RD, Bean TP, Macqueen DJ, Gundappa MK, Jin YH, Jenkins TL, et al. (2020). Harnessing genomics to fast-track genetic improvement in aquaculture. *Nature Reviews Genetics*, **21(7)**: 389-409.

Houston RD, Taggart JB, Cézard T, Bekaert M, Lowe NR, Downing A, et al. (2014). Development and validation of a high density SNP genotyping array for Atlantic salmon (*Salmo salar*). *BMC genomics* **15**: 90.

Joshi R, Árnyasi M, Lien S, Gjøen HM, Alvarez AT, Kent M (2018). Development and Validation of 58K SNP-Array and High-Density Linkage Map in Nile Tilapia (*O. niloticus*). *Frontiers in genetics*, **9**: 472.

Kamermans P, Galley T, Boudry P, Fuentes J, McCombie H, Batista FM, et al. (2013). *Blue Mussel Hatchery Technology in Europe*, Vol. 2009 Woodhead Publishing Limited

Knöbel L, Nascimento-Schulze JC, Sanders T, Zeus D, Hiebenthal C, Barboza FR, et al. (2021). Salinity Driven Selection and Local Adaptation in Baltic Sea Mytilid Mussels. *Frontiers in Marine Science*, **8:** 692078.

Koehn RK (1991). The genetics and taxonomy of species in the genus *Mytilus*. *Aquaculture* **94**: 125–45.

Koehn RK, Hall JG, Innes DJ, Zera AJ (1984). Genetic differentiation of *Mytilus edulis* in eastern North America. *Marine Biology* **79**: 117–26.

Koehn RK, Newell RIE, Immermann F (1980). Maintenance of an Aminopeptidase Allele Frequency Cline by Natural Selection. *Proceedings of the National Academy of Sciences* **77**: 5385–89.

Kranis A, Gheyas AA, Boschiero C, Turner F, Yu L, Smith S, et al. (2013). Development of a high density 600K SNP genotyping array for chicken. *BMC Genomics* **14**: 59.

Lapègue S, Harrang E, Heurtebise S, Flahauw E, Donnadieu C, Gayral P, et al. (2014). Development of SNP-genotyping arrays in two shellfish species. *Molecular Ecology Resources* **14**: 820–30.

Larraín MA, González P, Pérez C, Araneda C (2019). Comparison between single and multi-locus approaches for specimen identification in *Mytilus* mussels. *Scientific Reports* **9**: 19714.

Larraín MA, Zbawicka M, Araneda C, Gardner JPA, Wenne R (2018). Native and invasive taxa on the Pacific coast of South America: Impacts on aquaculture, traceability and biodiversity of blue mussels (*Mytilus* spp.). *Evolutionary Applications* **11**: 298–311.

Li H (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**: 2987–93.

Li H, Durbin R (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**: 1754–60.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)* **25**: 2078–79.

Liu R, Xing S, Wang J, Zheng M, Cui H, Crooijmans RPMA, et al. (2019). A new chicken 55K SNP genotyping array. *BMC Genomics* **20**: 410.

Mathiesen SS, Thyrring J, Hemmer-Hansen J, Berge J, Sukhotin A, Leopold P, et al. (2017). Genetic diversity and connectivity within *Mytilus* spp. in the subarctic and Arctic. *Evolutionary Applications* **10**: 39–55.

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**: 323–33.

Metzger MJ, Villalba A, Carballal MJ, Iglesias D, Sherry J, Reinisch C, et al. (2016). Widespread transmission of independent cancer lineages within multiple bivalve species. *Nature* **534**: 705–9.

Meuwissen THE, Hayes BJ, Goddard ME (2001). Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics* **157**: 1819 LP – 1829.

Michalek K, Ventura A, Sanders T (2016). *Mytilus* hybridisation and impact on aquaculture: A minireview. *Marine Genomics* **27**: 3–7.

Michelizzi VN, Wu X, Dodson M V, Michal JJ, Zambrano-Varon J, McLean DJ, et al. (2011). A global view of 54,001 single nucleotide polymorphisms (SNPs) on the Illumina Bovine SNP50 BeadChip and their transferability to water buffalo. *International journal of biological sciences* **7**: 18.

Nascimento-Schulze JC, Bean TP, Houston RD, Santos EM, Sanders MB, Lewis C, et al. (2021). Optimizing hatchery practices for genetic improvement of marine bivalves. *Reviews in Aquaculture*, **13(4)**: 2289-2304.

Nguyen TTT, Hayes BJ, Ingram BA (2014a). Genetic parameters and response to selection in blue mussel (*Mytilus* galloprovincialis) using a SNP-based pedigree. *Aquaculture* **420–421**: 295–301.

Nguyen TTT, Hayes BJ, Ingram BA (2014b). Genetic parameters and response to selection in blue mussel (*Mytilus galloprovincialis*) using a SNP-based pedigree Potential broodstock candidates. *Aquaculture* **420–421**: 295–301.

O'Connor BD, Van Auwera G (2020). Genomics in the cloud: using Docker, Gatk, and WDL in Terra.

Paggeot LX, DeBiasse MB, Escalona M, Fairbairn C, Marimuthu MPA, Nguyen O, et al. (2022). Reference genome for the California ribbed mussel, *californianus*, an ecosystem engineer. *Journal of Heredity* esac041.

Palti Y, Gao G, Liu S, Kent MP, Lien S, Miller MR, et al. (2015). The development and characterization of a 57K single nucleotide polymorphism array for rainbow trout. *Molecular Ecology Resources* **15**: 662–72.

Peñaloza C, Robledo D, Barría A, Trịnh TQ, Mahmuddin M, Wiener P, et al. (2020). Development and Validation of an Open Access SNP Array for Nile Tilapia (*Oreochromis niloticus*). *G3 (Bethesda, Md.)* **10**: 2777–85.

Perry A, Wachowiak W, Downing A, Talbot R, Cavers S (2020). Development of a single nucleotide polymorphism array for population genomic studies in four European pine species. *Molecular Ecology Resources* **20**: 1697–1705.

Popovic I, Matias AMA, Bierne N, Riginos C (2020). Twin introductions by independent invader mussel lineages are both associated with recent admixture with a native congener in Australia. *Evolutionary Applications* **13**: 515–32.

Potts RWA, Gutierrez AP, Penaloza CS, Regan T, Bean TP, Houston RD (2021). Potential of genomic

technologies to improve disease resistance in molluscan aquaculture. *Philosophical Transactions of the Royal Society B: Biological Sciences* **376**: 20200168.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics* **81**: 559–75.

Qi H, Song K, Li C, Wang W, Li B, Li L, et al. (2017). Construction and evaluation of a high-density SNP array for the Pacific oyster (*Crassostrea gigas*). *PloS one* **12**: e0174007–e0174007.

Quesada H, Warren M, Skibinski DO (1998). Nonneutral evolution and differential mutation rate of gender-associated mitochondrial DNA lineages in the marine mussel *Mytilus*. *Genetics* **149**: 1511–26.

Rawson PD, Agrawal V, Hilbish TJ (1999). Hybridization between the blue mussels *Mytilus galloprovincialis* and *M. trossulus* along the Pacific coast of North America: evidence for limited introgression. *Marine Biology* **134**: 201–11.

Rawson PD, Hayhurst S, Vanscoyoc B (2001). Species composition of blue mussel populations in the northeastern Gulf of Maine. *Journal of Shellfish Research* **20**: 31–38.

Rawson PD, Hilbish TJ (1995). Evolutionary relationships among the male and female mitochondrial DNA lineages in the *Mytilus edulis* species complex. *Molecular Biology and Evolution* **12**: 893–901.

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

Robledo D, Palaiokostas C, Bargelloni L, Martínez P, Houston R (2018). Applications of genotyping by sequencing in aquaculture breeding and genetics. *Reviews in Aquaculture* **10**: 670–82.

Romiguier J, Gayral P, Ballenghien M, Bernard A, Cahais V, Chenuil A, et al. (2014). Comparative population genomics in animals uncovers the determinants of genetic diversity. *Nature* **515**: 261–63.

Saarman NP, Pogson GH (2015). Introgression between invasive and native blue mussels (genus *Mytilus*) in the central California hybrid zone. *Molecular Ecology* **24**: 4723–38.

Simon A (2022). Three new genome assemblies of blue mussel lineages: North and South European *Mytilus edulis* and Mediterranean *Mytilus galloprovincialis*. *bioRxiv* 2022.09.02.506387.

Simon A, Arbiol C, Nielsen EE, Couteau J, Sussarellu R, Burgeot T, et al. (2020). Replicated anthropogenic hybridisations reveal parallel patterns of admixture in marine mussels. *Evolutionary Applications* **13**: 575–99.

Simon A, Fraïsse C, El Ayari T, Liautard-Haag C, Strelkov P, Welch JJ, et al. (2021). How do species barriers decay? Concordance and local introgression in mosaic hybrid zones of mussels. *Journal of Evolutionary Biology* **34**: 208–23.

Śmietanka B, Burzyński A (2017). Complete female mitochondrial genome of *Mytilus chilensis*. *Mitochondrial DNA Part B* **2**: 101–2.

Śmietanka B, Burzyński A, Wenne R (2010). Comparative genomics of marine mussels (*Mytilus* spp.) gender associated mtDNA: rapidly evolving atp8. *Journal of Molecular Evolution* **71**: 385–400.

Śmietanka B, Wenne R, Burzyński A (2016). Complete male mitochondrial genomes of European *Mytilus edulis mussels*. *Mitochondrial DNA Part A* **27**: 1634–35.

Stoffel K, van Leeuwen H, Kozik A, Caldwell D, Ashrafi H, Cui X, et al. (2012). Development and application of a 6.5 million feature Affymetrix Genechip® for massively parallel discovery of single position polymorphisms in lettuce (*Lactuca spp.*). *BMC Genomics* **13**: 185.

Stuckas H, Knöbel L, Schade H, Breusing C, Hinrichsen HH, Bartel M, et al. (2017). Combining hydrodynamic modelling with genetics: can passive larval drift shape the genetic structure of Baltic *Mytilus* populations? *Molecular Ecology,* **26(10)**: 2765-2782.

Subasinghe R (2017). *World Aquaculture 2015 : A Brief Overview*, Vol. 1140

Team RC (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. *http://www. R-project. org/*

Toro J, Innes DJ, Thompson RJ (2004). Genetic variation among life-history stages of mussels in a *Mytilus edulis–M. trossulus* hybrid zone. *Marine Biology* **145**: 713–25.

Väinölä R, Hvilsom MM (1991). Genetic divergence and a hybrid zone between Baltic and North Sea *Mytilus* populations (Mytilidae: Mollusca). *Biological Journal of the Linnean Society,* **43(2)**: 127-148.

van der Schatte Olivier A, Jones L, Vay L Le, Christie M, Wilson J, Malham SK (2020). A global review of the ecosystem services provided by bivalve aquaculture. *Reviews in Aquaculture* **12**: 3–25.

Vassilenko EI, Muttray AF, Schulte PM, Baldwin SA (2010). Variations in p53-like cDNA sequence are correlated with mussel haemic neoplasia: A potential molecular-level tool for biomonitoring. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis* **701**: 145–52.

Vendrami DLJ, De Noia M, Telesca L, Brodte E-M, Hoffman JI (2020). Genome-wide insights into introgression and its consequences for genome-wide heterozygosity in the *Mytilus* species complex across Europe. *Evolutionary Applications*, **13(8):** 2130-2142.

Wenne R, Bach L, Zbawicka M, Strand J, McDonald JH (2016). A first report on coexistence and hybridization of *Mytilus trossulus* and *M. edulis* mussels in Greenland. *Polar Biology* **39**: 343–55.

Wenne R, Zbawicka M, Bach L, Strelkov P, Gantsevich M, Kukliński P, et al. (2020). Trans-Atlantic distribution and introgression as inferred from single nucleotide polymorphism: Mussels *Mytilus* and environmental factors. *Genes* **11**: 530.

Wijesena HR, Rohrer GA, Nonneman DJ, Keel BN, Petersen JL, Kachman SD, et al. (2019). Evaluation of genotype quality parameters for SowPro90, a new genotyping array for swine1. *Journal of Animal Science* **97**: 3262–73.

Wilson J, Matejusova I, McIntosh RE, Carboni S, Bekaert M (2018). New diagnostic SNP molecular markers for the *Mytilus* species complex. *PLoS ONE* **13**: 1–13.

Yang J-L, Feng D-D, Liu J, Xu J-K, Chen K, Li Y-F, et al. (2021). Chromosome-level genome assembly of the hard-shelled mussel *Mytilus coruscus*, a widely distributed species from the temperate areas of East Asia. *GigaScience* **10**: giab024.

Yonemitsu MA, Giersch RM, Polo-Prieto M, Hammel M, Simon A, Cremonte F, et al. (2019). A single clonal lineage of transmissible cancer identified in two marine mussel species in South America and Europe. *eLife* **8**: e47788.

Zbawicka M, Sańko T, Strand J, Roman W (2014). New SNP markers reveal largely concordant clinal variation across the hybrid zone between *Mytilus* spp. in the Baltic Sea. *Aquatic Biology* **21**: 25–36.

Zbawicka M, Trucco MI, Wenne R (2018). Single nucleotide polymorphisms in native South American Atlantic coast populations of smooth shelled mussels: hybridization with invasive European *Mytilus galloprovincialis*. *Genetics Selection Evolution* **50**: 5.

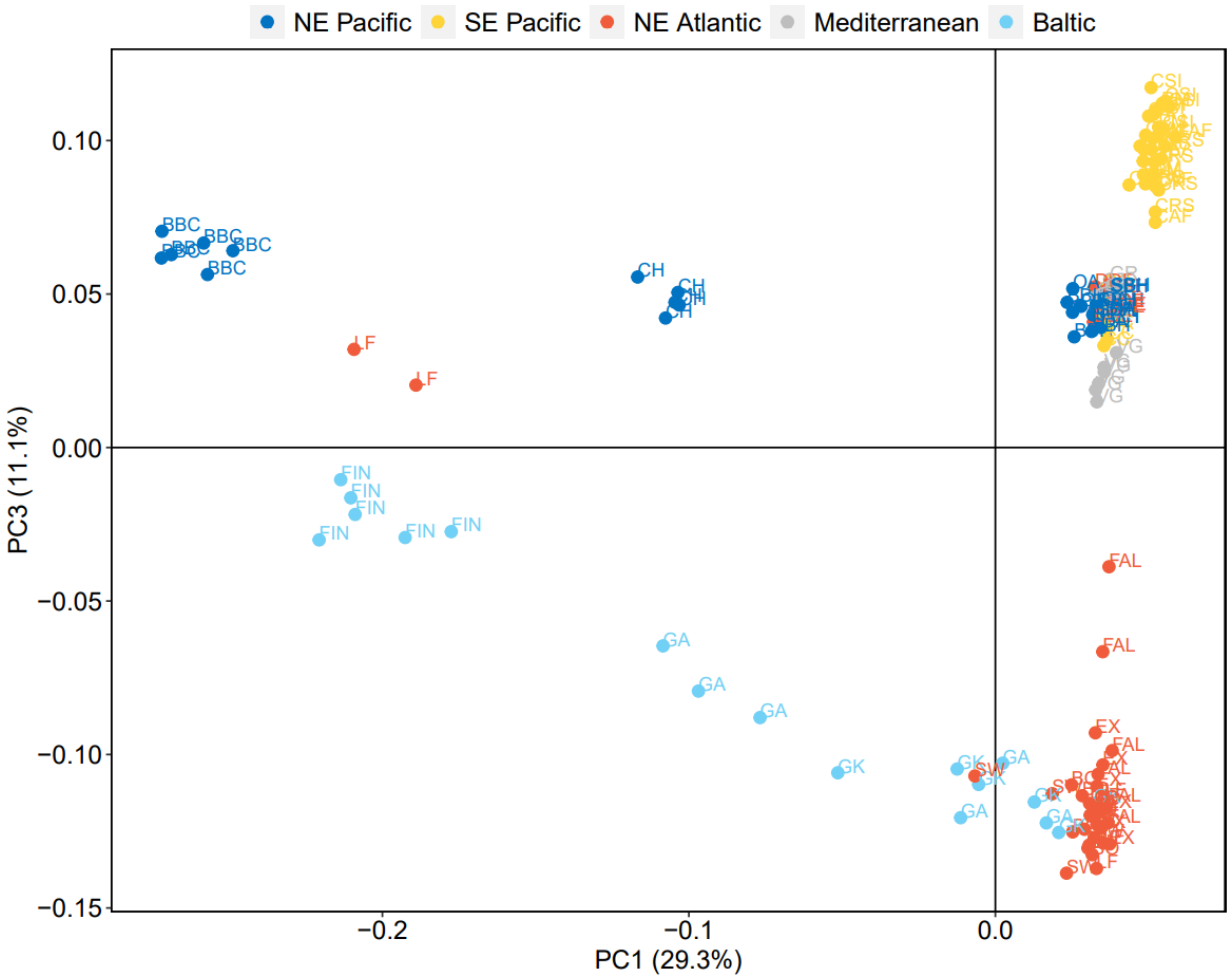
Zouros E, Oberhauser Ball A, Saavedra C, Freeman KR (1994). An unusual type of mitochondrial DNA inheritance in the blue mussel *Mytilus*. *Proceedings of the National Academy of Sciences of the United States of America* **91**: 7463–67.

Supplementary material

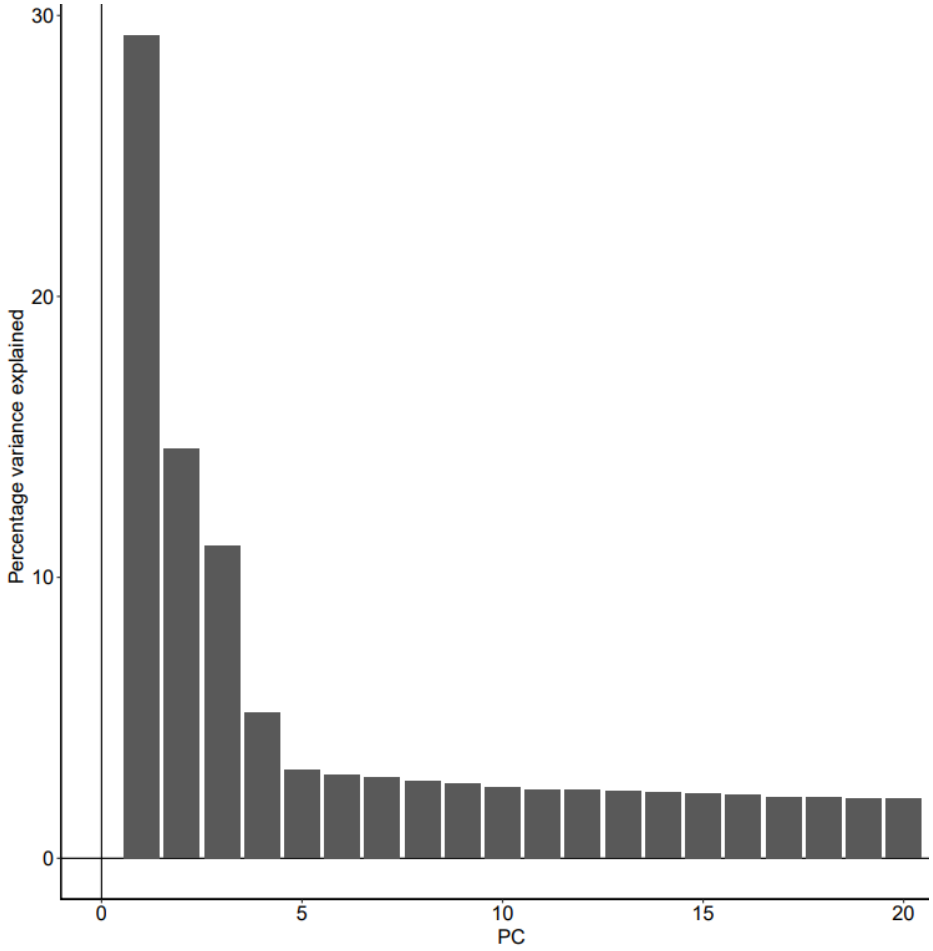
Supplementary Table 2.1: Summary of SNP quality assessment from Axiom Analysis Suite Software using different group of samples: all discovery population samples and samples from the four different Blue Mussel species used in the array. No QC and CR filters have been applied to these datasets.

Samples (DNA)	(n)	SNP conversion	SNPs (n, per category)	SNPs (Percentage)
Discovery populations	127	Other	50,017	83.1
		OTV	5,609	9.3
		PolyHighResolution	2,029	3.3
		NoMinorHom	1,785	3.0
		MonoHighResolution	702	1.2
		BestandRecommended	4,516	7.5
M. galloprovincialis	30	Other	38,777	64.5
		OTV	12,053	20.0
		PolyHighResolution	3,524	5.9
		NoMinorHom	2,873	4.8
		MonoHighResolution	2,915	4.8
		BestandRecommended	9,312	15.5
M. edulis	21	Other	42,590	70.8
		OTV	10,610	17.6
		PolyHighResolution	2,291	3.8
		NoMinorHom	3,138	5.2
		MonoHighResolution	1,512	2.5
		BestandRecommended	6,941	11.5
M. trossulus	14	Other	43,422	72.2
		OTV	9,937	16.5
		PolyHighResolution	1,361	2.3
		NoMinorHom	3,873	6.4
		MonoHighResolution	1,549	2.6
		BestandRecommended	6,783	11.3
M. chilensis	23	Other	43,261	71.9
		OTV	8,540	14.2
		PolyHighResolution	2,377	5.0
		NoMinorHom	2,955	4.9

MonoHighResolution	3,008	5.0
BestandRecommended	8,340	13.9



Supplementary Figure 2.1: Scatter plots of individual variation in PC 1 and 3 scores resulting from PCA. The amount of variation explained by each PC are given in percentages.



Supplementary Figure 2.2: Bar plot with the percentage of variance explained by each PC (1-20) resulting from the PCA. 2-7

3. Chapter III - Revisiting the Southwest England Blue Mussel Hybrid Zone Using a Genomic Approach

This chapter resulted from a collaborative work of myself, Charlie D. Ellis, Juliane Friedrich, Louisa Williams, Pamela Wiener, Tim P. Bean and Robert P. Ellis. I was responsible for designing the study, sampling and analysing the data as well as writing the manuscript.

Abstract

Blue mussels from the *Mytilus* genus possess a huge economic and ecological value, acting as a foundation species to the benthic environment. Species within this Blue mussel species complex (*M. edulis, M. galloprovincialis* and *M. trossulus*) easily hybridise wherever their distribution range overlaps, forming extensive hybrid zones. In the southwest (SW) coast of England, UK, two species of blue mussels, *M. galloprovincialis* and *M. edulis*, co-exists and hybridise. The distribution of genotypes in this region was previously described by analysing polymorphisms in few DNA nuclear markers. Here we utilised the recently developed blue mussel multi-species 60 K SNP-array to reassess genotype distribution across this region. Mussels were sampled from 29 intertidal locations across coastal Cornwall and Devon, UK, an Atlantic peninsula characterised by an exposed northern coast and more sheltered southern coast. A total of 16,732 SNPs were retained for genotyping. Our results suggest spatial stability of the hybrid zone by supporting the previously described distribution of *Mytilus* species, with *M. galloprovincialis* genotypes dominating the north coast of Cornwall, whilst *M. edulis* genotypes are more prevalent along the south coasts of Cornwall and Devon. Results from our hybrid index (h-index) analysis point towards the Lizard peninsula as the centre of the transition zone, where a shift in dominance between the two species occurs. The genotypic dominance of each species appears correlated to general exposure to wave action, warranting further investigation to understand the genotype-by-environment interactions and other geographical, biotic and abiotic factors modulating distribution of genotypes within this hybrid zone. Nonetheless, this is the first in-depth genomic characterisation performed in SW UK mussels, and can contribute to advancing our understanding of speciation in the region, as well as optimisation in aquaculture, providing farmers crucial knowledge on genetic background of mariculture stock and to the development and implementation of conservation efforts of these species.

Key-words: Hybrid zones, blue mussels, genotyping, SNPs, South West England

Background

Mussels of the genus *Mytilus* are a foundation species, found within the intertidal zone in temperate regions of the globe (Larraín *et al.*, 2019). They provide a number of essential ecosystem services including offering complex habitat structure and shelter for other species, nutrient cycling, and being an important component of the food chain (van der Schatte *et al.*, 2020). Species in this complex easily interbreed in areas of overlapping distribution, and hybrid zones have been extensively investigated (e.g. Bierne *et al.*, 2003; Daguin *et al.*, 2001; Gardner 1994a; Hilbish *et al.*, 2002; Miranda *et al.*, 2010; Rawson *et al.*, 1999; Riginos and Cunningham, 2005; Saarman and Pogson, 2015; Stuckas *et al.*, 2017; Zbawicka *et al.*, 2018). The complexities around genetic contribution of various parent populations to hybrid offspring, and the reasons for maintenance of species boundaries in spite of long standing genetic introgression, have not yet been fully resolved. Hybrid zones are formed between species in the *Mytilus* complex. In the Northern hemisphere this includes *M. edulis*, *M. galloprovinciallis* and *M. trossulus* (Bierne *et al.*, 2003; Daguin *et al.*, 2001; Gardner 1994a; Gardner *et al.*, 2016; Koehn *et al.*, 1984; Larraín *et al.*, 2018; Rawson *et al.*, 1999; Riginos and Cunningham, 2005; Skibinski *et al.*, 2008; Stuckas *et al.*, 2017, 2009; Zbawicka *et al.*, 2014). In the southwest (SW) of the UK an extensive hybrid zone of *M. galloprovincialis* and *M. edulis* has been identified between St Ives (northwest Cornwall) and Thurleston (southeast Devon) (Hilbish et al, 2002). Within this mussel hybrid zone, *M. galloprovincialis* populations have been described as dominating turbulent waters, whilst *M. edulis* are predominant in sheltered, lower-salinity habitats (Schmidt *et al.*, 2008). Moreover, in areas such as in Whitsand Bay, the frequency of *M. edulis* alleles present in populations appears to be negatively correlated with the age of individuals (Wilhelm and Hilbish, 1998). For the three markers used in this previous study, 51.3% of smaller individuals were homozygous for *M. edulis* alleles, dropping to 22.5% for larger animals. Wilhelm and Hilbish (1998) suggest this may relate to selection against these alleles, although more work was required to confirm this hypothesis. Subsequently, further analysis of DNA nuclear markers confirmed that *M. galloprovincialis* alleles at the Glu-5' locus were missing beyond the southeast limit of the hybrid zone (east of Start Point, Devon) (Hilbish, T. *et*

al., 2002). Nevertheless, more recent genotyping data generated by restricted site associated DNA sequencing approaches revealed the presence of *M. galloprovinciallis* alleles in mussel populations eastwards from this site (Vendrami *et al.*, 2020). These findings suggest that our understanding of species distribution in this area is, therefore, incomplete and/or outdated.

For sessile organisms, post-settlement processes such as phenotype-environment mismatches (i.e. non-random genotype mortality during post settlement development) act as biological barriers to connectivity (see review by Marshall *et al.*, 2010). The exact limits of intra- and inter- species distributions, therefore, depend on multiple conflicting environmental factors, which may themselves exert different pressures at each life stage. The natural geographical limits of species provide an ideal situation to study the interaction between environment and genetics in situ. As such, understanding the genetic structure of populations in these areas can provide clear indication as to the selective pressure that governs success of an individual. Further, observing changes in the distribution of genotypes can then help understand whether the environment is changing, or the species is adapting to allow new niches to become habitable.

Marine mussels are an important species to the aquaculture and fisheries sectors, being the most widely cultivated marine bivalve following Pacific oysters. Annual production of these taxa reached approximately 2 million tonnes globally in 2016, from which aquaculture is responsible for over 90% of landings (Avdelas *et al.*, 2021). Mussel farming is one of the most sustainable food production sectors (Hilborn *et al.*, 2018), delivering a highly nutritious product (Carboni *et al.*, 2019) with a lower ecological impact than most seafood, and among the lowest carbon footprint of any major animal protein (i.e. 0.02 - 0.1 CO2/kg; Seafood Carbon Emissions Tool, 2022). In 2016, the UK produced over 14,000 tonnes of blue mussels (Avdelas *et al.*, 2021), with the SW coast of England representing a key site for shellfish farming nationally (Ellis *et al.*, 2015; Laing and Spencer, 2006). However, between 2008 and 2016, UK blue mussel production has fallen by 61%, a much steeper decline compared to the decline in production observed across the rest of Europe in the same period (20%) (Avdelas *et al.*, 2021). This scenario has been linked to a number of environmental issues (e.g.

water quality, algal blooms, hybridisation with *M. trossulus*) and production factors (lack of adequate depuration facilities, product price, global market competition), and in some areas to seed availability (see Avdelas *et al.*, 2021 and references therein).

Despite the great economic and environmental relevance of these taxa, little is known about the genetic composition of populations being cultured in the UK, or how genetic diversity might impact production. Whilst Britain's largest offshore mussel farm is located in the SW UK region, off the south Devon coast, little effort has been afforded to the investigation of the genetic background of the cultured stock in this region. Consequently, little information on any ongoing genotype-by-environment (*G*x*E*) interactions is available for these taxa. Even less is known concerning if, and how, recent environmental changes and shifts in production have influenced the local patterns of species distribution in this region. Characterising the genetic background of mussel populations is crucial to provide clarity around a number of key questions, including i) providing a clear picture of the species distribution and which species are being cultured in farms and ii) advancing our understanding of the GxE interactions ongoing in the different habitats (e.g. rocky intertidal zone, estuaries). Such knowledge will facilitate the development of selective breeding approaches to these taxa. It will also contribute to development of management practices that mitigate declines in wild populations through restoration programmes. Most importantly, understanding the genetic composition and diversity of mussel populations is crucial to ensure the sustainable optimisation of this key food sector. Understanding why specific genotypes outperform others in different environments will allow farmers to select an appropriate grow-out site to their stock, which will favour commercially relevant traits (e.g. growth and survival), consequently enhancing production. Such advances can be fast-tracked with the aid of genomic tools (Houston *et al.*, 2020). Thus, there is an imminent need to increase efforts in characterising the genetic background of wild and commercial bivalve populations in the UK and globally (Nascimento-Schulze *et al.*, 2021).

In this study, our first aim was to assess the distribution of species (*M. galloprovincialis*, *M. edulis* and their hybrids) in the blue mussel hybrid zone located along the SW coast of the UK (Figure

3.1). For this, we sampled blue mussels at 29 intertidal locations along the SW UK coast and analysed individuals using a recently developed genomic tool, the blue mussel 60K SNP array (Nascimento-Schulze in press), updating the previous description of species distribution in this hybrid zone (Hilbish *et al.*, 2002). Our second aim was to determine whether distribution of genotypes in this species complex is related to habitat preferences. For this, we sampled populations from outside and within three estuaries around the SW peninsula, as well as from different tidal heights (high water mark, spring low water mark) at three separate locations across the sampling range. Understanding species dynamics across these gradients (time, tidal height and estuarine) is crucial to shed light on ongoing speciation and evolutionary processes in this species complex within these microhabitats.

Methods

Sample collection and processing

During spring tides of October-December 2021, mussels were sampled from 29 intertidal locations on the southwest peninsula of England. Samples were collected from the intertidal zone, within a 0-3h window around the time of the lowest tide-height. Three biometric parameters were recorded from 20 individual mussels per population: shell length (SL), shell height (SH) and total weight (TW). Where possible, samples were restricted to individuals with SL > 15 mm, which was only not possible for samples from Hallsands (Suppl. Fig. 3.1).

Site selection

In this study, we focused on mussel populations from the southwest peninsula of England, which constitute a *M. edulis* X *M. galloprovincialis* hybrid zone. Sampling sites chosen in this study were largely inspired by a previous study describing this area (Hilbish *et al.*, 2002). Nonetheless, in some locations mussel populations were no longer found and nearby corresponding sites were sampled as an alternative. The sampling locations are Budleigh Salterton (BDL), Exmouth (EXM), Starcross (STC), Torquay (TRQ), Hallsands (HLS), Millbay Marina (MIL), Whitsand Bay (WTS), Polkeris (PLK), Pentewan (PTW), Porthluney (PLU), Feock (FEO), St Anthony's Head (STA), Helford (HEL),

Porthleven (PLEV), Mousehole (MHO), Porthmeor (PME), Godrevy (GOD), St Agnes (STAGN), Newquay (NWQ), Porthcotan Bay (PCT), Trevone (TRE), Padstow/Camel (PAD), Port Gaverne (PORTGAV), Tintangel Castle (TIC), Bude (BUD)We sampled three estuarine populations: Starcross (River Exe, South Devon), Feock (River Fal, South Cornwall) and Padstow (River Camel, North Cornwall), as well as sampling individuals residing both within the high tide (individuals sampled close to the high tide mark) and low tide zones (nearest individuals to low tide mark during spring tides) for three locations: Exmouth, Whitsand Bay and Port Gaverne, resulting in 29 interdental populations (Figure 3.1).

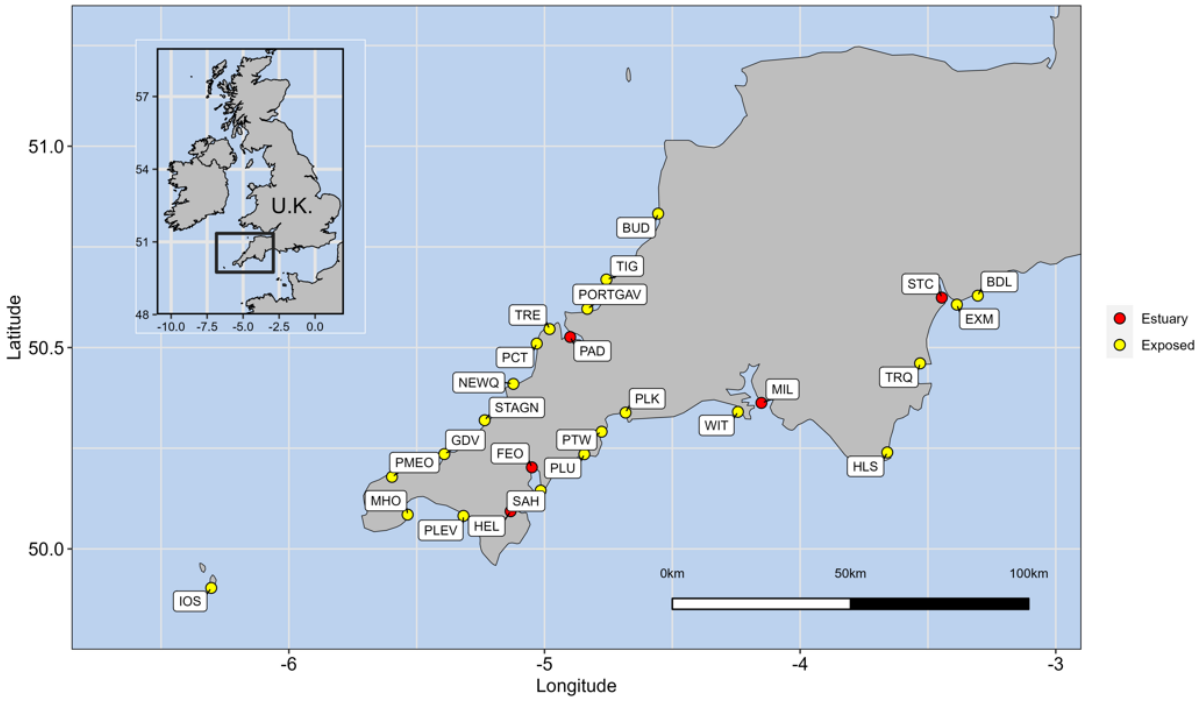


Figure 3-1: Locations used for sampling mussels in the SW coast of England. Three populations: Port Gaverne, Exmouth and Whitsand Bay contain low tide and high tide representatives, resulting 29 locations.

DNA extraction and genotyping

High quality genomic DNA was extracted from 12 individuals from each of the 29 populations. For each individual, 30 mg of adductor muscle tissue was stored in 98% ethanol and sent to Identigen (Identigen, Ireland, UK) for DNA extraction following a standard protocol. All samples were genotyped with the blue mussel multi species 60K array developed by Affymetrix (Thermo Fisher Scientific), described in chapter II of the thesis (Nascimento-Schulze *et al.* in press).

SNP filtering

Genotypes of the 348 individuals from the 29 sampling locations were analysed using the Axiom Analysis Suit software (Affymetrix, Thermo Fisher Scientific). Markers with a call rate (CR) below 95% were excluded from further analyses. In addition, markers with a minor allele frequency (MAF) < 0.01, and individuals with a CR < 90% were excluded using 'Plink' v1.9 (Purcell *et al.* 2007). We removed loci made monomorphic following the filtering of individuals by applying the 'isPoly' function from the R package 'adegenet' v2.1.3 (Jombart 2008).

Investigating Population Structure and Ancestry in the Southwest hybrid zone

To explore introgression present in individuals collected across the 29 sampling locations, we conducted admixture analysis (Admixture v1.3) (Alexander & Lange 2011), using the default settings. In this analysis, the best fitting number of *k* clusters that represents the ancestry of a population is represented by the *k* with the lowest cross-validation error. Seven additional populations of known species composition were included in the study (Nascimento-Schulze *et al.*, in press). These additional samples included individuals known to be Mediterranean *M. galloprovincialis* (from Delta del Ebro, Spain (DDE) and Eleutherón, Nea Peramos, Greece (GR)), Atlantic *M. galloprovincialis* (from Vigo, Spain, (VG)), Atlantic *M. edulis* (from Straumsvik, Iceland (IC) and Bodø, Norway (BO)) and *M. trossulus* from both the northeast Pacific (Bodega Bay Coast, California) and the Baltic Sea (Tvärminne, Finland (FIN)). These positive control samples were included to support species determination of SW UK derived samples. We tested *k* values between 2 and 10. Following evaluation of cross-validation error for each tested *k*, the most likely number of clusters in our dataset was *k*=3 (Suppl. Table 3.1, Suppl. Fig. 3.2). This analysis allowed us to infer the species ancestry from SW England populations, further supporting results from Hilbish *et al.* (2002).

