Determinants and relative weight of genetic and epigenetic variation in the resistance of the oyster Crassostrea gigas to the Pacific Oyster Mortality Syndrome

Janan Gawra

THÈSE

Pour obtenir le grade de

Docteur

Délivré par

UNIVERSITE DE PERPIGNAN VIA DOMITIA

Préparée au sein de

l'école doctorale ED305 Energie Environnement

Et de l'unité de recherche

UMR 5244 Interactions Hôtes Pathogènes Environnements (IHPE)

Spécialité : **Biologie**

Présentée par

Janan GAWRA

Determinants and relative weight of genetic and epigenetic variation in the resistance of the oyster *Crassostrea gigas* to the Pacific Oyster Mortality Syndrome

Soutenue le : 23 Mai 2022, devant le jury composé de

Mme. Christina Richards, HDR, University of South Florida	Rapporteur
Mme. Sylvie Lapègue, HDR, MARBEC	Rapporteur
M. Fabrice Pernet, HDR, LEMAR	Examinateur
M. Thierry Lagrange, HDR, LGDP	Examinateur
M. Christoph GRUNAU, HDR, UPVD	Directeur de thèse
M. Jeremie Vidal-Dupiol, HDR, IHPE	Co-directeur de thèse

Dedicated to

To my family

To the pure soul of my beloved father

I dedicate this thesis to all my family, thanks for all the support you gave me. Thank you my mother for all your prayers.

Acknowledgments

To start I want to show my warmest appreciation to **Jeremie Vidal-Dupiol**, thank you in supervising me, believing in me and always guiding me in the right direction. I am also grateful for the thorough reviews of my thesis writing; your feedback helped me a lot to improve my writing.

I send my appreciation to **Jean-Baptiste Lamy** that unfortunately left science while I was in my second year, so we could not work further together. I hope you enjoying in building your house.

I would like to further thanks **Christoph Grunau** for the fruitful discussion and advice during the journey of PhD and committee meetings.

My sincere appreciation to each of the PhD committee members **Benoit Pujol, Clemintine Vitte and Fabrice Roux**. I deeply thank you for all the effort, comment and advice you gave during my PhD. Furthermore, I am very grateful for **Christina Richards** and **Sylvie Lapègue** for accepting to be my thesis rapporteurs and taking the time to evaluate my work. I am also very thankful for **Fabrice Pernet** and **Thierry Lagrange** for accepting to be the examiners for my thesis.

I dedicate this work to all the IHPE lab members (the current and the ones who left) who I spent the last three and half years together. A special thanks to the le team **Zouz** (as **Maurine** called it).First of all a big thanks to **Jean** (you see a big thanks and not only thanks as you did) for all your support during my thesis, thanks for the code and bioinformatics tricks you thought me. A big thanks to **Maurine** for the help in R scripts and fruitful discussion we had as a neighbour. My thanks to **Daniel**, **Etienne** and **Pierre-Louis**, with you I had great during the beginning of my thesis.

I would like to thank each of the new PhD students that I will leave behind in the office, hope that you will soon have amazing results and finish your thesis. Thanks to **Aurèlie, Noémie, Laura, Louis and Sébastian**.

Thank you to all the people of the IHPE, **Delphine, Yannick, Caroline, Guillaume M., Guillaume Ch., Julien, Jean-Michel, Erika, Chin, Viviane, Agnès, Emanuel, Marko, Philip, Enola, Eric, Juliette, Oceane et Gaëlle** (merci à tous, merci pour la discussion en français, cela m'a beaucoup aidé pour apprendre le français). Thanks to all the master students who come and go during the

last three years. Thanks to the people in the IHPE Perpignan, especially **Sylvie Cabal** for your help in my administration work. I would to thanks everyone who participated in the releasing all the different step needed for this project. Additionally, thanks to **Diane**, **Anne** and **Camille** for your great work of managing the administrative work.

A special thanks to my lab partner **Mathilde**, thank you for your help in grinding, extracting DNA and all the bench work.

My great appreciation for **Alejandro** for all the support you gave me during the last five months and all the feedback in my thesis writing.

I also thank everyone who I met during the journey of the thesis, especially those who I met outside of the work. I special thanks to **Augustin** and **Najoua** for you lovely friendship and kindness, I am grateful to have meet you.

I dedicate this thesis to all my friends living in Iraq and worldwide (especially **Daniel**, **Johnny**, **Haval**, **Mustafa** and others).

I will end by a very special thanks to ma cherie **Claire** who support me during my thesis journey, thank you for your patient and supporting my late working. Yes, finally it is done, it is vacation time.

Table of content

List of Figures

List of Tables

Chapter 1

General Introduction

“Nothing in Biology Makes Sense

Except in the Light of Evolution”

Theodosius Dobzhansky

Chapter 1 : General introduction

1.1 The theory of evolutionary biology

Evolutionary biology is a branch that investigates evolutionary process to understand diversity of life being on the earth. Throughout the history, from philosophers to scientist living in different eras have been searching the origin of the enormous diversity and life history. However, these theories were repeatedly countered or opposed due to the lack of tangible evidence or religious beliefs. The notion of evolution only starts to expanse in the early 18th century through the work done by Lamarck, Darwin, Mendel and many others.

1.1.1 Lamarck

At the beginning of the 19th century Jean-Baptiste de Lamarck (1744-1829) pointed out the theory of "Transmutation of species", which gave arguments for the diversity of the living world. By comparing shell fossils and recollecting living invertebrates, Lamarck was convinced that the phenomena of species divergence was a result of the constant adaptation to successive fluctuations on their environment (Lamarck, 1809; Mayr, 1982). Additionally, Lamarck is considered as one of the first who proposed a theory for the evolutionary species (Danchin et al., 2019). Lamarck is known for his theory of "Inheritance of Acquired Characteristics" or "soft inheritance" where lineages change over the life time as a result of the use and disuse of organs. Thus, adaption is followed by the acquisition of new traits, which will be transmitted to the next generations (Lamarck, 1809). Therefore, the traits transmission from one generation to another transforms the living organisms and give rises to new species better adapted to their environment. Lamarck never claimed that the idea of "Inheritance of Acquired Characteristics" is his own idea, but only presented it as a self-evident idea (Burkhardt, 2013). Additionally, he did not look further for the mechanisms of transmission of characteristics from one generation to the next (Burkhardt, 2013).

1.1.2 Darwin

During the mid of the 19th century, many naturalists, especially Charles Darwin (1809-1882), were investigating the basic principles of the foundation of evolutionary biology. They were eager to answer two main questions: i) what is the origin of the enormous diversity and life

history; ii) what are the mechanisms behind the apparent match between the form and function in biological organism (Pigliucci, 2007). Darwin gave substantial arguments to these questions by his inspiration from works of others colleagues proposing that: the life's history and biodiversity is a consequence of the pattern of modification in a lineage, and the natural selection is only the mechanism responsible for the form-function dilemma (Pigliucci, 2007). Natural Selection is one of Darwin's theories that encountered the most of the resistance from naturalist and scientist, including those who supported Lamarckism theory. Darwin theory of Natural Selection was only accepted at the beginning of 1940th (Mayr, 1996). However, the problem of heredity remained as a matter of concern to Darwin. Thus, Darwin goes back to work of Lamarck and endorse some elements from Lamarck. That later Darwin proposed his theory of "blending inheritance" that the offspring get something intermediate from mother and some other things from father (Pigliucci, 2007). While, laws of inheritance remained unknown, Darwin proposed that in addition to inheritance of acquired characteristics, living organism also reproduce with variation on which natural selection can act (Danchin et al., 2019).

1.1.3 Mendel

In the end, neither of Darwin nor Lamarck presented explanation of the mechanisms behind the inheritance of traits. During the same period, Gregor Mendel (1822-1884), in 1866, published his work on pea plants about the fundamental laws of inheritance that he independently established. This occurred in the time of Darwin theory and even Darwin cite Mendel in one of his book, although he did not read it (Pigliucci, 2007). Years later after the first Mendel's publication, his theory were accepted among scientist community (beginning the 20th century). Latter on and through the combination of Mendel's laws and Darwin' natural selection led to development of Neo-Darwinism.

1.1.4 The Modern Synthetic theory

After Darwin, the debate between the biologists went on whether to keep some sort of Lamarckian elements or completely refute it in the inheritance system. Darwin, Thomas Henry Huxley (1825-1895), and George Romanes (1848-1894) proposed several times to keep some of Lamarckian elements, while Alfred Russel Wallace (1823-1913) and August Weismann (1834-1914) were completely blocked by the idea of Lamarck. In the opinion of Weismann,

adding any Lamarckian elements in the heredity empirically was refuted and theoretically dreadful (Pigliucci, 2007).

Weismann tested the Lamarck hypothesis, by cutting mice tails and reproducing them for several generations. Weismann observed that mice offspring still got long tails, so he concluded with this evidence a denial of the inheritance of acquired characteristics (Gauthier, 1990). Later, Weismann come with his doctrine of separation between the soma and the germ lines (Pigliucci, 2007), after his observations of cell division and meiosis. In Weismann theory of "Germ-Plasm", the organisms are compose of germ cells that transmit the hereditary information and somatic cells that perform the life functions (Winther, 2001). Thus, what the somatic cells experience during the life does not influence the germ cells (Haig, 2007).

Subsequently, Lamarck's "Transmutation of species" and "Inheritance of Acquired Characteristics" theories were rejected by Weismann, thus the term of "Neo-Darwinism" was born to refer to the doctrine of Weismann (Haig, 2007). However, others like James Mark Baldwin (1861-1934), Thomas Hunt Morgan (1866-1945), and Henry Fairfield Osborn (1857-1935), suggested that in the absence of hereditary variations that is needed to overcome a new challenge, acquired characteristics (non-heritable) at the individual level could help a population to survive (Pigliucci, 2007). Thus, maintaining the population until the apparition of new hereditary variations to face the challenge and be selected through natural selection; In other word these suggest the inheritance of acquired characters (Danchin et al., 2019; Pocheville & Danchin, 2017; Simpson, 1953).

At beginning of the 20th century and with the rediscovery of Mendel's results, the term "genetics" was born referring to the study of heredity. Then followed by the term "gene", a small element that represented a medium of heredity needed for the transmission of the physical characteristics from one generation to another (Gayon, 2016). Mendelian inheritance allowed other geneticists to discover the involvement of chromosomes in heredity. In same period, Ronald Fisher (1890-1962), John Burdon Sanderson Haldane (1892-1964) and Sewall Wright (1889-1988) followed the Weisman' pace as and embraced a similar view concluding that inheritance was absolutely genetic and totally separated from the environment. Before the discovery of the DNA structure (James Watson (Born in 1928) and Francis Crick (1916-2004)), Julian Huxley (1887-1975) succeeded in combining the natural selection and heredity with Mendelian genetics and population genetics to form a single theory of evolution called

the “Modern Synthesis”, constituting today the current paradigm of evolutionary biology (Pigliucci, 2007). This was further conceptualised and developed with the discovery of the DNA structure (Watson & Crick, 1953).

1.1.5 The Inheritance paradigm: toward a shift in evolutionary biology

During the last two decades an ongoing controversial debates have been rising for reforming the modern theory of evolution into an “extended” or “inclusive” theory of evolutionary synthesis (R. Bonduriansky, 2012; Danchin, 2013; Danchin et al., 2011; Huneman & Whalsh, 2017; Laland et al., 2014; Pennisi, 2008; Pigliucci & Müller, 2010; Wray et al., 2014). This debate generally goes around the concepts of heredity and particularly the question of the existence of other non-genetic heredity (Danchin et al., 2019). Our understanding of heredity has thoroughly changed since Lamarck come up with the theory of inheritance of acquired traits. With the Modern Synthesis, this theory was transformed in a paradigm assuming that inheritance across generations is only mediated by the DNA sequence.

Since 19th century, several questions have been raised about the source of heritable phenotypic variations. Then, with the beginning of the 20th century, the advances in plant breeding and production of isogenic lines by self-pollinations gave birth to experiment separating the organism’s observable nature from their inheritance system. These led to the development of the Genotype-Phenotype (G-P) concept (Johannsen, 2014). In the late 1940th, Haldane and others geneticist extended the G-P concept to what is known as the G x E concept (E= environment) (Bowman, 1972; Cosseau et al., 2017; Haldane, 1946). This concept refers to a phenotype that result from the interactions between the genotype and the environment. In addition, the gene was considered as the basic unit, and the importance of genetic was further structured as follow: **Phenotype = Genotype x Environment** (Bowman, 1972)**.** Bowman’s equation reflects a highly weighted importance of genetics in Modern Synthesis.

However, since the 90th several experiments renew the Lamarckian inheritance hypothesis: whether characters are acquired during the life of an individual can be inherited to the offspring (Danchin *et al.*, 2011; Mameli, 2004; West-Eberhard, 2005; Johannes *et al.*, 2008; Danchin and Wagner, 2010; Helanterä and Uller, 2010; Jablonka and Lamb, 2010).

There is an increasing empirical evidence of the existence of non-genetic inheritance. This non-genetic mechanisms could operate along with the genetic inheritance, allowing the

inheritance of acquired traits (R. Bonduriansky, 2012). For long time the acceptance of the non-genetic inheritance was strongly controversial, but it becomes more and more recognised. This recognition come from several evidence found in different organisms ranging from unicellular, plants and animals (including humans) (Danchin et al., 2019). Thus, it is more accepted that the DNA sequence transmission (or genetic) is not the only mechanism of parent-offspring resemblance, but also can be mediated through different interacting non-genetic mechanisms such as epigenetics, cultural, ecological and parental effects (Danchin, 2013; Jablonka & Raz, 2009). In this context, three types of non-genetic inheritance (transmission between parent-offspring) have been proposed. First, there is the "intergenerational effects", where the resemblance is only demonstrated between the F0 and F1. Second, there is the "multigenerational effects", where the resemblance has been extended to the F2 generation. Third, there is the "transgenerational effects", where the resemblance go beyond the F2 generation (Danchin et al., 2019; Y. Wang et al., 2017). For example, in mammal it has been reported that exposure to toxins (Dolinoy et al., 2006) and nutrition behaviour (Weaver et al., 2004) could be responsible for different phenotypes that are due to non-genetic modifications. There are evidence that non-genetic changes can be inherited transgenerationally (Gavery & Roberts, 2017), in vertebrates (Guerrero-bosagna et al., 2010; Knecht et al., 2017), in invertebrates (Klosin et al., 2017) and in plants (Hauser et al., 2011).

1.1.6 Beyond the phenotype-genotype-environment (P > G x E)

Different authors have been proposing to extend the inheritance system by adding the non-genetic inheritance. Thus, proposing that both genetic (DNA) and non-genetic processes can be inherited and therefore are fuel for evolution (R. Bonduriansky, 2012; Danchin et al., 2011; Danchin & Wagner, 2010). Recently, several terms have been proposed to combine the genetic and non-genetic heritability, "general heritability" (Mameli, 2004), "inclusive heritability" (Danchin & Wagner, 2010); and "inheritance system" (Cosseau et al., 2017). All these terms are synonymous and aim to encircle all elements of the inheritance. In order to incorporate the genetic and non-genetic inherited component, Danchin (2013) adopted an information-driven approach to evolution. The new approach led to the redefinition of the evolution "the process by which the frequencies of variants in a population change over time" (Bentley et al., 2004; Danchin et al., 2011; Danchin & Wagner, 2010). In this case, "variant"

has been used instead of the word “gene”, and this variant include all type of forms of inherited information from genetic to non-genetic inheritance (Figure 1.1).

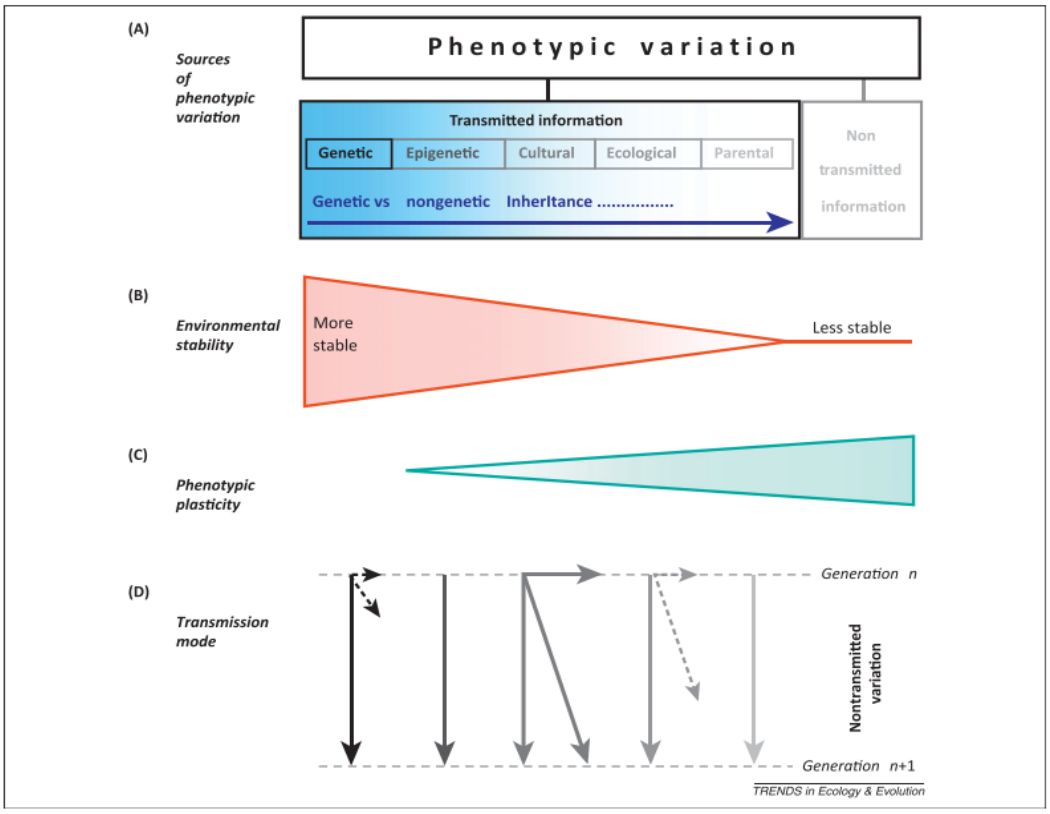


Figure 1.1: Genetic and non-genetic sources of phenotypic variation.