We explored sub-structuring and divergence within genetic clusters, using discriminant analysis of principal components (DAPC) (Jombart *et al.*, 2010). This analysis was executed using the adegenet v2.1.1 package (Jombart & Ahmed, 2011) in R v4.2.1 (R Core Team, 2022). Files were transformed to the genind format using vcfR (Knaus & Grünwald 2017). To choose the optimal number

of principal components (P-axes) to retain for DAPC models, we applied the 'adegenet' cross-validation function 'xvalDapc' across 500 replicates. We limited the maximum number of P-axes to one less than the number of sampling sites, to mitigate potential overplotting and subsequent overestimation of spatial structuring in the tested populations (Thia 2022). Visualisation of DAPC model results were plotted using the R package ggplot2 v4.0.0 (Wickham, 2016).

Differences between ecotypes (i.e. low vs high tide, estuarine vs. coastal) were investigated with a principal component analysis (PCA) (Plink v1.9) (Purcell *et al.* 2007) and Basic Statistics of Genetic Variation, calculated with the adegenet (Jombart, 2008) and hierFstat (Goudet, 2005) R packages. We used 6 individual analysis, three for estuarine vs. coastal populations (Exmouth vs Starcross, St Anthony's head vs. Feock and Hellford and Trevone vs Camel), and three for high vs low tide populations (Exmouth HT vs Exmouth LT, Port Gaverne HT vs Port Gaverne LT and Whitsand Bay HT vs Whitsand Bay HT). Basic statistics estimated included gene diversity (D_{ST}) and corrected gene diversity (DSTP) among individuals, the overall gene diversity (H_T) and corrected gene diversity (H_{TP}) among populations, the fixation index (F_{ST}) and corrected (F_{STP}) based on population, and the inbreeding coefficient (F_{IS}) per overall loci. The observed heterozygosity (H_o) and genetic diversity (H_s) within population. F_{ST} and corrected F_{ST} (F_{STP}). Whereas for neutral SNP, a low degree of genetic variation is expected, a high genetic divergence is observed in SNP subject to directional selection.

Sample grouping for hybrid index analysis

Hybrid index analysis are performed by grouping samples into two divergent ancestral populations (i.e. source populations S0 and S1), whilst the hybrids between these ancestral populations are test populations. In our study, we used genotyping results from the Admixture analysis using k = 2, allocating individuals with an ancestry proportion higher than 0.99 for one of the two clusters to either S0 or S1 populations. All other samples (hybrids) were grouped as test samples (Suppl. Table 3.2). In addition, for this analysis we grouped together all samples from low tide (LT) and high tide (HT) zones from the same geographical location.

Hybrid index analysis

H-index was estimated using the R package 'gghybrid' (Richard Bailey, 2022), applying the function 'esth'. Variants with the greatest allele frequency difference between individuals were removed using the 'data.prep' function, and those overlapping 95% credible intervals for allele frequencies (AF) in S0 and S1, with the option "AF.CIoverlap=FALSE", were kept to increase the probability that AF differences between ancestral populations were true. In order to reduce computational requirements, h-index estimation was run using a burn-in of 1000 iterations ('burnit') and 2000 subsequent iterations ('nitt'). Filtering of false discovery rates was done by testing the widely applicable information criterion (WAIC) onto the singular statistical model.

Results

Fine scale genetic divergence using discriminant analysis of principal components (DAPC)

DAPC analysis was first run with genotypes from all 16,732 SNPs across all 346 (excluding the two samples with a CR < 90%, supplementary table 3.1) individuals from the 29 sampled populations.

Table 3-1: Individuals removed from admixture analysis given their sample call rate (CR) was lower than 90%

Sample name	Sample CR (%)
Porthleven_3	89.01
Bude_11	84.87

The first Linear Discriminant axis (LD1) accounted for 77.8% of the total variation, whilst LD2 and LD3 explained 11.3% and 2.7% of the variation, respectively. Along the LD1 axis, we observed 3 clusters of samples (Figure 3.2). Individuals from the southern coasts of Devon and Cornwall were clearly separated from those of the northern Cornish coast. Individuals from locations in western Cornwall, which connects these two coasts as the tip of the peninsula, typically formed an intermediate cluster between those of the northern and southern coasts, and most were further

separated from both along LD2. The clustering arrangement was typical of that expected from two species (the distinct north and south coast populations) and an intermediate zone of hybridisation (the west coast population) along an open coastal environment.

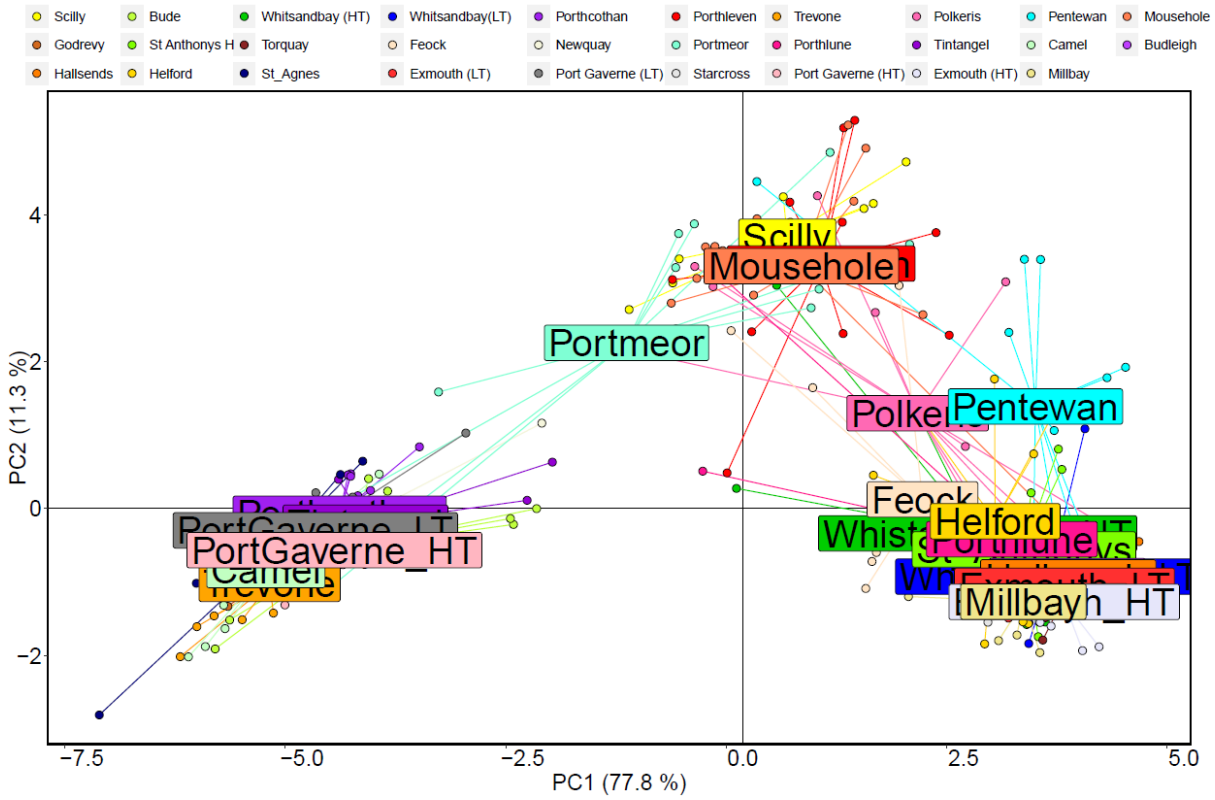


Figure 3-2: DAPC plot of linear discriminant arrangement of individuals (dots) and sample means (labels), both coloured as per the key at the head of the figure, using 6 populations, three from high tide zones (label ending with "(HT)") and three from low tide zone (label ending with "(LT)"). This analysis was generated using the 16,732 markers kept after quality control filtering.

We used PCA analysis to identify differences in genetic composition between high tide and low tide mussel populations (Figures 3.3, 3.4 and 3.5). Port Gaverne is located in the north coast of Devon, whilst Exmouth and Whitsand Bay are populations from the south coast of Devon. No clear segregation between low and high tide mussel populations was observed in any of the PCA's performed in this study. Furthermore, the variance explained by each PC was generally low, not surpassing 11.8% (PC1, Figure 3.3). Sampled populations comprise of hybrids between the two mussel

species from this area. Therefore, outlier individuals are likely a hybrid with higher ancestry proportion from the less abundant species in the sampling location.

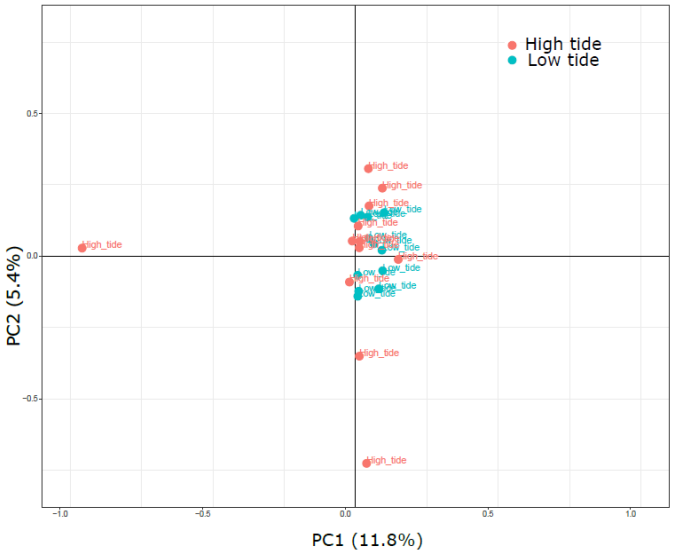


Figure 3-3: Scatter plots of individual variation in PC 1 and 2 scores resulting from PCA applied to the SNP-array genotyping dataset using SNPs with a CR > 95% and MAF < 0.01 including individuals with a CR > 90%. This analysis compared intertidal Exmouth mussels from the high tide (label ending with "(HT)") zone with individuals collected from the low tide (label ending with "(LT)").

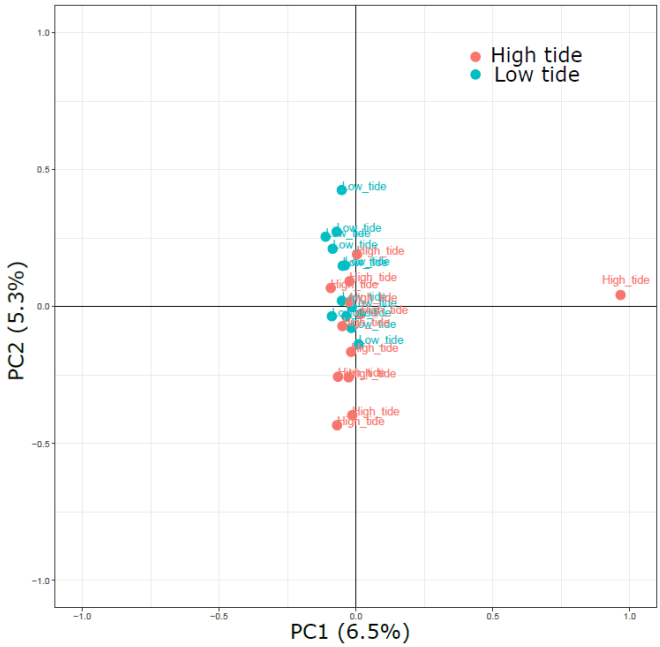


Figure 3-4: Scatter plots of individual variation in PC 1 and 2 scores resulting from PCA applied to the SNP-array genotyping dataset using SNPs with a CR > 95% and MAF < 0.01 including individuals with a CR > 90%. This analysis compared intertidal Port Gaverne mussels from the high tide (label ending with "(HT)") zone with individuals collected from the low tide (label ending with "(LT)").

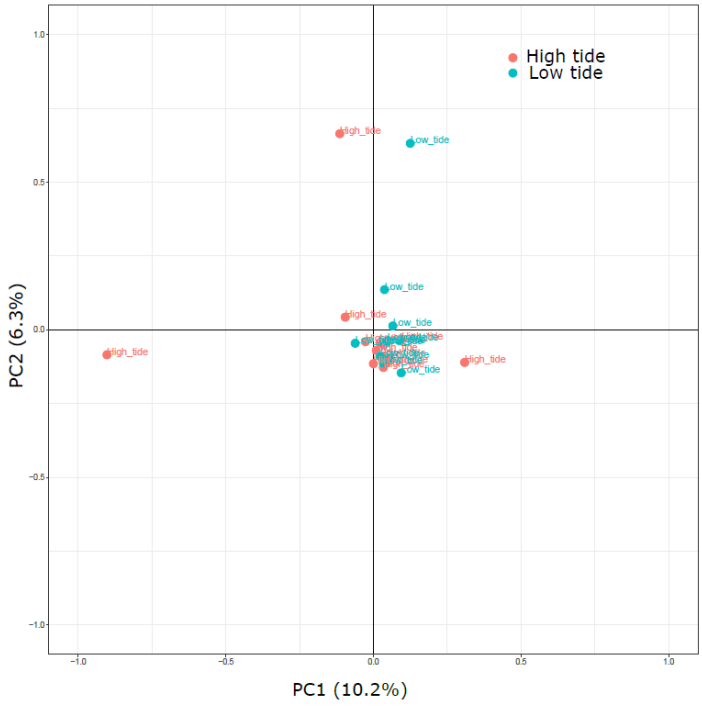


Figure 3-5: Scatter plots of individual variation in PC 1 and 2 scores resulting from PCA applied to the SNP-array genotyping dataset using SNPs with a CR > 95% and MAF < 0.01 including individuals with a CR > 90%. This analysis compared intertidal Whitsand Bay mussels from the high tide (label ending with "(HT)") zone with individuals collected from the low tide (label ending with "(LT)").

Principal component analysis was used to further investigate if estuarine populations, Feock and Helford (River Fall, South Cornwall), Starcross (River Exe, South Devon) and Camel (River Camel, North Cornwall), were genetically distinct from their reciprocal coastal populations, St Anthony's Head, Exmouth and Trevone, respectively (Figure 3.6, 3.7 and 3.8). No clear clustering was observed between coastal (Exmouth) and estuarine (Starcross) populations in the Ex river, with PC1 and PC2 explaining 13.1 and 5.5%, respectively (Figure 3.5).

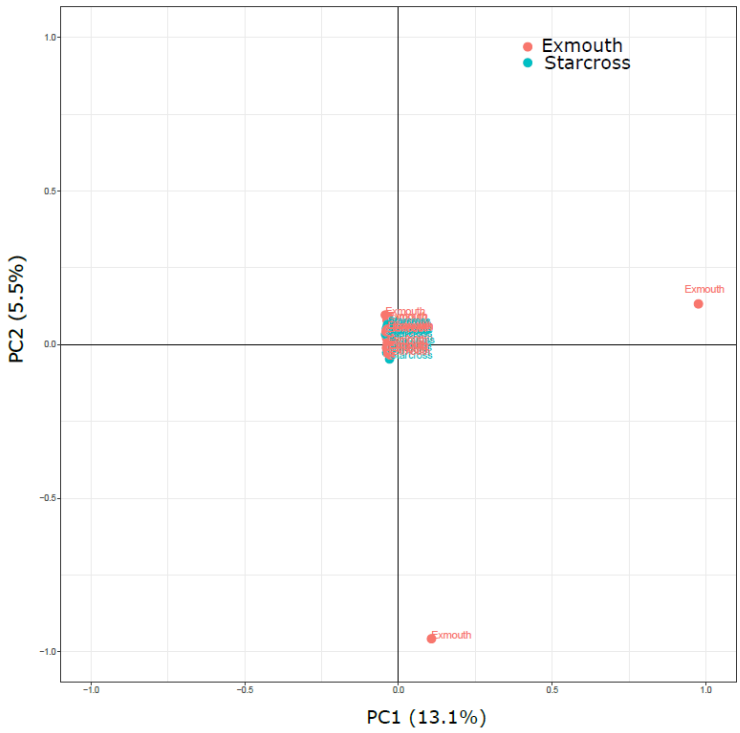


Figure 3-6: Scatter plots of individual variation in PC 1 and 2 scores resulting from PCA applied to the SNP-array genotyping dataset using SNPs with a CR > 95% and MAF < 0.01 including individuals with a CR > 90%. This analysis compared coastal mussels from Exmouth to estuarine mussels from Starcross.

Samples in the Fal and Helford estuary show a slight clustering apart from the coastal population of St Anthony's Head population, but most samples from the three population still cluster together in the PCA (Figure 3.5). Furthermore, estuarine mussels from Feock and Helford partially segregate from each other. Nonetheless, the amount of variance explained by PC1 and 2, 8,6% and 6,4%, respectively, is low.

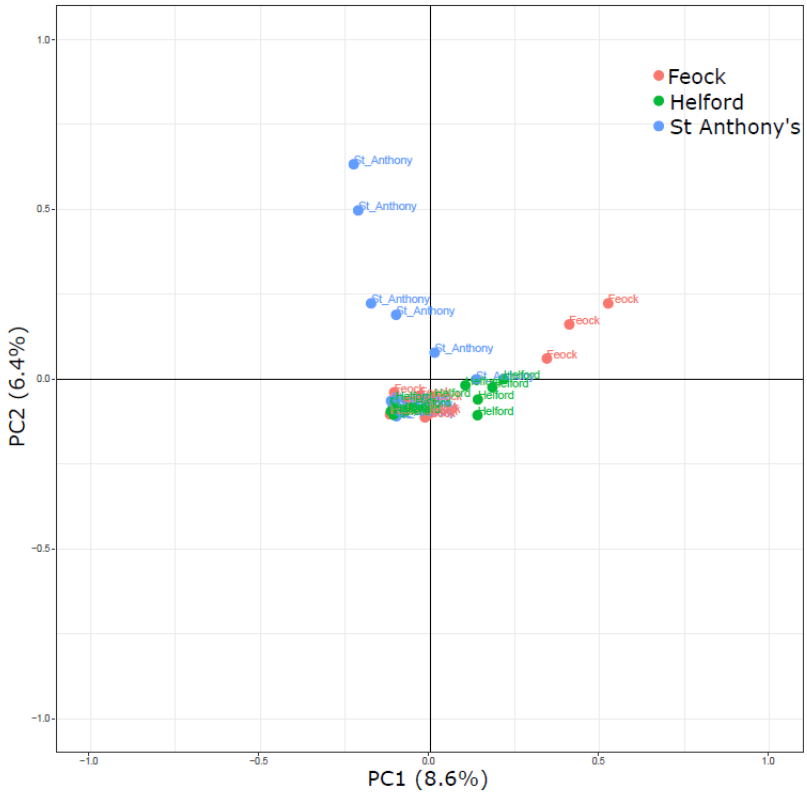


Figure 3-7: Scatter plots of individual variation in PC 1 and 2 scores resulting from PCA applied to the SNP-array genotyping dataset using SNPs with a CR > 95% and MAF < 0.01 including individuals with a CR > 90%. This analysis compared coastal mussels from St Anthony's Head to estuarine mussels from Helford and Feock.

Finally, PCA comparing the estuarine population from Camel to the coastal Trevone provided weak results; PC1 and 2, both, explained 5.5% of the variance (Figure 3.6). However, it is noticeable that Camel individuals grouped together, whilst Trevone mussels spread across the PC1 axis.

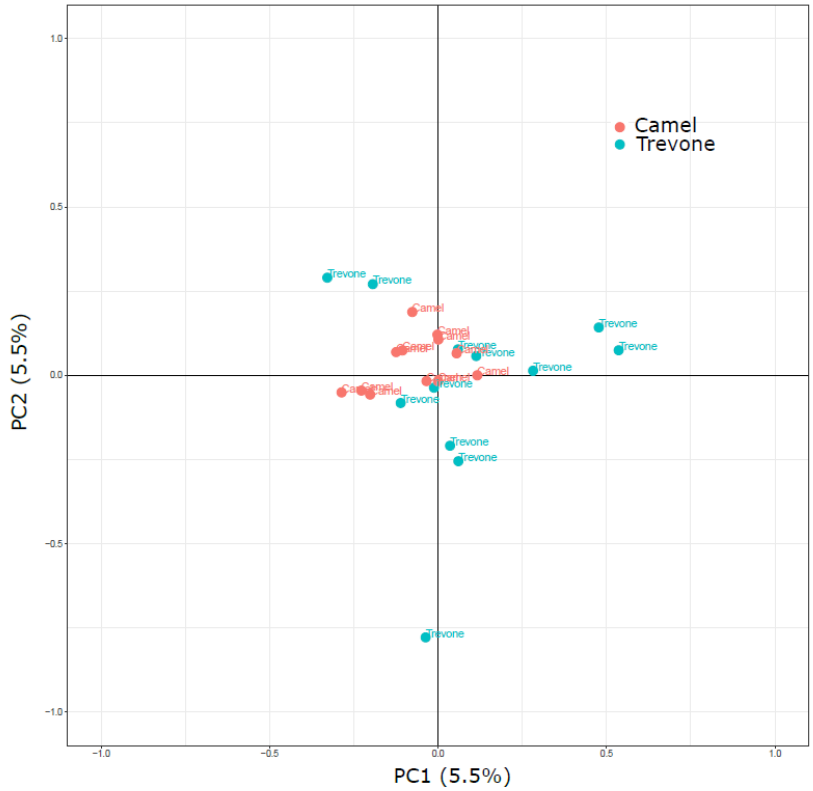


Figure 3-8: Scatter plots of individual variation in PC 1 and 2 scores resulting from PCA applied to the SNP-array genotyping dataset using SNPs with a CR > 95% and MAF < 0.01 including individuals with a CR > 90%. This analysis compared coastal mussels from St Anthony's Head to estuarine mussels from Helford and Feock.

Genetic diversity was analysed between the populations sampled in contrasting environments (Table 3.2). The observed heterozygosity within the populations (H_T) was between 0.22 and 0.24, whilst heterozygosity within compared populations ranged from 0.2 to 0.21 in all comparisons. Furthermore, in all comparisons, low values of fixation index (F_{ST}), and diversity among individuals (D_{ST}), indicating that the populations were observed, indicating that there was no significant differentiation between the populations used in the comparison. The negative inbreeding coefficient (F_{IS}) observed in all comparisons indicate avoidance of inbreeding.

Table 3-2: Genetic diversity statistics of each of the pairwise comparison data

Population	Habitat comparison	H_O	H_S	H_T	D_{ST}	H_{TP}	D_{STP}	F_{ST}	F_{STP}	F_{IS}	D_{EST}
Exmouth	High tide / Low tide	0.24	0.20	0.20	0.00	0.20	0.00	0.00	0.00	-0.20	0.00
Port Gaverne	High tide / Low tide	0.22	0.20	0.20	0.00	0.20	0.00	0.00	0.00	-0.09	0.00
Whitsand Bay	High tide / Low tide	0.24	0.20	0.20	0.00	0.20	0.00	0.00	0.01	-0.18	0.00
Exmouth (LT/HT), Starcross	Estuary / Coastal	0.24	0.20	0.20	0.00	0.20	0.00	0.00	0.00	-0.21	0.00
Trevon, Camel	Estuary / Coastal	0.22	0.20	0.20	0.00	0.20	0.00	0.00	0.00	-0.09	0.00
St Anthony's, Feock, Helford	Estuary / Coastal	0.24	0.21	0.21	0.00	0.21	0.00	0.00	0.01	-0.15	0.00

Ancestry in mussel populations along the SW coast of England

Ancestry estimation was explored across the 29 populations using Admixture software (Alexander and Lange, 2011). A total of 16,732 markers were kept for genotyping after applying a sample CR ≥ 90 %, marker CR of ≥ 95% and a MAF > 0.01 in all individuals genotyped by the AxaS software. Following evaluation of cross-validation error for each tested value of *k*, the most likely number of clusters in our SW genotyping dataset was *k* = 2 (Table 3.3). Results from the admixture coefficients (Q) inferring the two clusters are presented in Figure 3.9.

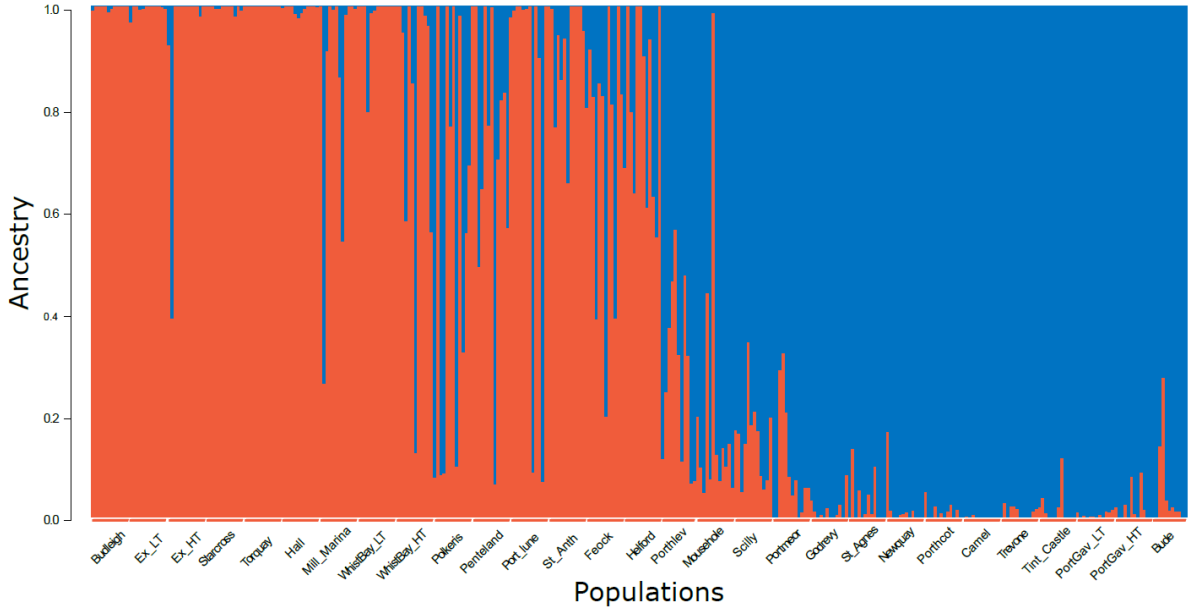


Figure 3-9: Results of genetic admixture analysis and ancestry inference. Cluster membership coefficients (Q = 2) where each individual is represented by a column partitioned into segments of different colour, the lengths of which indicate the estimated proportion of membership to each cluster. Solid bars represent an individual from a single species background.

Table 3-3: Cross validation errors generated by admixture analysis to infer the best number of ancestral populations in dataset. In this analysis, we tested k's from a 2 – 10 range. The k with the lower cross validation error is considered the most representative number of ancestral populations for a given dataset.

K	cross validation error
2	0.33704
3	**0.33921**
4	**0.34022**
5	**0.34662**
6	**0.35158**
7	**0.35400**
8	**0.36093**
9	**0.36576**
10	**0.37051**

In this study, assignments of clusters to specific species was the confirmation from the dataset generated by the array genotypes (Suppl. Fig. 3.2). Along the SW coast of England, southern mussel populations were mainly dominated by *M. edulis* ancestry. However, low levels of *M. galloprovincialis* ancestry were observed in all sampled populations. In addition, *M. galloprovincialis* dominates the northern coast of Cornwall and Devon.

Samples with a proportion > 0.99 of the respective ancestries in Admixture with $k = 2$ were grouped into the ancestries S0 (*M. galloprovincialis,* North Cornwall) and S1 (*M. edulis*, South Devon and Cornwall). The remaining samples were used as test samples to calculate the proportion of admixture in these hybrid populations (Figure 3.10). The h-index (genome-wide cline) for all samples is shown in Figure 3.8. A total of 112 samples were used in the genomic cline analysis as source *M. galloprovincialis* (S0) and 81 samples as source *M. edulis* (S1). A greater proportion of test individuals (n=73) had high proportions of *M. galloprovincialis* ancestry (h-index ≤ 0.2); high proportions of *M. edulis* ancestry (h-index ≥ 0.8) were found in 35 individuals.

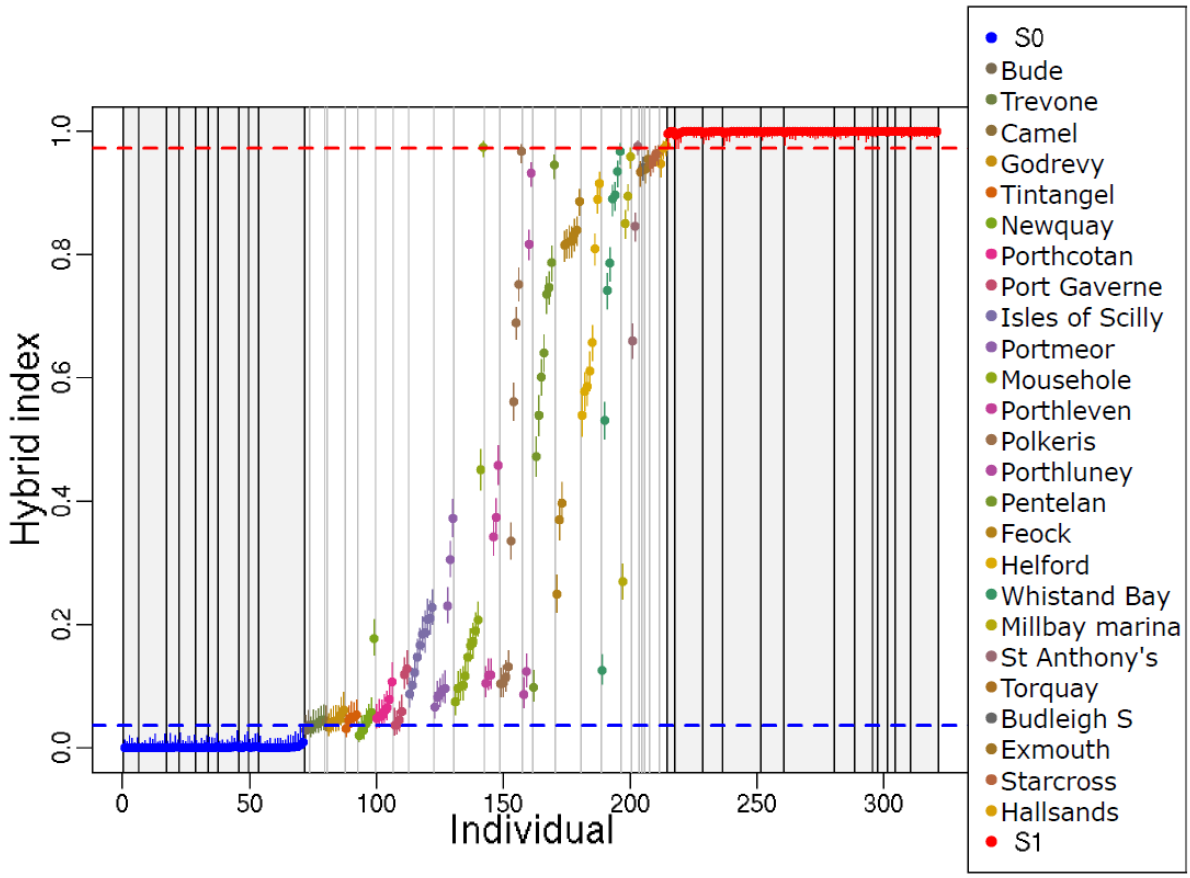


Figure 3-10: Distribution of hybrid-indices of blue mussels in the Southwest coast of England. Hybrid-indices are shown for S1 = *M. edulis* (red) and S0 = *M. galloprovincialis* (blue). Admixed mussel samples are coloured according to their population and ordered with increasing h-index.

Finally, we plotted the mean proportion of S0 and S1 for populations according to their h-index for each of the 29 populations against a map representing their geographical location (Figure 3.11). In this plot, we observed that populations from southeast Devon have higher frequencies of *M. edulis* genotypes. On the southern west coast of Devon and southeast coast of Cornwall, populations are admixed between *M. edulis* and *M. galloprovincialis*, with higher levels of *M. edulis* ancestry. West from the Lizard Peninsula, at the southwest coast of Cornwall (Porthleven and Porthmeor), a shift towards *M. galloprovincialis* genotypes can be observed. Moving towards the north coast and continuing further east, *M. galloprovincialis* genotypes dominate ancestry in sampled populations. In this figure, we also observed that populations positioned with increasing distance from Budleigh

Salterton, the most easterly south coast population we sampled, have decreasing frequencies of S1 (*M. edulis*) genotypes.

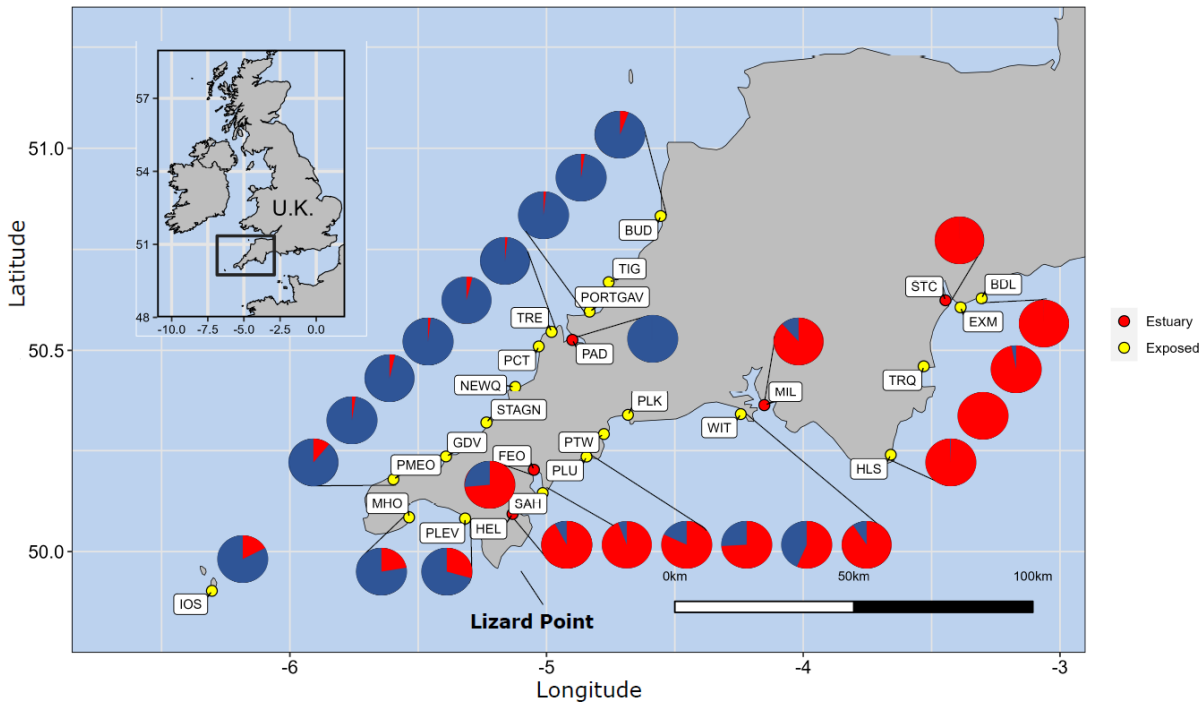


Figure 3-11: Geographical location of the intertidal mussel populations sampled in 2021 from the South West coast of England. Each of the populations has a correspondent pie-chart representing their mean h-index, with blue filling representing the proportion of hybridisation allocated to *M. edulis* genotypes, and orange, to *M. galloprovincialis*. In the smaller square plotted on the on the top left side of the figure orients the sampling area in the UK map.

Discussion

In this study, we investigate population structure in the blue mussel hybrid zone located along the South West coast of England, UK, formed by two commercially relevant species, *M. edulis* and *M. galloprovincialis*. For this, 29 locations were sampled along a ~500 km coast line extension, and 12 individuals per population were genotyped using the blue mussel multi-species 60K SNP-array developed by Affymetrix (Thermo Fisher Scientific) (Nascimento-Schulze in press). A total of 16,732 variants were used to assess levels of admixture across these populations as well as their h-index. Our findings highlight that *M. galloprovincialis* genotypes dominate the north coast Cornwall, as well as the extreme west of the Cornish peninsula. East from the Lizard peninsula, we observe a shift from *M.*

galloprovincialis to *M. edulis* dominated populations along the southern coasts of Cornwall and Devon. Furthermore, we saw no strong differentiation in species composition between estuarine populations and their corresponding coastal populations, and no strong genetic structural patterns differentiating low tide and high tide individuals.

Our observations are primarily in line with the earlier description of this hybrid zone by Hilbish *et. al.* (2002), with *M. galloprovincialis* being the main species on the northern coasts of Devon and Cornwall, and *M. edulis* dominating the southern coast. We assessed the mean h-index proportions of the 29 populations we sampled and compared it to the *M. edulis* allele frequency observed in the previous study. In this study, however, we could locate the transition zone between both species at Lizard point.

Several factors may contribute to the observed distribution of genotypes in the SW UK mussel hybrid zone observed in this study. Firstly, oceanic currents prevailing around the southern west coast of the UK likely impact larval drift and connectivity of *M. galloprovincialis* from north to south Cornwall and Devon (Hoch & Garreau 1998). The pelagic larval phase of the mussel life cycle typically lasts between two to four weeks, (Bayne 1965). Occasionally, this planktonic phase can prevail for longer as a result of a lack of appropriate conditions for development and/or settlement (Brenko & Calabrese 1969; Knöbel *et al.* 2021). Most individuals settle near the area they were spawned, although, free swimming larvae have the capacity to drift for dozens (Stuckas *et al.* 2017), or hundreds of kilometres (Demmer *et al.* 2022). Oceanic currents and phenomena which modulate their dynamics such as tidal activity and weather events (e.g. wind driven currents) therefore contribute to the dispersal of these planktonic larvae (Demmer *et al.* 2022; Stuckas *et al.* 2017). . Also likely is the potential for *M. galloprovincialis* genotypes to be fed into the southern coast of the UK from the Atlantic coast of France (Hoch & Garreau 1998). Secondly, the low frequency of *M. galloprovincialis* beyond the south eastern limits of this hybrid zone may suggest the existence of a strong barrier to larval drift from eastern to western populations. In this case, the Lizard peninsula may act as a biogeographical barrier, partially hindering connectivity across the hybrid zoneThirdly, a selection against *M. edulis* genotypes

in the north coast of Cornwall and Devon may be leading to the observed structure. Indeed, yearly patterns of sea surface temperature and plankton concentration differ between the north and south coasts of Cornwall and Devon (Hoch & Garreau 1998). It has been argued that *M. galloprovincialis* can withstand rougher conditions in comparison to *M. edulis*. Under such a scenario, genetic structure of these populations may be dictated by phenotype-environment mismatch that acts during larval or subsequent developmental stages (Hedgecock 1986). Genotype-dependent mortality, marked by a cline in the Lap^{94} allele, has been described in *M. edulis* populations living within the salinity gradient of an estuary zone in the eastern north Atlantic (Koehn *et al.* 1980). Evidence of genotype-dependent mortality leading to local adaptation has also been demonstrated in the Baltic Sea blue mussel hybrid zone (Knöbel *et al.* 2021), where individuals with high incidence of *M. trossulus* alleles dominate low salinities environments, whilst those characterised by *M. edulis* alleles thrive in higher salinities (Stuckas 2017). Sessile, benthic, marine populations demonstrate signs of local divergence and constrained gene flow (Marshall *et al.* 2010). Thus, processes such as local adaptation and differential species performance also play a role in the distribution of these taxa. Finally, across a finer geographic scale, the ability of larvae to actively choose settlement habitat may also contribute to shape the genetic structure of populations. Pre-settlement processes have been described as a key factor in maintaining genetic differentiation at the contact zone between mussel species *Myitlus galloprovinciallis* and *M. edulis* along the coast of Brittany, France (Bierne *et al.*, 2003). A mismatch in spawning events between species combined with a preference of *M. edulis* to settle in protected areas has been suggested to be driving fine-scale distribution in this area.

In a finer scale, however, we did not observe a strong genetic differentiation among populations inhabiting contrasting environments. The ability to withstand a set of contrasting environmental conditions can be associated with physiological adaptation and/or acclimation response, which are not only species but population specific. For example, a recent study suggests that in the Baltic Sea, *M. trossulus* genotypes dominate the benthic habitats in the low salinity portion of the Baltic Sea (Stuckas et al. 2017). In contrast, in the Newfoundland blue mussel hybrid zone on

the Atlantic Coast of Canada, mortality of *M. trossulus* genotypes is significantly higher than *M. edulis* ones in estuarine environments (Gardner & Thompson 2001). In the SW UK coast, it is proposed that the outperformance of *M. galloprovincialis* compared to *M. edulis* in exposed environments, may result in size-dependent selection against *M. edulis* alleles (e.g. Gardner, 1994b; Hilbish *et al.*, 1994; Kautsky *et al.*, 1990, Bierne *et al.* 2003) leading to the hybrid pattern demonstrated in this study, and previously. Strong declines in *M. edulis* allele frequencies in individuals > 30 mm inhabiting wave-exposed environments has also been reported in mussels from Whitsand Bay (Hilbish *et al.*, 2003). In the same study, no decline in *M. edulis* alleles was observed in the estuarine mussel populations in the Fal and Plymouth estuaries. Such findings support the hypothesis that *M. edulis* genotypes prefer sheltered environments whilst *M. galloprovincialis* are more robust and tolerate the exposed coasts. Our results, however, do not indicate a strong loss of *M. edulis* frequencies in individuals from Whitsand Bay, nor a lower frequency of *M. edulis* genotypes in comparison to the estuarine equivalent population in Millbay marina. Individuals sampled from Whitsand Bay were over > 20mm (high tide) and > 30 mm (low tide) in SL (Supplementary figure 3.1), which is comparable to the SL of mussels analysed in the previous study. Coastal and estuarine environments differ greatly from each other in terms of currents and wave exposure, as much as in terms of salinity and temperature regimes and chlorophyll-a concentration. These contrasting biotic and abiotic factors could likely influence the performance of mussels, favouring a specific phenotype. The weak structure detected in our analysis across these contrasting environments may be caused by the chosen study design. Low sample size can limit the ability to draw firm conclusions from an analysed dataset, nonetheless recent studies have demonstrated that accurate genomic structure can be revealed using samples sizes as low as n=3 per population (Li *et al.* 2020; Qu *et al.* 2020), far lower than the n=12 per population utilised here. Therefore, a plausible hypothesis is that the results acquired using the large set of genomic markers is not equivalent to the one described using allelic polymorphisms in a few selected markers. An alternative hypothesis is that the previous study sampled spat from a specific spawning event, which might have skewed the results towards a specific genotype, but might not be caused by ongoing

selection forces. Nonetheless, further investigation is required to understand if genotype segregation among environments is being underestimated in the present study.

We initially aimed to revisit the sampling sites used in Hilbish et al's (2002) study. However, it is important to stress that the absence of mussels in some of these original locations meant this was not always possible, and alternative locations had to be sampled instead. The decline in mussel populations was especially apparent along the southern coast, where mussels seemed to be less abundant in comparison to northern coast populations, and relatively harder to find. Moreover, some populations were composed of very few individuals (in the scale of hundreds). Such observations suggest that a strong decline of wild mussel populations, particularly in the southern coast of Cornwall and Devon, happened during the past decades, an alarming sign for the mussel aquaculture industry.

Conclusions and future perspectives

This study characterises the SW UK blue mussel hybrid zone using a genomic approach. The output of this study provides a greater understanding of the distribution of genotypes along this coast, and describes the dominance of *M. galloprovincialis* genotypes on the northern coasts of Cornwall and Devon, with an increasing frequency of *M. edulis* in the south coast from east of the Lizard peninsula. Our results did not point towards a strong genetic differentiation among populations inhabiting estuarine and coastal areas. Nonetheless, we suggest further investigation, using larger sample sizes for each population pair (coastal vs estuarine, or high tide vs low tide) or additional analogous paired population comparisons comprising of the same species. Blue mussels are the most produced marine bivalve in the UK. Negative interactions between genetics and environment impact the performance of individuals and can ultimately, affect shellfish production. Therefore, detecting genotypes that outperform in different environments of the SW UK coast will benefit the aquaculture industry. Furthermore, identifying genomic regions that are linked to performance will provide the basis to develop genomic selection in these taxa. Whilst genomic selection of blue mussels is not yet a viable option, climate change will impact the reproductive cycle and success of natural mussel

populations. Under such a scenario, a shift to a hatchery-based production system and the development of environmentally resilient lines might be inevitable to guarantee production of this highly sustainable source of animal protein. Our study provides fundamental knowledge that can contribute to such advances in the mussel aquaculture sector.

References

Alexander DH, Lange K (2011). Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics* **12**: 246.

Bayne BL (1965). Growth and the delay of metamorphosis of the larvae of *Mytilus edulis* (L.). *Ophelia,* **2(1)**: 1-47.

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

Brenko MH, Calabrese A (1969). The combined effects of salinity and temperature on larvae of the mussel *Mytilus edulis*. *Marine Biology* **4**: 224–26.

Demmer J, Robins P, Malham S, Lewis M, Owen A, Jones T, et al. (2022). The role of wind in controlling the connectivity of blue mussels (*Mytilus edulis* L.) populations. *Movement Ecology* **10**: 3.

Gardner JPA (1994). The structure and dynamics of naturally occurring hybrid *Mytilus edulis* Linnaeus, 1758 and *Mytilus galloprovincialis* Lamarck, 1819 (Bivalvia: Mollusca) populations: review and interpretation. *Arch Hydrobiol Suppl* **99**: 37–71.

Gardner JPA, Thompson RJ (2001). The effects of coastal and estuarine conditions on the physiology and survivorship of the mussels *Mytilus edulis, M. trossulus* and their hybrids. *Journal of Experimental Marine Biology and Ecology* **265**: 119–40.

Hedgecock D (1986). Is gene flow from pelagic larval dispersal important in the adaptation and evolution of marine invertebrates? *Bulletin of Marine Science* **39**: 550–64.

Hilbish T, Carson E, Plante J, Weaver L, Gilg M (2002). Distribution of *Mytilus edulis, M. galloprovincialis,* and their hybrids in open-coast populations of mussels in southwestern England. *Marine Biology* **140**: 137–42.

Hilbish TJ, Bayne BL, Day A (1994). Genetics of physiological differentiation within the marine mussel genus *Mytilus*. *Evolution* **48**: 267–86.

Hilbish TJ, Timmons J, Agrawal V, Schneider KR, Gilg MR (2003). Estuarine habitats protect hybrid mussels from selection. *Journal of Experimental Marine Biology and Ecology* **292**: 177–86.

Hoch T, Garreau P (1998). Phytoplankton dynamics in the English Channel: a simplified three-dimensional approach. *Journal of Marine Systems* **16**: 133–50.

Jombart T (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**: 1403–5.

Kautsky N, Johannesson K, Tedengren M (1990). Genotypic and phenotypic differences between Baltic and North Sea populations of *Mytilus edulis* evaluated through reciprocal transplantations. I. Growth and morphology. *Marine Ecology Progress Series* 203–10.

Knaus BJ, Grünwald NJ (2017). vcfr: a package to manipulate and visualize variant call format data in R. *Molecular Ecology Resources* **17**: 44–53.

Knöbel L, Nascimento-Schulze JC, Sanders T, Zeus D, Hiebenthal C, Barboza FR, et al. (2021). Salinity Driven Selection and Local Adaptation in Baltic Sea Mytilid Mussels. Frontiers in Marine Science, **8:** 692078.

Koehn RK, Newell RIE, Immermann F (1980). Maintenance of an Aminopeptidase Allele Frequency Cline by Natural Selection. *Proceedings of the National Academy of Sciences* **77**: 5385–89.

Li H, Qu W, Obrycki JJ, Meng L, Zhou X, Chu D, et al. (2020). Optimizing sample size for population genomic study in a global invasive lady beetle, *Harmonia axyridis*. *Insects* **11**: 290.

Marshall DJ, Monro K, Bode M, Keough MJ, Swearer S (2010). Phenotype-environment mismatches reduce connectivity in the sea. *Ecology letters*, **13(1)**: 128-140.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics* **81**: 559–75.

Qu W, Liang N, Wu Z, Zhao Y, Chu D (2020). Minimum sample sizes for invasion genomics: Empirical investigation in an invasive whitefly. *Ecology and evolution* **10**: 38–49.

Stuckas H, Knöbel L, Schade H, Breusing C, Hinrichsen HH, Bartel M, et al. (2017). Combining hydrodynamic modelling with genetics: can passive larval drift shape the genetic structure of Baltic *Mytilus* populations? *Molecular Ecology,* **26(10)**: 2765-2782.

Thia JA (2022). Guidelines for standardizing the application of discriminant analysis of principal components to genotype data. *Molecular Ecology Resources,* **23(3):** 523-538.

Supplementary material

Supplementary Table 3.1 Cross validation errors generated by admixture analysis to infer the best number of ancestral populations in dataset. In this analysis, we tested k's from a 2 – 7 range. The k with the lower cross validation error is considered the most representative number of ancestral populations for a given dataset.

K	cross validation error
2	0.26176
3	**0.25053**
4	0.25123
5	0.25298
6	0.25717
7	0.25869

Supplementary Table 3.2 Sample list and their allocation as either: a) ancestral for *M. galloprovincialis* (S0), b) ancestral for *M. edulis* (S1) and test populations (Hybrids) for the Hybrid Index and Cline analysis are found in the "Hybrid index" collumn. Samples were grouped into ancestry and test populations based on Admixture ancestry estimations with k = 2.

Sample Identification	Population	Ancestry S1	Ancestry S0	Hybrid Index
Bude_1	BUDE	0.00001	0.99999	S0
Bude_10	BUDE	0.00001	0.99999	S0
Bude_12	BUDE	0.00001	0.99999	S0
Bude_2	BUDE	0.00001	0.99999	S0
Godrevy_11	GOD	0.00001	0.99999	S0
Godrevy_3	GOD	0.00001	0.99999	S0
Godrevy_5	GOD	0.00001	0.99999	S0
Godrevy_7	GOD	0.00001	0.99999	S0
Godrevy_8	GOD	0.00001	0.99999	S0
Newquay_10	NEWQ	0.00001	0.99999	S0
Newquay_11	NEWQ	0.00001	0.99999	S0
Newquay_12	NEWQ	0.00001	0.99999	S0
Newquay_3	NEWQ	0.00001	0.99999	S0
Newquay_4	NEWQ	0.00001	0.99999	S0
Newquay_8	NEWQ	0.00001	0.99999	S0
Pinksoncreek_camel_1	CAMEL	0.00001	0.99999	S0
Pinksoncreek_camel_10	CAMEL	0.00001	0.99999	S0

Pinksoncreek_camel_11	CAMEL	0.00001	0.99999	S0
Pinksoncreek_camel_12	CAMEL	0.00001	0.99999	S0
Pinksoncreek_camel_3	CAMEL	0.00001	0.99999	S0
Pinksoncreek_camel_5	CAMEL	0.00001	0.99999	S0
Pinksoncreek_camel_6	CAMEL	0.00001	0.99999	S0
Pinksoncreek_camel_7	CAMEL	0.00001	0.99999	S0
Pinksoncreek_camel_8	CAMEL	0.00001	0.99999	S0
Pinksoncreek_camel_9	CAMEL	0.00001	0.99999	S0
Port_Gavirn_HT_11	PGAV	0.00001	0.99999	S0
Port_Gavirn_HT_12	PGAV	0.00001	0.99999	S0
Port_Gavirn_HT_2	PGAV	0.00001	0.99999	S0
Port_Gavirn_HT_3	PGAV	0.00001	0.99999	S0
Port_Gavirn_HT_5	PGAV	0.00001	0.99999	S0
Port_Gavirn_HT_8	PGAV	0.00001	0.99999	S0
Port_Gavirn_LT_2	PGAV	0.00001	0.99999	S0
Port_Gavirn_LT_4	PGAV	0.00001	0.99999	S0
Port_Gavirn_LT_7	PGAV	0.00001	0.99999	S0
Port_Gavirn_LT_9	PGAV	0.00001	0.99999	S0
Porthcotan_Bay_10	PCBAY	0.00001	0.99999	S0
Porthcotan_Bay_12	PCBAY	0.00001	0.99999	S0
Porthcotan_Bay_2	PCBAY	0.00001	0.99999	S0
Porthcotan_Bay_3	PCBAY	0.00001	0.99999	S0
Porthcotan_Bay_5	PCBAY	0.00001	0.99999	S0
Portmeor_1	PMEOR	0.00001	0.99999	S0
Portmeor_2	PMEOR	0.00001	0.99999	S0
Portmeor_9	PMEOR	0.00001	0.99999	S0
St_Agnes_1	STAGN	0.00001	0.99999	S0
St_Agnes_10	STAGN	0.00001	0.99999	S0
St_Agnes_11	STAGN	0.00001	0.99999	S0
St_Agnes_12	STAGN	0.00001	0.99999	S0
St_Agnes_3	STAGN	0.00001	0.99999	S0
St_Agnes_5	STAGN	0.00001	0.99999	S0
Tintangel_Castle_10	TINTC	0.00001	0.99999	S0
Tintangel_Castle_11	TINTC	0.00001	0.99999	S0
Tintangel_Castle_12	TINTC	0.00001	0.99999	S0
Tintangel_Castle_4	TINTC	0.00001	0.99999	S0
Tintangel_Castle_5	TINTC	0.00001	0.99999	S0
Tintangel_Castle_6	TINTC	0.00001	0.99999	S0
Tintangel_Castle_9	TINTC	0.00001	0.99999	S0
Trevone_1	TREV	0.00001	0.99999	S0
Trevone_10	TREV	0.00001	0.99999	S0
Trevone_3	TREV	0.00001	0.99999	S0
Trevone_7	TREV	0.00001	0.99999	S0
Trevone_8	TREV	0.00001	0.99999	S0
Trevone_9	TREV	0.00001	0.99999	S0
Porthcotan_Bay_7	PCBAY	0.000015	0.999985	S0
Port_Gavirn_LT_6	PGAV	0.001225	0.998775	S0
Port_Gavirn_LT_5	PGAV	0.001484	0.998516	S0

Pinksoncreek_camel_2	CAMEL	0.002261	0.997739	S0
Port_Gavirn_LT_3	PGAV	0.003866	0.996134	S0
Newquay_5	NEWQ	0.004342	0.995658	S0
Godrevy_4	GOD	0.004403	0.995597	S0
Pinksoncreek_camel_4	CAMEL	0.004463	0.995537	S0
Godrevy_9	GOD	0.004513	0.995487	S0
Port_Gavirn_LT_8	PGAV	0.004863	0.995137	S0
Newquay_6	NEWQ	0.005714	0.994286	S0
Port_Gavirn_HT_7	PGAV	0.005813	0.994187	S0
St_Agnes_8	STAGN	0.005874	0.994126	S0
St_Agnes_6	STAGN	0.005995	0.994005	S0
Tintangel_Castle_3	TINTC	0.008405	0.991595	S0
Porthcotan_Bay_6	PCBAY	0.008627	0.991373	S0
Port_Gavirn_LT_11	PGAV	0.009197	0.990803	S0
Port_Gavirn_LT_1	PGAV	0.009208	0.990792	S0
Newquay_7	NEWQ	0.009611	0.990389	S0
Portmeor_10	PMEOR	0.01053	0.98947	Hybrid
Porthcotan_Bay_8	PCBAY	0.010549	0.989451	Hybrid
Port_Gavirn_LT_10	PGAV	0.010774	0.989226	Hybrid
Bude_8	BUDE	0.010877	0.989123	Hybrid
Godrevy_2	GOD	0.011171	0.988829	Hybrid
Trevone_11	TREV	0.011707	0.988293	Hybrid
Bude_9	BUDE	0.012135	0.987865	Hybrid
Newquay_2	NEWQ	0.012573	0.987427	Hybrid
Bude_6	BUDE	0.013105	0.986895	Hybrid
Newquay_9	NEWQ	0.013225	0.986775	Hybrid
Port_Gavirn_LT_12	PGAV	0.014165	0.985835	Hybrid
Porthcotan_Bay_11	PCBAY	0.014334	0.985666	Hybrid
Port_Gavirn_HT_10	PGAV	0.01525	0.98475	Hybrid
Trevone_12	TREV	0.016058	0.983942	Hybrid
Trevone_6	TREV	0.016255	0.983745	Hybrid
Godrevy_6	GOD	0.01795	0.98205	Hybrid
Bude_7	BUDE	0.0189	0.9811	Hybrid
Port_Gavirn_HT_1	PGAV	0.019062	0.980938	Hybrid
Tintangel_Castle_1	TINTC	0.019347	0.980653	Hybrid
Tintangel_Castle_7	TINTC	0.020324	0.979676	Hybrid
Trevone_5	TREV	0.020521	0.979479	Hybrid
Trevone_4	TREV	0.020947	0.979053	Hybrid
Porthcotan_Bay_4	PCBAY	0.021304	0.978696	Hybrid
Porthcotan_Bay_9	PCBAY	0.024063	0.975937	Hybrid
Godrevy_10	GOD	0.024887	0.975113	Hybrid
Port_Gavirn_HT_4	PGAV	0.02514	0.97486	Hybrid
Trevone_2	TREV	0.028153	0.971847	Hybrid
Bude_5	BUDE	0.03263	0.96737	Hybrid
Godrevy_1	GOD	0.033423	0.966577	Hybrid
Tintangel_Castle_2	TINTC	0.037151	0.962849	Hybrid
Portmeor_7	PMEOR	0.043067	0.956933	Hybrid
St_Agnes_7	STAGN	0.044416	0.955584	Hybrid

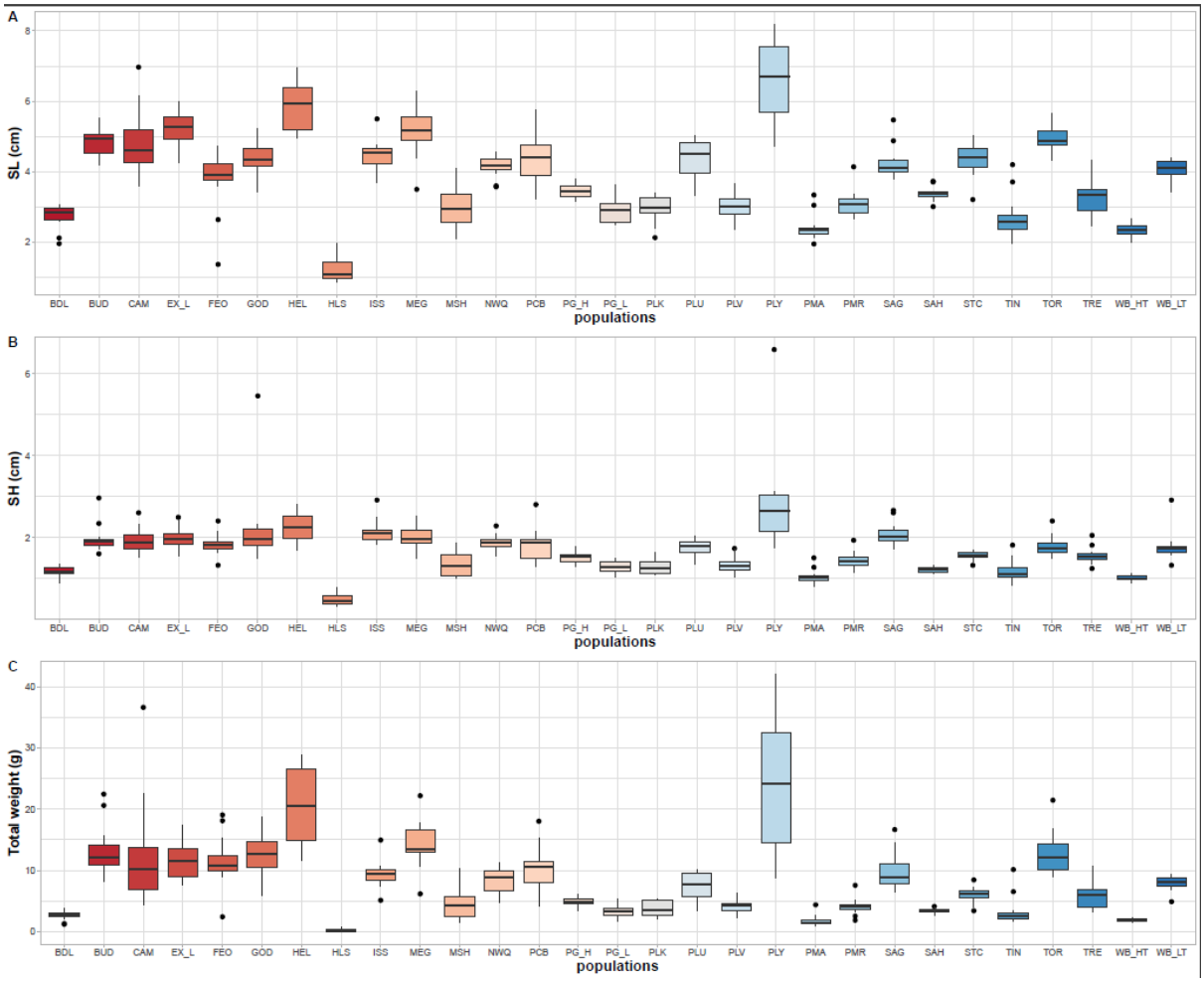
Mousehole_3	MOUSH	0.048324	0.951676	Hybrid
Porthcotan_Bay_1	PCBAY	0.049377	0.950623	Hybrid
Scilly_11	SCILLY	0.049651	0.950349	Hybrid
St_Agnes_4	STAGN	0.052175	0.947825	Hybrid
Scilly_7	SCILLY	0.053961	0.946039	Hybrid
Portmeor_12	PMEOR	0.057013	0.942987	Hybrid
Mousehole_12	MOUSH	0.058028	0.941972	Hybrid
Portmeor_11	PMEOR	0.058146	0.941854	Hybrid
Pentelan_Mevagissey_11	PENT	0.064008	0.935992	Hybrid
Porthleven_11	PLEV	0.065527	0.934473	Hybrid
Port_lune_11	PLUNE	0.069912	0.930088	Hybrid
Porthleven_12	PLEV	0.070203	0.929797	Hybrid
Mousehole_8	MOUSH	0.070747	0.929253	Hybrid
Portmeor_8	PMEOR	0.072384	0.927616	Hybrid
Scilly_8	SCILLY	0.073156	0.926844	Hybrid
Mousehole_5	MOUSH	0.074	0.926	Hybrid
Polkeris_1	POLK	0.077564	0.922436	Hybrid
Port_Gavirn_HT_6	PGAV	0.079848	0.920152	Hybrid
Portmeor_6	PMEOR	0.080097	0.919903	Hybrid
Scilly_6	SCILLY	0.080213	0.919787	Hybrid
Polkeris_3	POLK	0.083034	0.916966	Hybrid
Godrevy_12	GOD	0.083283	0.916717	Hybrid
Polkeris_4	POLK	0.08539	0.91461	Hybrid
Port_lune_8	PLUNE	0.086888	0.913112	Hybrid
Port_Gavirn_HT_9	PGAV	0.088181	0.911819	Hybrid
Mousehole_2	MOUSH	0.097385	0.902615	Hybrid
St_Agnes_9	STAGN	0.098816	0.901184	Hybrid
Mousehole_10	MOUSH	0.099396	0.900604	Hybrid
Polkeris_8	POLK	0.09995	0.90005	Hybrid
Porthleven_8	PLEV	0.109449	0.890551	Hybrid
Porthleven_1	PLEV	0.113485	0.886515	Hybrid
Tintangel_Castle_8	TINTC	0.115234	0.884766	Hybrid
Mousehole_7	MOUSH	0.123121	0.876879	Hybrid
Whistsandbay_HT_7	WHITS	0.126074	0.873926	Hybrid
St_Agnes_2	STAGN	0.133775	0.866225	Hybrid
Mousehole_9	MOUSH	0.135449	0.864551	Hybrid
Bude_3	BUDE	0.138731	0.861269	Hybrid
Scilly_12	SCILLY	0.14425	0.85575	Hybrid
Mousehole_11	MOUSH	0.144597	0.855403	Hybrid
Scilly_10	SCILLY	0.163519	0.836481	Hybrid
Newquay_1	NEWQ	0.167799	0.832201	Hybrid
Scilly_5	SCILLY	0.16885	0.83115	Hybrid
Scilly_1	SCILLY	0.170661	0.829339	Hybrid
Scilly_3	SCILLY	0.179702	0.820298	Hybrid
Scilly_9	SCILLY	0.194537	0.805463	Hybrid
Mousehole_1	MOUSH	0.196213	0.803787	Hybrid
Feok_7	FEOCK	0.19633	0.80367	Hybrid
Portmeor_5	PMEOR	0.204684	0.795316	Hybrid

Scilly_4	SCILLY	0.207669	0.792331	Hybrid
Porthleven_2	PLEV	0.244234	0.755766	Hybrid
Millbay_Marina_2	MILLBAY	0.261547	0.738453	Hybrid
Bude_4	BUDE	0.273708	0.726292	Hybrid
Portmeor_3	PMEOR	0.288796	0.711204	Hybrid
Porthleven_10	PLEV	0.316008	0.683992	Hybrid
Porthleven_7	PLEV	0.318545	0.681455	Hybrid
Portmeor_4	PMEOR	0.32109	0.67891	Hybrid
Polkeris_10	POLK	0.322527	0.677473	Hybrid
Scilly_2	SCILLY	0.342178	0.657822	Hybrid
Porthleven_4	PLEV	0.370896	0.629104	Hybrid
Feok_4	FEOCK	0.38688	0.61312	Hybrid
Feok_10	FEOCK	0.389398	0.610602	Hybrid
Exmouth_HT_2	EX	0.389845	0.610155	Hybrid
Mousehole_4	MOUSH	0.438303	0.561697	Hybrid
Porthleven_5	PLEV	0.462161	0.537839	Hybrid
Porthleven_9	PLEV	0.473435	0.526565	Hybrid
Pentelan_Mevagissey_6	PENT	0.489953	0.510047	Hybrid
Millbay_Marina_8	MILLBAY	0.540716	0.459284	Hybrid
Helford_11	HEL	0.548899	0.451101	Hybrid
Polkeris_11	POLK	0.556911	0.443089	Hybrid
Whistsandbay_HT_12	WHITS	0.558338	0.441662	Hybrid
Porthleven_6	PLEV	0.562754	0.437246	Hybrid
Pentelan_Mevagissey_3	PENT	0.565688	0.434312	Hybrid
Whistsandbay_HT_4	WHITS	0.580228	0.419772	Hybrid
Helford_8	HEL	0.605587	0.394413	Hybrid
Helford_10	HEL	0.627246	0.372754	Hybrid
Helford_4	HEL	0.633686	0.366314	Hybrid
Pentelan_Mevagissey_7	PENT	0.643343	0.356657	Hybrid
St_Anthonys_Head_7	STANT	0.653437	0.346563	Hybrid
Helford_1	HEL	0.68429	0.31571	Hybrid
Polkeris_12	POLK	0.6898	0.3102	Hybrid
Pentelan_Mevagissey_12	PENT	0.700721	0.299279	Hybrid
St_Anthonys_Head_3	STANT	0.762872	0.237128	Hybrid
Polkeris_6	POLK	0.765173	0.234827	Hybrid
Pentelan_Mevagissey_9	PENT	0.766782	0.233218	Hybrid
Helford_3	HEL	0.793644	0.206356	Hybrid
Whistsandbay_LT_4	WHITS	0.793803	0.206197	Hybrid
Feok_1	FEOCK	0.801612	0.198388	Hybrid
Feok_9	FEOCK	0.808631	0.191369	Hybrid
Pentelan_Mevagissey_1	PENT	0.816371	0.183629	Hybrid
Feok_3	FEOCK	0.823115	0.176885	Hybrid
Feok_6	FEOCK	0.82439	0.17561	Hybrid
Feok_12	FEOCK	0.827681	0.172319	Hybrid
Pentelan_Mevagissey_2	PENT	0.832331	0.167669	Hybrid
Feok_5	FEOCK	0.849505	0.150495	Hybrid
Whistsandbay_HT_6	WHITS	0.849875	0.150125	Hybrid
St_Anthonys_Head_5	STANT	0.856189	0.143811	Hybrid

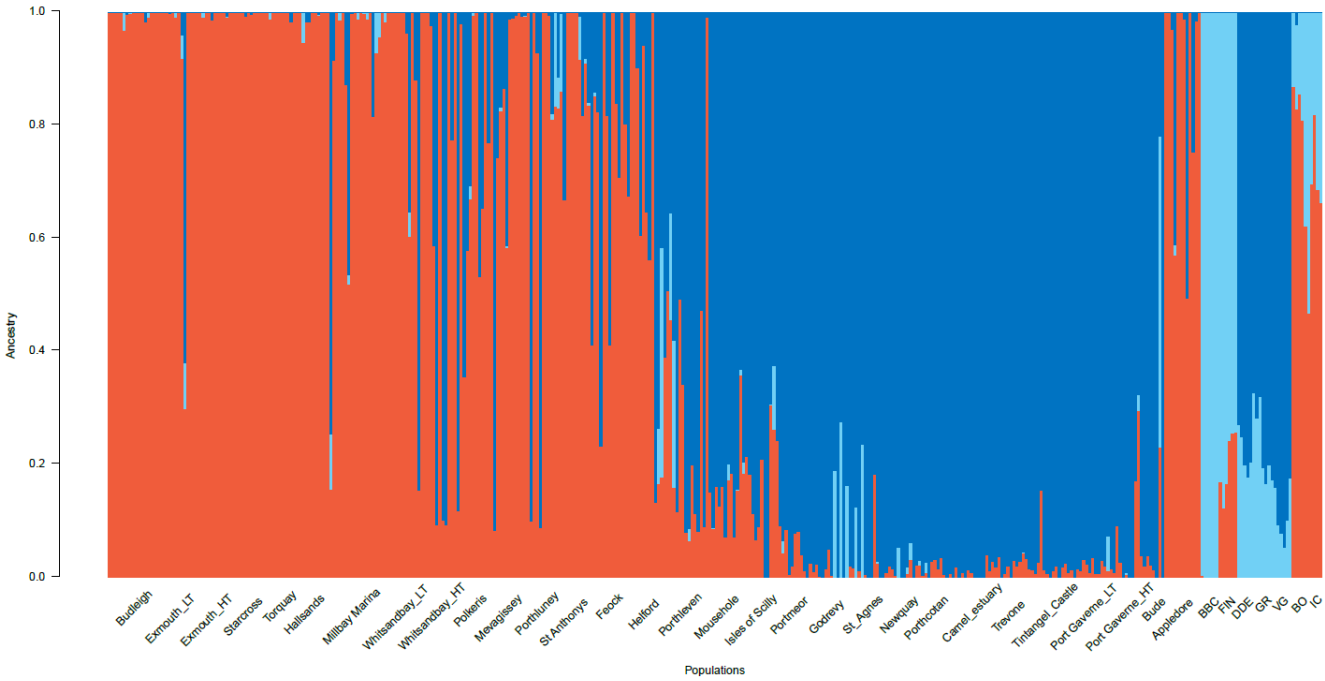
Millbay_Marina_7	MILLBAY	0.862124	0.137876	Hybrid
Port_lune_10	PLUNE	0.898932	0.101068	Hybrid
Helford_7	HEL	0.90356	0.09644	Hybrid
Millbay_Marina_3	MILLBAY	0.913035	0.086965	Hybrid
Feok_2	FEOCK	0.916669	0.083331	Hybrid
Exmouth_HT_1	EX	0.924761	0.075239	Hybrid
Helford_9	HEL	0.93619	0.06381	Hybrid
St_Anthonys_Head_6	STANT	0.936757	0.063243	Hybrid
St_Anthonys_Head_4	STANT	0.943957	0.056043	Hybrid
Whistsandbay_HT_3	WHITS	0.948672	0.051328	Hybrid
St_Anthonys_Head_12	STANT	0.953191	0.046809	Hybrid
Whistsandbay_HT_11	WHITS	0.961773	0.038227	Hybrid
Exmouth_LT_1	EX	0.969788	0.030212	Hybrid
Hallsends_north_6	HALL	0.977801	0.022199	Hybrid
Port_lune_1	PLUNE	0.97915	0.02085	Hybrid
Starcross_10	STAR	0.980928	0.019072	Hybrid
Exmouth_HT_11	EX	0.981386	0.018614	Hybrid
Whistsandbay_HT_10	WHITS	0.981586	0.018414	Hybrid
Polkeris_9	POLK	0.982507	0.017493	Hybrid
Millbay_Marina_9	MILLBAY	0.983623	0.016377	Hybrid
Hallsends_north_5	HALL	0.985437	0.014563	Hybrid
Hallsends_north_7	HALL	0.986904	0.013096	Hybrid
Whistsandbay_LT_5	WHITS	0.987914	0.012086	Hybrid
Mousehole_6	MOUSH	0.987984	0.012016	Hybrid
Budleigh_6	BUD	0.988733	0.011267	Hybrid
Starcross_12	STAR	0.991495	0.008505	S1
Budleigh_1	BUD	0.991646	0.008354	S1
Whistsandbay_LT_6	WHITS	0.992482	0.007518	S1
Port_lune_2	PLUNE	0.992593	0.007407	S1
Exmouth_LT_4	EX	0.993571	0.006429	S1
Millbay_Marina_5	MILLBAY	0.994006	0.005994	S1
Port_lune_5	PLUNE	0.994347	0.005653	S1
Hallsends_north_8	HALL	0.994742	0.005258	S1
Starcross_4	STAR	0.995232	0.004768	S1
Port_lune_6	PLUNE	0.995454	0.004546	S1
Millbay_Marina_12	MILLBAY	0.995574	0.004426	S1
Exmouth_LT_12	EX	0.995695	0.004305	S1
St_Anthonys_Head_2	STANT	0.995858	0.004142	S1
Exmouth_LT_5	EX	0.99589	0.00411	S1
Starcross_5	STAR	0.996001	0.003999	S1
Budleigh_7	BUD	0.996387	0.003613	S1
Hallsends_north_1	HALL	0.997286	0.002714	S1
Hallsends_north_12	HALL	0.998472	0.001528	S1
Exmouth_LT_11	EX	0.998529	0.001471	S1
Pentelan_Mevagissey_10	PENT	0.99962	0.00038	S1
Budleigh_10	BUD	0.99999	0.00001	S1
Budleigh_11	BUD	0.99999	0.00001	S1
Budleigh_12	BUD	0.99999	0.00001	S1

Budleigh_2	BUD	0.99999	0.00001	S1
Budleigh_3	BUD	0.99999	0.00001	S1
Budleigh_4	BUD	0.99999	0.00001	S1
Budleigh_5	BUD	0.99999	0.00001	S1
Budleigh_8	BUD	0.99999	0.00001	S1
Budleigh_9	BUD	0.99999	0.00001	S1
Exmouth_HT_10	EX	0.99999	0.00001	S1
Exmouth_HT_12	EX	0.99999	0.00001	S1
Exmouth_HT_3	EX	0.99999	0.00001	S1
Exmouth_HT_4	EX	0.99999	0.00001	S1
Exmouth_HT_5	EX	0.99999	0.00001	S1
Exmouth_HT_6	EX	0.99999	0.00001	S1
Exmouth_HT_7	EX	0.99999	0.00001	S1
Exmouth_HT_8	EX	0.99999	0.00001	S1
Exmouth_HT_9	EX	0.99999	0.00001	S1
Exmouth_LT_10	EX	0.99999	0.00001	S1
Exmouth_LT_2	EX	0.99999	0.00001	S1
Exmouth_LT_3	EX	0.99999	0.00001	S1
Exmouth_LT_6	EX	0.99999	0.00001	S1
Exmouth_LT_7	EX	0.99999	0.00001	S1
Exmouth_LT_8	EX	0.99999	0.00001	S1
Exmouth_LT_9	EX	0.99999	0.00001	S1
Feok_11	FEOCK	0.99999	0.00001	S1
Feok_8	FEOCK	0.99999	0.00001	S1
Hallsends_north_10	HALL	0.99999	0.00001	S1
Hallsends_north_11	HALL	0.99999	0.00001	S1
Hallsends_north_2	HALL	0.99999	0.00001	S1
Hallsends_north_3	HALL	0.99999	0.00001	S1
Hallsends_north_4	HALL	0.99999	0.00001	S1
Hallsends_north_9	HALL	0.99999	0.00001	S1
Helford_12	HEL	0.99999	0.00001	S1
Helford_2	HEL	0.99999	0.00001	S1
Helford_5	HEL	0.99999	0.00001	S1
Helford_6	HEL	0.99999	0.00001	S1
Millbay_Marina_1	MILLBAY	0.99999	0.00001	S1
Millbay_Marina_10	MILLBAY	0.99999	0.00001	S1
Millbay_Marina_11	MILLBAY	0.99999	0.00001	S1
Millbay_Marina_4	MILLBAY	0.99999	0.00001	S1
Millbay_Marina_6	MILLBAY	0.99999	0.00001	S1
Pentelan_Mevagissey_4	PENT	0.99999	0.00001	S1
Pentelan_Mevagissey_5	PENT	0.99999	0.00001	S1
Pentelan_Mevagissey_8	PENT	0.99999	0.00001	S1
Polkeris_2	POLK	0.99999	0.00001	S1
Polkeris_5	POLK	0.99999	0.00001	S1
Polkeris_7	POLK	0.99999	0.00001	S1
Port_lune_12	PLUNE	0.99999	0.00001	S1
Port_lune_3	PLUNE	0.99999	0.00001	S1
Port_lune_4	PLUNE	0.99999	0.00001	S1

Port_lune_7	PLUNE	0.99999	0.00001	S1
Port_lune_9	PLUNE	0.99999	0.00001	S1
St_Anthonys_Head_1	STANT	0.99999	0.00001	S1
St_Anthonys_Head_10	STANT	0.99999	0.00001	S1
St_Anthonys_Head_11	STANT	0.99999	0.00001	S1
St_Anthonys_Head_8	STANT	0.99999	0.00001	S1
St_Anthonys_Head_9	STAR	0.99999	0.00001	S1
Starcross_1	STAR	0.99999	0.00001	S1
Starcross_11	STAR	0.99999	0.00001	S1
Starcross_2	STAR	0.99999	0.00001	S1
Starcross_3	STAR	0.99999	0.00001	S1
Starcross_6	STAR	0.99999	0.00001	S1
Starcross_7	STAR	0.99999	0.00001	S1
Starcross_8	STAR	0.99999	0.00001	S1
Starcross_9	STAR	0.99999	0.00001	S1
Torquay_1	TORQ	0.99999	0.00001	S1
Torquay_10	TORQ	0.99999	0.00001	S1
Torquay_11	TORQ	0.99999	0.00001	S1
Torquay_12	TORQ	0.99999	0.00001	S1
Torquay_2	TORQ	0.99999	0.00001	S1
Torquay_3	TORQ	0.99999	0.00001	S1
Torquay_4	TORQ	0.99999	0.00001	S1
Torquay_5	TORQ	0.99999	0.00001	S1
Torquay_6	TORQ	0.99999	0.00001	S1
Torquay_7	TORQ	0.99999	0.00001	S1
Torquay_8	TORQ	0.99999	0.00001	S1
Torquay_9	TORQ	0.99999	0.00001	S1
Whistsandbay_HT_1	WHITS	0.99999	0.00001	S1
Whistsandbay_HT_2	WHITS	0.99999	0.00001	S1
Whistsandbay_HT_5	WHITS	0.99999	0.00001	S1
Whistsandbay_HT_8	WHITS	0.99999	0.00001	S1
Whistsandbay_HT_9	WHITS	0.99999	0.00001	S1
Whistsandbay_LT_1	WHITS	0.99999	0.00001	S1
Whistsandbay_LT_10	WHITS	0.99999	0.00001	S1
Whistsandbay_LT_11	WHITS	0.99999	0.00001	S1
Whistsandbay_LT_12	WHITS	0.99999	0.00001	S1
Whistsandbay_LT_2	WHITS	0.99999	0.00001	S1
Whistsandbay_LT_3	WHITS	0.99999	0.00001	S1
Whistsandbay_LT_7	WHITS	0.99999	0.00001	S1
Whistsandbay_LT_8	WHITS	0.99999	0.00001	S1
Whistsandbay_LT_9	WHITS	0.99999	0.00001	S1



Supplementary Figure 3.1 Boxplot including A) Size in Shell length (SL), B) Shell Height (SH) and C) Total weight (TW) of individuals from all 29 blue mussel populations collected from the Southwest coast of England used in this study measured in centimetres. SL and SH data are presented in centimetres, TW is presented in grams.