Adapted figure from Danchin (2013). To read from the top to the bottom and from left to right. A) Phenotypic variation sources. B) The environmental stability gradient, from more stable on the left to the more variable on right. C) Phenotypic plasticity, D) Transmission mode from one generation to the next generation, the longer the arrow the more important is the transmission mode and the broken versus unbroken arrows represent the rare versus frequent processes.

In the example of the “inheritance system”, the authors attempt to further expand the concept of the ***P -> G x E*** to add other elements of inheritance, for example, addition of the epigenetic in the above equation. With the advance in molecular biology field, the concept of *G x E* is being updated. If we consider adding epigenetics in the equation mentioned above, we should consider the G as the inheritance system. Then, this inheritance system would be composed from several elements such as the Genotype (G), Epigenotype (I), and in some other cases the cytoplasmic heritable elements and symbionts (Cosseau et al., 2017). For the purpose of my thesis work, we are focusing on the two elements of (G) and (I) which interact in a certain environment to result in a phenotype (the new equation is ***P -> (G x I) x E***). For epigenotype, here we are focusing on one mechanism of epigenetic called DNA methylation (explained in the next section).

1.1.7 Genetic heritability

1.1.7.1 DNA structure

The Deoxyribonucleic acid (DNA) molecule is generally a double helix (composed of two complementary strands wrapped around each other) present in the nucleus of all the living organism cells. Each strand is made of succession of bases (nucleotide) that are precisely ordered to make up information necessary for the carrying out of the functions of an organism. These nucleotides are one of four nitrogenous bases among adenine (A), cytosine (C), guanine (G) or thymine (T), and are covalently connected to each other via a phosphate. The linking between the two double helix is brought by the pairing of nitrogenous bases through hydrogen bonds (two hydrogen bonds between A and T; three hydrogen bonds between C and G).

1.1.7.2 DNA transmission

The DNA information constitutes the hereditary material of life that is transmitted from parents to offspring. Thus, in sexual reproduction each parent transmits a part of its genetic characteristics to the offspring via the gametes (sperm and eggs) and through the meiosis. A key step in transmitting the genetic information is the cell division. On the other side, during the organism development, the mitosis is the process of making new cells identical to mother cell through cell division. During cell division DNA replicates to make two exact copies by the replication mechanism. During replication and DNA repair it is known that error can occur at

a variable rate (depending on the organism) that constitutes the variations we see in the genome. These error rates ranges from $0.26x10^{-9}$ mutations per base pair per replication in *Escherichia coli*, and may even be as low as $0.06x10^{-9}$ in eukaryote germline (Lynch, 2010).

1.1.7.3 Gene expression

Certain parts of DNA sequences are commonly known as genes, and carry information for performing a certain function. Each three nucleotides are considered as a codon that codes for an amino acids. This is called the "genetic code" and it is used to build blocks of proteins (Crick et al., 1961). In order to carry this function the gene needs to be expressed through the transcription (mRNA) and translation mechanism of mRNA into protein (the central dogma).

1.1.7.4 Single nucleotide polymorphisms (SNPs)

However, the incidence of mutations, called Single Nucleotide Polymorphisms (or SNPs) when it is restricted to one base can lead DNA sequence changes and therefore to protein structural changes depending where in the gene they are located. These mutations have stronger effect when they introduce a change in the amino acid sequence (called non-synonymous mutation). By opposition a synonymous mutation does not add a change in the amino acid and therefore have no effect on the protein.

1.1.8 Epigenetic and its mechanisms

The definition of epigenetic could vary according to the research fields. In this thesis, the epigenetic is defined as "the study of heritable but reversible change in gene expression that do not rely on a modification of the underlying DNA sequence" (Bird, 2007; Bossdorf et al., 2008; E. J. Richards, 2006).

Epigenetic rely on four bearers of information: DNA methylation, histone tail modifications, Non-coding RNA, and chromatin localization (Figure 1.2). Those mechanisms are not independent from each other and are often in interaction to regulate gene expression (Berger, 2007; Bossdorf et al., 2008; Eirin-Lopez & Putnam, 2019; Grant-Downtown & Dickinson, 2005). Recently, it is more commonly accepted along with the empirical data that epigenetic modification induced by an environmental stimuli can mediate phenotypic changes (Granada et al., 2018).

The most studied epigenetic mechanism is the DNA methylation, which is a modification induced by the addition of a methyl group to the 5th carbon of a Cytosine to form 5-methyl-cytosine (5mC). The process of methylation is catalysed by DNA methyltransferases (DNTMS) (Zhu, 2011). This methylation often occurs in Cytosine-Guanine dinucleotide (CpG) context, and less frequently in a CHG or CHH context (H= C, T or A). In each Cytosine, the DNA methylation values are reported from no methylation (zero) to fully methylated (100).

The second epigenetic mechanism is the histone tail modifications. These modifications happen at the histone amino acid tails through the (de)acetylation, (de)methylation and (de)phosphorylation of specific amino acids (Bannister & Kouzarides, 2011). In turns, these modifications lead to interactions between the positive and negative charges of histone protein and DNA respectively which can lead to different chromatin states: i) a condensed and non-permissive to gene expression state (heterochromatin) or ii) relaxed and permissive to gene expression state (euchromatin).

The third epigenetic modification is non-coding RNA. Within genome, only few percentages of the total DNA molecules code for proteins and the remaining is referred as junk region. This junk region is now partly characterized as non-coding RNA (ncRNA) which play an important role in gene expression regulation and are categorised into two major classes, i) Long ncRNA (lncRNA) more than 200 nucleotides and ii) small ncRNA less than 200 nucleotides including

micro RNA (miRNA) short interfering RNA (siRNA) and PIWI-interacting RNA (piRNA). They are characterized by little or no coding ability. These lnRNAs are further classified into three groups: i) long intergenic noncoding RNAs (lincRNAs); ii) long noncoding natural antisense transcripts (lncNATs); iii) long intronic noncoding RNAs and overlapping lncRNAs.

The fourth epigenetic modification is the chromatin localization within the nucleus. For example, the presence of the heterochromatin within the nucleus periphery can affect and/or regulate the gene expression.

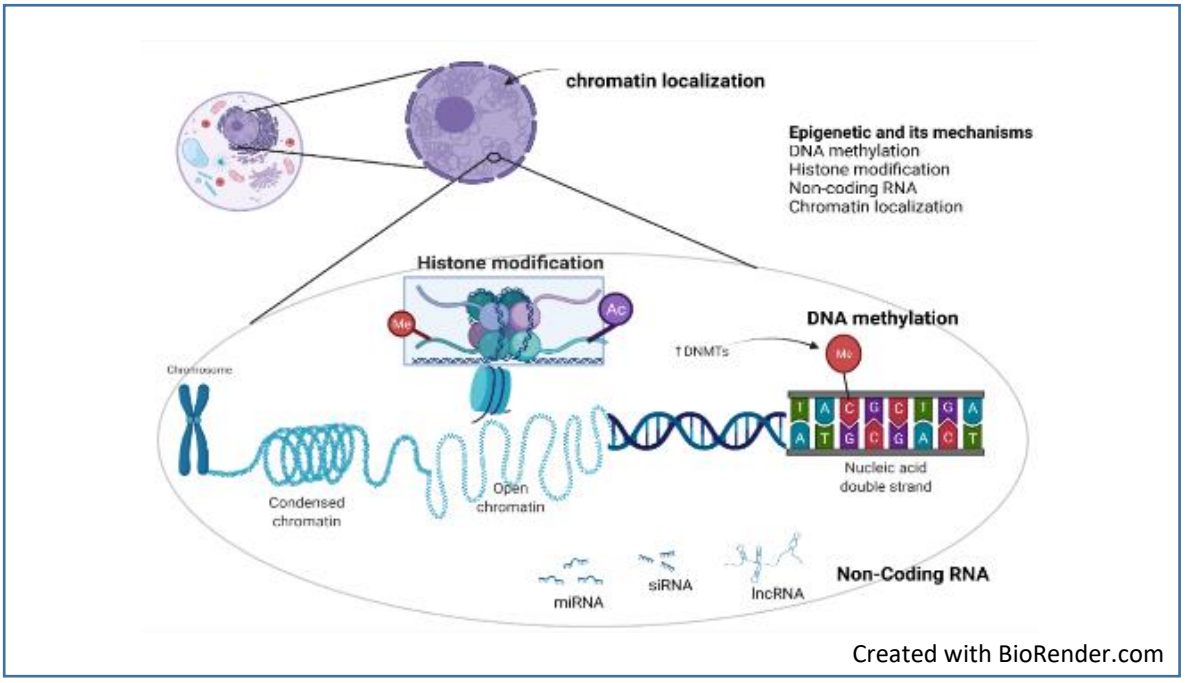


Figure 1.2: The four mechanisms of epigenetic.

In the nucleus of eukaryotic cells, the chromosomes are made up of DNA that is coiled around the histone. Each group of eight histones (composed of two pairs of H2A, H2B, H3 and H4 histone proteins) are called nucleosome. Each histone has a tail where several modifications can occur (all over comprise the epigenetic mechanism that is called histone modification). These nucleosomes comprise the repeating units that bunch of them are called chromatin. The location of the chromatin in the nucleus is another type of epigenetic mechanism. This chromatin is either closed (called heterochromatin) or open (called euchromatin). The open can been read by the transcription machinery to produce the messenger RNA. On these sequences the DNA methylation mechanism could occur by addition of the methyl group to the cytosine. Another mechanism of epigenetic is the non-coding RNA, which is comprised of the miRNA, siRNA and lncRNA. These non-coding RNAs bind to mRNA and inhibit the translation of target genes.

1.2 Biological model of the study

1.2.1 *Crassostrea gigas* or *Magallana gigas*?

The Pacific oyster or Japanese oyster *Crassostrea gigas* (or *Magallana gigas*; Thunberg, 1973) is one of the most important species belonging to Mollusca phylum and the class of Bivalvia. The Pacific oyster belongs to the family *Ostreidae* within Ostreida SuperFamily (Figure 1.3A). Ostreidae contains four sub-families *Crassostreinae*, *Ostreinae*, *Saccostreinae* and *Striostreinae* (Salvi & Mariottini, 2017). Previously, the Pacific oyster belonged to the *Crassostrea* genus within *Crassostreinae* subfamily. However, recently it has been reclassified in *Magallana* genus (Figure 1.3B), therefore this species can be found as *Magallana gigas* (Salvi & Mariottini, 2017). The new name *Magallana gigas* is not frequently used and the validity of these new affiliation still under discussion between scientists (Bayne et al., 2017; Salvi & Mariottini, 2021), therefore we will use the *Crassostrea gigas* (or Pacific oyster) in all along this thesis.

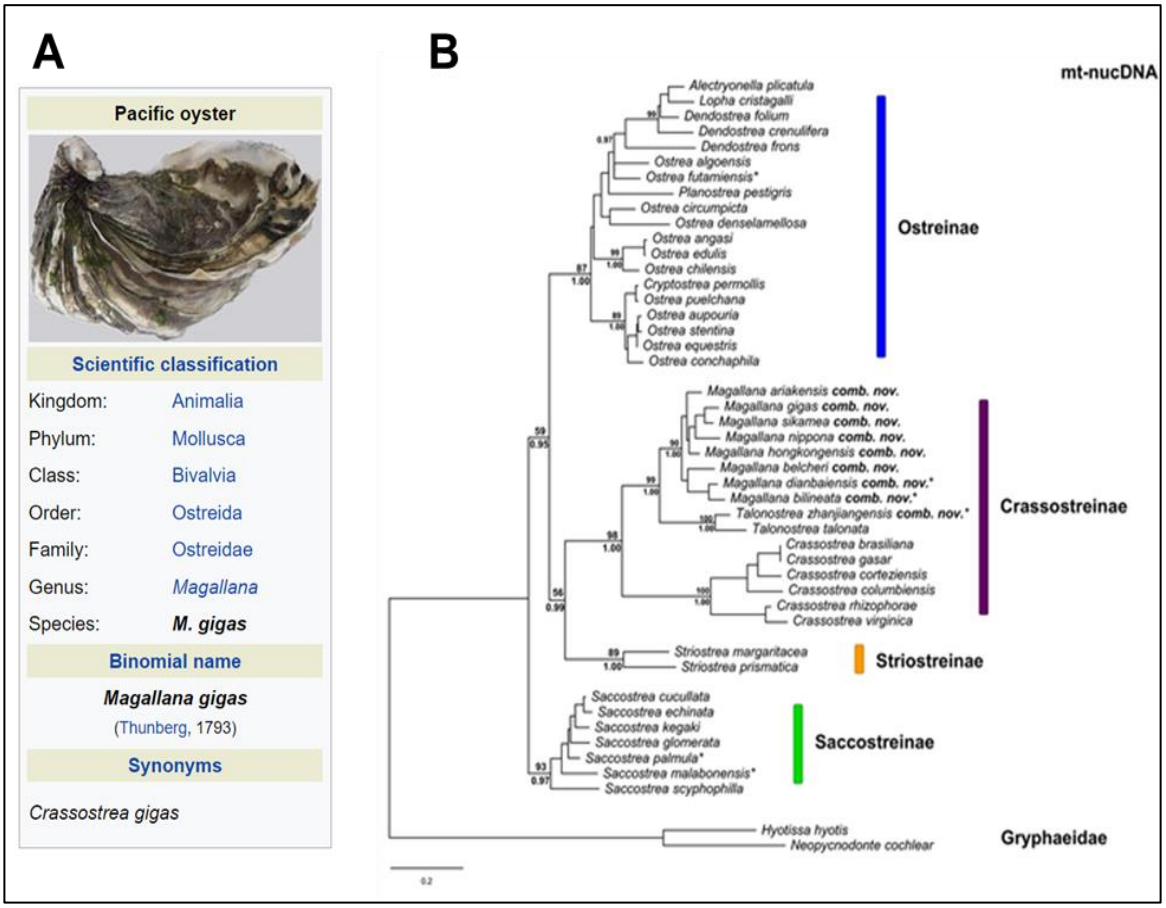


Figure 1.3: Classification of Pacific oyster *Crassostrea gigas* or *Magallana gigas*.

A) A scientific classification (@ Wikipedia). B) The phylogenetic analysis of family Ostreidae based on maximum-likelihood phylogenetic tree (Salvi & Mariottini, 2017).

1.2.2 Anatomy and physiology

The oyster shell colour, size and shape vary according to the environment they settle. Yet, in the case of farmed oysters, it depends on the technique of farming (Mizuta & Wikfors, 2019). The Pacific oyster consists of a soft body enclosed with a shell of two asymmetrical valves. The top valve is normally flat in shape (or slightly convex) and the bottom is quite deep and cup shaped. The two valves protect the internal soft bodies from the predator and potential sudden threat. These valves are attached by a ligament at the hinge close to the anterior and by an adductor muscle. The contraction of the adductor muscle allows them to open and close the valve during the emerging period and during feeding. Oyster valves are made of calcium carbonate that is produced by the mantle (Quayle, 1988). A mantle is a layer of tissue (containing the muscles, nerves and blood vessels) located at the border of the shell. The mantle also serves as a sensory organ for environmental changes. On the mantle lies the oyster viscera, which contains the mouth, gills, heart, and digestive system (Figure 1.4).

Oysters are filter feeders species, where the water flow provide them support in food intake. Oysters have two large crescent-shaped gills located just above the mantle. Gills are used for respiration but also to filter the food particles through their lamella. The retained particles are transported to the mouth by ciliated gill palps. The mouth is just near the hinge and surrounded by the labial palps (that sort and transport the suspended particles to the mouth). The mouth of the oyster has an irregularly shaped, narrow and curbed opening like an inverted U-shaped. Once the food particles are ingested, the particles travel through the entire digestive system (i.e. oesophagus, stomach, rectum and anus). The oesophagus is connected to the stomach directly, which is made of a hollow chambered sac and surrounded by the digestive glands. The stomach is followed by the midgut and then by ascending intestine, descending intestine and rectum. The rectum goes dorsally over the adductor muscle and ends by the anus, which is located in the cloacal chamber.

Oysters have a semi-open circulatory system, where the heart comprising a ventricle and an atrium ensures the blood circulation and the haemolymph circulates in the arteries, veins and sinuses to all the tissues of the oyster.

Oysters are poikilothermic species, their internal temperature varies according to the surrounding, as they cannot regulate their internal temperature.