Supplemntary Figure 3.2 Results of genetic admixture analysis and ancestry inference. Cluster membership coefficients (Q) where each individual is represented by a column partitioned into segments of different colour, the length of which indicate the posterior probability of membership in each cluster. Solid bars represent an individual from a single species background. In this figure we plot the k=3 for the dataset including all 29 sampling locations. Additionally, we investigate ancestry of the 5 control populations. The red cluster can be inferred to *M. edulis*, as this species dominates populations in IC and BO. The light-blue cluster can be inferred to *M. trossulus*, as genotypes from this species are highly frequent in finish mussel populations (FIN). Finally, the dark blue cluster can be inferred to *M. galloprovincialis*, as this is the main species colonising VG and GR.

4. Chapter IV - Thermal Selection Shifts Genetic Diversity and Performance of Blue Mussel Juveniles

This chapter resulted from a collaborative work of myself, Jahangir Vajedsamiei, Lisa Frankholz, Tim P Bean, Frank Melzner and Robert P Ellis. I was responsible for designing and carrying the 25-day experiment, analysing the data and writing the manuscript.

Abstract

Mussels from the genus *Mytilus* are key inhabitants of the benthos, and considered a foundation species to this habitat as they provide a number of key ecosystem services. Moreover, mussels are of high value for the aquaculture industry, and one of the most sustainable sources of animal protein. Interestingly, species within the blue mussel species complex easily hybridise whenever their geographic distributions overlap. The blue mussel hybrid zone in the Baltic Sea is mainly formed by two species, *M. edulis* and *M. trossulus*, but a low levels of introgression with *M. galloprovincialis* has been recently discovered. Much recent research has highlighted that blue mussels are susceptible towards anthropogenically induced climate change, being ectothermic calcifying intertidal invertebrates. Given their economic and ecological relevance, this study therefore aimed to investigate the natural capacity of these organisms to adapt to new thermal conditions, what could be used as a mechanism to fast-tracked environmental resilience in these taxa. For this, two cohorts of mussels from the Baltic Sea (Kiel, Germany) were developed by collecting juvenile individuals from wild and either exposing them to a short warming event (30^{o}C for ~48h) or keeping the individuals naïve to this stressor, under control conditions (17°C). In this area, genotypes of the three blue mussel species existing within the *Mytilus* species-complex occur. Subsequently, both cohorts were re-challenged to second thermal challenge, being exposed to a set of experimental temperatures lying at the upper thermal limit of these populations (21^{o}C, 22^{o}C, 23^{o}C, 24^{o}C, 25^{o}C and 26^{o}C) for 25 days. Fifty individuals from each of the cohorts were sampled for genotyping using the newly developed blue mussel multi-species 60K SNP-array. Results from this study suggest that selection towards thermal resilience can potentially impact the whole organism performance of mussels, measured as a gain in dry tissue weight (g). Changes in the genetic diversity of these cohorts were also identified, with the selected cohort shifting towards a dominance of *M. edulis* genotypes in comparison to the naïve one. Furthermore, four genetic markers were highly divergent between these populations (F_{ST} value > 0.15). Such findings are key as the foundation to contribute towards

futureproofing mussel aquaculture against climate change. Furthermore, results from this study take a first step towards the development of genomic selection in these taxa.

Keywords: SNP-array, blue mussels, genomic selection, global warming, Baltic sea

Background

Shallow water marine ecosystems are vital both for human livelihoods, providing essential services to society (e.g. fishing, goods, human transport and leisure), and nature, allowing for nutrient cycling and storage as well as maintaining biodiversity (Pörtner et al. 2019). Ongoing climatic changes in the ocean, occurring as shifts in relatively stable physicochemical regimes (e.g. mean sea water salinity, temperature, acidity and oxygen content), will accelerate over the coming decades (IPCC 2022). Such climatic transformations will be more prominent for shallow estuarine and coastal waters, which have less environmental inertia due to their limited depth (water volume) and their proximity to densely populated areas and land masses (Halpern et al. 2008; Lotze et al. 2006). The Baltic Sea is one such semi-enclosed shallow sea, which is connected to the North Atlantic through the narrow Danish strait. Being one of the most rapidly warming seas (Belkin 2009), an increase of 1.5 - 4^{o}C in mean sea surface temperature (SST) is predicted to occur in the Baltic by 2100 (Gräwe et al. 2013; Meier et al. 2012). This is occurring alongside an expansion of low-oxygen zones, increased riverine inflow and an increased frequency of extreme weather events (Gräwe et al. 2013; Rahmstorf & Coumou 2011). Because of its unique environmental characteristics, the Baltic Sea is already lower in biodiversity compared to other fully marine areas (Elmgren & Hill 1997; Ojaveer et al. 2010). Shifting temperature regimes, either through modified mean seasonal temperature or isolated shorter extreme events, can contribute to community shifts and set off a chain reaction of biodiversity loss, pushing this ecosystem towards an alternate state (Johannesson et al. 2011; Pansch et al. 2018).

Blue mussels of the *Mytilus* genus are a foundation and ecosystem engineer species, given the number of ecosystem services they provide by increasing habitat complexity providing spatial structure associated species and through their role in nutrient cycling (Johannesson et al. 2011; van

der Schatte Olivier et al. 2020). In 2016, global blue mussel production was approximately 2 million tonnes, with 94% resulting from aquaculture activity, which generated approximately $3.8 billion USD (Avdelas et al. 2021). Baltic Blue mussel populations are a highly introgressed swarm composed by the three species *M. edulis*, *M. trossulus* and *M. galloprovinciallis* (Vendrami et al. 2020). In this area, the distribution of species-specific alleles is correlated to the salinity gradient of the Baltic Sea: *M. edulis* genotypes dominate the benthic populations in western waters of the semi-enclosed basin where salinity is higher (> 16), transitioning towards *M. trossulus* alleles in the eastern portion where salinity is under 10 (Stuckas et al. 2017). These populations must cope with some of the fastest and strongest changes in SST globally, yet the genotypic and phenotypic variation present in these populations in response to such heat stress remains understudied.

Temperatures that are above optimal for an organism, or physiological process, disturb its energy balance (Pörtner, 2012). The pervasive implications of temperature, from molecular to whole-animal levels, ultimately affect the fitness and abundance of individuals (Hoschaka & Somero 2002). Subsequently, extreme temperatures have been associated with mass mortality events in blue mussel, a phenomenon likely to increase in prevalence with further warming (Jones et al. 2010; Seuront et al. 2019). One response to overcome thermal stress, which has been observed in mussels in response to increased SST, is a shift in the geographical range distribution (Jones et al. 2010) or a shift in the vertical extent of mussel beds (Harley 2011). Furthermore, intra and inter species variation on thermal tolerance is shown to exist within the species complex, which may influence the response of different species to warming and offer a mechanism for populations to adapt to warming regimes. For example: *M. trossulus*, *M. edulis* and *M. galloprovincialis* each maintain a positive Scope for Growth of the species under warming, but to a maximum temperature of 17^{o}C, 23^{o}C and 30^{o}C, respectively (Fly & Hilbish 2013), highlighting the differential thermal sensitivity of these organisms. Blue mussels demonstrate a plastic response towards a broad range of environmental temperatures, with their upper thermal tolerance being highly dependent on population background (see Zippay and Helmuth, 2012 and references therein).

A recent study described a ~2–3 fold increase in standard metabolic rate in mussels acclimated to 15°C compared to those acclimated to 10°C, highlighting the impact of elevated temperature on basal physiological and biochemical reactions (Matoo et al. 2021). Nevertheless, response mechanisms exist in mussels that are able to trigger trade-offs with other processes at above-optimal temperatures (Fitzgerald-deHoog, 2012; Pörtner, 2012). The magnitude and longevity of such responses have critical implications for the energy balance, fitness and ultimately survival of the organism. In mussels, such responses, whether observed in submerged or air-exposed animals, include deceleration of high energy demanding processes such as feeding, growth and/or a switch from aerobic to anaerobic metabolism (Anestis et al. 2007; Braby & Somero 2006; Tagliarolo & McQuaid 2015; Vajedsamiei et al. 2021b). Such an adaptive response minimises the difference between supply and demand of metabolic energy (Pörtner 2012). Consequently, individuals that can maintain lower metabolic rates when facing high environmental temperatures are more likely to survive under those conditions. Accordingly, in Pacific oyster (*Crassostrea gigas*) larvae, respiration rates are shown less sensitive to temperature increase than protein synthesis (Pan *et al.*, 2021). Their temperature sensitivity analysis showed that pacific oyster larvae metabolic rates have a Q_{10}value of 2.0 ± 0.15. On the other hand, they found that the Q_{10} value for protein synthesis, a major energy consuming process is 2.9 ± 0.18, and is disproportionally affected by temperature. Moreover, variation in growth rates noted among larval cohorts developed under similar rearing conditions was attributed to the differential potential for protein synthesis among families (Pan *et al.*, 2018). Such findings highlight that there is a genetic component underlying the response to thermal stress in these taxa. The genetic basis of physiological traits which allow organisms to cope with environmental stress, however, remains poorly investigated. A deeper understanding of this topic is thus required to clarify the adaptation/acclimation mechanisms of marine ectothermic species in response towards volatile climatic conditions.

To address this knowledge gap, we developed two separate cohorts of mussel spat (i.e. recently settled juveniles) from a single natural settlement event within Kiel Fjord: one cohort (half of

the settled individuals) was selected for thermal resilience, being exposed for ~48h to temperatures of 30°C., The second cohort was kept in control temperature conditions (17°C) and remained naïve to this stressor. The genetic composition of these two cohorts was then compared using a 60K SNP-array recently developed for the blue mussel species complex (Nascimento-schulze in press). Subsequently, both cohorts were exposed to various constant summer temperature scenarios over a 25-day period, with physiological performance of individual spat compared across multiple end points. We hypothesised that temperature selection would create a genetically distinct cohort, as non-resilient genotypes would have been excluded from the selected population following the initial thermal stress event. In addition, we hypothesised that thermally resilient cohort would outperform naïve individuals when re-exposed to a thermally challenging scenario.

Material and Methods

Development of selected and non-selected spat cohorts

Ca. 500,000 juvenile mussel spat (i.e. settled individuals) were collected from subtidal Kiel Fjord (Baltic Sea, Germany) in July 28, 2021 by hand from ca. 0.5m depth and gently sorted using a 1-mm mesh size plastic sieve suspended in a 10 L bucket of seawater. In July 29, around fifty thousand spat were evenly distributed between five separate 10 L tanks (~10,000 spat per tank) and kept at 17°C (temperature-controlled room at GEOMAR) with seawater wit salinity of 15 filtered through a 0.5 μm mesh (non-selected cohort). For the next 8 days, husbandry practices involved 50% water exchanges (WE) daily and the addition of 100 mL of *Rhodomonas salina* (~1-2 x 10^6 cells mL^{-1}) as food.

For selecting the thermally resilient cohort, the remaining individuals were evenly distributed among 16 x 2 L tanks (ca. 25,000-31,000 per tank) which were filled with 600 mL of seawater with salinity of 15, filtered with a 0.5 μm mesh (FSW), and maintained for 30 minutes at a non-lethal elevated temperature (26.5°C), with constant aeration supplied through 5 mL plastic pipette tips. The experimental tanks were placed into four steal water baths that maintained target temperatures with a deviation of less than 0.1°C (model Haake SWB25, Thermo Scientific). In order to initiate the thermal

selection challenge, a further 1 L of FSW (at 30°C) was added to each tank, with the temperature of the water baths subsequently increased to the same value. Target temperatures of 30°C were reached in roughly 15 minutes. Spat were then kept at this temperature for 47 hours, as we estimated survival rate would be close to our target rate of 10-20% based on preliminary small-scale experiments conducted directly before the main experiment. During this period, ~20 spat were sub-sampled three times, each time from a randomly selected tank. We observed 100%, ca. 90%, and ca. 40% survival rates by the 26th, 35th, 43th hours of the experiment. After 47 hours, heat exposure was terminated and four random sub-samples from six tanks were collected into small replicate 40 mL vials to measure survival rates after 24h recovery at 17°C. At this time point, post-heat stress survival was 16.5% (11% standard deviation).

On July 31, all heat-treated spat (dead or alive) were carefully washed in FSW to remove tissue of dead mussels and then transferred from the 16 x 2L tanks into 4 x 10L tanks with FSW (salinity 16, temperature 17°C, constant aeration) located adjacent to the tanks housing the naïve animals in a temperature-controlled room. For the first 2 days post transfer, water changes (~95% of volume) were undertaken daily to remove organic remains of dead animals. Following water changes 200 mL of *R. salina* (~1-2 x 10^6 cells mL^{-1}) were added to each tank. Over the subsequent nine days, husbandry practices shifted to 50% WE daily and the addition of 100 mL of *R. salina*. During this 9-day period heat-selected spat were stained with calcein (50 mg L^{-1} of on day 1, and 25 mg L^{-1} on day 2-9), added immediately after the water changes, to fluorescently label live individuals.

In August 11, selected spat from all tanks were mixed and three 0.5 mL sub-samples of mussels were pipetted into a 0.5 mL Eppendorf tube and the survival rates were checked by counting the percentage of empty shells present in the subsamples (8% ± 0.15). A total of 600 selected living spat were sorted from dead individuals (i.e. empty shells) using a stereomicroscope. Due to the calcein labelling procedure, living spat appeared bright green under the stereomicroscope equipped with standard GFP-filter sets (Leica MDG41). About 600 spat were randomly sorted from the ~50,000 naive spat kept in the five maintenance tanks at constant temperatures of 17°C. No mortalities were

observed in the naïve group. Each 600 naïve or selected spat were equally distributed among four 50 mL containers (150 individuals in each container, 4 containers per spat line), which were closed with mesh (300µm). Subsequently, two containers, one containing selected and one containing the naïve spat were kept in one 10L aquarium for 14 days, fed every day with *R. salina* with an initial concentration of ca. 7000 cells mL^{-1}, to ensure comparable condition between spat from the two different cohorts for subsequent experiments. In the following sections of this manuscript, spat surviving the thermal stress exposure are referred to as 'selected' (S), whilst spat not exposed to thermal stress are termed 'non-selected' (NS).

Thermal challenge in indoor benthocosm facilities

Experimental design

Following the initial development of S and NS spat cohorts, 960 individuals (480 S and 480 NS) were randomly selected for a follow-up experiment, in which we assessed the implications of thermal stress on juvenile performance. Beginning on August 20, spat were exposed to 5 different temperature treatments (21, 22, 23, 24, 25 and 26°C) over a 25-day period using the indoor Kiel Indoor Benthocosms system (Pansch & Hiebenthal 2019) (Figure 4.1). The 26°C treatment represented an end-of-century average daily marine heatwave temperature in the study region (Gräwe et al. 2013), sub-lethal to mussels for periods of <1 month (Vajedsamiei *et al.*, 2021). The 21°C treatment represents a present-day maximum summer temperature (Wolf et al. 2020) which is in the optimal range for growth of Baltic Sea mussel populations (Vajedsamiei *et al.*, 2021). The intermittent temperatures within the experimental range (22, 23, 24, 25 °C) were selected to identify any subtle, but significant, alterations in thermal performance or thermal breakpoint performance in this study. In this facility, 12 x 600L polythelyene tanks were connected to an aquarium control system (GHL Gmbh), coupled with heating and chilling devices, which enabled the control and recording of water temperature remotely. Animals were placed into 14 L plastic cylindrical tanks, maintained inside a 300 µm mesh cage, allowing water flow and food uptake to occur. Twenty juveniles were kept in each of the cages, with each treatment (S and NS) having two replicate cages per experimental tank, with

cages tied to a 15 mL pipettes (used for aeration) at the same water column height, providing similar environmental conditions between replicates of each treatment. One additional cage per treatment (S, NS), containing 12 individuals per cage, was added to each of the cylindrical tanks, to be sampled halfway through the course of the experiment. Finally, each of the 14 L experimental tanks was placed inside an individual benthocosm unit, and 2 benthocosms were allocated to each temperature treatment.

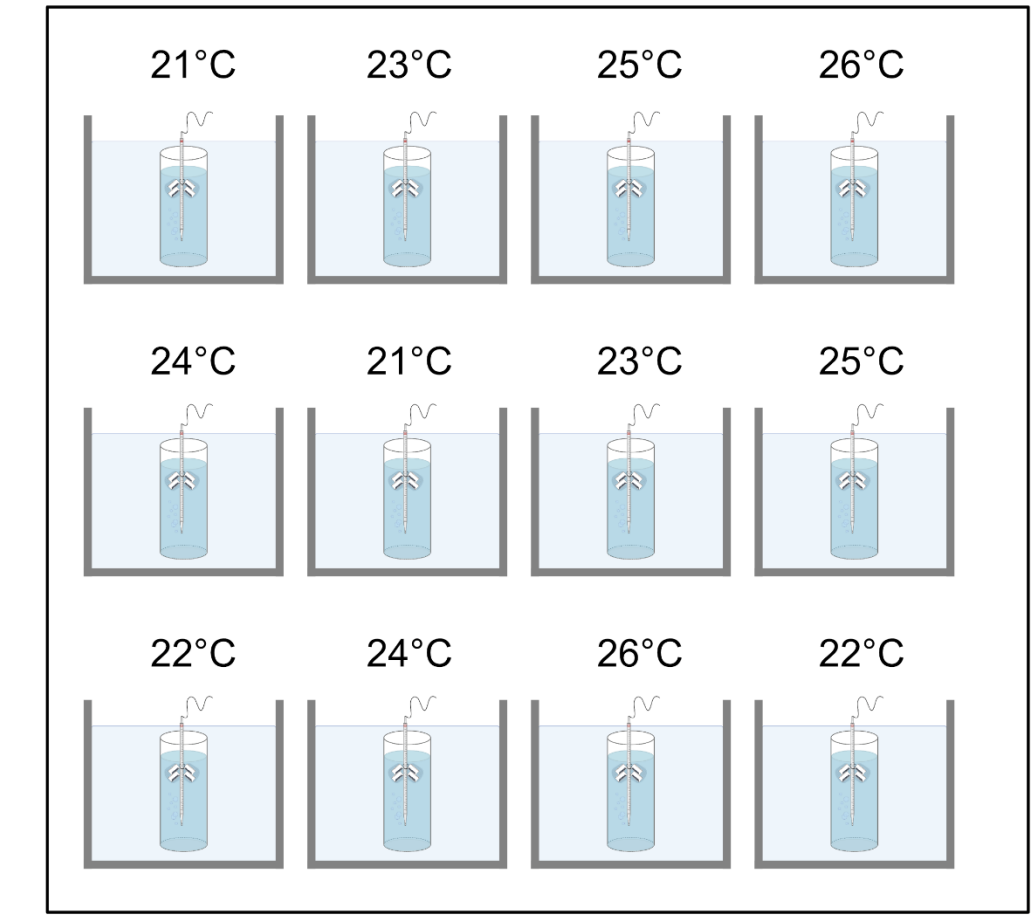


Figure 4-1: Scheme of the 25-day thermal challenge setup. In (A) an overview of the display of the 12 mesocosms used in this experiment in temperature-controlled room. Segment (B) of the figure represents the setup of an individual mesocosm controlling the temperature of a 14 L transparent plastic cylinder, with aeration provided by a 50 mL plastic pipette and four cages attached to it. In segment (C) a detailed demonstration of the individual cage, open with a 200 micron mesh to allow water and food circulation to spat placed inside the cage.

Husbandry

Throughout the length of this experiment, spat were fed with monocultures of *Rhodomonas salina*, added at a concentration of ~ 7,000 cells mL^{-1} (Riisgård et al., 2011) to the experimental cylinders after every water exchange. Cell concentration was measured daily (particle size 5-8 um) using a Coulter counter (Coulter Z2, Beckman Coulter GmbH), in all experimental tanks. If *R. salina* concentration in any of the experimental tanks dropped to threshold values in the range of 1,000 cells mL^{-1}, full water exchanges were performed in all the tanks in order to ensure that *R. salina* was constantly available in the water column at optimal concentrations. Water changes were conducted with FSW (filtered across a series of filters 50, 10 and 0.1 micron mesh), were undertaken every 3rd day (days 1-7), every other day (days 8-20), and subsequently daily after day 21 (Suppl. Table 4.1).

Growth, Dry Body Mass and Shell Mass

Initial shell length (SL) was measured in a sub-sample of 60 S and 60 NS individuals, which were removed from the conditioning tank before being distributed into experimental cages on day 0, the day prior to the start of the 25-day thermal exposure experiment. Following 13 days of exposure to different temperature regimes, the impact of elevated temperature on mussel spat was assessed by measuring individual's performance via growth as SL, dry tissue weight (TW) and dry shell weight (SW) in at least ten animals from each cage from those animals placed in the mid-experiment sampling cages. TW and SW were measured as described in Sanders *et al.* (2018) Briefly, sampled spat were killed in a microwave (400 watts, 30 s), and their soft tissue was removed from their shell under a stereomicroscope. Shell and soft tissue were individually placed into pre-weighed tin foil boats dried at 80°C for at least 72h and weighed to determine the total TW and SW (mg) by subtracting the value of the empty foil from the foil containing shell or soft tissue. Results are presented as mean ± standard deviation. Finally, the same measurements were taken at day 25, when the challenged ended, from at least 10 animals from each cage. Results are presented as mean ± standard deviation. Juvenile

performance in terms of DW SL and TW measured during mid experiment are presented in the supplementary material (Suppl. Fig. 4.1).

Assessing structure and genomic relatedness of the two cohorts

The initial genetic composition of S and NS cohorts was analysed using 50 S and 50 NS spat collected on day 0 of the experiment, the day prior to the start of the 25-day exposure challenge, and preserving them in 90% ethanol. Samples were sent to Identigen Ltd for DNA extraction and sequencing, using the 60K blue mussel SNP-array (Nascimento-Schulze et al. 2023).

A quality control filter for marker call rate (CR) of > 95% and sample CR >90% was applied using PLINK v1.9 (Purcell et al. 2007). We subsequently applied a minor allele frequency (MAF) filter < 0.01.

A principal component analysis (PCA) was applied to genotypes generated with the array, pruning out putatively linked loci using a window size of 50 Kpb, a step size of 10 Kbp and an r^2 threshold of 0.1 with PLINK v1.7 (Purcell et al. 2007), with no variants removed in this step. Subsequently, we explored introgression in spat by applying an admixture analysis (admixture v1.3) (Alexander & Lange 2011). In this analysis, the best fitting number of clusters representing the ancestry of a population is estimated where *k* has the lowest cross-validation error value. For this study, we tested *k* values between 2 and 10. In order to allocate a species to the ancestry cluster generated in this first admixture analysis, we ran an additional admixture analysis including 5 additional populations, Kiel (Baltic Germany, GK), Ahrenshoop (Baltic Germany, GA), Finland (Baltic Finland, FIN), Bude (North Devon coast of the UK) and Budleigh (South Devon coast of the UK). These populations have been genotyped with the blue species mussel multi-species 60K array in previous studies (Nascimento-Schulze et al. 2023; Nascimento-Schulze et al. unpublished data), and their ancestry is known to be *M. edulis/M. trossulus* hybrids for GK and GA and FIN populations, with an increased frequency of *M. trossulus* genotypes in FIN, *M. galloprovincialis* ancestry dominating Bude and *M. edulis* prevails in Budleigh. Admixture results of this data set are presented in Supplementary

Table 4.2 and visualised in Supplementary Figure 4.2, 4.3. All data visualisation was performed in R v4.0.0 (R Core Team 2021).

Following variant quality control, we used fixation index analysis (F_{ST}) to measure the proportion of genetic differentiation among the two cohorts (S and NS) caused by genetic structure. In this analysis, values of 0 imply no genetic differentiation among populations (panmixia), and values of 1 representing complete differentiation between populations. We analysed F_{ST} using plink v1.7 and –fst flag, filtering individual and marker CR of >10% using the options –mind and –geno.

Statistical analysis

An ANOVA analysis was applied to test for the effects of temperature selection on the SL of individuals at day-0, prior to the start of the 25-day thermal challenge. We applied a Shapiro–Wilk test to check that the data were normally distributed and the Levene's test to check for homogeneity of variances.

In addition, we investigated the effects of temperature on the TW, SW and SL of S and NS individuals following the 25-day thermal challenge previously described. The null hypotheses stated that observed variation in TW, SW, or SL growth rate would not be caused by the main effects of the *temperature* or *selection* treatments, or their interaction. These relations were modelled as generalised additive mixed models (GAMMs) using *gam* function from *mgcv* package (Wood 2006). In the GAMMs, *temperature* was defined as a smooth-effect predictor (with linear and/or nonlinear effects) and *selection* as an ordered factor. Then, using t-statistics and f-statistics (the Wald test), the intercept (or the mean) and the slope or nonlinearity (effective degrees of freedom, *edf*) of the reference level smoother (*non-selected*) were each compared to zero, respectively; and, the treatment level (*selected*) smoother's estimates were then compared to those of the reference level (Wood 2006). In addition, the random intercept effects of *water bath* and *cage,* as possible causes of residual dependence, were included, and for unbiased estimation of variance components, GAMMs were fitted using restricted maximum likelihood (Wood et al. 2016).

The models were first fitted with the *identity* link function and the assumption of residuals being gaussian distributed around the predicted means. The assumptions regarding the distributions of residuals were checked via the *DHARMa* package (Hartig 2018). As it was determined that the residual assumptions were violated, we fitted the models assuming the scaled t-distribution (*scat*) of residuals and found this assumption to be valid. For SL, the log_{10} link function was also needed to validate the residual assumption. In terms of the Akaike Information Criterion (AIC), the scat models were also superior to the gaussian models. We used the *predict* function from *car* package (Fox & Weisberg 2018) to predict the means and confidence intervals of responses to predictors. A p value of < 0.05 was assumed as the significance level for all analyses. All Analyses were applied in Rstudio v2022.07.1+554 (RStudio Team 2021) and R v4.2.1(R Core Team 2021).

Results

Impacts of increased temperature on Shell length, Tissue dry weight and Shell mass

No significant difference in SL was observed between S and NS lines at day-0, prior to the start of the 25-day thermal challenge (Suppl. Figure 4.4, Suppl. Table 4.3, $p > 0.05$).

Exposing spat to different experimental temperatures over a 25-day period did not result in a significant difference in SL between S and NS lines (Figure 4.1, Table 4.1, $p > 0.05$). The mean SL of S and NS lines was 2.33 ± 0.52 mm and 2.47 ± 0.67 mm, respectively. However, temperature had a significant impact on SL of NS spat (Figure 4.1, Table 4.1, $p = 0.0001$). The average SL of NS juveniles 26°C (2.04 mm ± 0.45) was significantly decreased in comparison to NS juveniles kept at 21°C (2.37 ± 0.49 mm). Average SL of S individuals was also significantly affected by temperature (Table 4.1, $p > 0.05$). The form of the temperature effect (represented by effective degree of freedom, *edf*) on SL of two spat lines was similar. Besides, a random-intercept effect of cage was present in the SL of individuals.

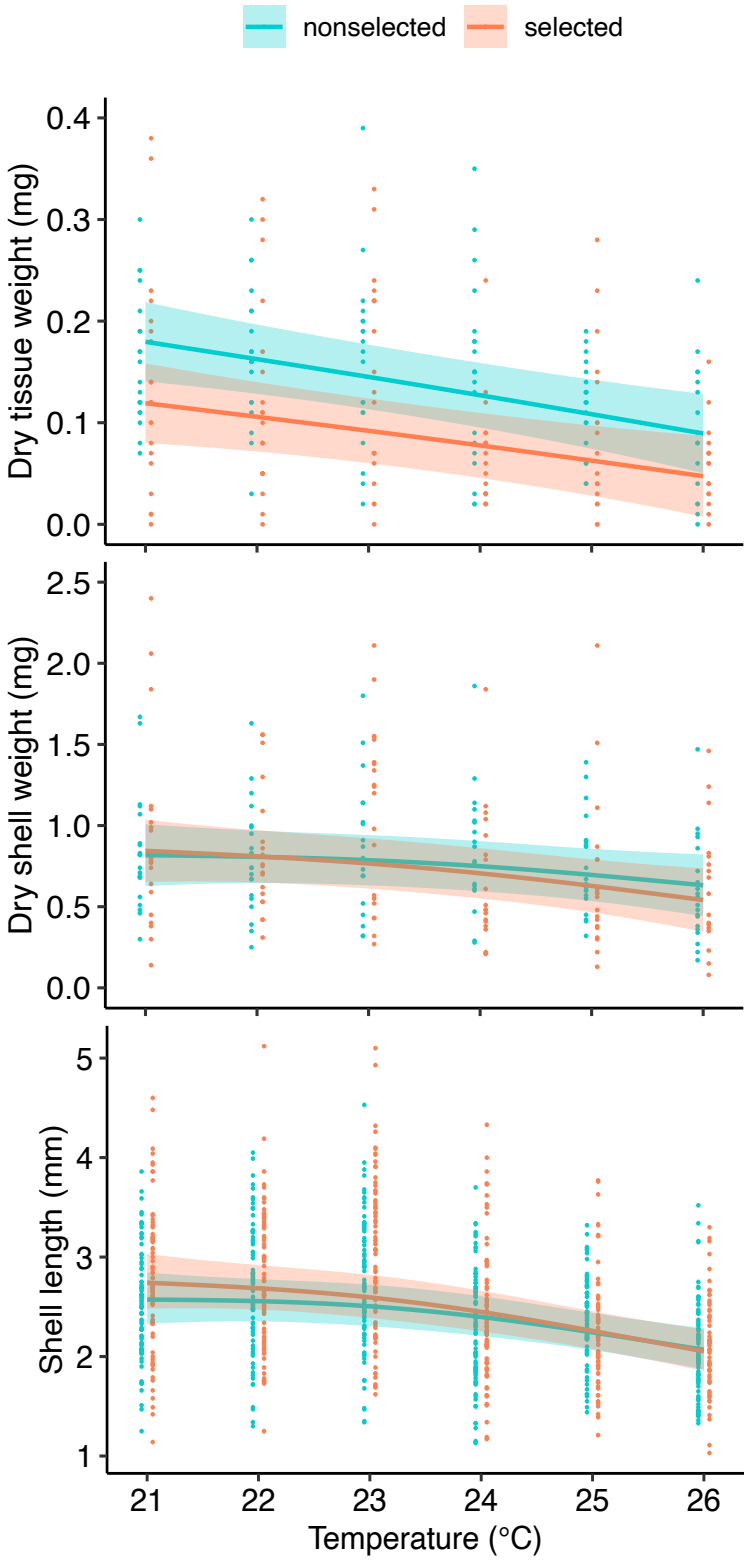


Figure 4-2: Final dry tissue weight, dry shell weight, and shell length of selected (S, red) and non-selected (NS, cyan) spat across experimental temperatures (21-26°C), following the 25 days of exposures to different temperatures. Each of the data points represents the measurement of an individual spat. These relations were modelled as additive mixed models assuming scaled t-distributions of residuals. Solid lines represent predicted means and shades represent 95% confidence intervals.

A significant difference in TW among spat lines was found: on average TW of S individuals (0.10 ± 0.09 mg) was significantly lower than those measured in NS ones (0.14 ± 0.10 mg) (Figure 4.1, Table 3.1, p = 0.0001). At 21 °C, larvae NS showed a TW of 0.16 ± 0.06, mg whilst at 26 °C the average TW was lower (0.06 ± 0.10 mg), indicating a decrease in DW with increasing temperatures. A similar form of temperature effect (*edf*) was observed for S individuals. Though, for S spat, DW of 0.14 ± 0.11 mg in 21 °C decreased to 0.05 ± 0.15 mg in 26°C (Table 4.1, p > 0.05). Selected individuals acclimated to lower temperatures showed a TW of 0.5

Finally, neither spat line nor temperature factors significantly impacted SW of spat in this experiment (Figure 4.1, Table 4.1. p > 0.05), with SM of S spat (0.79 ± 0.5 mg) and NS spat of (0.77 ± 0.36 mg).

Table 4-1: main and interactive effects of the temperature and selection treatments on dry tissue weight (TW), shell weight (SW) or shell length (SL) growth rates of *Mytilus* spat. These relations were modelled as additive mixed models assuming scaled t-distributions of residuals. The intercept (or the mean) and the slope or nonlinearity (effective degrees of freedom, edf) of the reference level smoother (naive) were compared to zero, respectively; and, the treatment level (selected) smoother's estimates were then compared to the reference level ones. In addition, the random intercept effects of cage were tested. The significant impact is considered for p-value ≤ 0.05.

	A. parametric coefficients	**Estimate**	**Std. Error**	**t-value**	**p-value**
	Intercept (non-selected)	0.1418	0.0087	16.3570	< 0.0001
	Intercept (selected)	-0.0459	0.0115	-3.9794	0.0001
TW	**B. smooth terms**	**edf**	**Ref.df**	**F-value**	**p-value**
	s(temperature, non-selected)	1.6596	1.8153	9.6515	0.0001
	s(temperature, selected)	1.3257	1.5142	0.2626	0.5926
	s(cage)	4.6894	22.0000	0.2925	0.1261
	A. parametric coefficients	**Estimate**	**Std. Error**	**t-value**	**p-value**
	Intercept (non-selected)	0.7492	0.0392	19.0895	< 0.0001
	Intercept (selected)	-0.0315	0.0510	-0.6189	0.5360
SW	**B. smooth terms**	**edf**	**Ref.df**	**F-value**	**p-value**
	s(temperature, non-selected)	1.4395	1.6167	2.0979	0.1726
	s(temperature, selected)	1.0001	1.0002	0.6174	0.4320
	s(cage)	6.0661	22.0000	8.6526	0.0832
	A. parametric coefficients	**Estimate**	**Std. Error**	**t-value**	**p-value**
	Intercept (non-selected)	0.3777	0.0081	46.3824	< 0.0001
	Intercept (selected)	0.0126	0.0068	1.8364	0.0663
SL	**B. smooth terms**	**edf**	**Ref.df**	**F-value**	**p-value**
	s(temperature, non-selected)	1.7442	1.7886	13.3916	0.0006
	s(temperature, selected)	1.0005	1.0010	2.3425	0.1260
	s(cage)	16.8297	23.0000	80.1398	< 0.0001

Genotyping

From the 15,049 markers with a CR > 95%, a total of 11,288 variants were kept for the analysis of genetic diversity between selected and non-selected lines (PCA, Admixture and F_{ST} analysis) following the quality control filtering steps.

The PCA analysis revealed no obvious stratification between S and NS cohorts (Figure 2). In this analysis, PC1 explained 20.7% of the variance, whilst PC2 and PC3 explained 17.5% and 7.87%, respectively.

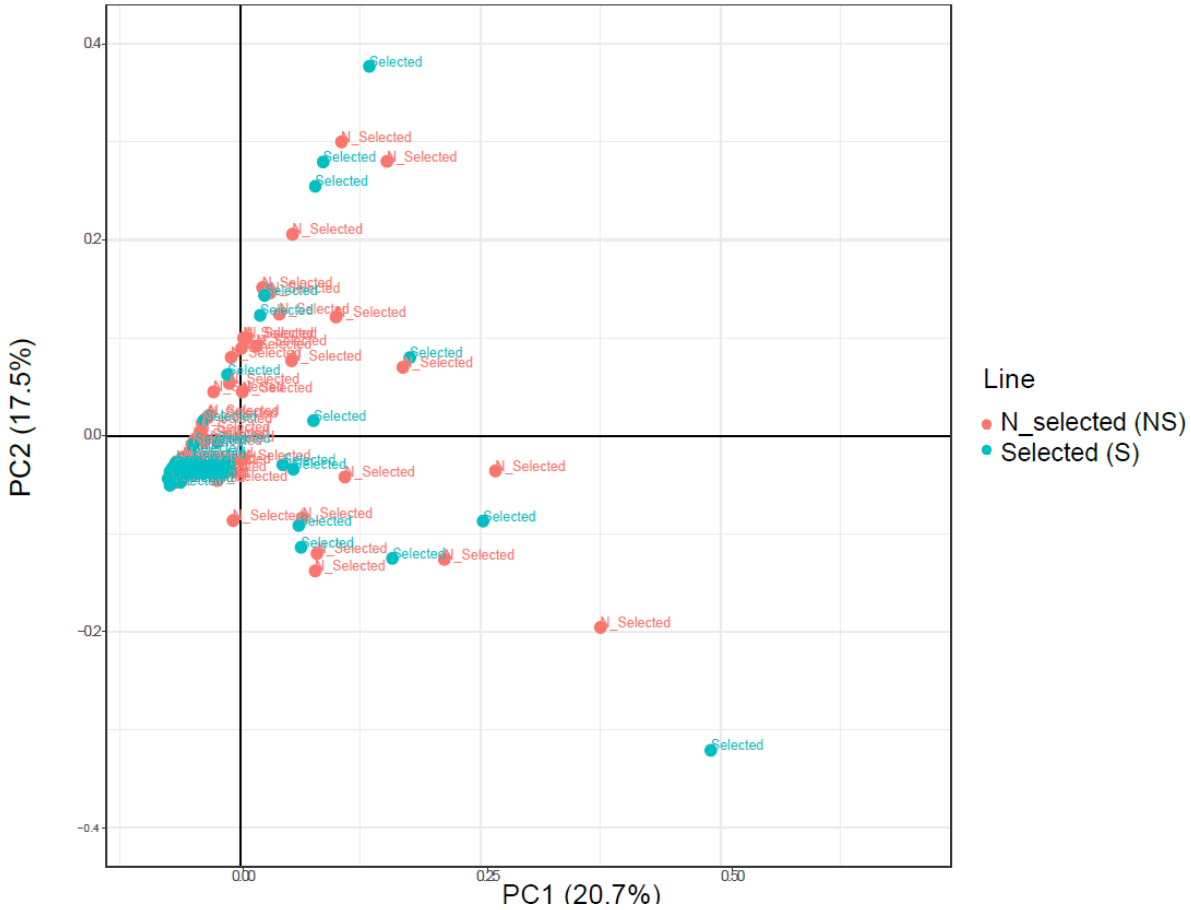


Figure 4-3: Scatter plots of individual variation in PC 1 and 2 scores resulting from PCA applied to the blue mussel juveniles from Kiel, Germany. Selected (S) and Non-selected (NS) individuals were genotyped with the Blue mussel multispecies 60K SNP array. The proportion of overall variation explained by each PC are given in percentages.

Ancestry was explored in spat from each of the two cohorts with admixture (Figure 3). Following evaluation of cross-validation error for each tested *k*, the most likely number of ancestry

clusters in our dataset was three, corresponding to the three *Mytilus* species previously described in the area (i.e. *M. edulis*, *M. trossulus* and *M. galloprovincialis*) (Vendrami et al. 2020).

Results from the admixture coefficients (Q) inferring the three clusters are presented in Figure 3 where each column represents a different individual, and the proportions of the different colours indicate the probabilities of being admixed between two (or more) clusters. Based on the analysis including the control populations (Suppl. Fig 2,3) we could infer the species for each of the three clusters present in the S and NS cohorts, with dark-blue cluster representing *M. edulis* ancestry, light-blue *M. trossulus* and yellow *M. galloprovincialis*.

In this analysis, we observe that most of the genotypes are allocated to *M. edulis*, followed by *M. trossulus* and finally *M. galloprovincialis*. However, the contribution of *M. edulis* to the selected cohort is greater than to the non-selected group, with a higher number of individuals being entirely allocated to this cluster. Hybrids between *M. edulis* and *M. trossulus*, and/or *M. galloprovincialis*, were more prevalent in the non-selected line.

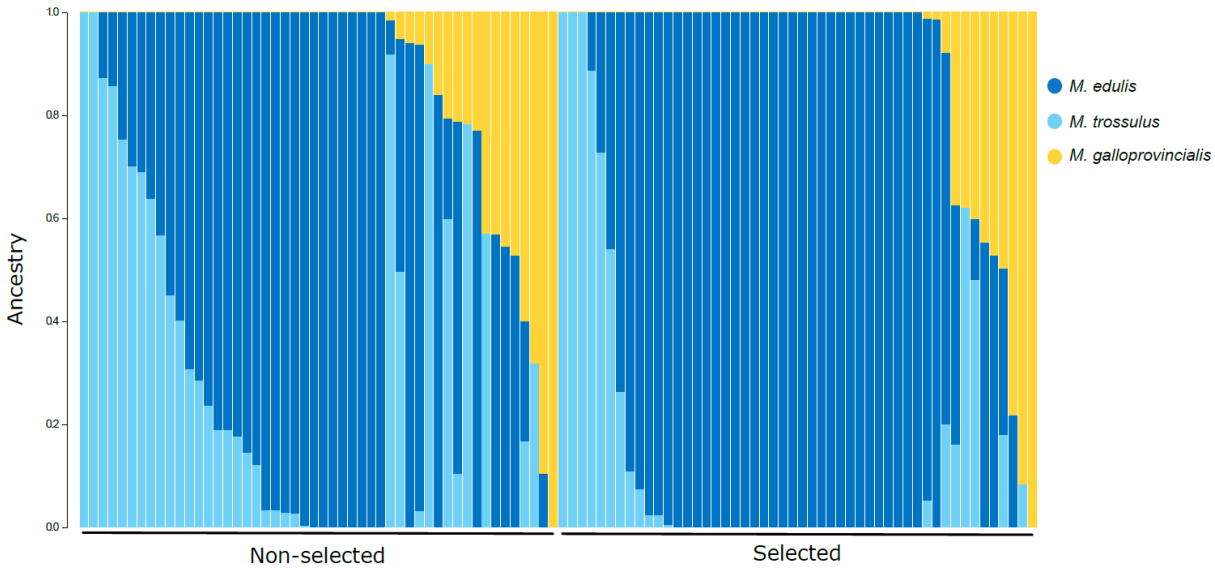


Figure 4-4: Results of genetic admixture analysis and ancestry inference. Cluster membership coefficients (Q) where each individual is represented by a column partitioned into segments of different colour, the length of which indicate the posterior probability of membership in each cluster, varying from 0 to 1. Solid bars represent an individual from a single species background.

Finally, to assess similarity between these two cohorts, we explored the F_{ST} levels of the variants retained following quality check. The average F_{ST} between S and NS cohorts was 0.006 meaning that the genetic differentiation among these two lines explained approximately 0.6% of the variation in allele frequencies across all markers. This analysis also allows the identification of outlier markers which strongly deviate between these two cohorts. Four SNPs, AX-603613715, AX-603257004, AX-603349307 and AX-603466751 (Suppl. Figure 4.5), had an F_{ST} value higher than 0.15, meaning that their allele frequencies are strongly differing between these populations. Fourteen SNPs had an F_{ST} value higher than 0.1.

Discussion

In this study, we aimed to assess the impact of elevated summer seawater temperatures on the performance of lines of high-temperature selected and non-selected Baltic juvenile *Mytilus spp*. We hypothesized that early exposure to high temperatures would select individuals more resilient to this stressor. Two lines of mussel spat were developed: spat were either selected via a simulated extreme heat wave (30 ^{o}C for 47h) (S), or were naïve to this stressor and exposed to an average summer temperature scenario (17 ^{o}C) (NS). These two lines were then exposed to range of constant high temperatures over a 25-day period (21, 22, 23, 24, 25 and 26 ^{o}C). TW, SL and SM measurements were used as a proxy for fitness. The second aim of this study was to investigate whether thermal selection could generate genetically distinct cohorts. Here we hypothesized that selection towards thermal stress would generate two genetically distinct cohorts, as weak genotypes, would be removed from populations during the selection event.

Impacts of increased temperature on Tissue dry weight, Shell length and Shell mass

Our results provide insights that pre-selection via a temperature challenge in *Mytilus spp* spat may lead to differential performance between the two developed mussel lines, in terms of TW, when those were subsequently exposed to elevated temperatures. At 13 days following the start of thermal,

no significant difference was observed between S and NS lines in terms of TW. However, at day 25, TW of NS individuals (0.14 ± 0.10 mg) was significantly higher than S ones (0.10 ± 0.09 mg).

Temperature is a major factor regulating the physiology of ectotherms (Hochachka and Somero, 2002). In these organisms, the relationship between physiological performance and body temperature can be described by an asymmetric thermal performance function (Angilletta et al. 2002). Positive effects of increased temperature on performance, such as increased growth, are typically observed when those temperatures fall within the tolerance range of a species/population, because of positive thermodynamics on enzymatic activity. Consequently metabolic rate rises, and an organism increases its respiration rate and feeding rate to supply the enhanced energy demand. Organisms possess the ability to adapt (or acclimate) to their thermal surroundings: with organisms living in tropical environments possessing a lower resting metabolic rate in comparison to those living in colder environments (Clarke 2003). However, when environmental temperatures exceed the limits of thermal tolerance of a species/population, the theory of oxygen and capacity limited tolerance (OCLTT) proposed by Pörtner (2010) suggests this stressor impacts the metabolic scope of an individual (i.e. difference between basal and maximal metabolic rate), limiting the supply of oxygen to tissues. As temperature rises, the mismatch of oxygen supply and demand enhances, unbalancing the metabolism and generating stress. Over longer periods, this temperature-induced mismatch between metabolic supply and demand ultimately hinders performance (Pörtner, 2012; Ritchie, 2018). In marine molluscs, metabolic suppression is an alternative mechanism, which enables individuals to reduce their energetic requirement, lowering the burden of energetically costly processes (Vajedsamiei et al. 2021b). This approach reduces the mismatch between energy uptake and demand, allowing organisms to cope with higher temperatures, as individuals allocate surplus energy to the maintenance of basic functions (Schulte et al., 2011). The significantly lower TW in S lines suggests that the thermal stress experiment selected organisms which have a lower metabolic rate. Furthermore, these organisms were not significantly different in terms of SL in the start of the experiment. If this is the case, according to the OCLTT theory, S individuals are better prepared to

control their resource uptake and internal energy allocation than the NS lines. Consequently, S lines may be better able to survive for prolonged periods under sub-optimally elevated temperatures than their NS counterparts, but at the expense of reduce growth performance. To confirm this hypothesis, longer-term exposures would be required to elucidate larval survival of both cohorts under warming scenarios. Moreover, initial TW was not measured in this experiment as the number of S and NS individuals for running the thermal challenge was limited. The lack of this measurement hinders our ability to prove that significant differences in TW between the two lines are caused by selection and the 25 day challenge rather than possible handling during the 11 day period between the selection event and the start of the challenge. However, measurements taken at day 13 of the thermal challenge show that there was no significant difference in TW between both lines.

In our study, no significant difference in SL between S and NS individuals was observed during the 25-day challenge. However, temperature had a significant impact on SL of NS individuals, in which this measurement decreased with increasing temperature. The costs of calcification in juvenile mussels are high, accounting for approximately 32% of the energy budget in organisms living in 18^{o}C (Sanders et al. 2018). The OCLTT theory (Pörtner, 2010) suggests that the mismatch between metabolic supply and demand of metabolic energy, caused when an individual faces unfavourable temperature conditions, hinders its performance. Environmental stress, such as osmotic stress, can modify the dynamic energy budget of this species (Maar et al. 2015). Accordingly, thermal stress has been shown to modify branchial metabolism, respiration rates and standard metabolic rates of mussels (Matoo et al. 2021). Furthermore, increasing temperatures can limit the energy supply of individuals, represented by the limitation of oxygen supply, leading to anaerobic metabolic accumulation, increased partial pressure of CO2 in haemolymph of mussels (Zittier et al. 2015). Consequently, less energy is available to non-vital processes, such as shell growth. Potentially, S lines in this study are metabolically suppressed and better able to cope with the higher experimental temperatures used in this study (Vajedsamiei et al. 2021b). Therefore, the negative effect of temperatures in the SL becomes more obvious in the NS line, which reach more rapidly their thermic

threshold. Nonetheless, both lines showed decreased feeding behaviour when facing increasing experimental temperatures (data nor presented). Perhaps, significant differences in SL could be observed in both lines under a prolonged exposure to thermal stress, given S individuals would require a longer period to grow. Conversely, calcification might be prioritised over soft tissue production, a pattern observed in response to parasitic infection (Bommarito et al. 2022).

The measured response to elevated temperature in larval mussels demonstrated here suggest that these levels of environmental thermal stress exceed the thermal optima for this species, with individuals occupying the lower end of their performance curve (Angilletta et al. 2002). As previously mentioned, these are temperatures which the Baltic benthic community are likely to face during summer months in the near future (Pansch et al. 2018). Blue mussels are mass spawners, and temperature is a major factor regulating gametogenesis and spawning events in these taxa (Gosling 2021). Consequently, an increase in mean SST, as well as extreme weather events, will have major impacts on the reproductive cycle of mussels. In other Baltic taxa, warming can lead to a strong shift in their annual reproductive cycle (Sommer & Lengfellner 2008). Furthermore, larval and spat development occurs during the warm months of the year, starting during late spring and carried throughout summer. As a result, blue mussels will be highly likely to experience the highest annual mean SSTs, or heat waves events, during their early developmental stages. Our findings suggest that the projected changes to environmental parameters in the Baltic Sea can hinder the performance of juveniles. Such impacts will likely have consequences for recruitment and subsequent distribution of blue mussels in this area.

Impacts of increased temperature on the genetic diversity of mussel spat

We additionally investigated whether exposure to thermal stress can modify the genetic structure of juvenile blue mussel populations. Here, we compared the genetic composition of these two spat cohorts using a recently developed multi-species 60K SNP-array for blue mussels (Nascimento-Schulze et al. 2023) to assess if pre-selection could generate genetically distinct cohorts.

Both cohorts originated from the same spat fall event, and were collected during July 2021 in Kiel Fjord.

Whilst the PCA output does not present any clear segregation among populations, results from the admixture analysis are in accordance with previously publish data in the Baltic Sea, with *M. edulis* genotypes dominating mussel populations in the Kiel Fjord, and the introgression with *M. trossulus* genotypes (Knöbel et al. 2021; Stuckas et al. 2017; Vendrami et al. 2020). In addition, some level of introgression with *M. galloprovincialis* genotypes has been recently described (Vendrami et al. 2020). Nonetheless, the contribution of *M. galloprovincialis* genotypes in spat analysed in this chapter is considerably higher than has been previously described in this region (Vendrami et al. 2020). In contrast to the mentioned study, which genotyped adult mussels, in our study, young spat (approximately two-months old) were genotyped. Therefore, one possibility to explain the loss of *M. galloprovincialis* genotypes in adult mussels in Kiel Fjord is that post-settlement selection pressures are acting resulting in a loss of alleles from this species, so that *M. edulis* and *M. trossulus* hybrids become dominant in adult populations. Larval selection due to low salinity has been demonstrated for hybrids between *M. edulis* and *M. trossulus* collected in Ahrenshoop, at the Southeast coast of the Baltic Sea (Knöbel et al. 2021). In this study, larvae were exposed to different salinity conditions (7, 9 and 11 PSU), and the frequency of *M. trossulus* genotypes was higher in settled spat acclimated to low salinity in comparison to those acclimated to higher salinity treatments. Although comparing different life stages, such observations point out that environmental factors can play a role in genotype frequency and thus, species distribution in these taxa. The Baltic Sea is a unique environment due to its salinity gradient and limited water admixture with the North Atlantic through the Danish Straight. *M. galloprovincialis* genotypes might be fed into this region either by water currents and/or carried by ballast water or attached to other recreational vessels. It is possible that environmental conditions in this region do not allow *M. galloprovincialis* genotypes to thrive, with these being lost during development.

Whilst less is known concerning the effects of temperature on blue mussel species distribution, our results suggest that S individuals presented a contribution of *M. edulis* genotypes of 73 ± 39 %, followed by *M. trossulus* (16 ± 30 %) and *M. galloprovincialis* (11 ± 25 %) to its ancestry. In the NS cohort, ancestry of *M. edulis* genotypes was lower (58 ± 37 %), with an increased contribution of *M. trossulus* genotypes (29 ± 34 %) and *M. galloprovincialis* (13 ± 24 %). Intra and inter species variation in thermal tolerance exists within this species complex (Fly and Hilbish, 2013), and the upper thermal tolerance is partially dependent on the genetic background of populations (Zippay and Helmuth, 2012). Although our genetic analysis did not demonstrate any strong patterns of genetic structure being impacted by selection in this experiment, a subtle shift towards *M. edulis* genotypes (or against *M. trossulus* genotypes) was observed in the S lines. Such an observation might be an indicator that this species can better cope with high temperatures, negatively impacting *M. trossulus* genotypes. *M. galloprovincialis* is known to cope well with high temperatures, colonising the Mediterranean Sea, an environment with high annual levels of salinity and temperature conditions. The loss of *M. galloprovincialis* genotypes observed in this study may result from the synergistic impact of low salinity and high temperature stressors in the spat. Furthermore, four markers were shown to be significantly divergent among the two lines (F_{ST} > 0.15), and a higher proportion of *M. trossulus* could be observed in the NS lines. Whilst these differences were not strong enough to assume a significant genetic segregation among these two lines, they provide insights on the impacts of rising temperatures on the genetic structure of Baltic mussels. Such factors can underlie the differential performance observed between S and NS lines.

Other factors might have influenced the lack of genetic differentiation between both cohorts observed in this study. First, spat sampled in our study were collected directly from the wild, with no control of relatedness among individuals, and no information on parental genotypes. A more controlled study design, with parental genotypes being recorded, or possibly the development of "thermally resilient lines" over generations, would enable the in depth investigation of the impacts of thermal stress on the genetic structure of these populations, minimising the noise generated by other

genetic and environmental variables present in wild populations. Second, only a low number of individuals were genotyped (50 individuals per line). A greater number of samples would enable a reliable genome wide association study (GWAS) to be undertaken, as GWAS depend on large sample sizes (thousands of individuals) to correctly identify reproducible genome-wide significant associations with a given trait (Uffelmann et al. 2021). Therefore a slightly altered study design would allow the relationship between thermal resilience among the different species of the complex to be better elucidated, as well as the presence of SNPs which contribute significantly to such resilience to be identified with greater confidence. Third, it is important to highlight that this study was the first attempt to apply the recently developed blue mussel SNP array on a natural population to investigate possible signatures to environmental selection in these taxa. The results presented in this chapter are, therefore, preliminary results, although not less relevant in their significance. In light of climate change, I hope to inspire future research investigating genomic processes underlying environmental plasticity and adaptation in mussels. I hereby provide the first indication that temperature might play an important role in genotype distribution of blue mussels in Baltic Sea populations. Nonetheless, future studies could benefit from using a greater number of selected generations, of samples per replicate and a longer duration of the experiment to understand how facing environmental stress in early developmental stages influences adult performance.

Conclusions and Future perspectives

Our results show that selected individuals are equally able to grow shells in comparison to NS individuals, but suggest that these individuals may be less able to gain tissue weight. Besides, this is the first study to characterise in depth the genetic composition of spat in the Kiel, Southern coast of the Baltic Sea, using a genomic approach. Whilst the results of this study are largely preliminary, as the findings are based on the performance of a single generation of hybrids between three mussel species. Differences in the performance and the genetic composition of selected and non-selected cohorts indicates an important role of temperature in this species complex that requires further

investigation. Climate change is projected to significantly modify oceanic physiochemical parameters globally, which have been relatively stable for millennia, with knock on implications for marine ecosystems. The Baltic Sea has been identified as a region that will very likely be severely impacted by such changes. Future studies are necessary to provide in depth understanding and quantification of the physiological and genetic mechanisms underlying the response of blue mussels to thermal stress and their ability to acclimate and/or adapt to such stressor. Blue mussels are species of high-economic and ecologic value, thus, guaranteeing that both natural and commercial stocks can thrive under new environmental conditions is a matter of high relevance. Combining the use of a blue mussel genomic toolbox with knowledge of species-specific physiological mechanisms could fast-track the development of thermally resilient lines.

References

Ahmed HGM-D, Naeem M, Zeng Y, Rashid MAR, Ullah A, Saeed A, et al. (2022). Genome-wide association mapping for high temperature tolerance in wheat through 90k SNP array using physiological and yield traits. *Plos one* **17**: e0262569.

Alexander DH, Lange K (2011). Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics* **12**: 246.

Anestis A, Lazou A, Pörtner HO, Michaelidis B (2007). Behavioral, 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**: R911–21.

Angilletta MJ, Niewiarowski PH, Navas CA (2002). The evolution of thermal physiology in ectotherms. *Journal of Thermal Biology* **27**: 249–68.

Avdelas L, Avdic-Mravlje E, Borges Marques AC, Cano S, Capelle JJ, Carvalho N, et al. (2021). The decline of mussel aquaculture in the European Union: causes, economic impacts and opportunities. *Reviews in Aquaculture* **13**: 91–118.

Bangera R, Correa K, Lhorente JP, Figueroa R, Yáñez JM (2017). Genomic predictions can accelerate selection for resistance against Piscirickettsia salmonis in Atlantic salmon (*Salmo salar*). *BMC genomics* **18**: 1–12.

Belkin IM (2009). Rapid warming of Large Marine Ecosystems. *Progress in Oceanography* **81**: 207–13.

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

Bommarito C, Khosravi M, Thieltges DW, Pansch C, Hamm T, Pranovi F, et al. (2022). Combined effects of salinity and trematode infections on the filtration capacity, growth and condition of mussels. *Marine Ecology Progress Series* **699**: 33–44.

Braby CE, Somero GN (2006). 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**:

2554–66.

Carboni S, Evans S, Tanner KE, Davie A, Bekaert M, Fitzer SC (2021). Are Shell Strength Phenotypic Traits in Mussels Associated with Species Alone? *Aquaculture Journal*, **1(1)**: 3-13.

Clarke A (2003). Costs and consequences of evolutionary temperature adaptation. *Trends in Ecology & Evolution* **18**: 573–81.

Corrochano-Fraile A, Davie A, Carboni S, Bekaert M (2021). Evidence of multiple genome duplication events in *Mytilus* evolution. *bioRxiv* 2021.08.17.456601.

Elmgren R, Hill C (1997). Ecosystem function at low biodiversity—the Baltic example. *Marine biodiversity: patterns and processes* 319–36.

FAO (2019). *The State of the World's Aquatic Genetic Resources for Food and Agriculture. Commission on Genetic Resources for Food and Agriculture.*

Fly EK, Hilbish TJ (2013). Physiological energetics and biogeographic range limits of three congeneric mussel species. *Oecologia* **172**: 35–46.

Fox J, Weisberg S (2018). *An R Companion to Applied Regression* Sage publications

Gallardo-Escárate C, Valenzuela-Muñoz V, Nuñez-Acuña G, Valenzuela-Miranda D, Tapia F, Yévenes M, et al. (2022). The native mussel *Myitlus chilensis* genome reveals adaptative molecular signatures facing the marine environment. *bioRxiv* 2022.09.06.506863.

Gardner JPA (1994). The *Mytilus edulis* species complex in Southwest England: Multi-locus heterozygosity, background genotype and a fitness correlate. *Biochemical Systematics and Ecology* **22**: 1–11.

Gardner JPA (1996). The *Mytilus edulis* species complex in southwest England: effects of hybridization and introgression upon interlocus associations and morphometric variation. *Marine Biology* **125**: 385–99.

Gardner JPA, Skibinski DOF (1991). Biological and physical factors influencing genotype-dependent mortality in hybrid mussel populations. *Marine Ecology Progress Series* 235–43.

Gardner JPA, Thompson RJ (2001). The effects of coastal and estuarine conditions on the physiology and survivorship of the mussels *Mytilus edulis, M. trossulus* and their hybrids. *Journal of Experimental Marine Biology and Ecology* **265**: 119–40.

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**: 3182–95.

Garner JB, Douglas ML, Williams SR, Wales WJ, Marett LC, Nguyen TTT, et al. (2016). Genomic selection improves heat tolerance in dairy cattle. *Scientific reports* **6**: 1–9.

Gerdol M, Moreira R, Cruz F, Gómez-Garrido J, Vlasova A, Rosani U, et al. (2019). Massive gene presence/absence variation in the mussel genome as an adaptive strategy: first evidence of a pan-genome in Metazoa. *bioRxiv* 781377.

Gerdol M, Moreira R, Cruz F, Gómez-Garrido J, Vlasova A, Rosani U, et al. (2020). Massive gene presence-absence variation shapes an open pan-genome in the Mediterranean mussel. *Genome Biology* **21**: 275.

Gosling E (2021). *Marine Mussels: Ecology, Physiology, Genetics and Culture* John Wiley & Sons.

Gräwe U, Friedland R, Burchard H (2013). The future of the western Baltic Sea: two possible scenarios. *Ocean Dynamics* **63**: 901–21.

Gutierrez AP, Matika O, Bean TP, Houston RD (2018). Genomic Selection for Growth Traits in Pacific Oyster (*Crassostrea gigas*): Potential of Low-Density Marker Panels for Breeding Value

Prediction. *Frontiers in Genetics* **9**: 391.

Halpern BS, McLeod KL, Rosenberg AA, Crowder LB (2008). Managing for cumulative impacts in ecosystem-based management through ocean zoning. *Ocean & Coastal Management* **51**: 203–11.

Harley CDG (2011). Climate Change, Keystone Predation and Biodiversity Loss. *Science* **334**: 1124–27.

Hartig F (2018). DHARMa: residual diagnostics for hierarchical (multi-level/mixed) regression models. R package v. 0.2. 0. *Regensburg: University of Regensburg*

Hilbish T, Carson E, Plante J, Weaver L, Gilg M (2002). Distribution of *Mytilus edulis, M. galloprovincialis*, and their hybrids in open-coast populations of mussels in southwestern England. *Marine Biology* **140**: 137–42.

Hoschaka PW, Somero GN (2002). *Biochemical Adaptation: Mechanism and Process in Physiological Evolution* Oxford University Press.

Johannesson K, Smolarz K, Grahn M, André C (2011). The Future of Baltic Sea Populations: Local Extinction or Evolutionary Rescue? *AMBIO* **40**: 179–90.

Jones SJ, Lima FP, Wethey DS (2010). Rising environmental temperatures and biogeography: poleward range contraction of the blue mussel, *Mytilus edulis* L., in the western Atlantic. *Journal of Biogeography* **37**: 2243–59.

Knöbel L, Nascimento-Schulze JC, Sanders T, Zeus D, Hiebenthal C, Barboza FR, et al. (2021). Salinity Driven Selection and Local Adaptation in Baltic Sea Mytilid Mussels. Frontiers in Marine Science, **8:** 692078.

Lotze HK, Lenihan HS, Bourque BJ, Bradbury RH, Cooke RG, Kay MC, et al. (2006). Depletion, degradation, and recovery potential of estuaries and coastal seas. *Science* **312**: 1806–9.

Maar M, Saurel C, Landes A, Dolmer P, Petersen JK (2015). Growth potential of blue mussels (*M. edulis*) exposed to different salinities evaluated by a Dynamic Energy Budget model. *Journal of Marine Systems,* **148:** 48-55*.*

Mathiesen SS, Thyrring J, Hemmer-Hansen J, Berge J, Sukhotin A, Leopold P, et al. (2017). Genetic diversity and connectivity within *Mytilus* spp. in the subarctic and Arctic. *Evolutionary Applications* **10**: 39–55.

Matoo OB, Lannig G, Bock C, Sokolova IM (2021). Temperature but not ocean acidification affects energy metabolism and enzyme activities in the blue mussel, *Mytilus edulis*. *Ecology and Evolution* **11**: 3366–79.

Meier M, Eilola K, Gustavsson BG, Kuznetsov I, Neumann T, Savchuk OP (2012). *Uncertainty Assessment of Projected Ecological Quality Indicators in Future Climate* SMHI

Melzner F, Stange P, Trübenbach K, Thomsen J, Casties I, Panknin U, et al. (2011). Food Supply and Seawater pCO2 Impact Calcification and Internal Shell Dissolution in the Blue Mussel *Mytilus edulis*. *PLOS ONE* **6**: e24223.

Meuwissen THE, Hayes BJ, Goddard ME (2001). Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics* **157**: 1819 LP – 1829.

Nascimento-Schulze JC, Bean TP, Peñaloza C, Paris JR, Whiting JR, Simon A, et al. (2023). SNP discovery and genetic structure in blue mussel species using low coverage sequencing and a medium density 60 K SNP-array. *Evolutionary Applications* **16**: 1044–60.

Nguyen TTT, Bowman PJ, Haile-Mariam M, Pryce JE, Hayes BJ (2016). Genomic selection for tolerance to heat stress in Australian dairy cattle. *Journal of dairy science* **99**: 2849–62.

Nguyen TTT, Hayes BJ, Ingram BA (2014). Genetic parameters and response to selection in blue mussel (*Mytilus galloprovincialis*) using a SNP-based pedigree Potential broodstock candidates.

Aquaculture **420–421**: 295–301.

Ojaveer H, Jaanus A, MacKenzie BR, Martin G, Olenin S, Radziejewska T, et al. (2010). Status of biodiversity in the Baltic Sea. *PLoS one* **5**: e12467.

Paggeot LX, DeBiasse MB, Escalona M, Fairbairn C, Marimuthu MPA, Nguyen O, et al. (2022). Reference genome for the California ribbed mussel, *californianus*, an ecosystem engineer. *Journal of Heredity,* **113(6):** 681-688..

Pan FTC, Applebaum SL, Manahan DT (2021). Differing thermal sensitivities of physiological processes alter ATP allocation. *Journal of Experimental Biology* **224**: jeb233379.

Pan T-CF, Applebaum SL, Frieder CA, Manahan DT (2018). Biochemical bases of growth variation during development: a study of protein turnover in pedigreed families of bivalve larvae (*Crassostrea gigas*). *Journal of Experimental Biology* **221**: jeb171967.

Pansch C, Hiebenthal C (2019). A new mesocosm system to study the effects of environmental variability on marine species and communities. *Limnology and Oceanography: Methods* **17**: 145–62.

Pansch C, Scotti M, Barboza FR, Al-Janabi B, Brakel J, Briski E, et al. (2018). Heat waves and their significance for a temperate benthic community: A near-natural experimental approach. *Global Change Biology* **24**: 4357–67.

Pörtner H-O (2010). Oxygen- and capacity-limitation of thermal tolerance: a matrix for integrating climate-related stressor effects in marine ecosystems. *Journal of Experimental Biology* **213**: 881–93.

Pörtner H-O, Roberts DC, Masson-Delmotte V, Zhai P, Tignor M, Poloczanska E, et al. (2019). The ocean and cryosphere in a changing climate. *IPCC Special Report on the Ocean and Cryosphere in a Changing Climate.*

Pörtner HO (2012). Integrating climate-related stressor effects on marine organisms: *Marine Ecology Progress Series* **470**: 273–90.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics* **81**: 559–75.

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2012.

Rahmstorf S, Coumou D (2011). Increase of extreme events in a warming world. *Proceedings of the National Academy of Sciences* **108**: 17905–9.

Regan T, Bean TP, Ellis T, Davie A, Carboni S, Migaud H, et al. (2021). Genetic improvement technologies to support the sustainable growth of UK aquaculture. *Reviews in Aquaculture* **13**: 1958–85.

Robledo D, Matika O, Hamilton A, Houston RD (2018a). Genome-wide association and genomic selection for resistance to amoebic gill disease in Atlantic salmon. *G3: Genes, Genomes, Genetics* **8**: 1195–1203.

Robledo D, Palaiokostas C, Bargelloni L, Martínez P, Houston R (2018b). Applications of genotyping by sequencing in aquaculture breeding and genetics. *Reviews in Aquaculture* **10**: 670–82.

RStudio Team (2021). RStudio: integrated development for R. RStudio, PBC, Boston, MA. 2020.

Ruggieri V, Calafiore R, Schettini C, Rigano MM, Olivieri F, Frusciante L, et al. (2019). Exploiting genetic and genomic resources to enhance heat-tolerance in tomatoes. *Agronomy* **9**: 22.

Saarman NP, Pogson GH (2015). Introgression between invasive and native blue mussels (genus *Mytilus*) in the central California hybrid zone. *Molecular Ecology* **24**: 4723–38.

Sanders T, Schmittmann L, Nascimento-schulze JC, Melzner F (2018). High Calcification Costs Limit Mussel Growth at Low Salinity. **5**: 1–9.

Seuront L, Nicastro KR, Zardi GI, Goberville E (2019). Decreased thermal tolerance under recurrent heat stress conditions explains summer mass mortality of the blue mussel *edulis*. *Scientific Reports* **9**: 17498.

Simon A (2022). Three new genome assemblies of blue mussel lineages: North and South European *Mytilus edulis* and Mediterranean *Mytilus galloprovincialis*. *bioRxiv* 2022.09.02.506387.

Simon A, Fraïsse C, El Ayari T, Liautard-Haag C, Strelkov P, Welch JJ, et al. (2021). How do species barriers decay? Concordance and local introgression in mosaic hybrid zones of mussels. *Journal of Evolutionary Biology* **34**: 208–23.

Sommer U, Lengfellner K (2008). Climate change and the timing, magnitude, and composition of the phytoplankton spring bloom. *Global Change Biology* **14**: 1199–1208.

Stuckas H, Knöbel L, Schade H, Breusing C, Hinrichsen HH, Bartel M, et al. (2017). Combining hydrodynamic modelling with genetics: can passive larval drift shape the genetic structure of Baltic *Mytilus* populations? *Molecular Ecology* **26(10)**: 2765-2782.

Symonds JE, King N, Camara MD, Ragg NLC, Hilton Z, Walker SP, et al. (2018). New Zealand aquaculture selective breeding: from theory to industry application for three flagship species. *Proceedings of the World Congress on Genetics Applied to Livestock Production* **11**: 1035.

Tagliarolo M, McQuaid CD (2015). Sub-lethal and sub-specific temperature effects are better predictors of mussel distribution than thermal tolerance. *Marine Ecology Progress Series* **535**: 145–59.

Thudi M, Upadhyaya HD, Rathore A, Gaur PM, Krishnamurthy L, Roorkiwal M, et al. (2014). Genetic dissection of drought and heat tolerance in chickpea through genome-wide and candidate gene-based association mapping approaches. *Plos one* **9**: e96758.

Tsai H-Y, Hamilton A, Tinch AE, Guy DR, Gharbi K, Stear MJ, et al. (2015). Genome wide association and genomic prediction for growth traits in juvenile farmed Atlantic salmon using a high density SNP array. *BMC Genomics* **16**: 969.

Uffelmann E, Huang QQ, Munung NS, de Vries J, Okada Y, Martin AR, et al. (2021). Genome-wide association studies. *Nature Reviews Methods Primers* **1**: 59.

Vajedsamiei J, Melzner F, Raatz M, Morón Lugo SC, Pansch C (2021a). Cyclic thermal fluctuations can be burden or relief for an ectotherm depending on fluctuations' average and amplitude. *Functional Ecology* **35**: 2483–96.

Vajedsamiei J, Wahl M, Schmidt AL, Yazdanpanahan M, Pansch C (2021b). The higher the needs, the lower the tolerance: Extreme events may select ectotherm recruits with lower metabolic demand and heat sensitivity. *Frontiers in Marine Science* **8**: 264.

van der Schatte Olivier A, Jones L, Vay L Le, Christie M, Wilson J, Malham SK (2020). A global review of the ecosystem services provided by bivalve aquaculture. *Reviews in Aquaculture* **12**: 3–25.

Vendrami DLJ, De Noia M, Telesca L, Brodte E-M, Hoffman JI (2020). Genome-wide insights into introgression and its consequences for genome-wide heterozygosity in the *Mytilus* species complex across Europe. *Evolutionary Applications* **13(8)**: 2130-2142.

Ventura A, Schulz S, Dupont S (2016). Maintained larval growth in mussel larvae exposed to acidified under-saturated seawater. *Scientific Reports* **6**: 1–10.

Wang J, Zhang G, Fang X, Guo X, Li L, Luo R, et al. (2012). The oyster genome reveals stress adaptation and complexity of shell formation. *Nature*

Wenne R, Bach L, Zbawicka M, Strand J, McDonald JH (2016). A first report on coexistence and

hybridization of *Mytilus trossulus* and *M. edulis* mussels in Greenland. *Polar Biology* **39**: 343–55.

Wilson J, Matejusova I, McIntosh RE, Carboni S, Bekaert M (2018). New diagnostic SNP molecular markers for the *Mytilus* species complex. *PLoS ONE* **13**: 1–13.

Wolf F, Bumke K, Wahl S, Nevoigt F, Hecht U, Hiebenthal C (2020). High resolution water temperature data between January 1997 and December 2018 at the GEOMAR pier surface. *Bremen PANGAEA* 10.

Wood SN (2006). *Generalized Additive Models: An Introduction with R* chapman and hall. CRC press.

Wood SN, Pya N, Säfken B (2016). Smoothing parameter and model selection for general smooth models. *Journal of the American Statistical Association* **111**: 1548–63.

Yang J-L, Feng D-D, Liu J, Xu J-K, Chen K, Li Y-F, et al. (2021). Chromosome-level genome assembly of the hard-shelled mussel *Mytilus coruscus*, a widely distributed species from the temperate areas of East Asia. *GigaScience* **10**: giab024.

Zbawicka M, Sańko T, Strand J, Roman W (2014). New SNP markers reveal largely concordant clinal variation across the hybrid zone between *Mytilus* spp. in the Baltic Sea. *Aquatic Biology* **21**: 25–36.