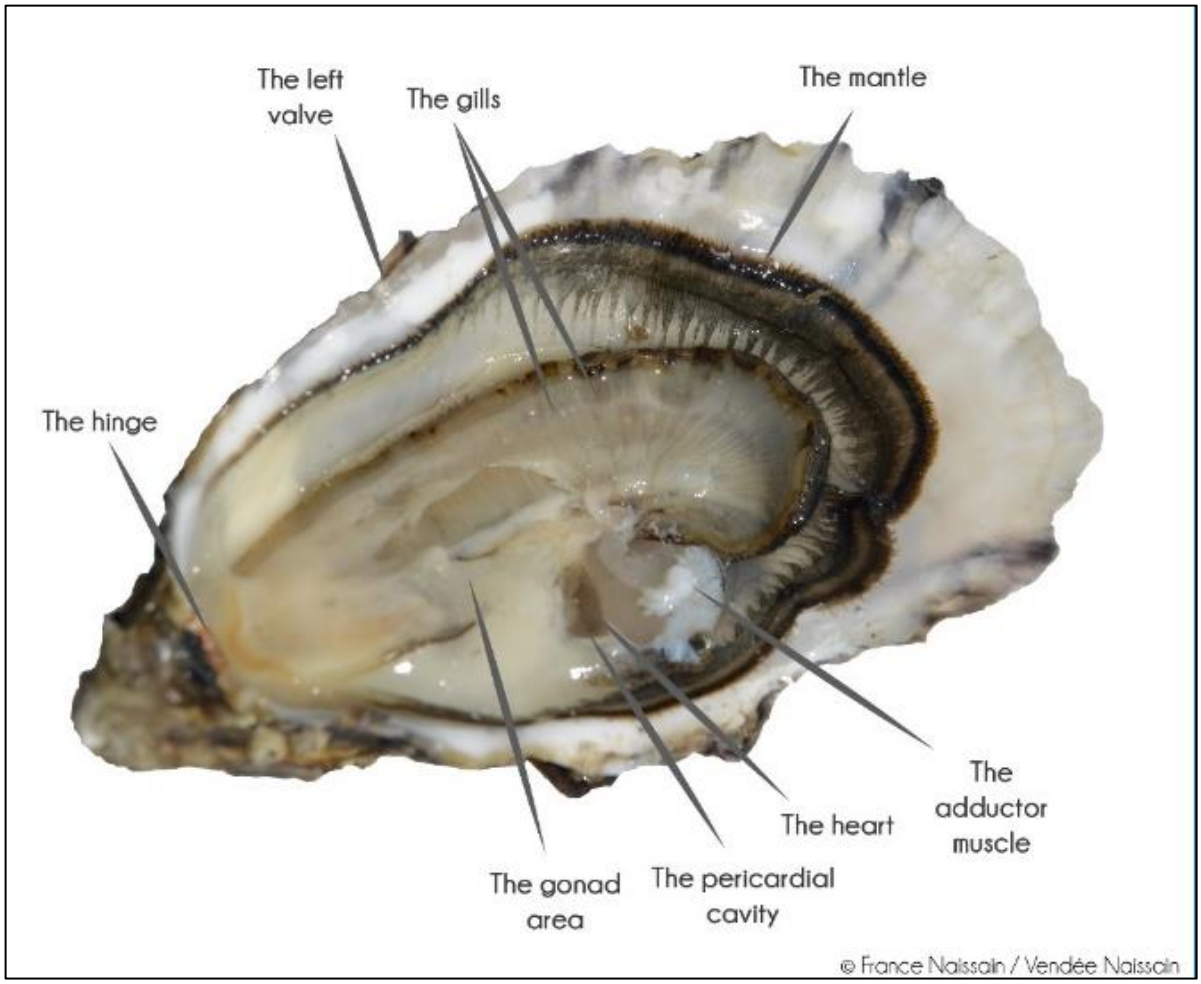


Figure 1.4: The Anatomy of the Pacific oyster showing the main part of oyster body. Figure adapted from (https://www.francenaissain.com).

1.2.3 Reproduction and life cycle

The Pacific oyster is a protandrous hermaphrodite species, they first develop as male and later in the following years, they may change sex to female (X. Guo et al., 1998). Oysters are oviparous with high level of fecundity. Every single individual (size of 8-15 cm) can spawn on average 50 to 200 million gametes. The gametogenesis is temperature dependent, which happens in summer when the temperature is between 18-22 °C (Enríquez-Díaz et al., 2009; X. Guo et al., 1998). Fertilization is external and occurs when the gametes (sperm and eggs) meet in the water column.

Finally, oysters are broadcast spawners and the life cycle of the oyster consist of two phases, first a larval pelagic phase and second a benthic adult stage. The benthic phase begins with the fixation of the spat to hard substrate (Figure 1.5; Bagusche, 2013). Once the eggs are fertilized, they undergo a spiral cleavage (early embryo cell divide) then within 12 hours, they transform into a trochophore (ciliated and motile planktotrophic larva). Around 24 hours post-fertilization, a D-shape larva is formed which can disperse over large distances. This D larva then metamorphoses in a veliger larva. After two weeks, the pediveliger larva undergoes metamorphosis with the development of the foot that allow them to crawl in the bottom until they find a suitable subtract. Once the suitable subtract is found, it attach to it permanently and preferentially onto hard or rocky surfaces. Once settled, oysters are called spat or juveniles and then they grow to adults.

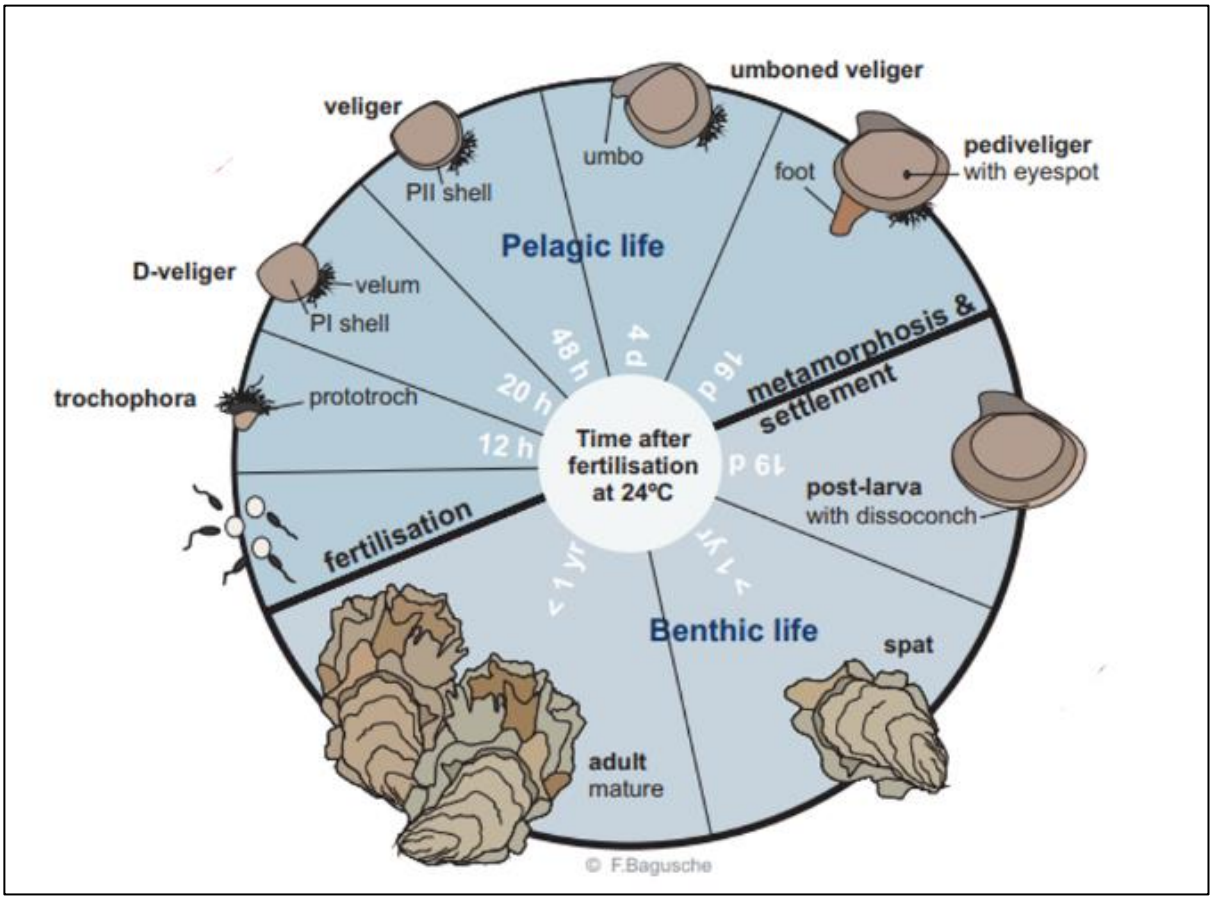


Figure 1.5: Life cycle of Pacific oyster.

The life cycle comprise two phases: the pelagic and the benthic phase.
Figure taken from (Bagusche, 2013).

1.2.4 Distribution and ecology

As from its name the Pacific or Japanese oyster is originally from northeast Asia and is endemic to Japan. It was intentionally introduced worldwide such as in Australia, New Zealand, Europe, North and South America for aquaculture purposes (Figure 1.6) (CIESM, 2000). *C. gigas* is an estuarine species preferring hard bottom substrates but that can be found in mud or sand-mud bottoms (Baggett et al., 2014). Oysters can be found from the intertidal zones up to 40 meters depth. It also can be found in brackish water thanks to its salinity range tolerance from very low such as 10 to high up to 45‰ with optimal ranging between 20-25‰. It has a wide range of temperature tolerance from zero to up to 30 °C (Miossec et al., 2009).

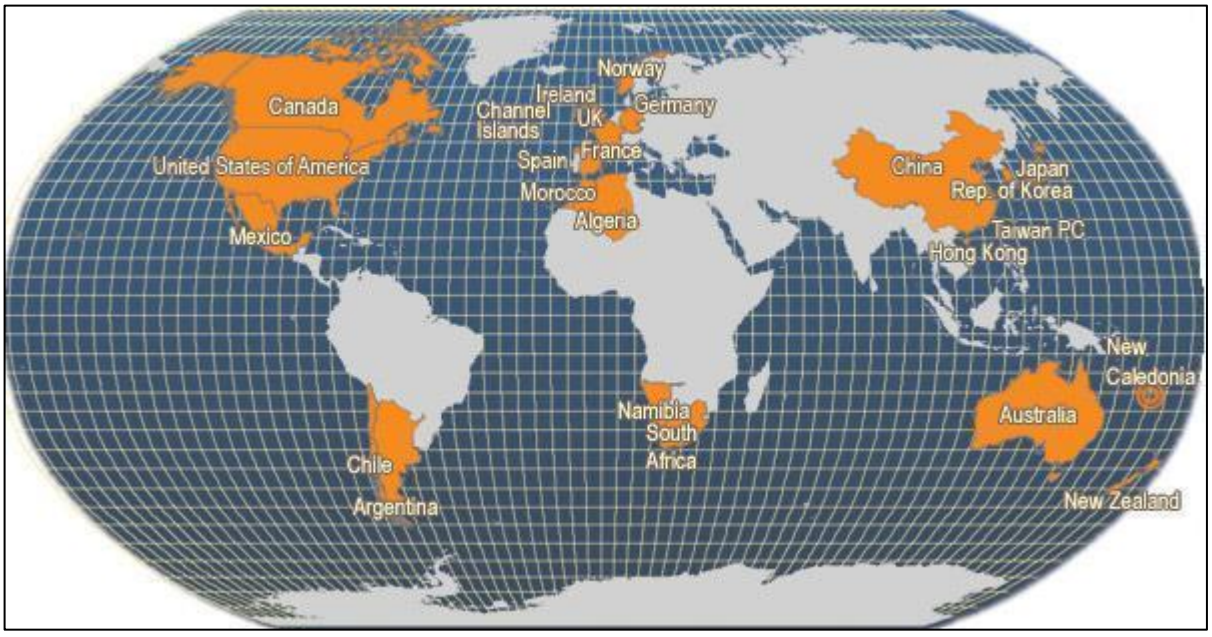


Figure 1.6: Pacific oyster distribution and main producing countries.
(FAO Fishery statistics, 2006)

1.3 Genome and epigenome of Pacific oyster

1.3.1 Genome

C. gigas is a diploid species, its karyotype contains 10 pairs of chromosomes (2N=20) (Bouilly et al., 2010). During the last decade, the range of genetic and genomic resources has been increasing significantly for the Pacific oyster. Most importantly, the genome was sequenced and assembled in 2012 for the first time, the assembly comprised of an estimated size of 559 megabases (Mb) with a contig N50 size of 19.4 kilobases (kb) and a scaffold N50 size of 401 kb (G. Zhang et al., 2012). Annotation of the genome predicted 28,027 genes of which 96.1 % have transcriptional activity. The genome is rich in repeated sequences (represent 36% of the total genome). The genome of oysters is highly polymorphic, with 2.3 % higher than that in most studied animal genomes (G. Zhang et al., 2012).

The fact that oysters have a sessile life and their ability to adapt to fluctuation and selective marine environment might be explained by its enormous gene repertoire. Many genes are characterised by high sequence, structural and functional diversity. This gene repertoire is composed of many gene families involved in stress response and immunity (G. Zhang et al., 2012; L. Zhang et al., 2015), thus revealing the existence of a complex immune system in oysters (X. Guo et al., 2015).

The accessibility of genome facilitates the genetic architecture studies that aim in identifying the underlying basis of a phenotype such as disease resistant or growth traits. Additionally, with the annotation information it is possible to design desired probes or primers to perform target approach studies (such as exome capture).

Nevertheless, more recently the genomic resource are increasing. For example, in 2014, the first Single-nucleotide polymorphisms (SNPs) panel was designed (Lapègue et al., 2014) and in 2017 two SNP array have been developed for Pacific oyster, thus allowing for genome-wide association studies, and for genomic selection (Gutierrez et al., 2017; Qi et al., 2017). In addition, in 2021, two new chromosome-level reference genomes of Pacific oyster have been published. The first has a final size of 586.8 Mb (Qi et al., 2021) with annotation of 30,078 protein-coding genes and the second has 647 Mb (Peñaloza et al., 2021) with 30,724 protein-coding genes (Figure 1.7). The availability of new genome assemblies along with previous one and SNP arrays provide support for further implementation of genetic and epigenetic studies.

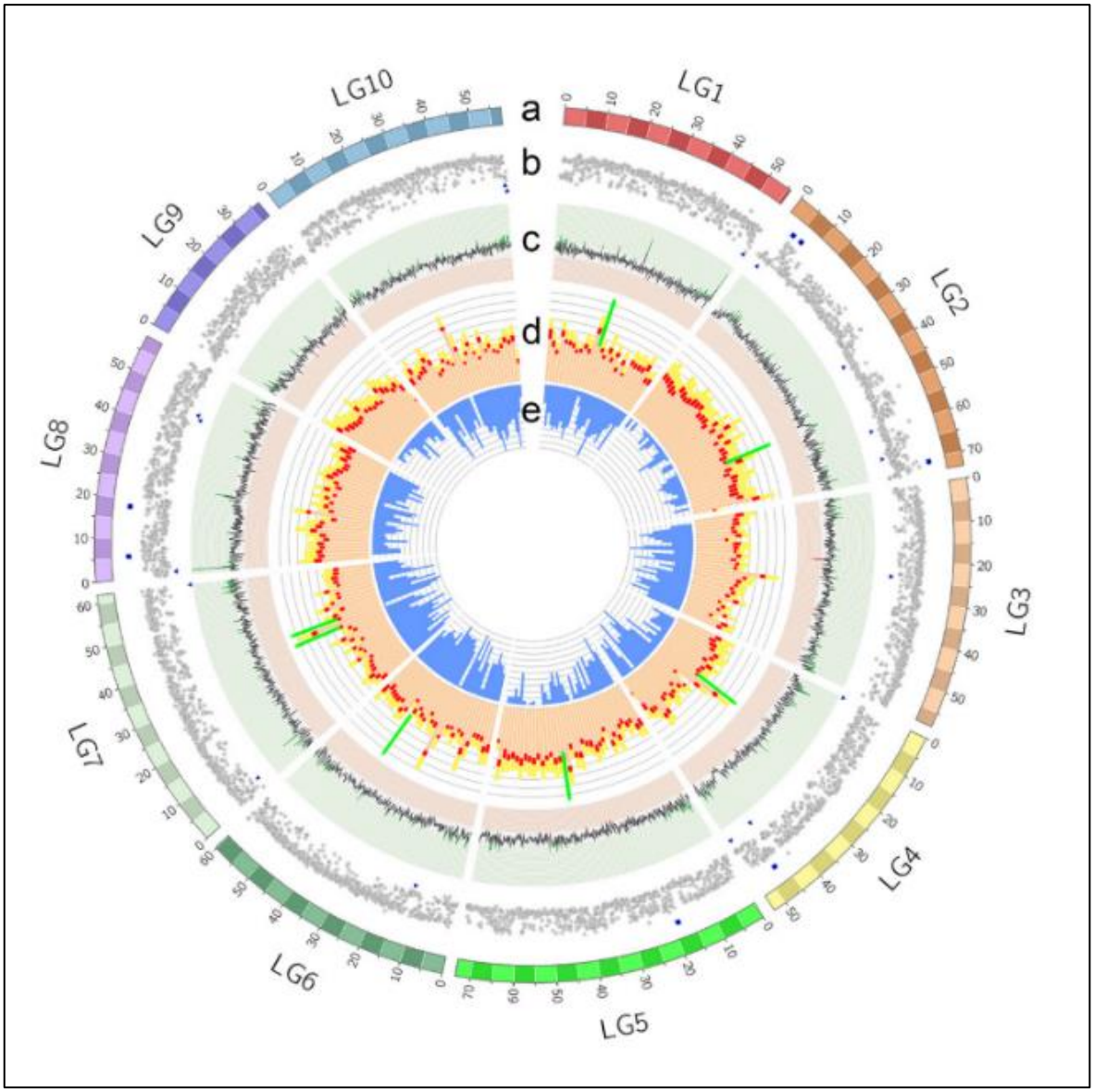


Figure 1.7: Pacific Oyster genome.