Zenger KR, Khatkar MS, Jones DB, Khalilisamani N, Jerry DR, Raadsma HW (2019). Genomic Selection in Aquaculture: Application, Limitations and Opportunities With Special Reference to Marine Shrimp and Pearl Oysters. Frontiers in genetics, **9:** 693.

Zippay ML, Helmuth B (2012). Effects of temperature change on mussel, *Mytilus*. *Integrative Zoology* **7**: 312–27.

Zittier ZMC, Bock C, Lannig G, Pörtner HO (2015). Impact of ocean acidification on thermal tolerance and acid–base regulation of *Mytilus edulis* (L.) from the North Sea. *Journal of Experimental Marine Biology and Ecology* **473**: 16–25.

Supplementary material

Supplementary Table 4.1: Measurements of *R. salinas* concentration (cells mL^{-1}) over a 72-hour period in 14 L experimental tanks containing no juvenile mussels.

Mesocom	*R. salinas* 0h	*R. salinas* 24h	*R. salinas* 36h	*R. salinas* 72h
M1	8768	7095.5	6912.5	6500
M2	8021	7140.5	6825	5241
M3	7650	7417	6660.5	5504.5
M4	7737.5	7797	7048.5	5519
M5	8278	8257	8301	7356.5
M6	7637.5	8477.5	7497	6488.5
M7	9090.5	8159	7695	6138
M8	7743	7534.5	6799.5	4943.5
M9	7685.5	8114.5	8798	7502.5
M10	8278	8019.5	8689.5	6641.5
M11	8589	9010.5	9154	7791.5
M12	7643	7918.5	8008.5	6659

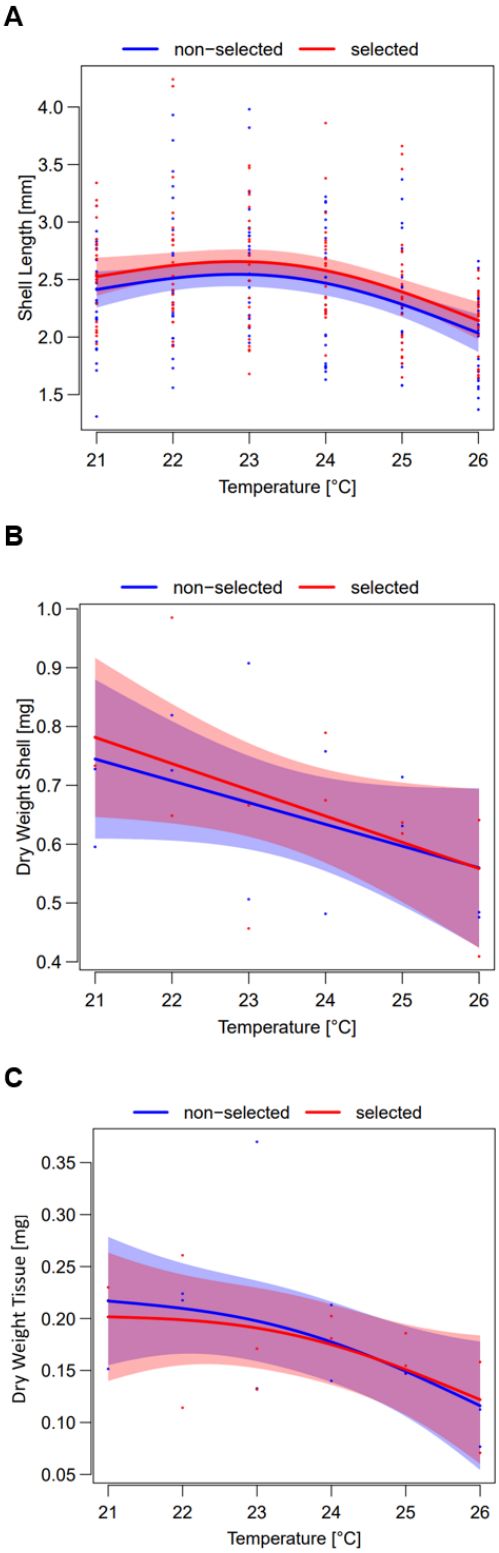
Supplementary table 4.2: Cross validation errors generated by admixture analysis to infer the best number of ancestral populations in dataset containing the selected and non-selected cohorts and the five additional control populations. In this analysis, we tested k's from a 2 – 10 range. The k with the lower error output is considered the most representative number of ancestral populations for a given dataset.

Number of K	Cross validation	Number of iterations
2	0.28594	15
3	**0.28224**	**20**
4	0.28204	25
5	0.29493	53
6	0.29915	34
7	0.30983	60
8	0.31806	139
9	0.32591	64
10	0.34396	55

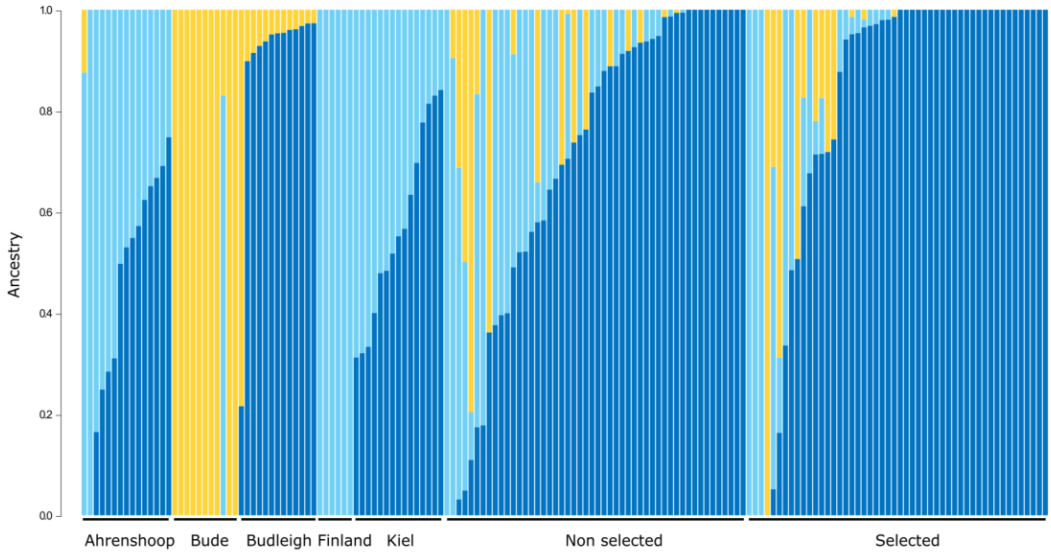
Supplementary Table 4.3: ANOVA analysis output of impacts of temperature in the shell length (SL) of juvenile mussels either selected (S) or non-selected (NS) to thermal resilience at day-0, prior to the start of the 25-day thermal challenge. In this analysis, a p value lower than 0.05 is considered significant.

	DF	Sum Sq	Mean Sq	F Value	Pr (>F)
Treatment	1	0.018	0.1725	0.155	0.695
Residuals	118	13.355	0.11318		

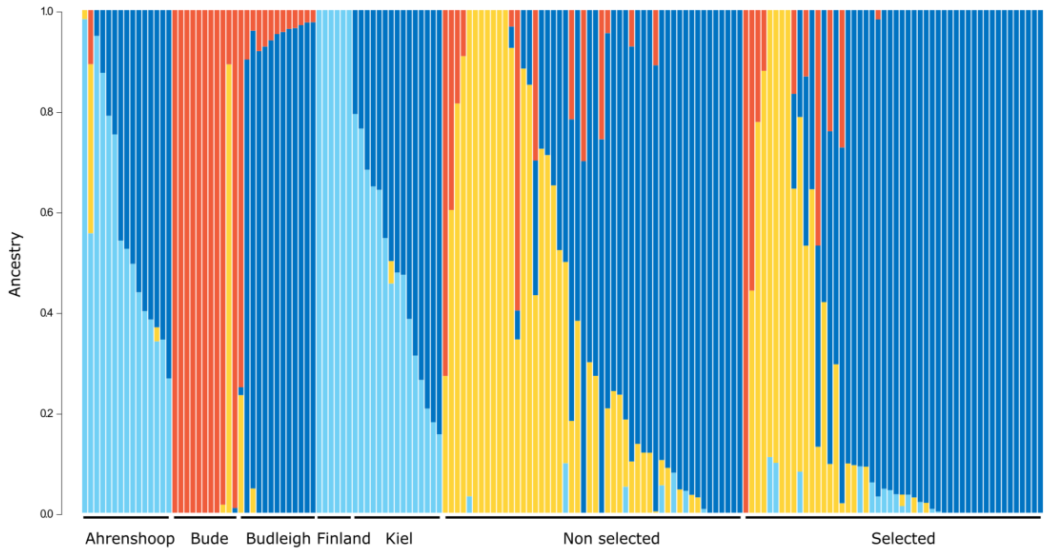
Supplementary Figure 4.*1:* Final dry tissue weight, dry shell weight, and shell length of selected (S, red) and non-selected (NS) spat across experimental temperatures (21-26°C), following the 13 days of exposures to different temperatures. Each of the data points represents the measurement of an individual spat. These relations were modelled as additive mixed models assuming scaled t-distributions of residuals. Solid lines represent predicted means and shades represent 95% confidence intervals.



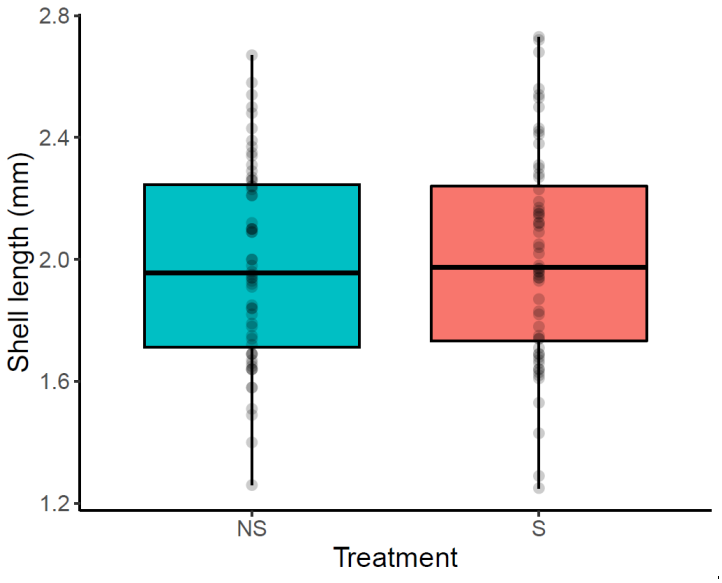
Supplementary Figure 4.2*:* Results of genetic admixture analysis using 11,288 SNPs. Each individual is represented by a column partitioned into segments of different colour, the length of which indicate the posterior probability of membership in each cluster. Solid bars represent an individual from a single species background. In this figure we plot a) admixture results of k=3 results for the dataset including both the cohorts (selected and non-selected) in this study and the 5 control populations. The dark-blue cluster can be inferred to *M. edulis*, this is the species dominating populations in Budleigh, UK. The light-blue cluster can be inferred to *M. trossulus*, genotypes from this species are highly frequent in finish mussel populations. The yellow cluster can be inferred to *M. galloprovincialis*, as this is the main species colonising Bude, UK.



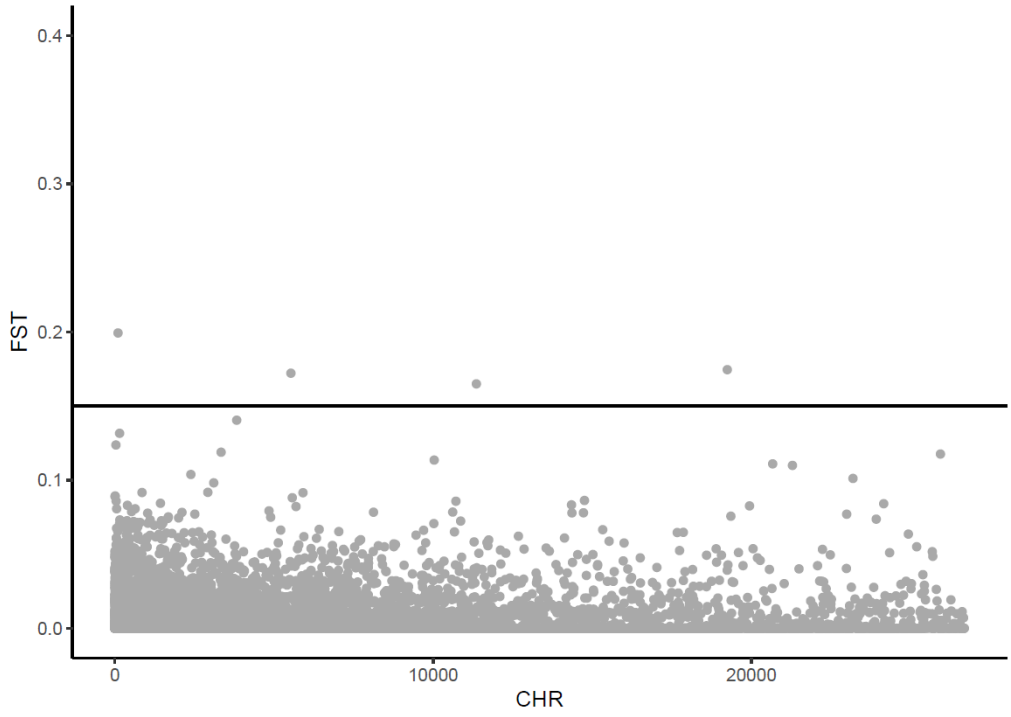
Supplementary Figure 4.3*:* Results of genetic admixture analysis using 11,288 SNPs. Each individual is represented by a column partitioned into segments of different colour, the length of which indicate the posterior probability of membership in each cluster. Solid bars represent an individual from a single species background. In this figure we plot a) admixture results of k=4 results for the dataset including both the cohorts (selected and non-selected) in this study and the 5 control populations. The dark-blue cluster can be inferred to *M. edulis*, this is the species dominating populations in Budleigh, UK. The light-blue cluster can be inferred to *M. trossulus*, genotypes from this species are highly frequent in finish mussel populations. The yellow cluster can also be inferred to *M. trossulus* sampled 4 years later, as this is the second most frequent species expected in the Selected/Non-selected populations in this study. The red cluster can be inferred to *M. galloprovincialis*, as this is the main species colonising Bude, UK.



Supplementary Figure 4.4: Boxplot of initial shell length (SL, mm) of selected (S, red) and non-selected (NS, blue) spat used in the thermal stress experiment. Each circle represents an individual's spat SL measurement in millimetres.



Supplementary Figure 4.5: F_{ST} values for the 11,288 markers kept for genetic analysis following quality check filtering steps. The central line is at the height y = 0.15. Four markers, had a F_{ST} value higher than 0.15.



5. Chapter V - General Discussion and Main Conclusions

General discussion

The overall aim of this thesis was to identify the main biological issues that currently hinder the sustainable development of shellfish aquaculture, and to assess whether and how such issues may hinder the implementation of genomic selection approaches in these taxa. I begin my thesis by outlining the existing issues throughout the main introduction (Chapter I), additionally proposing a set of actions that may contribute to their resolution. Subsequently, in each individual chapter (Chapters II to IV), I address one of the outlined actions required, seeking to elucidate unresolved questions whilst discussing possible avenues for how these actions can further the sustainable development of the sector. First, in Chapter II of this thesis, I describe the development of a genetic tool, a multi-species 60 K SNP-array, for the four blue mussel species that represent the vast majority (by volume) of commercially produced mussels across the globe: *M. edulis*, *M. galloprovincialis*, *M. trossulus* and *M. chilensis*. The blue mussel array can be applied for the consistent genotyping of these taxa across their entire geographic range globally. Consequently, this tool enables the investigation of genetic structure of mussel populations, as well as analysing the genetic architecture of traits of interest. In this same chapter, I validate this tool, in the process identifying its strengths and limitations. I propose the ways in which this SNP-array may contribute to the development of increasingly complex and targeted genomic tools for mussels in the future. In addition, I discuss how this tool may contribute expanding our understanding of genomic processes ongoing in this species group and how it might be used for breeding of mussels. In the following chapter (Chapter III), I assess the genetic structure of blue mussel populations in the southwest coast of England. In this area, two mussel species, *M. galloprovincialis* and *M. edulis*, form a mosaic hybrid zone. Accordingly, the SW coast of England is historically a key region of study that has contributed extensively to our understanding of hybridisation within these taxa. In this chapter, I aimed to investigate the utility of the SNP-array developed in the previous chapter for assessing population structure and admixture in a natural population. The SW UK hybrid zone had been previously described based on the polymorphism at one loci (Glu-5) (Hilbish *et al.* 2002). The existence of such a study allowed me to compare the previous

description of genetic structure in this hybrid zone to that generated by the newly developed SNP-array, which genotypes animals across many thousands of loci. Consequently, this chapter further validates the array performance. A final goal of this thesis was to understand how climate change will impact the performance of mussels across economically and ecologically relevant traits (e.g. growth through shell length, dry weight and shell mass). Such a question is a first step to quantify the viability of selective breeding for environmental resilience in mussels, and its answer provides crucial knowledge to mitigate the impacts of climate on the blue mussel industry. Therefore, in the last experimental chapter of this thesis (Chapter IV), I investigate how warming environments impact the performance of two different mussel lines, the first consisting of individuals which were selected for thermal resilience, and the second of which were naïve to such a stressor. In this same chapter, I also assess the implications of selection towards thermal resilience on the genetic structure of the selected populations.

In the following sections of this discussion chapter, I recapitulate and summarise the main findings of each of the experimental chapters of this thesis (Chapters II, III and IV). Furthermore, I discuss how each chapter has contributed to the proposed thesis aims, how the results contribute to our understanding of mussel genomics generally and highlight the benefits the main findings might bring to the sustainable optimisation of the aquaculture industry, through activities such as selective breeding.

Development of the blue mussel multi-species 60K SNP-array: a genomic tool to study genetic diversity and structure of blue mussel populations and its contributions to the aquaculture industry

The blue mussel multi-species medium density 60 K SNP-array, developed in Chapter II, is the first high throughput genotyping array for the four main cultured species in the *Mytilus* genus. The SNPs included in this platform were called from the alignment of sequences generated by lcWGS of

138 individual mussels, originating from 23 wild, globally-distributed populations, to the *M. galloprovincialis* genome (Simon 2022). The selected populations came from different regions including the Mediterranean Sea, North East Atlantic, Baltic Sea and North and South East Pacific. Following the removal of markers with a CR equal or lower than 95% and a MAF < 0.01 , the final array design retained 23,252 SNPs applicable for genotyping *M. edulis* individuals, 22,165 for *M. chilensis*, 20,504 for *M. galloprovincialis* and 20,149 for *M. trossulus*. The validation process described in this thesis demonstrated that this tool is useful to assess genetic structure of mussel populations, with populations structure and segregation comparable to the one generated by the lcWGS dataset (Figure 5.1). Therefore, this new array, in which genotyping at a density of thousands of genomic markers simultaneously is possible, represents a significant advance in capability compared to the comparatively low cost, high throughput and easily accessed low density genotyping platforms currently available (Carboni et al. 2021; Gardner et al. 2016; Mathiesen et al. 2017; Nguyen et al. 2014; Saarman & Pogson 2015; Simon et al. 2021; Wenne et al. 2016; Wilson et al. 2018; Zbawicka et al. 2014). High density panels facilitate the development of GWAS approaches to identify QTLs, as well as facilitate the analysis of levels of inbreeding and parental contribution in wild and farmed populations, key for enabling and fast-tracking genetic improvement programmes in mussels.

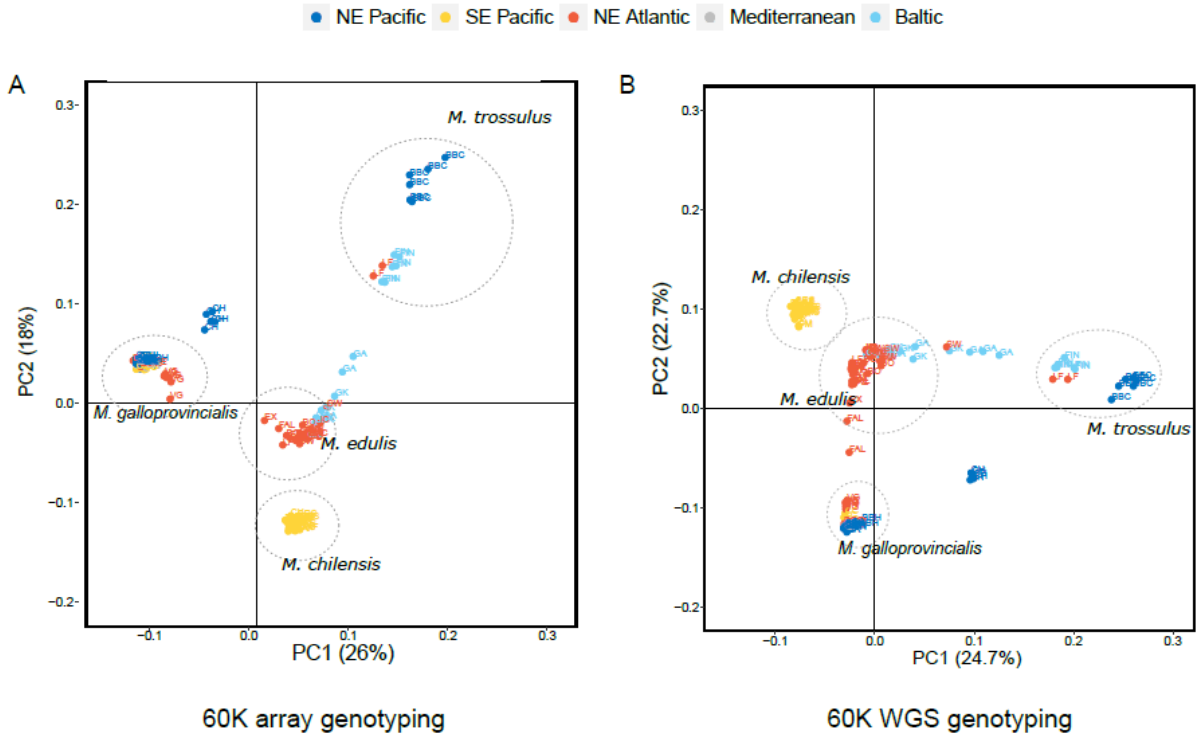


Figure 5-1: Principal component analysis for the different genotyping datasets: i) 60K SNPs with a CR > 95% from the microarray genotyping ii) genotypes called from the lcWGS dataset at the same 60K SNPs implemented in the blue mussel

Nonetheless, there was a high frequency of missing genotypes across individuals in the lcWGS dataset (>70%). The high levels of data missingness could be a result of multiple factors. First, samples from different species, as well as their hybrids, were sequenced using a low coverage approach of 3X to 5X coverage. These samples were all aligned to the *M. galloprovincialis* genome (Simon 2022), which at the time was assembled to a contig level. It is well established that the genomes of marine bivalves are highly variable, not only within populations but also among species (Gerdol *et al.* 2019, 2020; Wang *et al.* 2012). High levels of duplication and the expansion of gene families, especially those related to immune function, have been described for Pacific oysters (Wang *et al.* 2012). A pan genome, with a set of dispensable genes (20K out of 65K genes) widespread across the genome, was also recently described for the *M. galloprovincialis* species (Gerdol *et al.* 2020). Not surprisingly, this huge genomic variability, including presence absence variability (PAV), might have contributed to the

observed amount of seemingly missing data. In addition, samples from a *M. galloprovincialis* background had, in general, lower frequency of missing data when compared to other species. It is possible therefore that the alleles defined either as 'reference' or 'alternate' in our SNP discovery pipeline have been biased towards this species, missing suspected polymorphisms in flanking regions of SNPs within the sister species. Consequently, genotyping the sister species with the array results in higher levels of missing genotype data, in comparison to *M. galloprovincialis* samples. Nonetheless, the landscape of mussel genomics having changed rapidly in the past decade since the onset of this project, with new genomes having been assembled for *M. edulis*, *M. galloprovincialis, M. coruscus, M. chilensis and M. californianus* (Corrochano-Fraile *et al.* 2021; Gallardo-Escárate *et al.* 2022; Gerdol *et al.* 2020; Paggeot *et al.* 2022; Simon 2022; Yang *et al.* 2021). As a result, in the past few years alone, our understanding of this species complex wide genome diversity has also advanced. Such advances will facilitate greater understanding of the performance of our array across multiple species, enable the association of specific SNPs with genes of interest, as well as increase our ability to develop additional novel probes for sister species, which will enable refinement of the tool for subsequent versions.

Breeding of aquatic species for commercial purposes has conventionally evolved along a specific trajectory, starting with the development of species-specific hatchery systems, followed by the application of different breeding methods such as phenotypic selection or marker assisted selection, with the development of a species specific genomic tool-box being the final step of the process. Contrary to the process followed for other species, we have developed this blue mussel SNP-array prior to the widespread development and utilisation of hatchery systems for mussel aquaculture globally. As previously mentioned in this chapter, no high density genotyping tool including genome wide spread markers was yet available for these taxa. The development of the blue mussel multi species 60 K SNP-array provides the opportunity for genomic approaches to be implemented right at the beginning of the mussel production process. It also facilitates a greater understanding and control of the genetic background of individuals, the control of the levels of inbreeding and parental lineages

and seafood traceability. Therefore, if mussel hatcheries become available in the future, coupling such technology with the usage of the 60 K blue mussel SNP array can potentially contribute to enhance mussel aquaculture production.

Genomic selection has been highlighted as crucial for a fast and sustainable optimisation of the aquaculture industry (FAO 2019). A great benefit of genomic selection is that it facilitates the screening of thousands of genetic markers spread across the genome. Consequently, selection of multiple traits and especially of polygenic ones becomes possible (Meuwissen *et al.* 2001). One major step in the development of genomic selection in any farmed species is to create a tool which allows for rapid, cost effective scanning of a huge number of individuals, and genotyping of variants. SNP arrays are a great tool in this sense: they provide consistent genotyping of individuals using a high number of variants (up to the scale of hundreds of thousands) spread across the genome. The density of markers contained in arrays guarantees a high genomic coverage. In addition, as genotyping occurs in predefined loci, outputs generated by the platform are consistent, reproducible and comparable among studies (Robledo *et al.* 2018b). SNP arrays have previously contributed to the fast-tracking of selective breeding in cultured species (Zenger *et al.* 2019). Furthermore, GWAS studies can be developed to identify markers related to traits of interest (Bangera et al. 2017; Gutierrez et al. 2018; Robledo et al. 2018a; Tsai et al. 2015). This is especially important with environmental conditions projected to undergo rapid and intense modulation within the coming decades as a consequence of climate change (Pörtner *et al.* 2019). Consequently, the genomic tool developed here can contribute to quicken the development of environmentally resilient lines.

To summarise, the blue mussel multi-species 60 K SNP array, which the development of is described in this thesis, is of great utility to the blue mussel scientific and aquaculture community, where it can be applied for the rapid characterisation of wild and commercial mussel populations. Understanding the structure of wild populations, identifying novel variants and quantifying the baseline genomic background on which selective breeding efforts must build upon and be cognisant of, is a critical step in the development the blue mussel industry. Most importantly, this tool is a

platform that generates accessible and comparable data for mussel genomic studies. The tool is currently being applied in multiple ongoing projects: i) hybridisation and species invasion in blue mussel populations along the coast of Australia and New Zealand, ii) population composition in blue mussels along the coast of China, iii) identification of organisms contaminated with transmissible cancer globally and iv) validation of the use of the tool for identification of an individual's sex, further demonstrating its value to the mussel research community globally.

The utility of the genomic tool for assessing genetic diversity across blue mussels genotypes

The genetic diversity and population structure within the SW UK hybrid zone has been extensively investigated in the past decades (Gardner & Skibinski 1991; Gardner 1994, 1996; Hilbish et al. 2002). However, in these earlier studies, speciation and levels of hybridisation were estimated using a lower number of markers. In contrast to these studies, I used the blue mussel multi-species 60 K SNP-array, for which the development is detailed in Chapter II of the thesis, to investigate the structure of this hybrid zone. The existence of this historic understanding, now over 20 years old, enabled me to compare the output of both technologies, as well as to update any major modification in the species distribution across this hybrid zone.

This was the first application of the blue mussel 60 K array to investigate genetic structure of twenty-nine wild mussel populations along a ~500 km coast line extension. A total of 16,732 variants, over 25 % of the markers in the array, were used to assess levels of admixture across these populations, together with their hybrid index. Comparable to findings by Hilbish et al (2002), my study confirmed that the populations in this region were comprised of two different species, *M. edulis* and *M. galloprovincials*, and their hybrids. In addition, the results are in accordance to the overall genotype distribution of these species in this region, as described by Hilbish et. al. (2002). *M. galloprovincialis* genotypes are most abundant within populations situated along the northern coast of Devon and Cornwall, whilst southern populations show a higher presence of *M. edulis* genotypes. My data does

not enable the cause of this regional shift in speciation from *M. edulis* to *M. galloprovincialis* genotypes, either due to a geographical barrier hindering larval dispersal or by any other environmental factor that might be leading to post-settlement selective pressures, to be elucidated. Nonetheless, the outcomes of this chapter raise important questions for future research.

The results described in Chapter III did not point towards a clear habitat-specific distribution of genotypes as suggested by previously published literature (Bierne et al. 2003; Gardner 1994; Gardner & Thompson 2001). In previous studies authors have described a counter selection of *M. edulis* genotypes in wave-exposed environments, occurring along the SW coast of England and along the Atlantic coast of France. However, I found no strong segregation between low tide and high tide populations or within and out of estuaries. Nonetheless, my study largely focussed on a single size class of individuals, targeting adult individuals, whilst previous studies have sampled multiple size classes of mussels. Therefore, I suggest that an in-depth investigation of genotype specific post-settlement mortality could be carried in the SW UK mussel hybrid zone, sampling a set of size-classes, together with a higher number of individuals at specific locations. This would enable a better understanding of whether post-settlement selection results in dynamic species composition over the full life-cycle of cohorts within any given environment, as a result of incompatibility of genotypes to prevailing environmental conditions. Such an approach would provide fundamental knowledge on interactions between genetics and environment in this species group. Furthermore, such an experimental approach would address the yet unresolved question of whether *M. edulis* are counter-selected in wave-exposed environments along the north Devon and Cornish coasts in the SW of England, potentially allowing the identification of markers of relevance for environmental resilience.

Blue mussel aquaculture in the UK is an economically important activity, representing the second largest aquaculture sector by volume after salmon (~22,000 tonnes, £36M) (Regan *et al.* 2021). Reproductive technologies are well advanced for some marine bivalve species. Oyster hatcheries are well established worldwide, whilst for mussels, an outstanding example is the development of hatchery technologies for the Green Lipped mussel, in New Zealand (Symonds *et al.* 2018). Such

systems provide greater control over environmental conditions, enabling the optimisation of larval growth and survival. Hatchery based larval production systems are a major step forward in mussel production, which maximise the likelihood that spat is available to farmers all year-round. Besides, farmers will not have to depend on natural spat-fall, which poses a major challenge when settlement volumes are highly variable between years. Most importantly, hatcheries allow a greater control of the genetic composition of commercial stock. Nonetheless, global blue mussel production still relies on the capture of wild genotypes. Therefore, no control of the genetic background or phenotype of individuals is currently being applied in blue mussel aquaculture. In the wild, mussels from the *Mytilus edulis* species complex can easily hybridise wherever the sister species encounter each other, forming extensive hybrid zones. Furthermore, studies have indicated that intra and inter species variation across environmental gradients occurs within the genus (Knöbel *et al.* 2021; Sanders *et al.* 2018). These findings suggest that performance is dictated not only by the species itself, but also by genetic processes such as local adaptation and/or the acclimation potential of individuals. Under this scenario, characterising the genetic background of wild populations is a first step to unravel which of these processes are ongoing in these taxa. Such understanding is also crucial to maximise the success and productivity of selectively bred mussels in the future.

In providing the first characterisation of the SW UK mussel hybrid zone using a genomic approach, this third chapter of my thesis therefore significantly advances our understanding of speciation in this historically important region for mussel research. Such characterisation yields fundamental knowledge on the current status of this hybrid zone, updating its previous version which was published over 20 years ago (Hilbish *et al.* 2002). The research presented in this chapter also emphasises the value of the newly developed blue mussel multi-species 60 K SNP-array as a tool for investigating the genetic aspects of wild mussel populations, significantly advancing the capability of the global mussel research community.

5.3 Assessing the implications of environmental stress on the genetics and performance of blue mussel spat: a first step for selection towards environmental resilience

In Chapter IV of this thesis, I investigated how selection towards thermal resilience during early development affected the physiological responses of blue mussel juveniles. In addition, I assessed if any changes in genetic diversity occurred in the selected cohorts. Results from Chapter IV demonstrate that juvenile mussels selected under thermal stress differ significantly to naïve organisms in terms of their dry weight mass. More specifically, juveniles selected to resist high temperature events had a significantly lower dry weight in comparison to non-selected ones.

One possible hypothesis to explain such a difference could be OCLTT theory (Pörtner 2010). This theory proposes that temperature impacts the metabolic scope of individuals, limiting the oxygen supply to tissues. Such a mismatch increases with temperature until its limit is reached. Over longer term exposures, or when facing acute thermal stress, individual performance is subsequently affected. According to the OCLTT theory, individuals with initial lower metabolic rates seem better able to cope with higher temperatures, demonstrating an enhanced ability to suppress their metabolism (Vajedsamiei et al. 2021). An explanation is that the organisms which have a decreased metabolic demand are less likely to suffer from the heat-induced supply and demand mismatch of energy demand. While highly speculative, if this is the case in this study, the outcome suggests a selection process through which juveniles with a lower metabolic rate resisted the thermal challenge. This hypothesis could explain the significantly lower dry weight of selected individuals in comparison to the non-selected ones. Alternatively, the lower mean dry weight of selected individuals could be caused directly by detrimental impacts of the thermal stress, the impacts from which organisms were not able to recover. Nonetheless, further investigation is required to determine the reason why selected individuals in this study have a significantly lower dry weight in comparison to naïve ones.

Within the coming decades, climate change is expected to severely modify biotic and abiotic conditions across the oceans (Pörtner *et al.* 2019). Mussels are susceptible to changes in their

environment, with many studies having been published that highlight severe implications of environmental stress on their performance (Jones *et al.* 2010; Knöbel *et al.* 2021; Melzner *et al.* 2011; Ventura *et al.* 2016). It is unclear, however, whether early exposure to an environmental stressor might select for genotypes of individuals which are more resilient to environmental stress. Furthermore, if such selection occurs, it is also unclear what impacts this process may have on the genetic diversity of populations. Results from this thesis suggests that, although not significant, subtle changes in genetic diversity occurred as a consequence of selection to stress. Early exposure to high temperatures resulted in an increase in the introgression of *M. edulis* alleles within the genotyped individuals (or a selection against *M. trossulus* genotypes). The desalination of the Baltic sea is expected to favour *M. trossulus* genotypes (Knöbel *et al.* 2021), but results from this chapter point out that rising temperatures may do the opposite. This raises important questions on the future of Baltic mussel populations and their response to multi-factorial climate change that deserves further investigation.

Nonetheless, the observed changes in genetic composition of selected and non-selected individuals described in this thesis are subtle. Contrary to the admixture analyses, the PCA results presented in Chapter IV do not show a clear segregation between the cohorts. In addition, only four loci had an F_{ST} value > 0.15 and none > 0.2, with further investigation needed to understand if these loci are involved in key thermal adaptation processes. Selection towards thermal resilience has been proven possible in different taxa including cattle (Garner *et al.* 2016; Nguyen *et al.* 2016) and different species of crop (Ahmed *et al.* 2022; Ruggieri *et al.* 2019; Thudi *et al.* 2014). However, GWAS studies normally rely on a large number of samples, normally on the scale of thousands. Future experiments including higher number of genotyped individuals and, perhaps, thermal selection over generations, could facilitate the detection of changes in genetic structure as a consequence of thermal selection in blue mussels. Such results will be of great relevance to enable the mussel aquaculture industry to thrive under warming oceanic conditions.

The study described in Chapter IV of the thesis was the first to attempt to select mussels to thermal resilience, and subsequently to investigate shifts in performance of individuals when those where re-exposed to this stressor and to simultaneously combine a genomic approach to search for changes in genetic structure caused by selection. Although the results from this Chapter are subtle in terms of significance, they represent a great advance in mussel research capability and demonstrate that breeding environmentally resilient lines could be possible in these taxa.

Conclusions

1) The Blue mussel multi-species 60K SNP-array was developed and can be applied for the genotyping of the four main cultured species *M. edulis*, *M. galloprovincialis*, *M. trossulus* and *M. chilensis*. This array will contribute to the generation of comparable data across studies. In addition, this tool will facilitate the usage of genomic analysis in studies due to its accessible workflow for data generation (i.e. requirements for sample preparation and sample quality) and analysis.

2) The blue mussel multi-species 60 K SNP-array can be applied as a tool to investigate structure and diversity of wild mussel populations. Compared to other existing genotyping platforms for these taxa, this tool contains the highest density of genomic markers to perform such studies, particularly as it works flexibly across at least four sister species, and their hybrids.

3) The genetic diversity of the SW UK hybrid zone, historically a key area for blue mussel farming in the UK, has been characterised with a genomic approach, the blue mussel multi-species 60K SNP-array, for the first time.

4) The results of this thesis agree with previously published data demonstrating that *M. galloprivincialis* genotypes dominate the north coast of SW England, whilst *M. edulis* is more abundant

in the south coast of Cornwall and Devon. These results provide a greater clarity and depth of understanding on the current distribution of genotypes in this hybrid zone.

5) Selection towards thermal resilience significantly affected performance in dry weight of juvenile blue mussels. In addition, signatures of selection were apparent in the genetic composition of the challenged population. These results represent a key first step towards selection for environmentally resilient mussel lines.