Figure adapted from Peñaloza et al. (2021) showing the genome features of the 10 Pacific oyster chromosomes by a circos plot. (a) The 10 chromosomes (Linkage group (LG) 1 to 10 on a Megabases scale). (b) Short-read coverage plot. (c) GC content percentage. (d) Distribution of repeat elements (e) Gene density.

1.3.2 Epigenome

The availability of a reference genome for Pacific oyster, provided an important source for understanding the epigenetic mechanisms in molluscs. Oysters are an excellent model to study the epigenetic mechanisms in the adaptation to stress and to the harsh environment they are in contact with such as fluctuation in temperature and presence of pathogens.

DNA methylation of Pacific oyster has been intensively studied in comparison to other Molluscs species. The existence of the DNA methylation in *C. gigas* was revealed by Gavery & Roberts (2010), suggesting a regulatory role in gene expression and particularly in gene families involved in developmental processes, stress and environmental response. In another study, the authors showed an association between the methylated genes and the transcript abundance (Gavery & Roberts, 2013). The genes encoding the DNA methyltransferase 1 (DNMT1), DNMT2 and DNMT3 have been identified (Xiaotong Wang et al., 2014) in Pacific oyster. These genes are important genes involved in DNA methylation machinery.

The DNA methylation pattern in oysters is similar to the pattern found in other invertebrates with a mosaic type of methylation (Sarda et al., 2012). The methylation presents all along the genome, with areas of highly methylated cytosine separated by large blocks of unmethylated cytosine. The DNA methylation is mostly located in CG context and mainly intragenic (within genes; exons and introns). However, repeated elements and intergenic regions are methylated at a lower level. Studies have shown that the level of DNA methylation (based on Mantle and gamete tissues) in the Pacific oyster correlates with the size of the genes as well as expression rates (Olson & Roberts, 2014; Xiaotong Wang et al., 2014). Within each specific gene, high level of DNA methylation (from mantel and gamete tissues) have been found in higher in internal exons then followed by last and first exons (Song et al., 2017). Ubiquitously expressed genes related to housekeeping functions are hypermethylated while environmentally responsive genes (regulated depending on the context) are hypomethylated (Dixon et al., 2018; Gavery & Roberts, 2010).

Furthermore, DNA methylation plays a major role in the development of *C. gigas*. Thus, genes encoding the DNA methylation machinery are overexpressed in gonadal tissues compared to somatic tissues. During larval development, an overall methylation increase occurs in the exon relative to other genomic constituents (Riviere et al., 2017). Furthermore, tissue and stage-

specific expression of DNA methylation machinery' genes have been observed (Xiaotong Wang et al., 2014).

Likewise, the presence of histones in *C. gigas* have been demonstrated by the identification of histone H3 (Bouilly et al., 2010). In addition, the role of the JmjC (Jumonji C) histone demethylases family have also been identified and were shown to be regulated at the mRNA levels during gametogenesis and embryogenesis (Fellous et al., 2014). Moreover, the expression of this demethylase was also shown to be influenced by temperature changes during the early development of the oyster (Fellous et al., 2015). Altogether, this suggests an important functional outcomes of histone methylation in the developmental trajectories of *C. gigas* (Fellous et al., 2019).

The involvement of non-coding RNAs (ncRNA) in Pacific oysters has been little studied. In Pacific oyster, a total of 11,668 lincRNAs (long intergenic noncoding RNAs) have been identified. Furthermore, another study, demonstrated a co-expression relationship between 14 differentially expressed lncRNAs and 17 differentially expressed immune-related-mRNAs. Altogether, these results suggest a potential role of lncRNAs in immune-related functions.

On the other side, the miRNA are composed of RNA sequence of around 22 bp long. These miRNAs are able to target specific mRNA by complementary base pair binding at the 3' untranslated region (3'UTR), where it reduces the translation and stability of the targeted mRNAs (Ha & Kim, 2014). Additionally, the involvement of miRNA in the developmental stages of marine bivalve such as in *Crassostrea gigas*, has already been shown (Rosani et al., 2016). Moreover, miRNAs overexpressed during immune challenge have also been identified and potentially are involved in the immune response (Zhi Zhou et al., 2014). The expression of some miRNAs also varies in response to osmotic stress revealing the important role of these miRNAs in stress response and salinity tolerance (Zhao et al., 2016).

1.4 Immune system of oyster

We saw previously that oyster are distributed in the intertidal zones and are filter feeders with a sessile life. Living in such habitat put them in constant and direct contact with the surrounding environmental pressures and with many microorganisms (inside the oyster and outside). Inside the oysters, there are microbiota that could be mutualistic, opportunistic and pathogenic and most of these associations are under a fine control involving the oyster immune system. In oyster, the first line of defense is the shell and the mucus covering the soft body and which acts as physical barriers against the infectious agents. Once these barriers are crossed and the pathogens enter the oysters' tissues, the next line of defense is the innate immune system (Green et al., 2015; G. Zhang et al., 2012).

1.4.1 Innate immune response in oyster

For their immune response oysters depend on a series of cascades of reactions to eliminate pathogens. Once the pathogen pass the first line of defence (the physico-chemical barriers; the shell and the mucus) the second line (innate immune response) is activated and rely on cellular and molecular defence mechanisms to prevent the further proliferation of pathogens.

1.4.1.1 The haemocyte

Oysters have a semi-open circulatory system containing the haemolymph (analogous to blood) which permits the circulation of oxygen and nutrients (Figure 1.8A). The haemolymph is composed of plasma and different circulating cell called haemocytes. Haemocytes are multipotent and immunocompetent cells circulating in the sinuses, vessels and heart and also found in different organs of oysters such as the mantle and gills (Bachère et al., 2004). In bivalve the main cellular mediator of defence system are the haemocyte (Schmitt, Duperthuy, et al., 2012). However, haemocytes along its immune functions can perform other functions such as wound healing and shell repair (Canesi et al., 2002). Phagocytosis is one of best designated immune functions of haemocytes, which lead to eliminate the microorganisms recognized as non-self. They are also involved in many metabolic mechanisms such as respiration, nutrient transport, digestion and excretion (S. Y. Feng, 1988). The haemocyte cells are distinguished based on morphological classification either by microscopy or flow cytometry. Two types of populations were described, the agranulocytes (blast-like cells and hyalinocytes) and the granulocytes (Schmitt, Duperthuy, et al., 2012) (Figure 1.8B). A

difference between granulocytes and agranulocytes (hyalinocytes) is the presence of granules in the cytoplasm of granulocyte cells. The blast-like cells are small cells containing a central ovoid or spherical nucleus and surrounded by cytoplasm (Bachère et al., 2015; Hine, 1999).

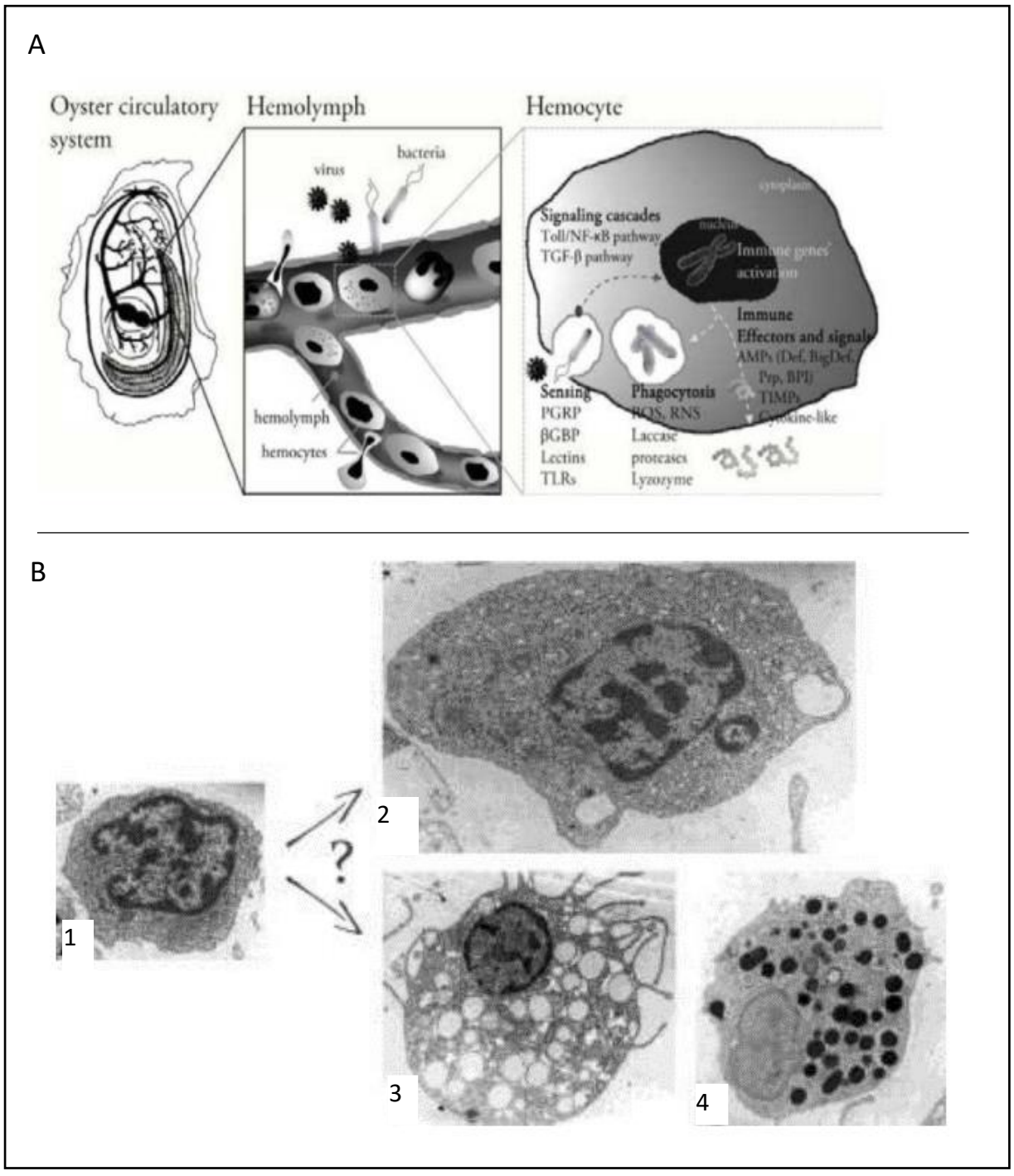


Figure 1.8: The circulatory system and haemocyte types of Pacific oyster.

A) The circulatory system of Pacific oyster with a focus on the role of haemocytes in defense mechanism. B) Transmission electron microscopy showing the three main populations of oyster haemocytes: blast-like cells (1), agranular haemocytes (2), and granulocytes (3 and 4). Figure (A and B) adapted from haemocyte (Schmitt, Duperthuy, et al., 2012)..

Haemocyte are capable of recognising 'self' from 'non-self' through the recognition of small molecules usually called Pathogen-associated molecular patterns (PAMPS) or opsonins (Bachère et al., 2015). Consequently, it trigger a cell-mediated response mechanisms characterized by the aggregation, the phagocytosis, the apoptosis or the encapsulation. Additionally, haemocytes are capable of the secretion of microbicidal compounds such as antimicrobial peptides, DNA extracellular traps, hydrolytic enzymes and reactive oxygen species following the activation of different cascades of immune signalling pathways (Bachère et al., 2015). In the event of injury or infection, haemocytes will migrate through the haemolymph and aggregate at the site of infection and deliver essential elements to repair or fight the infection (e.g. secretion of calcium for shell repair or defence molecules) (S. Y. Feng, 1988).

The movement of haemocytes and its aggregation at the infection event is the first step of the response. This helps to control the infection and avoid its spread to other cells and tissues. Upon recognition of the non-self (pathogens), haemocytes are able to respond by activating a series of immune reactions. These include i) phagocytosis (with the production of e.g. reactive oxygen species (ROS), reactive nitrogen species (RNS) and proteases). ii) and through the activation of signalling pathways such as the NF-κB pathway (nuclear factor kappa-light-chain-enhancer of activated B cells) the production of effectors and immune signals such as antimicrobial peptides (Defensin, BigDefensin, cytokine-like, etc.) (Bachère et al., 2015).

Phagocytosis is an evolutionary conserved cellular process (Jiang et al., 2016). It involves the process of internalization and elimination of foreign particles (such as bacteria, viruses or protozoan parasites). Additionally, it has a crucial role in the pathogen killing and removal as well as the nutrition uptake. More importantly, phagocytosis play an important role in immune defence of oysters (Duperthuy et al., 2011). It starts by the recognition of ligands (PAMPs) present on the foreign particles via receptors on haemocytes such as C-type lectins (CTLs), scavenger receptors (SRs) or integrins. After the recognition, the phagosome formation starts by the internalization of the recognized particle through the deformation of plasma membrane supported by changes in actin cytoskeleton to form a vacuole. Phagosome maturate to form the phagolysosome by fusion of phagosome to lysosome. The lysosome have a microbicidal and degradative activities that lead to lysis of the phagocytized particle. The lysosome release their enzymes into the phagosome leading to a series of processes such

as acidification, accumulation of toxic metals, production of reactive oxygen and nitrogen species (ROS/RNS) and the supply of antimicrobial peptides. All these processes will lead to the lysis of the ingested particle (Schmitt, Rosa, et al., 2012).

1.4.1.2 Non-self recognition molecules

Similar to other invertebrates, the defence mechanisms of the oyster rely on the innate immune system. In order to induce an effective response, its immune system need to be able to differentiate between molecules from the host organism (self) and molecules from outside organisms (non-self).

Oysters have proteins that allow them to recognise external agents (non-self). It relies on a limited number of pattern recognition receptors (PRRs) that can recognize the evolutionarily conserved structures of pathogens, called pathogen-associated molecular patterns (PAMPs) and activate the defence mechanisms (X. Guo et al., 2015). Microorganisms can present different patterns such as peptidoglycan, lipoteichoic acids in Gram-positive bacteria or lipopolysaccharides in Gram-negative bacteria. In case of viruses, this will mainly be DNA or RNA and viral glycoproteins (Mogensen, 2009). The sequencing of *C. gigas* genome in 2012 (G. Zhang et al., 2012) revealed that oysters have sophisticated repertoire of receptors that can recognise a broad range of microorganisms.

PRRs are classified into three types according to their functions. First are the endocytic receptors, which are present on the cell surface and functions in the recognizing and internalization of PAMPs. Second are the signalling receptors, which are responsible for recognizing and activating the intracellular signalling pathway. Third are the soluble molecules, which mediate the linking between the PAMPs and cells (Jeannin et al., 2008; L. Liu et al., 2011). These PRRs include the following:

1- Toll-like receptors (TLRs): TLRs are considered as key receptors with broad range pathogen recognition ability (by their PAMPs) and the activation of innate immune response. They have highly conserved structure (from the cnidarians to mammals) that consist of an extracellular domain of leucine-rich repeats (LLRs) and intracellular domain of Toll-interleukin-1-receptor (TIR). TLRs are activated once an external agent is detected. Then the signals are transduced downstream to activate major transcription factors in the regulation of the inflammatory response and effector mechanisms of innate immunity. A

total of 83 genes coding TLRs have been identified in *C. gigas* genome (L. Zhang et al., 2015). Studies have shown that at least six TLRs have been identified to participate in the immune response (including CgToll-1, CgTLR1, CgTLR2, CgTLR3, CgTLR4, CgTLR6) (L. Wang et al., 2018). In study of (de Lorgeril et al., 2020) found an overexpression of Toll-like receptor 13 (TLR13) in resistant oysters, suggesting that this receptor could be able to detect the viral and bacterial RNA and participate in TLR/NF-KB pathway.

2- Peptidoglycan recognition proteins (PGRPs): Peptidoglycan recognition proteins (PGRPs): PGRPs are another set of PRRs molecules in innate immunity essential for recognizing the peptidoglycan in cell wall of pathogenic bacteria and eliminating them. So far, about nine PGRPs have been identified in *C. gigas* genome (L. Zhang et al., 2015). Some are characterised by a short PGRP and a conserved PGRP/amidase domain in their C-terminus (L. Wang et al., 2018), others by including additional goose-type lysosome or defensin-like domain (Allam & Raftos, 2015).

3- Scavenger receptors (SRs): SRs are endocytic receptor having major role in non-opsonic phagocytosis through its ability to recognise various ligands present in the pathogens. They are also involved in vital functions to maintain host homeostasis and defence, this include apoptosis, autoimmunity, inflammation, and lipid metabolism. In *C. gigas* genome a set of 71 genes encoding the SRs were identified. Studies in oyster have shown an increase in SRs expression during summer mortalities (de Lorgeril et al., 2020; Elodie Fleury et al., 2010; Huvet et al., 2004).

4- Lectins: lectins are sugar-binding proteins with high a diversity. In animal, there are different group of lectin, such as c-type lectin, galectin, f-type lectin and rhamnose-binding lectin (Iiyama et al., 2021). These proteins play an important role in immune defence mechanisms. Lectins bind to glycoproteins, glycolipids or polysaccharides present on the pathogens surface, thus prohibiting them to adhere to cell surface of host or play the role of opsonins. In *C. gigas*, studies have identified different types of lectins, including gigalins, ficollins, c-type lectins, integrins, galectins and chitinase-like lectins (Badariotti et al., 2007; Duperthuy et al., 2011; Terahara et al., 2006; Yamaura et al., 2008).

    4.1 The C-type lectins (CTLs) are comprised of a superfamily of calcium (CA+2)-dependent carbohydrate-recognition proteins (presenting minimum one carbohydrate-recognition domain, CRD) (Cambi et al., 2005). CTLs can be found as soluble or transmembrane and play an important role in innate immunity and non-

self recognition. So far, a total of 266 genes encoding protein that containing CTLs domain have been identified in *C. gigas* genome (L. Zhang et al., 2015). Studies have shown that CTLs and galectins can be associated in the recognition of Gram-positive bacteria and/or can enhance the phagocytosis activity (Hui Li et al., 2015). An overexpression of CTLs have been found in *C. gigas* in response to Ostreid herpesvirus 1 virus (OsHV-1) infection (He et al., 2015).

4.2 The Mannose binding lectins (MBLs), which are soluble protein involved in lectin-mediated complement system activation (Holmskov et al., 2003). In *C. gigas*, studies have shown the MBLs along with CTLs are important for activation of complement system upon the infection with *Vibrio splendidus* (Hui Li et al., 2015).

4.3 Fibrinogen-related proteins (FREPs): These lectins are essential for coagulation in vertebrate, however in invertebrates there are not involved in coagulation but have an important function in host defence (Hanington & Zhang, 2011). They are highly diverse and 190 FREPs have been identified in Pacific oyster genome (G. Zhang et al., 2012).

5- The complement component 1q (C1q) is one of important proteins involved in the activation of the complement system via the classic pathway. In vertebrates, the complement system consists of three pathways, the classical, the Lectin and the alternative pathway. The classical pathway is initiated when the C1q binds to antibodies that are attached to pathogen surface. The lectin based activation by Mannose Binding Lectins pathway, is initiated when the MBLs encounters the conserved carbohydrate motifs found on the surface of the pathogens (Dunkelberger & Song, 2010). Finally, the alternative pathway is activated directly by pathogens (Nonaka & Miyazawa, 2002). In genome of *C. gigas*, 337 proteins coding the C1q domain-containing (C1qDC) have been identified (Gerdol et al., 2015).

6- RIG-like receptors (retinoic acid-inducible gene-I-like receptors, RLRs): These receptors are from family of DExD/H box RNA helicases (Yoneyama & Fujita, 2009). RLRs are intracellular PRRs, which can sense the presence of viruses through the recognition of PAMPs on the surface. They transmit the information to downstream transcription factor for the production of type 1 interferon (IFN) and expression of antiviral genes for controlling the virus infection by an intracellular immune response. There are three members in the RLRs; the RIG-1 (Retinoic acid-inducible gene; encoded by DDX58), MDA5 (melanoma

differentiation associated protein 5) and laboratory of genetics and physiology 2 (LGP2) (Kumar et al., 2009). The RIG-1 is capable to recognize the double strands RNA (dsRNA; short size around 300 bp) of virus (B. Huang et al., 2017). On the other hand, the MDA5 can recognise dsRNA with sizes bigger than 1000 bp (Reikine et al., 2014). There are 11 RLRs identified In *C. gigas* genome, (L. Zhang et al., 2015).

1.4.1.3 Signalling pathways involved in innate immune response

Different genes involved in the above mentioned pathways have been identified by comparative approach between oyster genome and those of vertebrate (Figure 1.9) (Escoubas et al., 1999; Green et al., 2015; Montagnani et al., 2004, 2008; G. Zhang et al., 2012; L. Zhang et al., 2015).

Once the PRRs recognize pathogens through PAMPs, it initiates different immune response pathways. These pathways are: 1) cell defense pathways such as autophagy and apoptosis; 2) Pathways initiated by different PRRs such as TLR/NF-kB, the RLRs/STING, and JAK/STAT; and 3) The RNA interference (RNAi) pathway (Figure 1.9). For example, These TLRs are essential for the activation of the transcription factor nuclear factor-kappaB (NF-κB). The TLR/NF-kB is a cell signalling pathway with significant similarities in species from mammals to cnidarian. This is one of the main pathways involved in innate immune system. It involve a family of Toll-like receptors (TLRs), which act as primary sensors detecting different component of pathogens and initiate an innate immune response. The TLRs send signal through the myeloid differentiation primary response 88 (MyD88) gene through the homophilic interactions between the TIR-TIR domains. Through the Death domain present on MyD88 it allow the association of Myd88 with death domain of the serine threonine protein kinase IL-1R-associated kinase (IRAK) gene. Thus leading to activation of IRAK and it return it further associate to Tumor Necrosis Factor receptor-associated factor (TRAF6). Then it is followed by TRAF6 oligomerization, which it led to the activation of IkappaB kinase (IKK), phosphorylation and degradation. The NF-kB factors are translocated into the nucleus, where it induce the transcription of target genes (Horng & Medzhitov, 2001). Thus, the NF-kB controls the expression of different inflammatory cytokine genes (Kawai & Akira, 2007).

1.4.1.3.1 Programed cell death

There are two types programed cell death (PCD) namely the apoptosis and autophagy. PCD are a fundamental process in innate immunity and homeostasis of the organisms (Green et al., 2015).

Apoptosis can be defined as the process of self-destruction of infected or defected cells. Apoptosis is induced by environmental changes (heat, salinity, heavy metals, hypoxia) and the presence of bacteria, viruses and parasites (L. Wang et al., 2018) (Figure 1.9). The characteristic features of apoptosis are shrinkage of cytoplasm, condensation of chromatin and DNA fragmentation (overall shrinking of cell). Apoptosis could be activated by two major pathways, which are the intrinsic (signal from inside the cell) and the extrinsic (signal form outside the cell) apoptotic pathways. These two pathways are regulated by different pro-apoptotic or anti-apoptotic proteins. In the case of intrinsic pathway, this is triggered by the initiator caspase-9 protein, while the extrinsic pathway involves initiator caspase-8 (L. Wang et al., 2018). So far, many pro-apoptotic and anti-apoptotic molecules have been found in oysters such as the B-cell lymphoma-2 (Bcl-2) or inhibitor of apoptosis proteins (IAP) (Qu et al., 2015; L. Zhang et al., 2011). An overexpression of Bcl-2 and IAP have been associated with the OshV-1 infection, suggesting the involvement of apoptosis in oyster response to virus (Green et al., 2014; Jouaux et al., 2013; Segarra et al., 2014)

Autophagy is another type of programmed cell death that has an important role in innate immunity and cell homeostasis. Autophagy function in the lysis of intracellular pathogens (such as the bacteria and virus) and cytosolic organelles resulting in the formation of a double membrane structure called autophagosome (Figure 1.9). The autophagosome will fuse with lysosome to form autolysosome that will degrade the contents by enzymatic reaction. Autophagy have been reported in oyster and many autophagy-related genes (ATG) have been identified in the oyster genome (L. Wang et al., 2018). In addition, the membrane-bound form of LC3 (LC3-II) have been identified in oyster, it is well documented and described that LC3 is associated with the autophagosome and the autophagy activation (Moreau et al., 2015). Interestingly, the inhibition of autophagy led to a decreased level of the survival to OshV-1 virus and *Vibrio aesturianus* bacteria (Moreau et al., 2015), indicating a strong functional importance of autophagy against the viral and bacterial infections.

1.4.1.3.2 Interferon-like system response

In vertebrate, once a virus is detected, the interferon (IFN) system functions as the first line of immune defence. Interferon system acts as an antiviral immune response by recognizing the virus (for example the recognition of double strand RNA and glycoproteins). The IFN system controls the viral infection by hindering its replication at different stages of the viral life cycle. The activation of IFN evoke an antiviral response by interacting with their equivalent receptors leading to the activation of Janus kinases/ Signal transducers and activators of transcription (JAK/STAT) pathway and subsequently, induce the expression of IFN-stimulated genes (ISGs) to render the viral replication (Lu et al., 2018) (Figure 1.9). IFN system was thought to be only found in vertebrate due to its absence in model organism genomes (i.e. Drosophila and mosquitoes) (Green et al., 2015; Loker et al., 2004; Robalino et al., 2004). However, in invertebrate, the interferon-like system was progressively identified and called IFN-like system components (Qiao, Wang, et al., 2021). In *C. gigas,* study revealed with the genomic sequence data the existence of many genes with similarity to ISGs, (Green et al. (2015). These genes include 2'-5-oligoadenylate synthetase (OAS), Mx protein, viperin, ADAR-L and IFI44 (Green et al., 2015; G. Zhang et al., 2012). In addition, in *C. gigas* there is accumulating evidence for the existence of interferon-like system. The evidence is the identification of several evolutionary conversed sensors of nucleic acid including TLRs and RLRs (Green et al., 2015). These sensors provide the ability to oyster to recognise non-specific nucleic acid (such as poly I:C) and therefore to induce an antiviral response that subsequently give protection against the OshV-1 infection (Green & Montagnani, 2013). Second, the identification of effectors of downstream signalling pathways including the different genes in involved in the TLR/NF-KB pathway (Kawai & Akira, 2007) such as IkappaB Kinase (IKK), Cg-Rel (REL Proto-Oncogene, NF-KB Subunit) or inhibitor of nuclear factor kappa B (IκB) (Escoubas et al., 1999; Montagnani et al., 2004, 2008). Third, the identification of interferon regulatory factors (IRFs) and stimulator of interferon (STING) that are known to be involved in interferon type I production. Finally, studies have identified IFN-like protein (CgIFNLP) IFN receptor (CgIFNR-3) and novel identification of CgIFNLPR-1, which after the knocked down of the CgIFNLPR-1 decreased notably the expression of ISGs (such as the CgMx, viperin and IFNIP-44) in the haemocytes cells (Qiao, Zong, et al., 2021; R. Zhang et al., 2015, 2016).

1.4.1.3.3 RNA interference (RNAi)

This mechanism was first discovered in plants and is known as a post-transcriptional gene silencing, RNA silencing or RNA interference (RNAi) mechanisms that widely occurs in many eukaryotic organism (Meister & Tuschl, 2004). RNAi is divided into three type according to their structure and function (Kingsolver et al., 2013):

i) Short interfering RNAs (siRNAs),

ii) MicroRNAs (miRNAs) and

iii) PIWI-interacting RNAs (piRNAs).

RNAi is able to regulate the gene expression in animal (Meister & Tuschl, 2004; Randall & Goodbourn, 2008), induce transcriptional response by interfering with interferon pathway and control different post-transcriptional gene process (Meister & Tuschl, 2004) (Figure 1.9). RNAi in anti-viral immunity is induced by the recognition of virus-derived intracellular dsRNA by the endoribonuclease Dicer (or Dicer) (Fire et al., 1998). Dicer is a ribonuclease enzyme that degrades the dsRNA into siRNA (short around 21 bp). Then via the argonauts (AGO), the siRNA is transported to the RNA-induced silencing complex (RISC), where the siRNA is unwound into two single strands, the sense strand is freed, and the antisense strand is used as a guide to find the complementary mRNA sequence to be degraded or silenced (Fire et al., 1998; X. B. Wang et al., 2010).

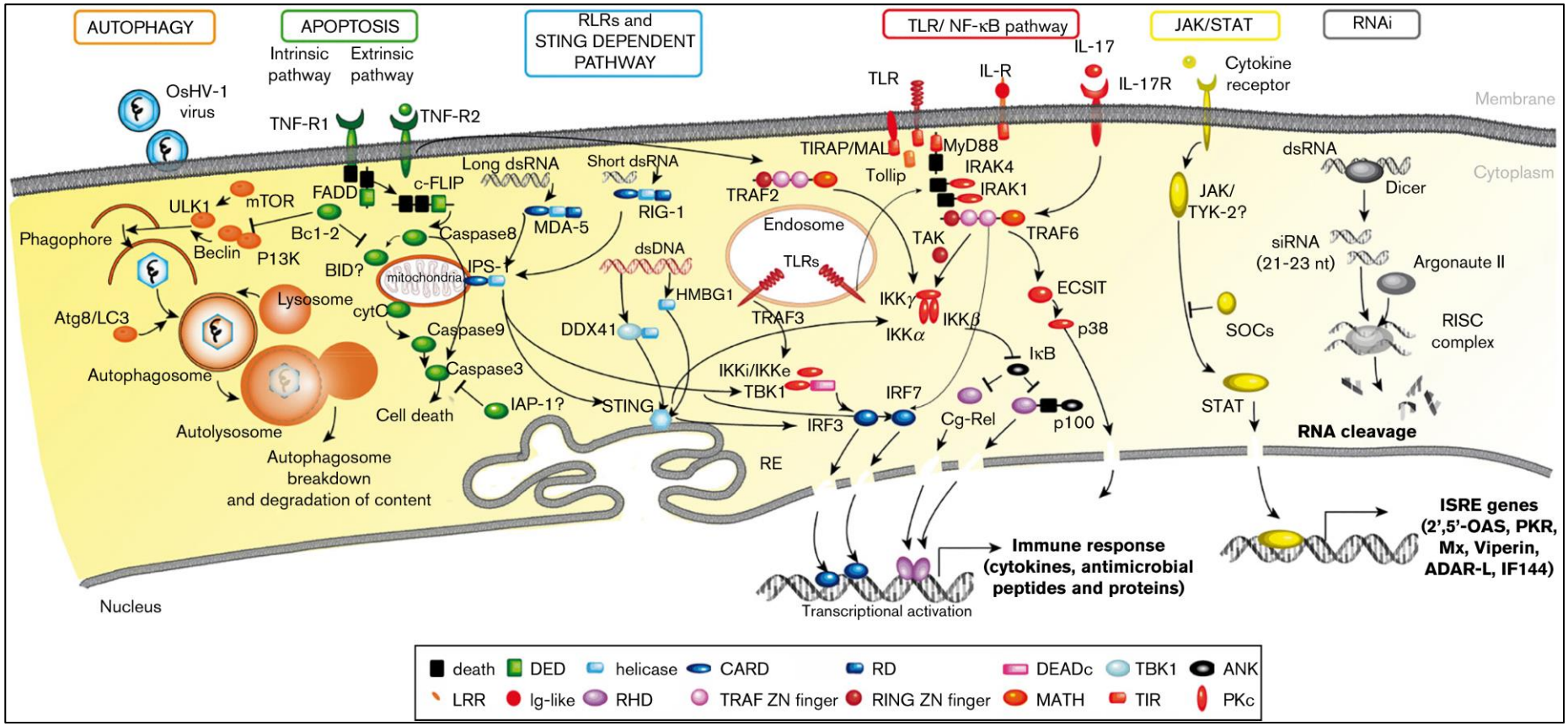


Figure 1.9: Conserved antiviral signalling pathways in the Pacific oyster.

Figure adapted from the (Green et al., 2015).

1.4.1.3.4 Immune effectors

Immune effectors are large group of molecules that are induced by the PRRs and associated signal transduction pathways. They are produced by various organs and epithelial cells (such as gills, mantle, digestive gland and intestine, which participate in the antimicrobial defense mechanisms). They are sensitive to the environmental changes and active against a wide range of pathogens. Therefore, they are important molecules for the immune system of oyster by limiting the invaders capacity and their elimination (X. Guo et al., 2015).

1.4.1.3.4.1 Plasma proteins

Oysters have a semi-open circulatory system, where the haemolymph provide a protective line between the immune system and the invaders (bacteria) that enter the oyster body. Different plasma proteins have been identified in the haemolymph including the extracellular metalloenzyme Superoxide Dismutases (SODs), dominin and cavortin. SODs have an antioxidant role in oyster and one member of cg-EcSODs have lipopolysaccharide (LPS)-binding properties and can act as an opsonin for fighting against *Vibrio splendidus* (Duperthuy et al., 2011; Gonzalez et al., 2005). It also acts as an opsonin displaying antibacterial responses again *Vibrio tasmaniensis* LGP32. This response is achieved by Cg-EcSOD recognizing the outer membrane protein (OmpU) a virulence factor on the LGP32 surface needed for the adhesion and invasion) and then followed by the phagocytosis (Duperthuy et al., 2011).

1.4.1.3.4.2 Antimicrobial peptides or proteins (AMPs)

AMPs represent a large and diverse group of chemically and structurally heterogeneous family of molecules. They can be distinguishable by their size (small molecules), cationic and amphipathic structures (Bachère et al., 2015). They have a microbicidal or bacteriostatic action; however, AMPs characterization is mainly abundant for bactericidal and bacteriostatic functions. For example, they can stop bacterial proliferation by preventing the synthesis of proteins or components of the bacterial wall process (Bulet et al., 2004). Or acting directly through bacterial lysis by forming pores on the bacterial membrane (Brogden, 2005). They can be engaged in phagocytosis to eliminate pathogens. In *C. gigas*, so far, many antimicrobial peptides and proteins have been characterised. They can be classified into 6 classes: defensins, big-defensins, proline-rich peptides (PRPs), bactericidal/permeability increasing proteins (BPIs), ubiquitins and molluscidins (Bachère et al., 2015; Schmitt, Rosa, et al., 2012). Their expression varies between

different groups depending on the infection agent or tissue type involved. For example, some could be constitutively expressed while others expressed only during an infection (Schmitt, Rosa, et al., 2012).