Marine bivalve molluscs are one of the most sustainable sources of animal protein produced globally. Farming of these organisms will contribute to guarantee the availability of high quality feed for the growing human population. However, the future of this practice is undoubtedly threatened by climate change. For this reason, investing efforts in developing environmentally resilient lines of marine bivalve molluscs is crucial to futureproof shellfish aquaculture against the upcoming environmental changes, assuring the continuity of this traditional practice. The application of genomic selection has been suggested as an essential approach to fast-track the development of such lines. During the past decades, significant advances in the marine bivalve mollusc genomic toolbox have occurred, expanding considerably our understanding of these taxa on a physiological and omic scale. However, further studies are necessary to unravel the extent to which genomic selection can contribute to develop shellfish lines able to cope with climate change, and if there is still enough time to implement such approaches throughout the production process. If positive results are obtained, significant effort from both government entities and farmers will be required to implement hatchery-based systems for spat production on a global scale. This thesis takes a small step towards the development of genomic selection in marine bivalve molluscs, providing scientific evidence that such approaches may be a plausible solution to guarantee production under a rapidly changing environment, developing a tool which can possibly be applied for such purposes and applying this tool

to assess genetic diversity of a mollusc species in an economically relevant area. Hopefully, this thesis will inspire future research and contribute to the development of shellfish aquaculture.

References

Ahmed HGM-D, Naeem M, Zeng Y, Rashid MAR, Ullah A, Saeed A, et al. (2022). Genome-wide association mapping for high temperature tolerance in wheat through 90k SNP array using physiological and yield traits. *Plos one* **17**: e0262569.

Bangera R, Correa K, Lhorente JP, Figueroa R, Yáñez JM (2017). Genomic predictions can accelerate selection for resistance against Piscirickettsia salmonis in Atlantic salmon (*Salmo salar*). *BMC genomics* **18**: 1–12.

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

Carboni S, Evans S, Tanner KE, Davie A, Bekaert M, Fitzer SC (2021). Are Shell Strength Phenotypic Traits in Mussels Associated with Species Alone? *Aquaculture Journal,* **1(1)**: 3-13.

Corrochano-Fraile A, Davie A, Carboni S, Bekaert M (2021). Evidence of multiple genome duplication events in *Mytilus* evolution. *bioRxiv* 2021.08.17.456601.

FAO (2019). *The State of the World's Aquatic Genetic Resources for Food and Agriculture*

Gallardo-Escárate C, Valenzuela-Muñoz V, Nuñez-Acuña G, Valenzuela-Miranda D, Tapia F, Yévenes M, et al. (2022). The native mussel *Mytilus chilensis* genome reveals adaptative molecular signatures facing the marine environment. *bioRxiv* 2022.09.06.506863.

Gardner JPA (1994). The *Mytilus edulis* species complex in Southwest England: Multi-locus heterozygosity, background genotype and a fitness correlate. *Biochemical Systematics and Ecology* **22**: 1–11.

Gardner JPA (1996). The *Mytilus edulis* species complex in southwest England: effects of hybridization and introgression upon interlocus associations and morphometric variation. *Marine Biology* **125**: 385–99.

Gardner JPA, Skibinski DOF (1991). Biological and physical factors influencing genotype-dependent mortality in hybrid mussel populations. *Marine Ecology Progress Series* 235–43.

Gardner JPA, Thompson RJ (2001). The effects of coastal and estuarine conditions on the physiology and survivorship of the mussels *Mytilus edulis, M. trossulus* and their hybrids. *Journal of Experimental Marine Biology and Ecology* **265**: 119–40.

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**: 3182–95.

Garner JB, Douglas ML, Williams SR, Wales WJ, Marett LC, Nguyen TTT, et al. (2016). Genomic selection improves heat tolerance in dairy cattle. *Scientific reports* **6**: 1–9.

Gerdol M, Moreira R, Cruz F, Gómez-Garrido J, Vlasova A, Rosani U, et al. (2019). Massive gene presence/absence variation in the mussel genome as an adaptive strategy: first evidence of a pan-genome in Metazoa. *bioRxiv*

Gerdol M, Moreira R, Cruz F, Gómez-Garrido J, Vlasova A, Rosani U, et al. (2020). Massive gene presence-absence variation shapes an open pan-genome in the Mediterranean mussel. *Genome Biology* **21**: 275.

Gutierrez AP, Matika O, Bean TP, Houston RD (2018). Genomic Selection for Growth Traits in Pacific Oyster (*Crassostrea gigas*): Potential of Low-Density Marker Panels for Breeding Value Prediction. *Frontiers in Genetics* **9**:391.

Hilbish T, Carson E, Plante J, Weaver L, Gilg M (2002). Distribution of *Mytilus edulis, M.*

galloprovincialis, and their hybrids in open-coast populations of mussels in southwestern England. *Marine Biology* **140**: 137–42.

Jones SJ, Lima FP, Wethey DS (2010). Rising environmental temperatures and biogeography: poleward range contraction of the blue mussel, *Mytilus edulis* L., in the western Atlantic. *Journal of Biogeography* **37**: 2243–59.

Knöbel L, Nascimento-Schulze JC, Sanders T, Zeus D, Hiebenthal C, Barboza FR, et al. (2021). Salinity Driven Selection and Local Adaptation in Baltic Sea Mytilid Mussels. *Frontiers in Marine Sciences,* **8:** 692078.

Mathiesen SS, Thyrring J, Hemmer-Hansen J, Berge J, Sukhotin A, Leopold P, et al. (2017). Genetic diversity and connectivity within *Mytilus* spp. in the subarctic and Arctic. *Evolutionary Applications* **10**: 39–55.

Melzner F, Stange P, Trübenbach K, Thomsen J, Casties I, Panknin U, et al. (2011). Food Supply and Seawater pCO2 Impact Calcification and Internal Shell Dissolution in the Blue Mussel *Mytilus edulis*. *PLOS ONE* **6**: e24223.

Meuwissen THE, Hayes BJ, Goddard ME (2001). Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics* **157**: 1819 LP – 1829.

Nguyen TTT, Bowman PJ, Haile-Mariam M, Pryce JE, Hayes BJ (2016). Genomic selection for tolerance to heat stress in Australian dairy cattle. *Journal of dairy science* **99**: 2849–62.

Nguyen TTT, Hayes BJ, Ingram BA (2014). Genetic parameters and response to selection in blue mussel (*Mytilus galloprovincialis*) using a SNP-based pedigree Potential broodstock candidates. *Aquaculture* **420**–**421**: 295–301.

Paggeot LX, DeBiasse MB, Escalona M, Fairbairn C, Marimuthu MPA, Nguyen O, et al. (2022). Reference genome for the California ribbed mussel, *californianus*, an ecosystem engineer. *Journal of Heredity* esac041.

Pörtner H-O (2010). Oxygen- and capacity-limitation of thermal tolerance: a matrix for integrating climate-related stressor effects in marine ecosystems. *Journal of Experimental Biology* **213**: 881–93.

Pörtner H-O, Roberts DC, Masson-Delmotte V, Zhai P, Tignor M, Poloczanska E, et al. (2019). The ocean and cryosphere in a changing climate. *IPCC Special Report on the Ocean and Cryosphere in a Changing Climate*

Regan T, Bean TP, Ellis T, Davie A, Carboni S, Migaud H, et al. (2021). Genetic improvement technologies to support the sustainable growth of UK aquaculture. *Reviews in Aquaculture* **13**: 1958–85.

Robledo D, Matika O, Hamilton A, Houston RD (2018a). Genome-wide association and genomic selection for resistance to amoebic gill disease in Atlantic salmon. *G3: Genes, Genomes, Genetics* **8**: 1195–1203.

Robledo D, Palaiokostas C, Bargelloni L, Martínez P, Houston R (2018b). Applications of genotyping by sequencing in aquaculture breeding and genetics. *Reviews in Aquaculture* **10**: 670–82.

Ruggieri V, Calafiore R, Schettini C, Rigano MM, Olivieri F, Frusciante L, et al. (2019). Exploiting genetic and genomic resources to enhance heat-tolerance in tomatoes. *Agronomy* **9**: 22.

Saarman NP, Pogson GH (2015). Introgression between invasive and native blue mussels (genus *Mytilus*) in the central California hybrid zone. *Molecular Ecology* **24**: 4723–38.

Sanders T, Schmittmann L, Nascimento-schulze JC, Melzner F (2018). High Calcification Costs Limit Mussel Growth at Low Salinity. **5**: 1–9.

Simon A (2022). Three new genome assemblies of blue mussel lineages: North and South European *Mytilus edulis* and Mediterranean *Mytilus galloprovincialis bioRxiv* 2022.09.02.506387.

Simon A, Fraïsse C, El Ayari T, Liautard-Haag C, Strelkov P, Welch JJ, et al. (2021). How do species barriers decay? Concordance and local introgression in mosaic hybrid zones of mussels. *Journal of Evolutionary Biology* **34**: 208–23.

6. Appendix

Genome-wide association and genomic prediction of growth traits in the European flat oyster (*Ostrea edulis*)

This work has been published in the Scientific Journal 'Frontiers in Genetics as Carolina Peñaloza, Agustin Barria, Athina Papadopoulou, Chantelle Hooper, Joanne Preston, Matthew Green, Luke Helmer, Jacob Kean Hammerson, **Jennifer C Nascimento-Schulze**, Diana Minardi, Manu Kumar Gundappa, Daniel J Macqueen, John Hamilton, Ross D Houston, Tim P Bean (2022). Genome-wide association and genomic prediction of growth traits in the European flat oyster (*Ostrea edulis*). *Front. genetics* (doi: https://doi.org/10.1101/2022.06.10.495672)

Work description

As part of my PhD studies, I have worked on the "AquaLeap: Innovation in Genetics and Breeding to Advance UK Aquaculture Production" project, investigating the impacts of Bonamiosis, a disease caused by the haplosporidean *Bonamia ostreae*, on European flat oysters (*Ostrea edulis*) at Cefas, Weymouth, UK.

During ten months, I contributed to the development of methods applied in the flat oyster project including: production of a bonamia plasmid for parasite quantification, parasite purification, as well as on the execution of the infection experiment and husbandry of oysters throughout the challenge. I was also involved in data collection at the end of experiments carried within the project.

The first output of the research developed in this project is the publication released in the scientific journal 'Frontiers in Genetics' led by Dr Carolina Penaloza. The publication is attached in the following pages of the appendix.

Symonds JE, King N, Camara MD, Ragg NLC, Hilton Z, Walker SP, et al. (2018). New Zealand aquaculture selective breeding: from theory to industry application for three flagship species. *Proceedings of the World Congress on Genetics Applied to Livestock Production* **11**: 1035.

Thudi M, Upadhyaya HD, Rathore A, Gaur PM, Krishnamurthy L, Roorkiwal M, et al. (2014). Genetic dissection of drought and heat tolerance in chickpea through genome-wide and candidate gene-based association mapping approaches. *Plos one* **9**: e96758.

Tsai H-Y, Hamilton A, Tinch AE, Guy DR, Gharbi K, Stear MJ, et al. (2015). Genome wide association and genomic prediction for growth traits in juvenile farmed Atlantic salmon using a high density SNP array. *BMC Genomics* **16**: 969.

Vajedsamiei J, Wahl M, Schmidt AL, Yazdanpanahan M, Pansch C (2021). The higher the needs, the lower the tolerance: Extreme events may select ectotherm recruits with lower metabolic demand and heat sensitivity. *Frontiers in Marine Science* **8**: 264.

Ventura A, Schulz S, Dupont S (2016). Maintained larval growth in mussel larvae exposed to acidified under-saturated seawater. *Scientific Reports* **6**: 1–10.

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

Wenne R, Bach L, Zbawicka M, Strand J, McDonald JH (2016). A first report on coexistence and hybridization of *Mytilus trossulus* and *M. edulis mussels* in Greenland. *Polar Biology* **39**: 343–55.

Wilson J, Matejusova I, McIntosh RE, Carboni S, Bekaert M (2018). New diagnostic SNP molecular markers for the *Mytilus* species complex. *PLoS ONE* **13**: 1–13.

Yang J-L, Feng D-D, Liu J, Xu J-K, Chen K, Li Y-F, et al. (2021). Chromosome-level genome assembly of the hard-shelled mussel *Mytilus coruscus*, a widely distributed species from the temperate areas of East Asia. *GigaScience* **10**: giab024.

Zbawicka M, Sańko T, Strand J, Roman W (2014). New SNP markers reveal largely concordant clinal variation across the hybrid zone between *Mytilus* spp. in the Baltic Sea. *Aquatic Biology* **21**: 25–36.

Zenger KR, Khatkar MS, Jones DB, Khalilisamani N, Jerry DR, Raadsma HW (2019). Genomic Selection in Aquaculture: Application, Limitations and Opportunities With Special Reference to Marine Shrimp and Pearl Oysters. *Frontiers in genetics*, **9:** 693.

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 15 July 2022
doi: 10.3389/fgene.2022.926638

Genome-Wide Association and Genomic Prediction of Growth Traits in the European Flat Oyster (*Ostrea edulis*)

Carolina Peñaloza[1], Agustin Barria[1], Athina Papadopoulou[2], Chantelle Hooper[2], Joanne Preston[3], Matthew Green[2], Luke Helmer[3,4,5], Jacob Kean-Hammerson[4], Jennifer C. Nascimento-Schulze[2,6], Diana Minardi[2], Manu Kumar Gundappa[1], Daniel J. Macqueen[1], John Hamilton[7], Ross D. Houston[8]† and Tim P. Bean[1]*†*

[1]The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Edinburgh, United Kingdom, [2]Centre for Environment, Fisheries and Aquaculture Science (Cefas), Weymouth Laboratory, Weymouth, United Kingdom, [3]Institute of Marine Sciences, University of Portsmouth, Portsmouth, United Kingdom, [4]Blue Marine Foundation, London, United Kingdom, [5]Ocean and Earth Science, University of Southampton, Southampton, United Kingdom, [6]College of Life and Environmental Sciences, University of Exeter, Exeter, United Kingdom, [7]Lochnell Oysters, Oban, United Kingdom, [8]Benchmark Genetics, Penicuik, United Kingdom

OPEN ACCESS

Edited by:
Nguyen Hong Nguyen,
University of the Sunshine Coast, Australia

Reviewed by:
Neil Thompson,
Agricultural Research Service (USDA), United States
Panya Sae-Lim,
MOWI Genetics, Norway

***Correspondence:**
Tim P. Bean
tim.bean@roslin.ed.ac.uk
Ross D. Houston
ross.houston@bmkgenetics.com

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: 22 April 2022
Accepted: 17 June 2022
Published: 15 July 2022

Citation:
Peñaloza C, Barria A, Papadopoulou A, Hooper C, Preston J, Green M, Helmer L, Kean-Hammerson J, Nascimento-Schulze JC, Minardi D, Gundappa MK, Macqueen DJ, Hamilton J, Houston RD and Bean TP (2022) Genome-Wide Association and Genomic Prediction of Growth Traits in the European Flat Oyster (Ostrea edulis). Front. Genet. 13:926638. doi: 10.3389/fgene.2022.926638

The European flat oyster (*Ostrea edulis*) is a bivalve mollusc that was once widely distributed across Europe and represented an important food resource for humans for centuries. Populations of *O. edulis* experienced a severe decline across their biogeographic range mainly due to overexploitation and disease outbreaks. To restore the economic and ecological benefits of European flat oyster populations, extensive protection and restoration efforts are in place within Europe. In line with the increasing interest in supporting restoration and oyster farming through the breeding of stocks with enhanced performance, the present study aimed to evaluate the potential of genomic selection for improving growth traits in a European flat oyster population obtained from successive mass-spawning events. Four growth-related traits were evaluated: total weight (TW), shell height (SH), shell width (SW) and shell length (SL). The heritability of the growth traits was in the low-moderate range, with estimates of 0.45, 0.37, 0.22, and 0.32 for TW, SH, SW and SL, respectively. A genome-wide association analysis revealed a largely polygenic architecture for the four growth traits, with two distinct QTLs detected on chromosome 4. To investigate whether genomic selection can be implemented in flat oyster breeding at a reduced cost, the utility of low-density SNP panels was assessed. Genomic prediction accuracies using the full density panel were high (> 0.83 for all traits). The evaluation of the effect of reducing the number of markers used to predict genomic breeding values revealed that similar selection accuracies could be achieved for all traits with 2K SNPs as for a full panel containing 4,577 SNPs. Only slight reductions in accuracies were observed at the lowest SNP density tested (i.e., 100 SNPs), likely due to a high relatedness between individuals being included in the training and validation sets during cross-validation. Overall, our results suggest that the genetic improvement of growth traits in oysters is feasible. Nevertheless, and although low-density SNP panels appear as a promising strategy for applying GS at a reduced cost, additional populations with different degrees of genetic relatedness should be assessed to derive estimates of prediction accuracies to be expected in practical breeding programmes.

Keywords: *Ostrea edulis*, oyster, GWAS, genomic selection, growth, aquaculture

INTRODUCTION

The European flat oyster (*Ostrea edulis*) was an abundant native bivalve species and an important fishery resource in much of Europe up to the 19th century (Pogoda, 2019). However, populations of *O. edulis* experienced a severe decline across their biogeographic range due to an array of factors including overfishing and habitat degradation (Thurstan et al., 2013), the subsequent invasion of non-native species (e.g., slipper limpet, *Crepidula fornicata*) (Helmer et al., 2019; Preston et al., 2020) and pathogenic diseases (Robert et al., 1991; Sas et al., 2020). The continuous decimation of native populations in the Atlantic and Mediterranean seas led to significant changes in oyster production, which progressively shifted towards farming (Korringa, 1976), and eventually to the cultivation of different species including *Crassostrea angulata* (Boeliguf, 2000) and the non-indigenous Pacific oyster (*Crassostrea gigas*) (Walne and Helm, 1979; Grizel and Héral, 1991). The Pacific oyster was introduced into Europe for aquaculture purposes owing to its favourable production traits, such as a faster growth rate and higher resistance to the main diseases affecting *C. angulata* and *O. edulis* (Grizel and Héral, 1991; Renault et al., 1995). Worldwide oyster production is now dominated by the Pacific oyster (97.7%), while the production of the European flat oyster remains stably low, constituting just ~0.2% of global production in 2002 (FAO, 2022). Despite the demand for shellfish continues to increase (Botta et al., 2020), the level of *O. edulis* production is stagnant. One of the main factors that hinders the growth of the industry is the lack of a substantial and steady supply of oyster seed (i.e., juveniles) (see Colsoul et al. (2021) for a review). Hence, the optimization of oyster larval production in hatcheries and spatting ponds is key for future European flat oyster aquaculture, as well as for restoration projects, which are also expected to rely on sustainable sources of juveniles for restocking (Pogoda et al., 2020). Importantly, the artificial propagation of flat oyster seed will facilitate the application of selective breeding programmes. Although selective breeding programmes are typically used to improve aquaculture production, they could also benefit the ecological restoration of *O. edulis*. If desirable traits such as disease resistance are found to have a strong genetic component, then increased resistance to life-limiting diseases — such as bonamiosis (Naciri-Graven et al., 1998; Culloty et al., 2004) — could potentially be achieved while maintaining the adaptive potential (i.e., genetic diversity) of restored populations.

Selective breeding in oysters has mainly focused on improving meat yield, disease resistance, survival and growth (Toro and Newkirk, 1990; Allen et al., 1993; Ragone Calvo et al., 2003; Ward et al., 2005; Dégremont et al., 2015; De Melo et al., 2016; Proestou et al., 2016; Camara et al., 2017; Zhang et al., 2019), with a recent interest in nutritional content and shell shape (Grizzle et al., 2017; Liu et al., 2019; Meng et al., 2019; Wan et al., 2020; He et al., 2022). Among these traits, growth is comparatively simple to assess and consequently select for using phenotypic information. Although the direct comparison of heritability estimates from different studies is difficult (e.g., due to intrinsic differences between populations), estimates for growth rate in oysters tend to be moderate (e.g., 0.31 and 0.55—Evans and Langdon (2006) and De Melo et al. (2016), respectively). As a result, fast-growing lines of oysters have been developed for some of the main commercial species, such as the Pacific (*C. gigas*) (Zhang et al., 2019), Portuguese (*C. angulata*) (Vu et al., 2020), Eastern (*C. virginica*) (Varney and Wilbur, 2020) and Sydney rock (*Saccostrea glomerata*) (Fitzer et al., 2019) oyster. Initial attempts to genetically improve the European flat oyster *O. edulis* resulted in an average 23% increase in growth rate compared to an unselected (control) line (Newkirk and Haley, 1982). This striking genetic response was not replicated in a second generation of selection, possibly due to unintentional inbreeding (Newkirk and Haley, 1983). Indeed, even relatively modest levels of inbreeding have been shown to significantly affect performance traits in oysters (Evans et al., 2004), highlighting the importance of an adequate management of genetic diversity in hatchery-derived stocks. Moreover, oysters and bivalves in general, appear to have a high genetic load [see for a review Plough (2016)] and, therefore, may be particularly susceptible to inbreeding depression. Hence, the incorporation of genomic tools into shellfish breeding schemes will be key for balancing genetic gain with population diversity in order to sustain the long-term progress for traits under selection.

A vast array of genomic tools and resources have become available for genetic research and breeding applications in oysters. For example, for economically relevant species, chromosome-level genome assemblies (Li et al., 2021; Modak et al., 2021; Peñaloza et al., 2021; Qi et al., 2021), SNP arrays (Lapegue et al., 2014; Gutierrez et al., 2017; Qi et al., 2017) and medium-density linkage maps (Jones et al., 2013; Wang et al., 2016; Gutierrez et al., 2020; Li et al., 2018; Yin et al., 2020) have been produced. These resources have been applied to examine the genetic basis of growth (Jones et al., 2014; Gutierrez et al., 2018; He et al., 2021), low salinity tolerance (McCarty et al., 2022), disease resistance (Gutierrez et al., 2020; Yang et al., 2022) and nutritional content (Meng et al., 2019). For the European flat oyster, high-quality genomes have recently been released (Boutet et al., 2022; Gundappa et al., 2022), which along with available high-throughput genotyping techniques (e.g., SNP arrays and genotype-by-sequencing approaches), provide the opportunity for gaining insight into the genomic architecture of relevant production traits. Most of the traits of economic importance in aquaculture species have a polygenic architecture (Zenger et al., 2019). For polygenic traits (i.e., those controlled by many loci), the application of predictive techniques such as genomic selection (GS) may enable a faster genetic gain than conventional pedigree-based selection. GS is a method based on genome-wide markers in which the effect of all loci are simultaneously used for predicting the estimated breeding values (EBV) of selection candidates (Meuwissen et al., 2001), and has shown major potential in aquaculture species, where it can be used to

characterise variation within and between large families of potential breeders. However, commercial application to aquaculture production is largely limited to the major finfish and crustacean species (e.g., salmonids, Nile tilapia, shrimp) (Zenger et al., 2019; Lillehammer et al., 2020; Boudry et al., 2021). Studies into the feasibility of applying genomic selection schemes in oyster breeding programmes have shown that for growth (Gutierrez et al., 2018; Vu et al., 2021b), edibility (Vu et al., 2021a), and disease resistance traits (Gutierrez et al., 2020; Vu et al., 2021a), greater genetic gains could be achieved through GS compared to traditional breeding. Nevertheless, the practical application of GS as a selection strategy will likely depend on how cost-effective it is compared to pedigree-based methods. The development of feasible alternatives for reducing genotyping costs, such as using affordable low-density genotyping tools that yield similar accuracies than higher-density panels, will be critical for the potential of GS to be realized by oyster breeding programmes.

In line with the increasing interest in supporting oyster culture and restoration through the breeding of stocks with enhanced performance, the overall aim of this study was to evaluate the potential of GS for the genetic improvement of growth and growth-related (morphometric) traits in the European flat oyster. First, the heritability of total weight, shell length, shell width and shell height was estimated for a hatchery-derived population genotyped using a ~15K SNP array. Second, a genome-wide association (GWAS) analysis was conducted to dissect the genetic architecture of the measured traits. Last, to evaluate whether GS may be an effective and cost-effective strategy for improving traits associated with oyster growth, the accuracy of genomic predictions using reduced density SNP marker panels was assessed.

MATERIALS AND METHODS

Field Experiment

The European flat oyster population used in this study was generated in a UK-based hatchery (Seasalter Morecombe hatchery) by mass spawning of approximately 40 broodstock parents over several spawning events. The resulting F1 generation was then deployed to Lochnell oysters (56.494°N, 5.459°W) and grown for 6 months in ortac grow-out cages. Next, animals were transferred to the Institute of Marine Sciences at the University of Portsmouth (UK), and maintained in a flow-through system until deployment. During this holding period, ~1,000 randomly selected oysters were individually tagged and their first phenotype measurements recorded (see "Phenotypes" section below). Prior to deployment, animals were cleaned of fouling, washed in fresh water and dried. Embossed plastic tags with unique identifier codes were attached with epoxy resin glue. Animals were returned to aquaria within the hour. Oysters were placed in Aquamesh® cages (L 0.55 m × W 0.55 m × D 0.4 m) at a density of 200 oysters per cage, and deployed 1 m below floating pontoons at Port Hamble Marina (MDL) in the River Hamble (50.861°N, 1.312°W) in January 2019. Mortalities were recorded monthly and dead oysters—i.e., those with empty or gaping shells—were removed from the experiment. General disease status was assessed on subsets of oysters throughout the experiment by histology and *in situ* hybridisation using an adaptation of available methods (Montagnani et al., 2001; Fabioux et al., 2004). In addition, the presence of *Bonamia ostreae*, a protozoan parasite that causes a lethal infection of flat oyster haemocytes (Pichot et al., 1979), was assessed by qPCR following Robert et al. (2009). The prevalence of *B. ostreae* infections was negligible; hence disease status had a minor influence on the assessment of growth traits in the experimental population. After 10 months of growth under field conditions, gill tissue was dissected from individuals alive at the end of the study and preserved in molecular grade absolute ethanol (Fisher Scientific) for genetic analysis.

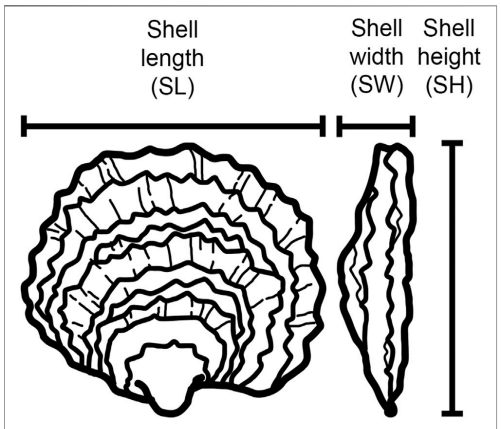


FIGURE 1 | Nomenclature of the growth-related morphometric traits measured in this study.

Phenotypes

Four growth-associated traits were measured at three time points over the course of 10 months: total weight (TW, the weight of an individual oyster including the shell), shell length (SL, the maximum distance between the anterior and posterior margins), shell height (SH, the maximum distance between the hinge to the furthermost edge), and shell width (SW, the maximum distance at the thickest part of the two shell valves) (**Figure 1**). Weight was recorded in grams up to one decimal place. Shell measurements were taken with traceable digital callipers (Fisher Scientific) with 0.02 mm precision. Oysters were cleaned and defouled before measurements were taken.

DNA Extraction

Total DNA was isolated from gill tissue following a CTAB (cetyltrimethylammonium bromide)-based extraction protocol [details in Gutierrez et al. (2017)]. The integrity of the extracted DNA was assessed by agarose gel electrophoresis, while DNA purity was verified on a Nanodrop ND-1000 (Thermo Fisher Scientific) spectrophotometer by checking the 260/280 and 260/230 ratios. All samples had 260/280 and 260/230 values ≥ 1.85 and ≥1.96, respectively.

SNP Genotyping and Quality Control

Whole-genome genotyping of ~15K SNPs was carried out by IdentiGEN (Dublin, Ireland) using the combined-species Affymetrix Axiom oyster SNP-array (Gutierrez et al., 2017). Signal intensity files were imported to the Axiom analysis Suite v4.0.3.3 software for quality control (QC) assessment and genotype calling. Genotypes were generated using the default parameter settings for diploid species, resulting in 11,808 SNPs typed for 870 individuals. To assess the reproducibility of genotype calls, five DNA samples from the same individual were genotyped independently on three different arrays, and their genotype concordance evaluated through an identity-by-state (IBS) analysis. The genotype concordance rate among replicates was 99.7%, demonstrating a high reproducibility of the genotyping assays. The flanking region of these markers were mapped to the *O. edulis* chromosome-level genome assembly (Gundappa et al., 2022). Of the 11,808 SNPs, 10,025 had uniquely mapping probes and were retained for downstream analysis. A total of 1,539 markers (15.4%) were monomorphic in the population under study. Additional QC was performed on markers and samples using Plink v2.0 (Chang et al., 2015). SNP variants were retained for further analysis if they had a call rate > 95% and a minor allele frequency (MAF) > 0.05. These filters removed 4,391 SNPs (leaving a total of 5,634 SNPs), of which the majority were filtered out based on the MAF threshold (i.e., were monomorphic or near-monomorphic in this population). Given that significant sub-clustering was detected in the data (**Supplementary Figure S1**), possibly due to a high variance in the reproductive success of broodstock parents and/or temporal variation in spawning, a k-means clustering method was used to assign individuals into groups. Deviations from Hardy-Weinberg Equilibrium (HWE) were tested separately in each of the three genetic clusters identified by the analysis. SNP markers showing significant deviations (HWE *p*-value < 1e-10) in two of the three clusters were excluded from the analysis. Sample QC included removing individual oysters with a missingness above 5% and high heterozygosity (i.e., more than three median absolute deviations from median). Finally, a principal component analysis (PCA) was performed using a set of ~3.5K SNPs for which no pair of markers within a window of 200 kb had a $r^2 > 0.5$. The top five PCs, which explain 47% of the variance (considering 20 PCs), were fitted in the model to account for the effect of population structure. The final dataset comprised 840 samples genotyped at 4,577 genome-wide SNPs.

Genetic Parameter Estimation

Genetic parameters for growth-related traits were estimated by fitting the following univariate linear mixed model in GEMMA v0.95alpha (Zhou and Stephens, 2012):

$$y = \mu + Xb + Zu + e \quad (1)$$

Where $\boldsymbol{y}$ is the vector of observed phenotypes; $\boldsymbol{\mu}$ is the overall mean of the phenotype in the population; $\mathbf{b}$ is the vector of fixed effects to be fitted (the first five principal components were included as covariates); $\mathbf{u}$ is the vector of the additive genetic effects; $\boldsymbol{X}$ and $\boldsymbol{Z}$ are the corresponding incidence matrices for fixed and additive effects, respectively; and e is a vector of residuals. The following distributions were assumed: $u \sim N(0, G\sigma_u^2)$ and $e \sim N(0, I\sigma_e^2)$. Where σ_u^2 and σ_e^2 are the additive genetic and residual variance, respectively, $\mathbf{G}$ is the genomic relationship matrix and $\mathbf{I}$ is the identity matrix. The heritability of growth-related traits was estimated as the ratio of the additive genetic variance to the total phenotypic variance.

Bivariate animal linear models were implemented to estimate the genetic (co)variance between TW, SL, SH and SW. Each bivariate analysis was fitted with the same top 5 PCs mentioned above. Subsequently, genetic correlations among traits were estimated as the ratio of the covariance of two traits to the square root of the product of the variance for each trait. Phenotypic correlations between traits were calculated using the Pearson correlation coefficient.

Genome-Wide Association Study

To identify SNPs in the flat oyster genome correlated with variation in growth-related traits, a GWAS was performed by implementing the same model described previously in the GEMMA software. SNPs were considered significant at the genome-wide level if their likelihood ratio test P-values surpassed a conservative Bonferroni-corrected significance threshold (α/4,577 = 1.09e-5). To derive a threshold for chromosome-wide (suggestive) significance, α was divided by the average number of SNPs per chromosome (α/457 = 1.09e-4). The single-marker P-values obtained from GEMMA were plotted against their chromosome location using the R package qqman v 0.1.4 (Turner, 2018). To assess the inflation of the association statistics, the genomic control coefficient lambda λGC was calculated following Devlin and Roeder (1999). Candidate genes were searched within 100 kb of the most significant SNP loci using BEDOPS v2.4.26 (Neph et al., 2012).

Genomic Prediction

To evaluate the accuracy of genomic selection, a 5-fold cross validation approach — animals split into training (80%) and validation (20%) sets—was used on a population of 840 oysters genotyped for 4,577 informative SNP markers. To reduce stochastic effects arising from individual sampling, each analysis was repeated 10 times. For each replicate, animals were randomly partitioned into five subsets (each subset contained 168 individuals). TW, SL, SH and SW phenotypes recorded in individuals allocated to one of the subsets (validation set) were masked. The breeding values of the validation set were then predicted based on the information from the remaining four subsets (training sets) using model (1). The model was fitted using the AIREMLF90 module from BLUPF90. The accuracy of genomic predictions was calculated as follows:

$$Accuracy = \frac{r_{gEBV,\,y}}{h}$$

where $r_{gEBV,y}$ is the correlation between the predicted and the actual phenotypes of the validation set, while h is the square root of the heritability of the trait estimated as described above.

Evaluation of the Effect of SNP Density on Genomic Predictions

To assess the effect of SNP density on the accuracy of genomic predictions of growth-related traits, SNP panels of varying sizes were randomly sampled from the final pool of QC-filtered array markers (n = 4,577 SNPs). Panels of the following densities were evaluated: 4, 3, 2, 1K, 500, 400, 300, 200 and 100 SNPs. To build the lower-density panels, markers were randomly sampled from the full QC-filtered SNP dataset in proportion to chromosome lengths using the R package CVrepGPAcalc v1.0 (Tsairidou, 2019; Tsairidou et al., 2020). To account for sampling bias, 10 SNP panels were generated for each of the SNP densities. The average genomic prediction accuracies of the different low-density panels were compared against the equivalent accuracy values obtained with the full panel.

RESULTS AND DISCUSSION

Growth Traits and Heritability

Improvement of growth rate is typically one of the first traits to be included as a selection target in breeding programmes across many farmed species. In this study, oyster growth rate was assessed in a hatchery-derived oyster population that was translocated to a growing site and monitored for 10 months. The experimental population had a lower genome-wide heterozygosity (Ho = 0.27; He = 0.22) compared to the values reported by Vera et al. (2019) (Ho ≥ 0.31) for a diverse set of flat oyster populations genotyped with the same array. An overall mortality of 14% was observed during the field trial, among which the majority (36%) occurred during a summer month (July). At the end of the experimental period, the *O. edulis* population had the following growth means and standard deviations: +15.7 g (SD = 5.8), 50.8 mm (SD = 7.3), 12.9 mm (SD = 2.3) and 45.8 mm (SD = 8.9), for TW, SH, SW, SL, respectively (**Table 1**). The phenotypic correlation was found to be the highest (r > 0.8) between two pairs of traits: (i) TW and SH, and (ii) TW and SL (**Figure 2**).

For the European oyster population under study, the heritability estimates of these growth-related traits were in the low to moderate range of 0.22 (for SW) to 0.45 (for TW) (**Table 2**). Consistent with similar studies carried out in related oyster species (Xu et al., 2017; Vu et al., 2020), heritability estimates based on SNP markers were higher for total weight than for growth-related morphometric traits (i.e., shell height, shell width and shell length). The estimation of heritability for total weight (TW) was similar to those reported for nine- month-old Portuguese oysters (h^2 = 0.45) and a 2-year old Pacific oyster strain (h^2 = 0.42) (Xu et al., 2017; Vu et al., 2021b). Total weight, as measured in this study, is a composite phenotype made up of the animal's shell and soft tissue weights, in addition to the weight of any pallial fluid - thus is not a direct reflection of meat yield. Nevertheless, in *C. angulata*, a positive genetic correlation (0.63) has been found between TW and soft tissue weight (Vu et al., 2021b), suggesting that selecting for TW—a trait easier to measure—could lead to improvements in meat yields. Such indirect improvements of correlated traits have been reported in a Portuguese oyster line selected for harvest weight. While the achieved average selection response for total weight at harvest was

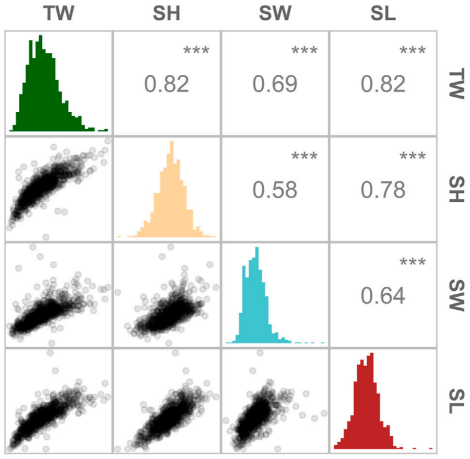


FIGURE 2 | Distribution and magnitude of the phenotypic correlations between growth-related traits in *Ostrea edulis*. Pearson's correlation between traits (above the diagonal), histogram of trait distribution (diagonal) and scatterplots comparing two traits (below the diagonal). TW (total weight), SL (shell length), SH (shell height) and SW (shell width). *** indicates p-values < 0.001.

TABLE 1 | Summary statistics of the phenotypic data (SD: Standard deviation; CV: coefficient of variation).

Trait	Unit	Mean	Min	Max	SD	CV (%)
Total weight	g	15.7	4.0	38.5	5.8	36.8
Shell height	mm	50.8	22.9	76.1	7.3	14.4
Shell width	mm	12.9	6.5	27.7	2.3	18.2
Shell length	mm	45.8	22.2	94.4	8.9	19.5

TABLE 2 | Estimates of heritability (h^2) and SE on the diagonal and pairwise genetic correlations (below the diagonal) for growth-related traits in a European flat oyster population.