1.4.2 Immune memory and Priming

The ability of immune memory of specific pathogen is the characteristic of vertebrates for adaptive immunity. Vertebrates have immune memory, which they can recognise the pathogen in secondary encounter or event and then be more effective in this second response minimizing the potential risk of infection. This immune memory is based on dendritic cells (DCs), specialized T cells and B lymphocyte cells that ensure adaptive immune responses (Netea et al., 2020). However, invertebrate organisms lacking lymphocytes have long been considered unable of responding specifically to pathogens considering the immune memory to be only exclusive for adaptive immunity. However, since last decades, it has been shown that innate immune cells show adaptive characteristics. Recent discoveries in vertebrate innate systems, showed a memory capabilities and a better adaptive response for a secondary infection. These capabilities are denominated “trained immunity” or “immune priming”. Recent literatures in plants and invertebrates showed that their immune system could be primed in an adaptive manner (e.g. it could be trained for better response for a secondary infectious attack) (Conrath et al., 2015; Milutinović & Kurtz, 2016).

Invertebrates were thought to be without immune memory. However, recent work in molluscs including *C. gigas* showed the existence of immune memory (Pinaud et al., 2016; Portela et al., 2013) (Green et al., 2016; Green & Montagnani, 2013; Lafont et al., 2017, 2020; Y. Li et al., 2017; C. Liu et al., 2016; T. Zhang et al., 2014); suggesting the possibility of antiviral and antibacterial immune priming. It was also showed the possibility of improving the immune capacities of oysters in the face of OshV-1 infection by the injection of poly (I:C), mimicking a viral infection (Lafont et al., 2020).

Overall, these studies are reinforcing the hypotheses raised in invertebrates on the existence of immune memory. It also highlights the potential implication of epigenetic remodelling in the establishment of this innate immune memory. Therefore, further studies are needed for identifying the role of epigenetic mechanisms in immune priming.

1.5 Oyster in aquaculture

1.5.1 Food demand and aquaculture

Along exponential growth of human population, there is an increasing demand for food supply. The human population is growing at fast rates since 1950th. During the last 70 years, the human population tripled from 2.5 billion to 7.5 billion. By 2030, the human population is expected to hit 8.5 billion and 9.7 billion in 2050 (Figure 1.10A; United Nations, DESA, 2019).

Consequently, aquaculture production has increased rapidly since the last 70 years. During the last 20 years the total production has tripled from 41 million tons to 120 million tons (Figure 1.10B; FAO, 2021). Aquaculture is one of sources to cope with this food demand. Aquaculture consist of different activities of breeding, rearing and harvesting of fish, algae, shellfish, crustacean and other organisms in different water environments.

To endure this increase in the production, high numbers of fish and shellfish species were domesticated and introduced all over the world for economic purposes and food supply-demand. However, these intensification of production and species transferred from different points led to the emergence of serious disease outbreaks (Rodgers et al., 2015). Introduction of new species was spatially and temporally concomitant with the appearance of diseases. The occurrence of diseases sometimes led to the collapse of aquaculture industries (Hill, 2002). To cope with this problematic issue, selective breeding would be an interesting way to enhance the quality of aquaculture species by improving their traits of economic importance such as disease resistance. In addition it would represent a useful tool to manage and control disease in farming areas that generally localized in the wild environment (Stear et al., 2001). Human has selected animal and plants displaying traits of high economic importance (Gjedrem, 1983). The selection was either, unconscious by domestication or intentional modification using various technics from breeding programs, polyploidisation and some genome editing tools (Dégremont, Garcia, et al., 2015).

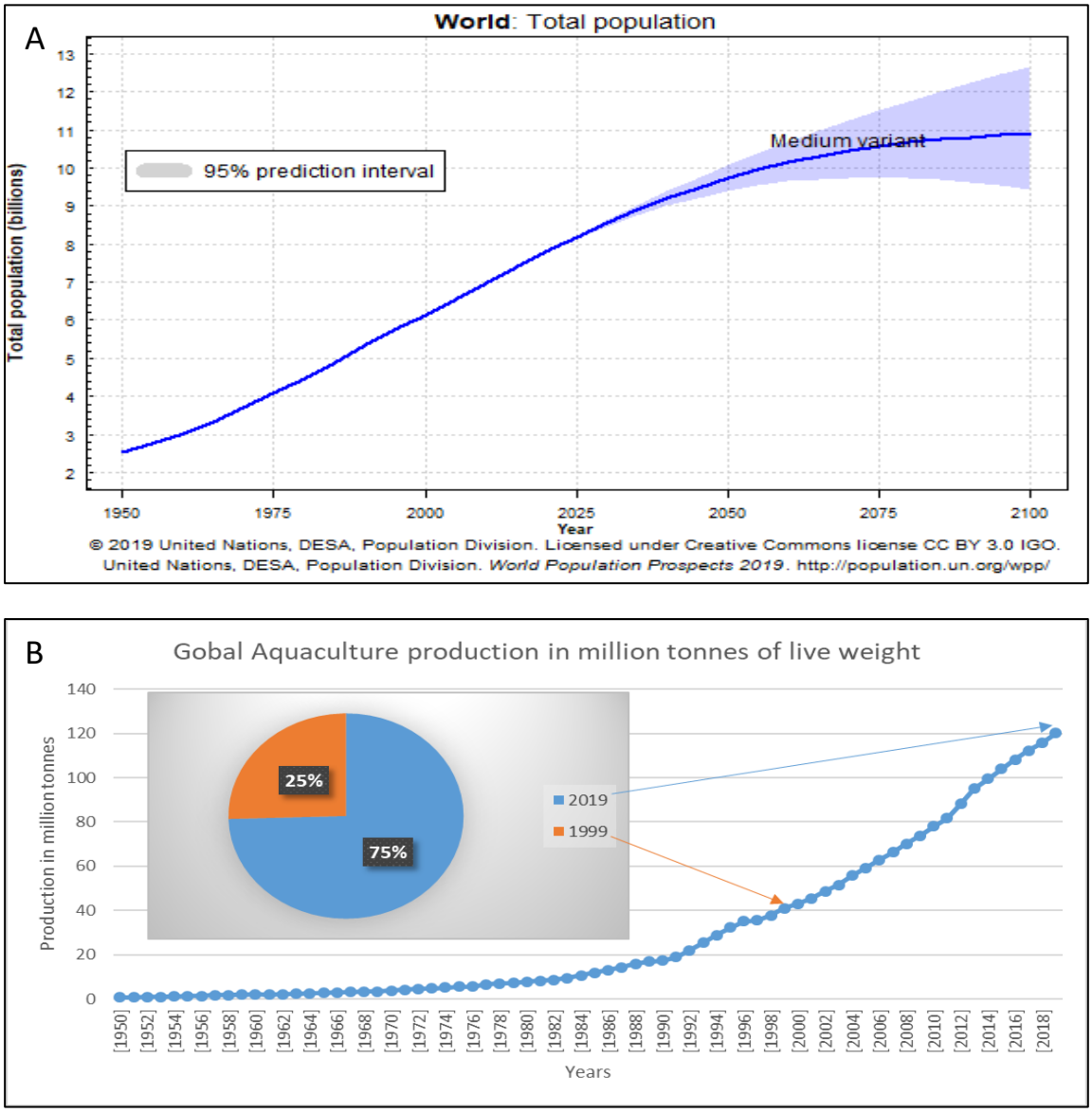


Figure 1.10: Human population growth and Aquaculture production.

A) Human population growth estimation up to 2100 (2019 United Nations, DESA). B) Global aquaculture production from 1950 to 2019 (FAO 2021).

1.5.2 Oyster aquaculture in France

Aquaculture is one of the rapidly growing food industries, and currently, molluscs' aquaculture is one of the biggest. Molluscs are the major group in aquaculture accounting for 14.6 % of global aquaculture production by weight in 2019 (FAO, 2021). Within molluscs, the oysters are the most important taxonomic group in terms of volume produced each year. Oysters are accounting for 34.8 % of total molluscs volume produced in 2019 (which was 6.1 million metric tonnes of a total 17.5 million metric tonnes) ahead of 'clams, cockles, arkshells' group and mussels (FAO, 2021). Pacific oyster represents the major player in this industry. France is the fifth largest producer of oysters (and fourth for Pacific oyster) after China, South Korea, Japan and United states of America (including the American oyster, *Crassostrea virginica*) (FAO, 2021). In Europe, France is the top producers of oysters.

Historically, since the 18th century, the oyster production in France has passed through a succession of different oyster species. This was primarily because the oyster production has collapsed several times after disease outbreaks (Figure 1.11). At the beginning of the 20th century, a massive and unexplained mortality of flat oysters (*Ostrea edulis*) was reported all over Europe. These mortalities led to the disappearance of this oyster from almost all the Atlantic coastline of France. During the 1950th, high mortalities of this species also occurred in the Mediterranean Sea, which have impacted the production of flat oyster. Later, during the 1970th a protozoan (*Marteilia refringens*) and a parasite (*Bonamia ostreae*) have declined the production at the Brittany coastlines (Buestel et al., 2009; Pernet et al., 2016). To compensate the loss of the flat oyster production, importation of Portuguese oyster (*Crassostrea angulata*) started during the middle of the 19th century. After the massive mortalities events that hit the flat oyster, the Portuguese oyster successfully replaced the flat oyster production. Then, during the 1960th - 1970th the Portuguese oyster became the main species cultivated in the Atlantic coasts (bays of Arcachon, Marennes-Oléron and Brittany coast, France). In 1967 a disease characterised by labial gill lesions (Comps, 1969; COMPS, 1970) affected the Portuguese oyster. Later, other symptoms were diagnosed such as the invasion of connective tissue by the blood cells. These symptoms accompanied by extreme mortalities during the years 1970-1973 resulted in the disappearance of the Portuguese oyster from the French coasts. This devastating disease was caused by an iridovirus (Comps, 1983; Comps et al., 1976). Nearly 5,000 oyster farmers were affected and the economic loss was estimated to be above 550 million francs (approx. 8.3 million Euros) with the

annual loss of 60,000 tonnes of oyster (Grizel & Héral, 1991). To circumvent this lost, during the late 1960s, the Pacific oyster was intentionally introduced from Japan to France and additionally it was imported to other European countries (Grizel & Héral, 1991; Rodgers et al., 2015; Rohfritsch et al., 2013). The Pacific oyster has successfully adapted to its new environment and the population is currently spreading over the coasts. There is no clear difference in the genetic structure between the French oyster population and the source population (Japan), an no lose in the genetic diversity (Gagnaire et al., 2018; Lapègue et al., 2020). It is well-known that multiple massive introductions of Pacific oyster could attribute to the absence the founder effect.

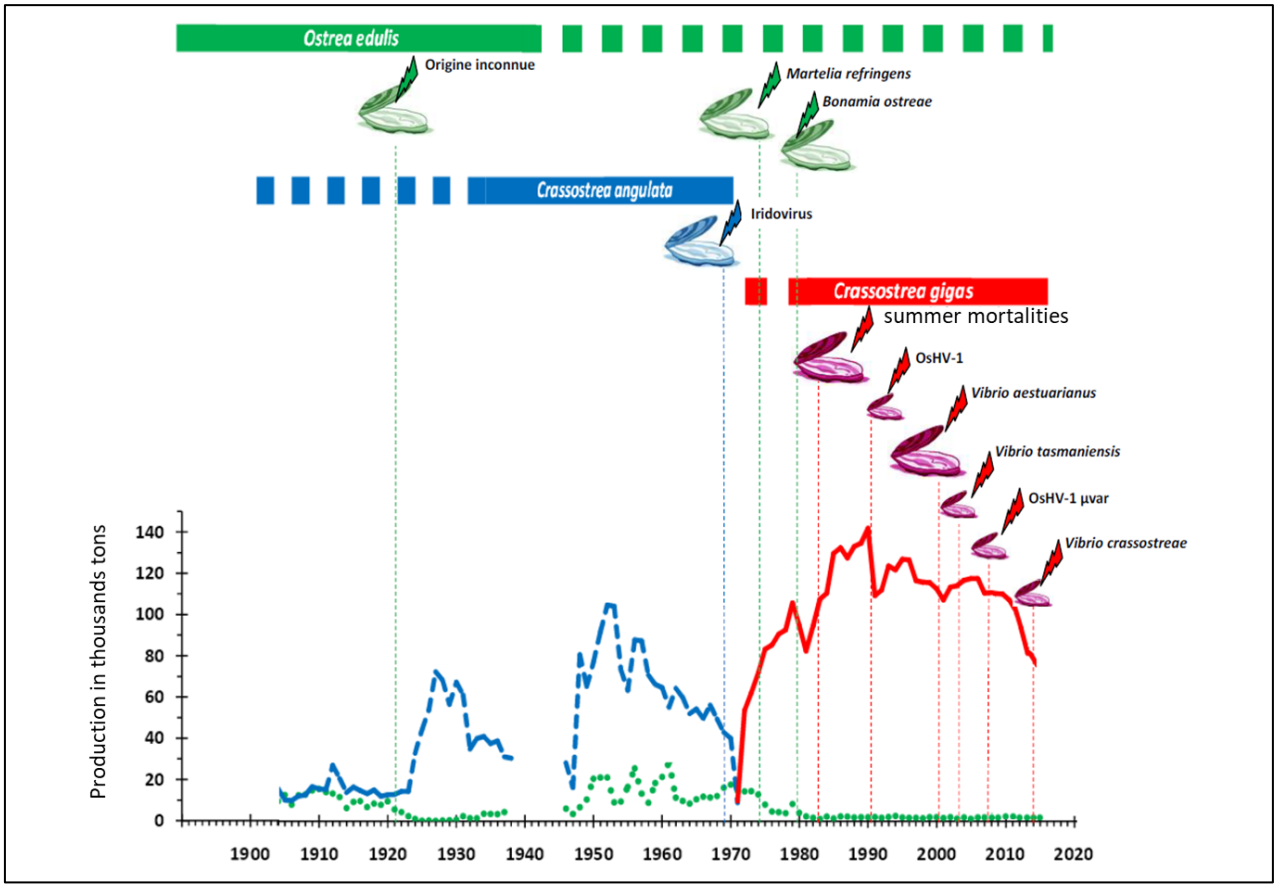


Figure 1.11: History of oyster production and mortalities events

Figure adapted from the thesis of (Lafont, 2017) and showing the history of oyster production and mortalities events associated with collapse in the production of oyster in France. *Ostrea edulis* (green), *Crassostrea angulata* (blue) and *Crassostrea gigas* (red).

1.6- Massive mortalities of pacific oyster *Crassostrea gigas*

1.6.1 Mortalities in Pacific oyster

Pacific oyster species is characterised by wide tolerance against many factors such as temperature, salinity and pathogen susceptibility. For example, the Pacific oysters are able to survive in temperature ranging from zero to > 30 °C (Bougrier et al., 1986; Diederich et al., 2005; Le Gall & Raillard, 1988; Quayle, 1988) and showed higher resistant to the iridovirus, the virus that affected the *C. angulata*. Pacific oyster recorded high yield production and growth rate was two times higher than those obtained for the Portuguese oyster (Bougrier et al., 1986; Héral et al., 1986; His, 1972).

However, since the introduction started, stressful culture conditions and displacement of spat from one farming area to another has resulted in the emergence of several diseases. Since 1991, high incident of spat and juveniles of *C. gigas* mortalities have been observed in different farming areas of the French coasts (Nicolas et al., 1992). These mortality events were called “summer mortalities” syndrome. This happens during the sexual maturation of oyster when the seawater temperatures reached around 19 °C. A herpes-like virus has been associated with recurrent summer mortalities, first in France (Renault et al., 1994, 2000) and later in the USA (C.A. Burge et al., 2006, 2007). These mortality events were recurrent and have been increasing since 2008. These mortalities count for decimating up to 90% of oyster production (depending on the year and region of production) and accounting for huge and considerable economic loss for the oyster farmers.

In 2008-2009 major mass mortality events occurred affecting one year old *C. gigas* all over French coasts when seawater temperatures reached 17 °C (Bédier et al., 2009) and mortalities ranged from 40% to 100%. The first mortalities normally start in April in the Mediterranean Sea (e.g. Thau lagoon), then in the Atlantic coast (such Marennes-Oléron Bay and Bay of Brest). In the Mediterranean sea, in Thau lagoon up to 85% mortalities occurred during the summer of 2008 (Pernet et al., 2010); while in the French Atlantic coast (e.g. Marennes-Oléron Bay) Dégremont (2011) reported that the mortality in 2009 was up to 50%. Such mortality events were also reported in Ireland and later in New Zealand and Australia (Paul-Pont et al., 2013). To avoid the negative market impact of using the term “herpesvirus” the acronym POMS (Pacific Oyster Mortality Syndrome) was used to describe these mass mortalities (Paul-Pont et al., 2013).

1.6.2 Breakthrough in understanding the Pacific Oyster Mortality Syndrome (POMS)

The POMS is of complex aetiology and in recent years it has become panzootic. It is present in all the coasts of France and in other worldwide countries (EFSA, 2015). For better understanding the POMS, worldwide research has made efforts to elucidate this syndrome. First, researchers focused on the viral aetiology of the POMS, because of the recurrent detection of Ostreid herpes virus (OsHV-1) variants in populations examined (Martenot et al., 2011; Segarra et al., 2010). Therefore, to better understand and enable a fast detection of the OsHV-1, a series of diagnostic assays have been developed including the polymerase chain reaction PCR, quantitative qPCR and in situ hybridization (Corbeil et al., 2015; Martenot et al., 2011; Pepin et al., 2008; Renault et al., 2012; Renault, Tchaleu, et al., 2014). Second, studies suggested the involvement of other pathogenic agents highly related with viral infections. The study of Petton et al. (2015), showed that with antibiotic treatment and the absence of bacteria, the viral load itself is not enough sufficient to induce mortality. The presence of bacterial strains from the genus Vibrio were in particular important in the POMS disease (Bruto et al., 2017, 2018). However, most of these studies lacked in their design the reproduction of a realistic infection and passing through natural route which may explain that the complex disease process was still unclear in 2016.

The development of several breakthroughs helped in deciphering the complexity of the POMS. For example, the development of new method of infection (ecologically realistic) helped to better understand the complexity of the disease (B Petton et al., 2019; Bruno Petton et al., 2013). This method is based on pathogen-free oysters that since born are reared in a bio-secured setting. Then, half of these oysters are naturally infected in the field and labelled as donor. Then donors are re-joined to the other half in the bio-secured settings (called recipient oysters or receptors) leading to the natural transmission of the disease from donors to the recipients oyster through cohabitation. Thus keeping the complexity of the infectious environment, mimicking the natural process of infection (B Petton et al., 2019; Bruno Petton et al., 2013) and permitting simultaneous disease transmissions to all the recipients.

Another important advance was the use of biparental families of oyster with resistant and susceptible phenotypes against the virus (Azéma et al., 2017). Thus, during the pathogen challenge the dynamics of the POMS can be monitored in both distinct phenotypes. The Use of integrative molecular approaches is one of the last breakthrough in understanding and deciphering the POMS. Thus, allowing in one experimental framework the surveillance of the

dynamics of host response to the disease and the changes in the microbiota composition (including the putative pathogens). All the above-mentioned efforts led to better design the experimental framework and decipher the POMS disease.

1.6.3 Pathogens agents involved in POMS

1.6.3.1 What is POMS

The POMS results from the complex interactions between the oyster and its pathogens; the virus OsHV-1 µVar, and opportunistic pathogen bacteria from several genera such as *Vibrio Tenacibaculum* and *Marinomonas*. (Davison et al., 2005; de Lorgeril et al., 2018; Friedman et al., 2005; Renault et al., 1994; Renault & Novoa, 2004; David Schikorski et al., 2011). The study of de Lorgeril et al. (2018) deciphered the POMS mechanisms using an ecologically realistic model of infection, through susceptible and resistant oyster families with known phenotype, and integrative molecular approaches (including the histology, dual RNAseq, 16S rDNA metabarcoding). Authors established that POMS is a polymicrobial disease and the presence of OsHV-1 is the primary step for onset of infection. While the virus could infect both the resistant and susceptible oysters, an intense virus replication was only determined in the susceptible oysters. This intense viral replication is needed for the disease development. The virus attacks the immune cells (haemocytes) which affects their expression of AMPs leading to their action suppression against the surrounding bacteria which ultimately enable the development of a lethal secondary infection (de Lorgeril et al., 2018).

In summary POMS onset requires both OsHV-1 and opportunistic bacteria. Inhibiting either the viral replication with poly-IC injection or bacterial proliferation with antibiotics, hinders the infectious process and prevents oyster death.

1.6.3.2 Virus: the main agent triggering the POMS

The POMS disease starts once the OsHV-1 µVar virus infects the oysters. Then for the disease development, the virus replicates intensively and reaches oyster haemocyte. For example, in the study of Martenot et al. (2017) authors detected the replication of the virus OshV-1 in the oyster haemocytes after that the viral suspension was injected in the adductor muscle. The same authors showed an increase in the expression of viral genes coding for membrane proteins (Open reading frame; ORF 25, 41 and 72) and a gene coding for an inhibitor of apoptosis (ORF 87). Additionally, the interaction between the virus and the haemocytes have been studied by

exposing healthy haemocytes with viral suspension (Morga et al., 2017). The virus was detected after one hour of exposure, and viral transcripts increase rapidly up to 24 hours. However, enveloped virus particles were not detected in the haemocytes suggesting that virus replicates also in other tissues of the oyster and that haemocytes are not a suitable tissue for the production of infectious virus particles. Another study has detected viral transcripts in the gills, mantle and heart 26 hours post infection (Segarra et al., 2016). However, the quantity of viral transcripts was higher in heart by comparison to other tissues suggesting that the viral replication cycle could start in oyster heart and could reach different tissues.

Recently, the POMS disease was deciphered using biparental families of oyster using a combination of transcriptomic and metabarcoding approaches. During the POMS event, the virus is capable of infecting both susceptible and resistant oysters with intense replication in susceptible oysters. The resistant oysters are capable of an early antiviral response while the susceptible oysters respond strong but late, resulting in the incapacity to control the viral replication. The resistant oysters are able to effectively respond to the viral infection by the induction of antiviral genes and effectors involved in antiviral signalling pathways. In contrast, in the susceptible oysters, concomitantly with the strong replication of the virus, oysters intensively expressed numerous genes that encode endogenous Inhibitors of Apoptosis Proteins (de Lorgeril et al., 2018). Though the mechanism by which OsHV-1 is able to induce endogenous IAPs expression is still unknown. However such mechanism has been described in human Gamma herpesvirus and Epstein–Barr virus (EBV). For example, once the EBV has infected cultured human umbilical vein endothelial cells, the virus were capable of an increase in the expression of IAP-2 gene that inhibits the apoptosis, therefore protecting the virus from apoptosis (Xiong et al., 2004). Some viral components are able to prevent the apoptosis mechanism in human cells. For example, viral proteins of certain viruses (such as herpes-like viruses, adenoviruses or papillomaviruses) are able to inhibit host cellular proteins (such as P53 protein or the Fas cell surface death receptor (FAS)) involved in the initiation of apoptosis (Teodoro & Branton, 1997). Other proteins can activate cellular proteins that inhibit apoptosis, such as the Bcl-2 protein (Krajcsi & Wold, 1998). Finally, viral proteins can directly inhibit apoptotic mechanisms by expression of exogenous IAPs (Krajcsi & Wold 1998). Remarkably, during the POMS event, the intense OsHV-1 replication is also associated with the over expression of exogenous IAPs of viral origin (de Lorgeril et al., 2018). These viral proteins contain the Baculovirus Inhibitor of apoptosis

protein Repeat (BIR) domain known to have anti-apoptotic activities that favour viral replication (Miller, 1999).

Altogether these results suggest that OsHV-1 virus could intensively replicate in the susceptible oysters by rendering their apoptosis mechanism through endogenous and exogenous anti-apoptotic processes (de Lorgeril et al., 2018). Concomitant with this, the virus attenuates the oyster immune cells (haemocytes), which it influences the haemocytes physiology and blocks the expression of AMPs either by a transcriptional regulation (directly) or by the induction of cell death or lytic processes (indirectly) (Martenot et al., 2017).

The increasing of Next Generation Sequencing (NGS) data in Pacific oysters (Abbadi et al., 2018; Bai et al., 2019; E. A.V. Burioli et al., 2017; Erika A.V. Burioli et al., 2018; Segarra et al., 2010) revealed the genetic diversity of the OsHV-1 µVar genotypes, which first detected in 2010 (Segarra et al., 2010). This has raised the question of what is the impact of the genetic diversity on the fitness of the virus and its consequence on the disease. Viruses can produce various genetically linked mutants (variants or also called as the viral populations). These viral populations are preserved by mutation-selection balance (Perales et al., 2015; Poirier & Vignuzzi, 2017), and have the possibility of generating a beneficial interactions that help in viral fitness and adaptability to its host (Brooke, 2017; Mao et al., 2007; Pfeiffer & Kirkegaard, 2005). In the context of POMS disease and OshV-1, the possibility of having distinct viral populations have been studied recently (Delmotte et al., 2020). The authors confronted different biparental families of oysters to different infectious environments. The study revealed a distinct viral populations of OsHV-1 associated to POMS in two distinct infectious environment. Moreover, within each environment, distinct viral populations were associated to different oyster families.

1.6.3.3 Bacteria: viral infection enable bacteraemia by opportunistic bacteria

Following infection of the haemocytes by the virus, the opportunistic bacteria colonise the tissues of susceptible oysters (de Lorgeril et al., 2018). So far, most studies used culture-based methods for identifying the bacterial component of POMS disease (Bruno Petton et al., 2021a). These studies showed a distinct association between bacteria belonging to Vibrio species and the occurrence of the POMS.

Such study efforts led to a broad characterization of the role and contributions of the Vibrio in the event of oyster mortality. For example, Bruto et al. (2017) characterised the population structure of Vibrio species found during a POMS episode using pathogen-free oyster spats and field based approaches. During the POMS event, in the diseased oysters different bacteria from the Splendidus clade have been systematically isolated (e.g. *Vibrio tasmaniensis, V. splendidus, V. cyclitrophicus, V. harveyi, V. aestuarianus, and V. crassostreae*) (Bruto et al., 2017; Saulnier et al., 2010; Segarra et al., 2010). Furthermore, Bruto et al. (2017) revealed that the Vibrio population structure is different according to seasons and contrasts in oysters affected by POMS. However, even in healthy oysters (in case of no mortality) Vibrio of the Splendidus clade are present (Saulnier et al., 2010) but it expresses its pathogenic potential very rarely (Oyanedel et al., 2020). Remarkably, *V. crassostreae* is predominant during mortality event (Bruto et al., 2017) and was shown to replace the resident Vibrio community during a POMS episode (Lemire et al., 2015). Interestingly, experimental infection in different experiments enabled the identification of factors contributing to Vibrio virulence. These factors are mostly found in the two Vibrio species (*V. crassostreae* and *V. tasmaniensis*; a facultative intracellular pathogens of oyster haemocytes) (Duperthuy et al., 2011). These two species have been isolated from the Atlantic region during POMS events (Bruto et al., 2017, 2018; Duperthuy et al., 2011; Lemire et al., 2015; Rubio et al., 2019; Vanhove et al., 2016). For example, Bruto et al. (2018) revealed that in *V. crassostreae*, the R5.7 gene is needed for the virulence and is ancestral within the Splendidus clade. As the R5.7 is not cytotoxic (Bruto et al., 2018), it only mediates the cytotoxicity by their physical contact with the haemocytes (Rubio et al., 2019).

In contrast, *V. tasmaniensis* has acquired a type VI secretion system (T6SS; upon their loss of ancestral R5.7 gene) by which intracellular cytotoxic effectors can be delivered to haemocytes. Altogether, the Vibrio cytotoxicity is a key factor of oyster colonization. This helps Vibrio to get away from cellular defence mechanisms leading to a systemic infection (Rubio et al., 2019).

Recently, using a 16S metabarcoding and a metatranscriptomics approaches, Lucasson et al. (2020) studied during an experimental POMS the structure of bacterial communities and functions expressed by different bacterial genera, respectively. In total, the study found five bacterial genera (Arcobacter, Marinobacterium, Marinomonas, Vibrio, and Pseudoalteromonas) which colonize oysters during POMS. These five genera (referred as POMS core pathobiome) were remarkably consistent between different oyster biparental families submitted to different infectious environment.

Altering the oyster's haemocyte physiology (Inhibition of antimicrobial peptide expression) by the OsHV-1 is the key determinant of the successful bacterial colonization. Without the OshV-1, the Vibrio fails to colonize the oysters and to express their pathogenic potential (de Lorgeril et al., 2018).

1.6.4 Factors involved in POMS permissiveness

Study efforts have shown that risk of POMS outbreaks is linked to subtle interactions, not only the oyster (host) and the pathogens (polymicrobial), but also environmental factors. Therefore, POMS is considered a multifactorial disease (Bruno Petton et al., 2021b). Factors that are known to modulate the POMS outbreaks are: i) Microbiota, 2) permissive factors (age of oysters and water temperature and food availability).

1.6.4.1 Microbiota

Previous studies showed that viral infection cause changes in the structure and diversity of the bacterial communities. These changes are a consequence of the stress and probably of the immune suppression. Many studies have observed a change in the microbiota composition during the oyster mortalities. For example, in the study of (Lokmer & Wegner, 2015), authors showed that thermal stress can disrupt microbial associations and lead to the development of pathogenic opportunistic bacteria. Another study has shown that transferring oysters (non-treated with antibiotics) to different environment increases mortalities, thus, suggesting that the change in bacterial communities in transferred and untreated oysters may be a consequence of interactions between the resident and external microbiota (Lokmer et al., 2016).These studies highlight the important role of the resident microbiota prior to a stress event.

Many studies have investigated the role of bacterial microbiota using 16S metabarcoding. During the POMS, studies have identified alterations in the composition of microbial community

(referred as dysbiosis). For example, the large-scale study of Lasa et al. (2019) analysed the microbiota of diseased oysters from three different sites in Europe (France, Ireland and Spain). This study identified dysbiosis in oysters infected with OsHV-1 virus and was characterized by the emergence of the pathobiota of opportunistic pathogens including Vibrio and Arcobacter species. Furthermore, in an integrative and ecologically realistic study showed that the viral infection alter the antibacterial defence system of oysters that subsequently results in the dysbiosis and colonization by opportunistic bacteria (de Lorgeril et al., 2018).

Interestingly, a correlation between microbial community composition and genetic relatedness of the oysters. Suggested a genotype specific composition of the microbial communities living in oyster (Wegner et al., 2013). Another study that used a set of 35 families of healthy oysters with different levels of susceptibility to POMS revealed a fundamental variance in composition of microbiota community (King et al., 2019). In this last study, the most susceptible oysters were significantly associated with the Operational Taxonomic Units (OTUs) of the *Photobacterium*, Vibrio, Aliivibrio, Streptococcus, and Roseovarius genera. Furthermore, using Vibrio-specific qPCR assay, they found significant increase in Vibrio load in disease-susceptible families.

More recently, the study of (Clerissi et al., 2020) found an association between the *Mycoplasmataceae*, *Rhodospirillaceae*, *Vibrionaceae* and *Photobacterium* genera and the low resistant oyster families. More interestingly, oyster families that survived in the field to the infectious period of POMS showed higher proportion of specific taxa including *Cyanobacteriaceae*, *Colwelliaceae*, and *Rhodobacteraceae*. In addition, POMS resistant oyster families revealed higher evenness of their microbiota, suggesting that opportunistic pathogens could colonise easier oyster with low microbial diversity (Clerissi et al., 2020).

1.6.4.2 Host and environmental factors influencing POMS

1.6.4.2.1 Age

Several studies have investigated the role of oyster age on mortality. Firstly, studies have reported that adult oysters have lower rates of POMS-induced mortality compared to spat and juvenile oysters (Peeler et al., 2012; Pernet et al., 2012; Bruno Petton, Boudry, et al., 2015). However, the experimental designs of these studies lack the ability to disentangle the age effect from the selection for the resistance. It is probable that adult oysters used in above-mentioned studies could have gone through POMS event (at least one POMS event) and eventually selected for resistance to POMS.

Therefore, to study the role of oyster age on mortality, oysters that never encounter the POMS (juveniles and adults) were subjected to the Marennes-Oléron Bay (France) infectious environment (Dégremont, 2013a). The study revealed that resistance to mortality increased with oyster age. Another study showed similar results in the same infectious environment (Azéma et al., 2017). In addition, these results were confirmed in another infectious environment (Brest Bay, France; Bruno Petton, Boudry, et al., 2015), where they confronted simultaneously oysters of different ages that never encounter the POMS to a similar infectious environment. The results showed an increase in survival for older oysters and acquiring the resistance after 24 months (Azéma et al., 2017; Bruno Petton, Boudry, et al., 2015). The mechanisms are currently investigated (ANR DECICOMP) but one of the hypothesis concern the advanced maturation of the immune system.

1.6.4.2.2 Water temperature

Temperature is another important factor correlated with oyster mortalities. One of key drivers of disease outbreaks is the increase in temperature (Harvell et al., 2002b). This increase is associated with climate change. In many invertebrates, temperature influences the host and/or pathogen physiology therefore affecting the outcome of interactions between the host and pathogen (Ben-Haim et al., 2003; Ittiprasert & Knight, 2012; Kimes et al., 2012; Vidal-Dupiol et al., 2011, 2014).

In the case of POMS disease, temperature is considered a key element in the infectious process and mortalities of oyster is associated to a window of permissiveness comprised between 16°C

and 24°C (Lionel et al., 2013; Pernet et al., 2012; Bruno Petton et al., 2013) (ECOSCOPA, 2021; E Fleury et al., 2020; Pernet et al., 2014).

The molecular mechanism by which temperature affects oyster mortalities is not yet elucidated, numerous studies have shed light on this link. For example, temperature could favour the activation of pathogens. Pernet and his colleagues revealed that the virus persists in oysters at low temperatures (10°C and 13°C) and that it is reactivated when the temperature is increased to 21°C (Pernet et al., 2015).

Similarly, in an experiment involving cohabitation between oysters exposed to infectious environment in the field, and unexposed oysters, Petton and his collaborators demonstrated that the transmission of the virus is achieved between 16°C and 22°C. Interestingly, no viral transmission occurred at 13°C (Bruno Petton et al., 2013). As for the age factor, the mechanisms behind this permissivity is currently studied through the DECICOMP project.