Trait	Total weight	Shell height	Shell width	Shell length
Total weight	**0.45 (0.06)**			
Shell height	0.99	**0.37 (0.06)**		
Shell width	0.96	0.90	**0.22 (0.05)**	
Shell length	0.95	0.93	0.88	**0.32 (0.06)**

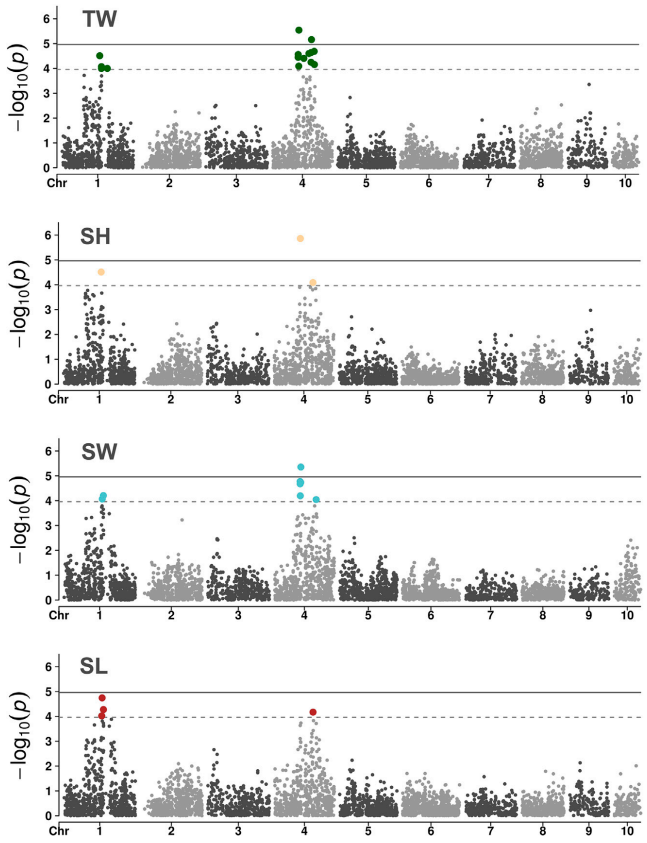


FIGURE 3 | Manhattan plots of the GWAS for growth-related traits in a European flat oyster population. Solid lines indicate the threshold value for genome-wide significance. Dashed lines indicate the threshold for a suggestive (chromosome-level) significance.

5.8% per generation, genetic gains were also observed for soft tissue weight, with indirect gains reaching a 1.2% increase per generation (Vu et al., 2020). For the shell-related traits examined in this study (SH, SW and SL), heritability estimates were in line with previous studies (Yuehuan et al., 2017; Gutierrez et al., 2018), and ranged from 0.22 to 0.37. Traditionally, the focus on shell morphometric traits was to improve oyster growth. Nevertheless, in recent years, oyster shell shape is increasingly being viewed as an attractive goal for selective breeding due to its growing importance for consumers (Mizuta and Wikfors, 2019). The perceived attractiveness of an oyster shell can be represented as a secondary trait derived from a ratio between primary (shell dimension) traits, such as the shell width index (Kube et al., 2011). Given that significant heritable variation was observed for the three examined morphometric traits, strategies for homogenizing particular shell shapes may be feasible in *O. edulis*.

Genome-Wide Association Analysis for Growth-Related Traits

A GWAS of ~4.5K SNPs passing the filtering criteria were genotyped on 840 oysters with phenotypic records to gain insight into the genetic basis of growth rate variation in *O. edulis*. Three of the four examined traits showed association signals surpassing the genome-wide level of significance (**Figure 3**). The genomic inflation factor lambda of the GWAS analysis were close to the desired value (λ = 1) (see **Supplementary Figure S2**), indicating that population structure was adequately accounted for by the model. For TW, the GWAS identified two putative quantitative trait loci (QTLs) on chromosome 4 associated with the trait. The presence of two separate QTLs is supported by the low linkage disequilibrium observed between the most significant SNPs at each locus

(pairwise $r^2 < 0.1$). An additional 13 suggestive loci were also identified, of which nine were located in the vicinity of the two abovementioned genome-wide hits and four were found on chromosome 1 (**Supplementary Table S1**). The SNP showing the strongest association with TW (AX-169174635) explained 3% of the phenotypic variance. This lead SNP was also found to be significantly associated with SH and SW. For SL, no SNP reached a genome-wide significance level, although a few of the same markers showing associations with TW, SH and SW surpassed the threshold for suggestive significance. The complete overlap of GWAS hits across the different traits suggests a high degree of shared genetic control among them, consistent with the high positive genetic correlations observed (**Table 2**). Overall, the GWAS results indicate that growth-related traits in *O. edulis* are influenced by many small-effect loci, exhibiting a polygenic architecture, but that two regions on chromosome 4 may have a moderate effect on these traits.

The marker showing the most significant association with TW, SH and SW is located in the exon of a gene annotated as a *N4BP2* (NEDD4 Binding Protein 2)-like protein (Gene ID: *FUN_017843*; Gundappa et al., 2022). The predicted protein product of this gene contains a highly conserved AAA domain, able to bind and hydrolyse ATP (Lupas and Martin, 2002). Proteins with these domains have been shown to be involved in several cell processes, including protein folding, proteolysis and membrane fusion. Further characterization of this *N4BP2*-like protein would help better understand the genetic component of growth variation in oysters. Nevertheless, considering that the candidate marker on *N4BP2* explained a small percentage of the phenotypic variance, independent oyster populations should first be evaluated to confirm the validity of the association signal. A second genome-wide significant association—detected only in the TW GWAS—was located in the exon of an uncharacterized gene (*FUN_018833*) whose product displays > 80% amino acid identity and 99% coverage (best BLASTp hit to NCBI's nr database) with similarly uncharacterized proteins in *C. gigas* and *C. virginica* (NCBI accession numbers XP_011433755.2 and XP_022325737.1, respectively). Additional genes within the two genomic regions (+/- 100 kb) showing significant associations with flat oyster growth traits are shown in **Supplementary Table S2**. Given that the SNPs identified in this study had a small effect on the traits in question, GS would be an effective approach for increasing genetic gains from selection.

Genomic Selection

The incorporation of genetic markers into breeding programmes requires a previous understanding of the genetic architecture of the targeted trait(s). In the *O. edulis* population under study, the genetic contribution to the observed variation in growth-related traits was largely polygenic in nature. For the improvement of polygenic traits, genomic selection has been shown to be superior to alternative marker-aided selection due to genome-wide markers capturing a higher proportion of the genetic variation in a trait compared to individual QTL-targeted markers. Consequently, by means of applying GS, higher predictions have been achieved for several production traits in a wide range of commercially important aquaculture species [reviewed in Houston et al. (2020)]. Despite GS not yet being widely operational in oyster breeding programmes

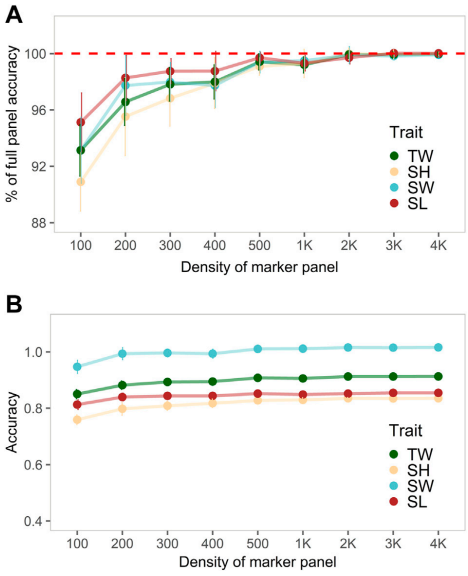


FIGURE 4 | Evaluation of the effect of SNP density on genomic predictions of growth related traits in a European flat oyster population. **(A)** Percentage of the maximum genomic prediction accuracy achieved using different lower density SNP panels. Values were calculated by dividing the mean accuracy (averaged over ten replicates) estimated at each nominal SNP density by the accuracy obtained using the full SNP dataset. **(B)** Average genomic prediction accuracy values obtained for oyster growth traits at different panel densities.

(Boudry et al., 2021), studies have demonstrated the potential of incorporating genome-wide information into selection schemes in these taxa. In the Pacific oyster, Gutierrez et al. (2018) showed that prediction accuracies for growth-traits increased ~25–30% when the genetic merit of individuals was estimated from SNP markers using the Genomic Best Linear Unbiased Prediction (GBLUP) model (VanRaden, 2008) compared to a classical pedigree-based approach (PBLUP). Similar results were reported in the Portuguese oyster, as prediction accuracies increased 7–42% for growth-related traits when EBVs obtained by GBLUP were compared to those obtained by PBLUP (Vu et al., 2021a). Since the flat oyster population under study derived from a mass-spawning event, the pedigree structure was unknown. Therefore, comparisons between pedigree and genome based methods for estimating breeding values (e.g., GBLUP and Bayesian approaches) could not be performed.

One of the major barriers of implementing GS is the high number of markers required to accurately predict EBVs and the cost of genotyping these markers (Goddard and Hayes, 2007). Therefore, the design of a strategy to reduce the cost of genotyping is critical for the extensive adoption of genomic prediction approaches in aquaculture breeding programmes.

experiment including sample collection and recording phenotype information. AP extracted the DNA used for genotyping. CP and AB performed the analysis. MG, DM and CH evaluated disease status. MKG and DJM provided intellectual input to the analysis. CP and AB wrote the manuscript with input from all authors.

FUNDING

The authors acknowledge funding from the Biotechnology and Biological Sciences Research Council (BBSRC), including Institute Strategic Programme grants (BBS/E/D/20002172, BBS/E/D/30002275 and BBS/E/D/10002070), a grant within the AquaLeap project (BB/S004181/1), funding from Sustainable Aquaculture Innovation Centre (SAIC), Blue Marine Foundation and National Fish and Wildlife Foundation (NFWF).

ACKNOWLEDGMENTS

The authors would like to acknowledge MDL Port Hamble Marina for allowing positioning of oyster cages within the Marina, and thank Eric Harris-Scott, Matthew Sanders, Monica Fabra, Tim Regan and Zenaba Khatir for their invaluable help with setup and sampling of the field experiment. For the purpose of open access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.926638/full#supplementary-material

Supplementary Figure S1 | PCA of the *O.edulis* population under study showing the dominance of single cluster comprised of highly related individuals (clst_0).

Supplementary Figure S2 | Quantile-quantile (Q-Q) plots showing the distribution of the expected (red dashed line) versus observed P-values of the GWAS of growth-related traits.

Supplementary Figure S3 | Example of the pairwise kinship coefficient between individuals in the training and validation set across randomly selected iterations of different 5-fold cross validation (CV) replicates (ten replicates in total). Boxplots show the distribution of values for pairs of individuals (one from the training set and one from the validation set) belonging to the same or different population clusters. Relatedness was inferred using the KING-robust method implemented in Plink v2.0.

REFERENCES

Allen, S. K., Gaffney, P. M., and Ewart, J. W. (1993). *Genetic Improvement of the Eastern Oyster for Growth and Disease Resistance in the Northeast*. Massachusetts: Northeastern Regional Aquaculture Center, 210.

Al-Tobasei, R., Ali, A., Garcia, A. L. S., Lourenco, D., Leeds, T., and Salem, M. (2021). Genomic Predictions for Fillet Yield and Firmness in Rainbow Trout Using Reduced-Density SNP Panels. *BMC Genomics* 22, 92. doi:10.1186/s12864-021-07404-9

Barría, A., Benzie, J. A. H., Houston, R. D., De Koning, D.-J., and de Verdal, H. (2021). Genomic Selection and Genome-wide Association Study for Feed-Efficiency Traits in a Farmed Nile Tilapia (*Oreochromis niloticus*) Population. *Front. Genet.* 12, 737906. doi:10.3389/fgene.2021.737906

Boeliguf, G. (2000). Present Status of the French Aquaculture. *Aquac. Sci.* 48 (2), 243–248.

Botta, R., Asche, F., Borsum, J. S., and Camp, E. V. (2020). A Review of Global Oyster Aquaculture Production and Consumption. *Mar. Policy* 117, 103952. doi:10.1016/j.marpol.2020.103952

Boudry, P., Allal, F., Aslam, M. L., Bargelloni, L., Bean, T. P., Brard-Fudulea, S., et al. (2021). Current Status and Potential of Genomic Selection to Improve Selective Breeding in the Main Aquaculture Species of International Council for the Exploration of the Sea (ICES) Member Countries. *Aquac. Rep.* 20, 100700. doi:10.1016/j.aqrep.2021.100700

Boutet, I., Monteiro, H. J. A., Takeuchi, T., Baudry, L., Bonnivard, E., Billoud, B., et al. (2022). Chromosomal Assembly of the Flat Oyster (*Ostrea Edulis* L.) Genome as a New Genetic Resource for Aquaculture. *bioRxiv [Preprint]*. Available at: https://doi.org/10.1101/2022.06.26.497643. (Accessed June 29, 2022).

Camara, M. D., Yen, S., Kaspar, H. F., Kesarcodi-Watson, A., King, N., Jeffs, A. G., et al. (2017). Assessment of Heat Shock and Laboratory Virus Challenges to Selectively Breed for Ostreid Herpesvirus 1 (OsHV-1) Resistance in the Pacific Oyster, *Crassostrea gigas*. *Aquaculture* 469, 50–58. doi:10.1016/j.aquaculture.2016.11.031

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., and Lee, J. J. (2015). Second-generation PLINK: Rising to the Challenge of Larger and Richer Datasets. *GigaSci* 4, 7. doi:10.1186/s13742-015-0047-8

Colsoul, B., Boudry, P., Pérez-Parallé, M. L., Bratoš Cetinić, A., Hugh-Jones, T., Arzul, I., et al. (2021). Sustainable Large-scale Production of European Flat Oyster (*Ostrea edulis*) Seed for Ecological Restoration and Aquaculture: A Review. *Rev. Aquac.* 13, 1423–1468. doi:10.1111/raq.12529

Correa, K., Bangera, R., Figueroa, R., Lhorente, J. P., and Yáñez, J. M. (2017). The Use of Genomic Information Increases the Accuracy of Breeding Value Predictions for Sea Louse (*Caligus rogercresseyi*) Resistance in Atlantic Salmon (*Salmo salar*). *Genet. Sel. Evol.* 49, 15. doi:10.1186/s12711-017-0291-8

Culloty, S. C., Cronin, M. A., and Mulcahy, M. F. (2004). Potential Resistance of a Number of Populations of the Oyster *Ostrea edulis* to the Parasite *Bonamia ostreae*. *Aquaculture* 237, 41–58. doi:10.1016/j.aquaculture.2004.04.007

Dégremont, L., Nourry, M., and Maurouard, E. (2015). Mass Selection for Survival and Resistance to OsHV-1 Infection in *Crassostrea gigas* Spat in Field Conditions: Response to Selection after Four Generations. *Aquaculture* 446, 111–121. doi:10.1016/j.aquaculture.2015.04.029

De Melo, C. M. R., Durland, E., and Langdon, C. (2016). Improvements in Desirable Traits of the Pacific Oyster, *Crassostrea gigas*, as a Result of Five Generations of Selection on the West Coast, USA. *Aquaculture* 460, 105–115. doi:10.1016/j.aquaculture.2016.04.017

Devlin, B., and Roeder, K. (1999). Genomic Control for Association Studies. *Biometrics* 55, 997–1004. doi:10.1111/j.0006-341x.1999.00997.x

Evans, S., and Langdon, C. (2006). Direct and indirect responses to selection on individual body weight in the Pacific oyster (*Crassostrea gigas*). *Aquaculture* 261, 522–534. doi:10.1016/j.aquaculture.2006.07.037

Evans, F., Matson, S., Brake, J., and Langdon, C. (2004). The Effects of Inbreeding on Performance Traits of Adult Pacific Oysters (*Crassostrea gigas*). *Aquaculture* 230, 89–98. doi:10.1016/j.aquaculture.2003.09.023

Fabioux, C., Huvet, A., Lelong, C., Robert, R., Pouvreau, S., Daniel, J. Y., et al. (2004). Oyster Vasa-like Gene as a Marker of the Germline Cell Development in *Crassostrea gigas*. *Biochem. Biophys. Res. Commun.* 320, 592–598. doi:10.1016/j.bbrc.2004.06.009

FAO (2022). *Cultured Aquatic Species Information Programme (CASIP)* Ostrea Edulis *Linnaeus 1758*. https://www.fao.org/fishery/en/culturedspecies/ostrea_edulis/en. (Accessed June 26, 2022).

Fitzer, S. C., McGill, R. A. R., Torres Gabarda, S., Hughes, B., Dove, M., O'Connor, W., et al. (2019). Selectively Bred Oysters Can Alter Their Biomineralization Pathways, Promoting Resilience to Environmental Acidification. *Glob. Change Biol.* 25, 4105–4115. doi:10.1111/gcb.14818

Fraslin, C., Yáñez, J. M., Robledo, D., and Houston, R. D. (2022). The Impact of Genetic Relationship between Training and Validation Populations on

Genomic Prediction Accuracy in Atlantic Salmon. *Aquac. Rep.* 23, 101033. doi:10.1016/j.aqrep.2022.101033

Goddard, M. E., and Hayes, B. J. (2007). Genomic Selection. *J. Anim. Breed. Genet.* 124, 323–330. doi:10.1111/j.1439-0388.2007.00702.x

Grizel, H., and Héral, M. (1991). Introduction into France of the Japanese Oyster (*Crassostrea gigas*). *ICES J. Mar. Sci.* 47, 399–403. doi:10.1093/icesjms/47.3.399

Grizzle, R. E., Ward, K. M., Peter, C. R., Cantwell, M., Katz, D., and Sullivan, J. (2017). Growth, Morphometrics and Nutrient Content of Farmed Eastern Oysters, *Crassostrea Virginica* (Gmelin), in New Hampshire, USA. *Aquac. Res.* 48, 1525–1537. doi:10.1111/are.12988

Gundappa, M. K., Peñaloza, C., Regan, T., Boutet, I., Tanguy, A., Houston, R. D., et al. (2022). A Chromosome Level Reference Genome for European Flat Oyster (*Ostrea Edulis* L.) *bioRxiv [Preprint]*. Available at: https://doi.org/10.1101/2022.06.26.497633. (Accessed June 29, 2022)

Gutierrez, A. P., Turner, F., Gharbi, K., Talbot, R., Lowe, N. R., Peñaloza, C., et al. (2017). Development of a Medium Density Combined-Species SNP Array for Pacific and European Oysters (*Crassostrea gigas* and *Ostrea edulis*). *G3 (Bethesda)* 7, 2209–2218. doi:10.1534/g3.117.041780

Gutierrez, A. P., Matika, O., Bean, T. P., and Houston, R. D. (2018). Genomic Selection for Growth Traits in Pacific Oyster (*Crassostrea gigas*): Potential of Low-Density Marker Panels for Breeding Value Prediction. *Front. Genet.* 9, 391. doi:10.3389/fgene.2018.00391

Gutierrez, A. P., Symonds, J., King, N., Steiner, K., Bean, T. P., and Houston, R. D. (2020). Potential of Genomic Selection for Improvement of Resistance to Ostreid Herpesvirus in Pacific Oyster (*Crassostrea gigas*). *Anim. Genet.* 51, 249–257. doi:10.1111/age.12909

He, X., Li, C., Qi, H., Meng, J., Wang, W., Wu, F., et al. (2021). A Genome-wide Association Study to Identify the Genes Associated with Shell Growth and Shape-Related Traits in *Crassostrea gigas*. *Aquaculture* 543, 736926. doi:10.1016/j.aquaculture.2021.736926

He, X., Wu, F., Qi, H., Meng, J., Wang, W., Liu, M., et al. (2022). Whole-genome Resequencing Reveals the Single Nucleotide Polymorphisms Associated with Shell Shape in *Crassostrea gigas*. *Aquaculture* 547, 737502. doi:10.1016/j.aquaculture.2021.737502

Helmer, L., Farrell, P., Hendy, I., Harding, S., Robertson, M., and Preston, J. (2019). Active Management is Required to Turn the Tide for Depleted *Ostrea edulis* Stocks from the Effects of Overfishing, Disease and Invasive Species. *PeerJ* 7, e6431. doi:10.7717/peerj.6431

Hickey, J. M., Dreisigacker, S., Crossa, J., Hearne, S., Babu, R., Prasanna, B. M., et al. (2014). Evaluation of Genomic Selection Training Population Designs and Genotyping Strategies in Plant Breeding Programs Using Simulation. *Crop Sci.* 54, 1476–1488. doi:10.2135/cropsci2013.03.0195

Houston, R. D., Bean, T. P., Macqueen, D. J., Gundappa, M. K., Jin, Y. H., Jenkins, T. L., et al. (2020). Harnessing Genomics to Fast-Track Genetic Improvement in Aquaculture. *Nat. Rev. Genet.* 21, 389–409. doi:10.1038/s41576-020-0227-y

Jones, D. B., Jerry, D. R., Khatkar, M. S., Raadsma, H. W., and Zenger, K. R. (2013). A High-Density SNP Genetic Linkage Map for the Silver-Lipped Pearl Oyster, *Pinctada maxima*: A Valuable Resource for Gene Localisation and Marker-Assisted Selection. *BMC Genomics* 14, 810. doi:10.1186/1471-2164-14-810

Jones, D. B., Jerry, D. R., Khatkar, M. S., Moser, G., Raadsma, H. W., Taylor, J. J., et al. (2014). Determining Genetic Contributions to Host Oyster Shell Growth: Quantitative Trait Loci and Genetic Association Analysis for the Silver-Lipped Pearl Oyster, *Pinctada maxima*. *Aquaculture* 434, 367–375. doi:10.1016/j.aquaculture.2014.08.040

Korringa, P. (1976). *Farming the Flat Oyster of the Genus Ostrea*. Amsterdam: ElsevierScientific.

Kriaridou, C., Tsairidou, S., Houston, R. D., and Robledo, D. (2020). Genomic Prediction Using Low Density Marker Panels in Aquaculture: Performance across Species, Traits, and Genotyping Platforms. *Front. Genet.* 11, 124. doi:10.3389/fgene.2020.00124

Kube, P., Cunningham, M., Dominik, S., Parkinson, S., Finn, B., Henshall, J., et al. (2011). *Enhancement of the Pacific Oyster Selective Breeding Program*. Australia: FRDC and Seafood CRC.

Lallias, D., Taris, N., Boudry, P., Bonhomme, F., and Lapègue, S. (2010). Variance in the Reproductive Success of Flat Oyster *Ostrea edulis* L. Assessed by Parentage Analyses in Natural and Experimental Conditions. *Genet. Res.* 92, 175–187. doi:10.1017/S0016672310000248

Lapegue, S., Harrang, E., Heurtebise, S., Flahauw, E., Donnadieu, C., Gayral, P., et al. (2014). Development of SNP-Genotyping Arrays in Two Shellfish Species. *Mol. Ecol. Resour.* 14, 820–830. doi:10.1111/1755-0998.12230

Li, C., Wang, J., Song, K., Meng, J., Xu, F., Li, L., et al. (2018). Construction of a High-Density Genetic Map and Fine QTL Mapping for Growth and Nutritional Traits of *Crassostrea gigas*. *BMC Genomics* 19, 626. doi:10.1186/s12864-018-4996-z

Li, A., Dai, H., Guo, X., Zhang, Z., Zhang, K., Wang, C., et al. (2021). Genome of the Estuarine Oyster Provides Insights into Climate Impact and Adaptive Plasticity. *Commun. Biol.* 4, 1287. doi:10.1038/s42003-021-02823-6

Lillehammer, M., Bangera, R., Salazar, M., Vela, S., Erazo, E. C., Suarez, A., et al. (2020). Genomic Selection for White Spot Syndrome Virus Resistance in Whiteleg Shrimp Boosts Survival under an Experimental Challenge Test. *Sci. Rep.* 10, 20571. doi:10.1038/s41598-020-77580-3

Liu, S., Li, L., Zhang, S., Wang, W., Yang, J., and Zhang, G. (2019). Heritability Estimates for Nutritional Quality-related Traits of the Pacific Oyster, *Crassostrea gigas*. *J. World Aquacult. Soc.* 50, 738–748. doi:10.1111/jwas.12588

Lupas, A., and Martin, J. (2002). AAA Proteins. *Curr. Opin. Struct. Biol.* 12, 746–753. doi:10.1016/s0959-440x(02)00388-3

McCarty, A. J., Allen, S. K., and Plough, L. V. (2022). Genome-wide Analysis of Acute Low Salinity Tolerance in the Eastern Oyster *Crassostrea virginica* and Potential of Genomic Selection for Trait Improvement. *G3 (Bethesda)* 12, jkab368. doi:10.1093/g3journal/jkab368

Meng, J., Song, K., Li, C., Liu, S., Shi, R., Li, B., et al. (2019). Genome-wide Association Analysis of Nutrient Traits in the Oyster *Crassostrea gigas*: Genetic Effect and Interaction Network. *BMC Genomics* 20, 625. doi:10.1186/s12864-019-5971-z

Meuwissen, T. H. E., Hayes, B. J., and Goddard, M. E. (2001). Prediction of Total Genetic Value Using Genome-wide Dense Marker Maps. *Genetics* 157, 1819–1829. doi:10.1093/genetics/157.4.1819

Mizuta, D. D., and Wikfors, G. H. (2019). Seeking the Perfect Oyster Shell: A Brief Review of Current Knowledge. *Rev. Aquacult.* 11, 586–602. doi:10.1111/raq.12247

Modak, T. H., Literman, R., Puritz, J. B., Johnson, K. M., Roberts, E. M., Proestou, D., et al. (2021). Extensive Genome-wide Duplications in the Eastern Oyster (*Crassostrea virginica*). *Phil. Trans. R. Soc. B* 376, 20200164. doi:10.1098/rstb.2020.0164

Montagnani, C., Le Roux, F., Berthe, F., and Escoubas, J.-M. (2001). Cg-TIMP, an Inducible Tissue Inhibitor of Metalloproteinase from the Pacific oyster *Crassostrea gigas* with a Potential Role in Wound Healing and Defense Mechanisms[1]. *FEBS Lett.* 500, 64–70. doi:10.1016/s0014-5793(01)02559-5

Naciri-Graven, Y., Martin, A.-G., Baud, J.-P., Renault, T., and Gérard, A. (1998). Selecting the Flat Oyster *Ostrea edulis* (L.) for Survival when Infected with the Parasite *Bonamia ostreae*. *J. Exp. Mar. Biol. Ecol.* 224, 91–107. doi:10.1016/S0022-0981(97)00171-8

Neph, S., Kuehn, M. S., Reynolds, A. P., Haugen, E., Thurman, R. E., Johnson, A. K., et al. (2012). BEDOPS: High-Performance Genomic Feature Operations. *Bioinformatics* 28, 1919–1920. doi:10.1093/bioinformatics/bts277

Newkirk, G. F., and Haley, L. E. (1982). Progess in Selection for Growth Rate in the European Oyster *Ostrea edulis*. *Mar. Ecol. Prog. Ser.* 10, 3. doi:10.3354/meps010077

Newkirk, G. F., and Haley, L. E. (1983). Selection for Growth Rate in the European Oyster, *Ostrea edulis*: Response of Second Generation Groups. *Aquaculture* 33, 149–155. doi:10.1016/0044-8486(83)90396-4

Peñaloza, C., Gutierrez, A. P., Eöry, L., Wang, S., Guo, X., Archibald, A. L., et al. (2021). A Chromosome-Level Genome Assembly for the Pacific Oyster *Crassostrea gigas*. *GigaScience* 10, giab020. doi:10.1093/gigascience/giab020

Pichot, Y., Comps, M., Tige, G., Grizel, H., and Rabouin, M. A. (1979). Recherches sur Bonamia ostreae gen. n., sp. n., parasite nouveau de l' huitre plate *Ostrea edulis* L. [France]. *Rev. Trav. Inst. Peches Marit.* 43, 131–140.

Plough, L. V. (2016). Genetic Load in Marine Animals: a Review. *Curr. Zool.* 62, 567–579. doi:10.1093/cz/zow096

Pogoda, B., Boudry, P., Bromley, C., Cameron, T. C., Colsoul, B., Donnan, D., et al. (2020). NORA Moving Forward: Developing an Oyster Restoration Network in Europe to Support the Berlin Oyster Recommendation. *Aquat. Conserv. Mar. Freshw. Ecosyst.* 30, 2031–2037. doi:10.1002/aqc.3447

Pogoda, B. (2019). Current Status of European Oyster Decline and Restoration in Germany. *Humanities* 8, 9. doi:10.3390/h8010009

Preston, J., Fabra, M., Helmer, L., Johnson, E., Harris-Scott, E., and Hendy, I. W. (2020). Interactions of Larval Dynamics and Substrate Preference Have Ecological Significance for Benthic Biodiversity and *Ostrea edulis* Linnaeus, 1758 in the Presence of Crepidula Fornicata. *Aquat. Conserv. Mar. Freshw. Ecosyst.* 30, 2133–2149. doi:10.1002/aqc.3446

Proestou, D. A., Vinyard, B. T., Corbett, R. J., Piesz, J., Allen, S. K., Small, J. M., et al. (2016). Performance of Selectively-Bred Lines of Eastern Oyster, *Crassostrea virginica*, across Eastern US Estuaries. *Aquaculture* 464, 17–27. doi:10.1016/j.aquaculture.2016.06.012

Qi, H., Song, K., Li, C., Wang, W., Li, B., Li, L., et al. (2017). Construction and Evaluation of a High-Density SNP Array for the Pacific Oyster (*Crassostrea gigas*). *PLoS One* 12, e0174007. doi:10.1371/journal.pone.0174007

Qi, H., Li, L., and Zhang, G. (2021). Construction of a Chromosome-level Genome and Variation Map for the Pacific Oyster *Crassostrea gigas*. *Mol. Ecol. Resour.* 21, 1670–1685. doi:10.1111/1755-0998.13368

Ragone Calvo, L. M., Calvo, G. W., and Burreson, E. M. (2003). Dual Disease Resistance in a Selectively Bred Eastern Oyster, *Crassostrea virginica*, Strain Tested in Chesapeake Bay. *Aquaculture* 220, 69–87. doi:10.1016/S0044-8486(02)00399-X

Renault, T., Cochennec, N., and Grizel, H. (1995). Bonamia Ostreae, Parasite of the European Flat Oyster, *Ostrea edulis*, Does Not Experimentally Infect the Japanese Oyster, *Crassostrea Gigas*. *Bull. Eur. Ass. Fish. Pathol.* 15, 78.

Robert, R., Borel, M., Pichot, Y., and Trut, G. (1991). Growth and Mortality of the European oyster *Ostrea edulis* in the Bay of Arcachon (France). *Aquat. Living Resour.* 4, 265–274. doi:10.1051/alr:1991028

Robert, M., Garcia, C., Chollet, B., Lopez-Flores, I., Ferrand, S., François, C., et al. (2009). Molecular Detection and Quantification of the Protozoan *Bonamia ostreae* in the Flat Oyster, *Ostrea edulis*. *Mol. Cell. Probes* 23, 264–271. doi:10.1016/j.mcp.2009.06.002

Sas, H., Deden, B., Kamermans, P., zu Ermgassen, P. S. E., Pogoda, B., Preston, J., et al. (2020). Bonamia Infection in Native Oysters (*Ostrea edulis*) in Relation to European Restoration Projects. *Aquat. Conserv. Mar. Freshw. Ecosyst.* 30, 2150–2162. doi:10.1002/aqc.3430

Thurstan, R. H., Hawkins, J. P., Raby, L., and Roberts, C. M. (2013). Oyster (*Ostrea edulis*) Extirpation and Ecosystem Transformation in the Firth of Forth, Scotland. *J. Nat. Conserv.* 21, 253–261. doi:10.1016/j.jnc.2013.01.004

Toro, J., and Newkirk, G. (1990). Divergent Selection for Growth Rate in the European Oyster *Ostrea edulis* Response to Selection and Estimation of Genetic Parameters. *Mar. Ecol. Prog. Ser.* 62, 219–227. doi:10.3354/meps062219

Tsai, H.-Y., Hamilton, A., Tinch, A. E., Guy, D. R., Bron, J. E., Taggart, J. B., et al. (2016). Genomic Prediction of Host Resistance to Sea Lice in Farmed Atlantic Salmon Populations. *Genet. Sel. Evol.* 48, 47. doi:10.1186/s12711-016-0226-9

Tsairidou, S., Hamilton, A., Robledo, D., Bron, J. E., and Houston, R. D. (2020). Optimizing Low-Cost Genotyping and Imputation Strategies for Genomic Selection in Atlantic Salmon. *G3 (Bethesda)* 10, 581–590. doi:10.1534/g3.119.400800

Tsairidou, S. (2019). CVrepGPAcalc. Available at: https://github.com/SmaragdaT/CVrep/tree/master/CVrepGPAcalc (Accessed March 15, 2020).

Turner, S. D. (2018). Qqman: An R Package for Visualizing GWAS Results Using Q-Q and Manhattan Plots. *J. Open Source Softw.* 3 (25), 731. doi:10.21105/joss.00731

Vallejo, R. L., Silva, R. M. O., Evenhuis, J. P., Gao, G., Liu, S., Parsons, J. E., et al. (2018). Accurate Genomic Predictions for BCWD Resistance in Rainbow Trout are Achieved Using Low-density SNP Panels: Evidence that Long-range LD is a Major Contributing Factor. *J. Anim. Breed. Genet.* 135, 263–274. doi:10.1111/jbg.12335

VanRaden, P. M. (2008). Efficient Methods to Compute Genomic Predictions. *J. Dairy Sci.* 91, 4414–4423. doi:10.3168/jds.2007-0980

Varney, R. L., and Wilbur, A. E. (2020). Analysis of Genetic Variation and Inbreeding Among Three Lines of Hatchery-Reared *Crassostrea virginica* Broodstock. *Aquaculture* 527, 735452. doi:10.1016/j.aquaculture.2020.735452

Vera, M., Pardo, B. G., Cao, A., Vilas, R., Fernández, C., Blanco, A., et al. (2019). Signatures of Selection for Bonamiosis Resistance in European Flat Oyster (*Ostrea edulis*): New Genomic Tools for Breeding Programs and Management of Natural Resources. *Evol. Appl.* 12 (9), 1781–1796. doi:10.1111/eva.12832

Vu, S. V., Knibb, W., Nguyen, N. T. H., Vu, I. V., O'Connor, W., Dove, M., et al. (2020). First Breeding Program of the Portuguese Oyster *Crassostrea angulata* Demonstrated Significant Selection Response in Traits of Economic Importance. *Aquaculture* 518, 734664. doi:10.1016/j.aquaculture.2019.734664

Vu, S. V., Gondro, C., Nguyen, N. T. H., Gilmour, A. R., Tearle, R., Knibb, W., et al. (2021a). Prediction Accuracies of Genomic Selection for Nine Commercially Important Traits in the Portuguese Oyster (*Crassostrea angulata*) Using DArT-Seq Technology. *Genes* 12, 210. doi:10.3390/genes12020210

Vu, S. V., Knibb, W., Gondro, C., Subramanian, S., Nguyen, N. T. H., Alam, M., et al. (2021b). Genomic Prediction for Whole Weight, Body Shape, Meat Yield, and Color Traits in the Portuguese Oyster *Crassostrea angulata*. *Front. Genet.* 12, 661276. doi:10.3389/fgene.2021.661276

Walne, P. R., and Helm, M. M. (1979). *Introduction of* Crassostrea gigas *into the United Kingdom*. Cambridge: MIT Press, 83–105.

Wan, S., Li, Q., Yu, H., Liu, S., and Kong, L. (2020). Estimating Heritability for Meat Composition Traits in the Golden Shell Strain of Pacific Oyster (*Crassostrea gigas*). *Aquaculture* 516, 734532. doi:10.1016/j.aquaculture.2019.734532

Wang, J., Li, L., and Zhang, G. (2016). A High-Density SNP Genetic Linkage Map and QTL Analysis of Growth-Related Traits in a Hybrid Family of Oysters (*Crassostrea gigas* × *Crassostrea angulata*) Using Genotyping-By-Sequencing. *G3 (Bethesda)* 6, 1417–1426. doi:10.1534/g3.116.026971

Ward, R. D., Thompson, P. A., Appleyard, S. A., Swan, A. A., and Kube, P. D. (2005). Sustainable Genetic Improvement of Pacific Oysters in Tasmania and South Australia. FRDC Project 2000/206. CSIRO Marine and Atmospheric Research.

Xu, L., Li, Q., Yu, H., and Kong, L. (2017). Estimates of Heritability for Growth and Shell Color Traits and Their Genetic Correlations in the Black Shell Strain of Pacific Oyster *Crassostrea gigas*. *Mar. Biotechnol.* 19, 421–429. doi:10.1007/s10126-017-9772-6

Yang, B., Zhai, S., Zhang, F., Wang, H., Ren, L., Li, Y., et al. (2022). Genome-wide Association Study toward Efficient Selection Breeding of Resistance to *Vibrio alginolyticus* in Pacific Oyster, *Crassostrea gigas*. *Aquaculture* 548, 737592. doi:10.1016/j.aquaculture.2021.737592

Yin, X., Arias-Pérez, A., Kitapci, T. H., and Hedgecock, D. (2020). High-Density Linkage Maps Based on Genotyping-By-Sequencing (GBS) Confirm a Chromosome-Level Genome Assembly and Reveal Variation in Recombination Rate for the Pacific Oyster *Crassostrea gigas*. *G3 (Bethesda)* 10, 4691–4705. doi:10.1534/g3.120.401728

Yoshida, G. M., Bangera, R., Carvalheiro, R., Correa, K., Figueroa, R., Lhorente, J. P., et al. (2018). Genomic Prediction Accuracy for Resistance against *Piscirickettsia salmonis* in Farmed Rainbow Trout. *G3 (Bethesda)* 8, 719–726. doi:10.1534/g3.117.300499

Yoshida, G. M., Lhorente, J. P., Correa, K., Soto, J., Salas, D., and Yáñez, J. M. (2019). Genome-Wide Association Study and Cost-Efficient Genomic Predictions for Growth and Fillet Yield in Nile Tilapia (*Oreochromis niloticus*). *G3 (Bethesda)* 9, 2597–2607. doi:10.1534/g3.119.400116

Yuehuan, Z., Wu, X., Qin, Y., Xiao, S., Ma, H., Li, J., et al. (2017). Sustained Response to Selection of Growth Traits to the Third Generation for Two Strains of Kumamoto Oyster *Crassostrea sikamea*. *J. Fish. Sci. China* 24, 1161–1167. doi:10.3724/SP.J.1118.2017.16350

Zenger, K. R., Khatkar, M. S., Jones, D. B., Khalilisamani, N., Jerry, D. R., and Raadsma, H. W. (2019). Genomic Selection in Aquaculture: Application, Limitations and Opportunities with Special Reference to Marine Shrimp and Pearl Oysters. *Front. Genet.* 9, 693. doi:10.3389/fgene.2018.00693

Zhang, J., Li, Q., Xu, C., and Han, Z. (2019). Response to Selection for Growth in Three Selected Strains of the Pacific Oyster *Crassostrea gigas*. *Aquaculture* 503, 34–39. doi:10.1016/j.aquaculture.2018.12.076

Zhou, X., and Stephens, M. (2012). Genome-wide Efficient Mixed-Model Analysis for Association Studies. *Nat. Genet.* 44, 821–824. doi:10.1038/ng.2310

Conflict of Interest: Author RH was employed by Benchmark Genetics. Author JH was employed by Lochnell oysters.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.