1.6.4.2.3 Food availability – growth – energy contents

The availability of food improve the physiological status of the host and by reduce the risk of infectious disease (Lochmiller & Deerenberg, 2000; Sheldon & Verhulst, 1996). On the other hand, a restricted food availability, slows down the growth and metabolism of the host, therefore limiting the resources needed for the virus (Ayres & Schneider, 2009; Civitello et al., 2018; Hall et al., 2009; V. H. Smith et al., 2005). Previous studies on oyster have revealed that starved oyster have lower risk of death compared to fed oysters (Evans et al., 2015; Moreau et al., 2015). Contradictory, other studies have shown that the food availability and increased energy content is associated with improved resistance or tolerance to OsHV-1 (Pernet et al., 2012, 2014, 2018). Further study by (Pernet et al., 2019) experimentally investigated the role of the food availability growth rate and energy content factors on oyster mortality risk. The study revealed that food availability and oyster growth were associated with a higher risk of mortality. However, energy content associated with a lower risk of mortality (Pernet et al., 2019). This phenomenon must be viewed as a fine balance between factor increasing resistance (high energy reserve and starving) *versus* factors increasing susceptibility (high food quantity and high metabolic activity). As for the age and temperature the mechanisms behind this permissivity is currently studied through the DECICOMP project.

1.6.5 Genetic factor associated to resistance

Oyster are benthic species living in open environment with constant fluctuations of its surroundings environment. The livestock loses is one of important problems caused by the OsHV-1 virus. So far, there is no treatment or control applicable in the natural open environment when the POMS occurs (e.g. no option for vaccination). In order to manage the oyster farming losses induced by the POMS, different strategies have been developed to reduce mortality. One of these strategies to improve the survival of oyster is through selective breeding programs. For example in France since 2001, large IFREMER project called ‘‘MORtality ESTival’’ (MOREST) was initiated and one of its objectives was selective breeding for better survival. Other selective program were carried on in different regions (see review; (Dégremont, Garcia, et al., 2015)). The MOREST program has resulted in a significant improvement in the resistance of oysters; the resistant families of oysters to OsHV-1 infection had a significant differences to the susceptible families in terms of survival (Lionel Degremont, 2003). The resistant phenotypes were further observed in the later generation, these resistant oysters were described in field and experimental challenges, the narrow sense heritability of survival ranged from 0.21 to 0.60 (Azéma et al., 2017; Camara et al., 2017; Dégremont, 2011, 2013b; Dégremont, Lamy, et al., 2015). These findings were based on selection programs produced by selecting families with desired phenotypes, which needs space, time and is costly.

Mass Selection (MS) can offer a simpler and less expensive way to select a desired trait (e.g. resistance or growth). In MS individuals are selected from a group of population without completely accounting for the family structure. In oysters, the MS approach has been implemented by selecting wild oysters surviving the OsHV-1 in the field and by breeding them to produce oysters’ lines with higher resistance to the disease. These mass selection program over four generations of selection indicated a gain in survival of 22%, 44%, 50% and 62%, respectively (Dégremont, Nourry, et al., 2015). These results suggested a positive response to selection and a gain in resistance to OsHV-1, further indicating a genetic basis of resistance to OshV-1 infections. Based on the line and the size of oysters when challenged with OsHV-1, the narrow sense heritability of survival ranged from 0.34 to 0.63 (Dégremont, Nourry, et al., 2015).

Although, the selective breeding offers a great deal of improving the trait of interest (here the resistance to OsHV-1), two main post selection limitation can arise (Sauvage et al., 2010). First, the selective breeding focuses on identifying the phenotype (here the dead vs alive phenotypes)

that is based on a set of oyster from a specific family (families) with potential limited genetic background, limited spatial condition and limited timing of the experimental infection in the field (Dégremont et al., 2005). Second, it is absolutely necessary to better understand the cause of mortality to identify the underlying factors behind the mortality. The phenotyped oysters need to be closely monitored, which in return would allow describing the physiological (from host and pathogen) and immunological status (Sauvage et al., 2010).

However, these limitations in selective breeding programs can be overcome by implementing marker-assisted selection (MAS) and genomic selection (GS), trough the aid of genetic markers such as Quantitative Trait Loci (QTL) and Single Nucleotide Polymorphisms (SNPs). MAS can be used for the traits with few SNPs or QTLs that have a large effect on the trait. On the other hand, the GS can be applied for the traits with polygenic nature that is controlled by many SNPs and QTLs that each have little effect (Nascimento-Schulze et al., 2021). With the availability of genomic resources and tools for Pacific oyster such as SNPs (Elodie Fleury et al., 2009; Sauvage et al., 2007) and microsatellites (Sauvage et al., 2009), it become possible in 2010 to develop the first QTL analysis and to identify five putative QTLs associated to survival and OsHV-1 load. These QTLs were spread out in linkage groups V, VI, VII and IX (Sauvage et al., 2010). However, the genes behind these QTLs were not evaluated. In addition, the study had a moderate accuracy rates due to the use of a low number of markers.

The availability of oyster reference genome and SNP arrays enabled the investigation of the genetic architecture of traits of interest (Gutierrez et al., 2017; Lapègue et al., 2014; Qi et al., 2017; G. Zhang et al., 2012). More recently, the updated chromosome level assemblies provide valuable resources for further genomic research (Peñaloza et al., 2021; Qi et al., 2021). These updated genomic tools can further facilitate and expands the study of the genetic underpinnings of *C. gigas* resistance to OsHV- 1 infection.

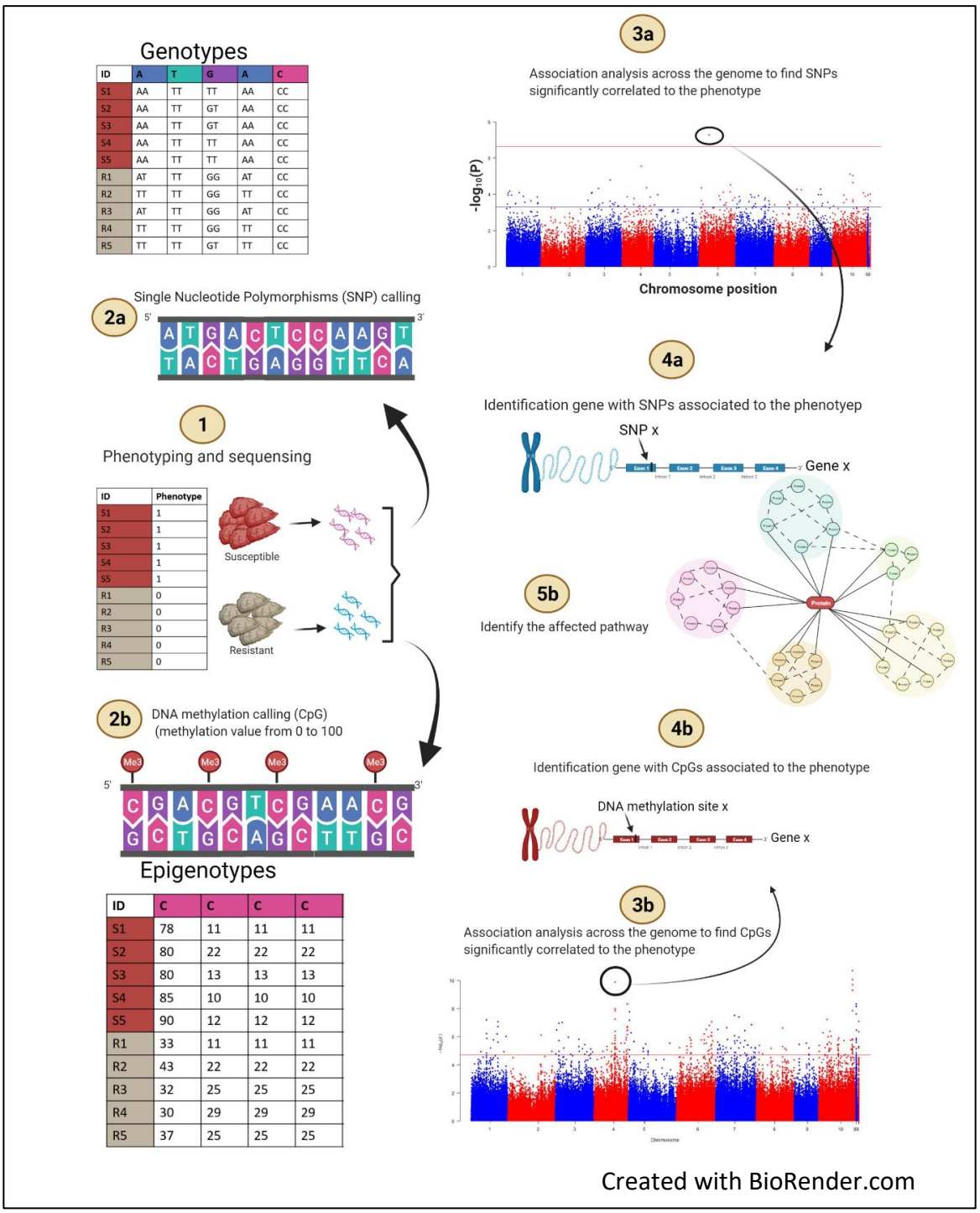
Genome-wide association studies (GWAS) are powerful approaches to study the genetic architecture of traits of ecological and economic importance (see Figure 1.12; for principles of GWAS). So far, several GWAS studies have been implemented in Pacific oyster, these studies were on shell growth, salinity adaptation and glycogen content traits (She et al., 2018; Meng et al., 2020; He et al., 2021). GWAS evaluates genomic regions associated to specific traits (such as resistance phenotype). For example, benefiting from the newly developed SNP array, a study was performed to further augment the accuracy of heritability estimation and identification of

genomic regions associated with survival and/or viral load (Gutierrez et al., 2018, 2020). The first study showed a significant but low to moderate estimates of heritability (range from 0.12 – 0.25) compared to previous heritability estimation (that ranged from 0.21 to 0.63). On the other hand, several significant and suggestive SNPs were identified and further located in or near several genes. However, the function of these genes is not perfectly understood (Gutierrez et al., 2018). In a second study, the authors identified more accurate heritability estimation using genomic prediction than rather pedigree prediction. The heritability estimation ranged from 0.25 to 0.37 for genomic prediction and pedigree prediction respectively (Gutierrez et al., 2020). Interestingly, these studies suggested a polygenic nature of Pacific oyster resistance to OsHV-1 (Gutierrez et al., 2018, 2020; Sauvage et al., 2010).

More recently, a basal transcriptomic study on naïve oysters (without infection state) from biparental families displaying contrasted level of resistance and susceptibility revealed differences in the basal expression of genes involved in stress response, protein modification, immune and antiviral pathways. Additionally, resistant families showed resemblances but also clear difference between different molecular pathways (de Lorgeril et al., 2020). All these results suggest that resistance phenotypes depend on several genes (polygenic) and it is likely that resistance to OsHV-1 µVar relies on several non-exclusive mechanisms.

However, the underlining determinants still remains unknown and need further investigations. Nevertheless, the common point between all the mentioned studies that tried to identify the heritability of the phenotypes that it was identified using the Mendelian laws. Recently, is more accepted that genetic and non-genetic could be involved in explaining phenotypes. Moreover, we know that GWAS studies have not been able to explain heritability estimation and that it explained only a small parts of phenotypic variance, a phenomenon known as “missing heritability” problem (Banta & Richards, 2018; Maher, 2008) (explained in the next section). Interestingly, and by opposition to genomic prediction, heritability values combine the effect of all the phenotypic determinants from which it is therefore impossible to disentangle the genetic and epigenetic effect (Banta & Richards, 2018).

Altogether, this suggests that other mechanisms could be important in explaining the resistant of the oyster to POMS. Many studies revealed that the epigenetic (DNA methylation) in particular would be taken into account as an inherited factor of phenotypic variation.



ID	A	T	G	A	C
S1	AA	TT	TT	AA	CC
S2	AA	TT	GT	AA	CC
S3	AA	TT	GT	AA	CC
S4	AA	TT	TT	AA	CC
S5	AA	TT	TT	AA	CC
R1	AT	TT	GG	AT	CC
R2	TT	TT	GG	TT	CC
R3	AT	TT	GG	AT	CC
R4	TT	TT	GG	TT	CC
R5	TT	TT	GT	TT	CC

ID	Phenotype
S1	1
S2	1
S3	1
S4	1
S5	1
R1	0
R2	0
R3	0
R4	0
R5	0

ID	C	C	C	C
S1	78	11	11	11
S2	80	22	22	22
S3	80	13	13	13
S4	85	10	10	10
S5	90	12	12	12
R1	33	11	11	11
R2	43	22	22	22
R3	32	25	25	25
R4	30	29	29	29
R5	37	25	25	25

Figure 1.12: principle of Genome/Epigenome wide association studies (GWAS/EWAS).

It involve several steps: 1) Phenotyping and sequencing samples (here exome capture with bisulfite conversion sequencing to obtain the genetic and epigenetic data). 2a) SNPs (Single Nucleotide Polymorphisms) calling and genotypes file preparation. 2b) CpGs (DNA methylation at CG context) calling and Epigenotypes file preparation 3a) association analysis between the phenotype and the SNPs represented by a Manhattan plot with pvalue as negative log10 for each SNPs. 3b) as for 3a but with CpG methylation level. 4a) Identification of genes displaying the most significant SNPs (or suggestive SNPs). 4b) as the 4a with CpG. 5) gene enrichment analysis and pathway identification.

1.6.6 Missing heritability

During the last two decades, many studies aimed at calculating the proportion rate of a heritable trait. They provided a valuable information about the genetic inheritance of phenotypes. The heritability can be defined as the total rate of a phenotypic variations explained by additive genetic factors. Such kind of studies have been implemented in oyster, and they estimates the genetic basis of resistance phenotype to POMS. Survival against OsHV-1 infection have been shown to display a significant additive genetic component and a heritability values ranging from 21% to 63% (Azéma et al., 2017; Dégremont, Nourry, et al., 2015).

Research in genetics has been able to make a great jump forward, especially with the era of new technologies, mass sequencing and genome assembly studies. GWAS are powerful tools to studying the genetic architecture of traits, which evaluates genomic regions responsible for important traits (such as resistance phenotype) by associating genetic variants (SNPs) to the phenotype of interest.

Many genetic variants have been identified by GWAS analysis. However, there is a mismatch between the heritability studies and GWAS studies. For example, studies of genetic variants linked to different traits (such as height, autism or schizophrenia) did not explained 100% of the heritability (Maher, 2008; Trerotola et al., 2015). Studies have identified many SNPs, but they only explained small parts of the phenotypic variance. Thus, the unexplained rate of phenotypic variation is called the “missing heritability” (Figure 1.13) (Maher, 2008).

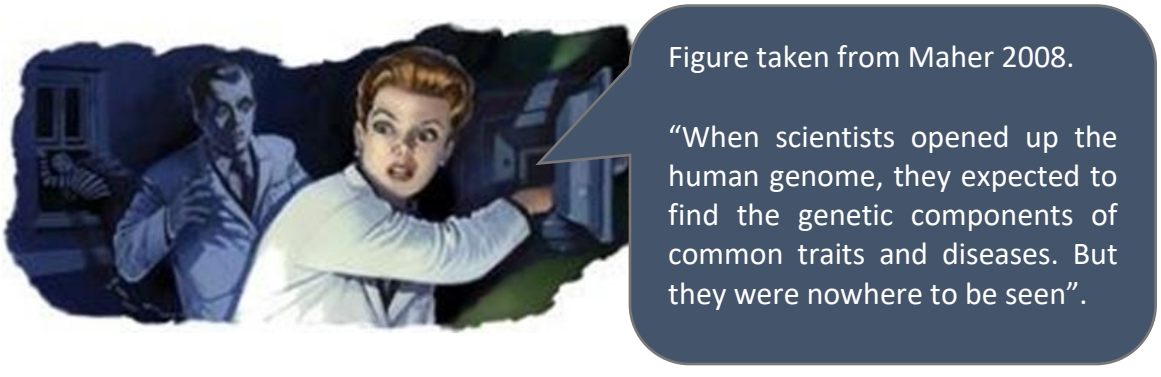


Figure 1.13: The Missing heritability case of human genome.

Thus, leading many authors to call for a revision of evolutionary theory to incorporate, in addition to genetics, other mechanisms that may play a role in the establishment of the phenotype and its transmission between generations (Danchin & Wagner, 2010; Jablonka & Noble, 2019; Laland et al., 2014; Pigliucci, 2007). Since the 90th several experiments renew the Lamarckian inheritance hypothesis: *e.g.* the characters acquired during the life of an individual can be inherited. Thus, give rise to the new modern synthesis proposing that both the genetic (DNA) and non-genetic processes (non-DNA) can be inherited and therefore are fuel for evolution (R. Bonduriansky, 2012; Danchin et al., 2011; Danchin & Wagner, 2010). Different terms have been proposed to combine the genetic and non-genetic heritability, "general heritability" (Mameli, 2004), "inclusive heritability" (Danchin & Wagner, 2010) and "'inheritance system" (Cosseau et al., 2017); these terms are synonymous and aim to encircle all elements of the inheritance. Non-genetic inheritance can be mediated through different interacting mechanisms such as epigenetics, cultural, ecological and parental effect (Danchin, 2013). Epigenetic have been proposed to be a solution for the missing heritability (for a review see Banta & Richards, 2018).

1.6.7 Evidence of epigenetic association to POMS

Studies have shown a link between the expression of traits that are commercially important in aquaculture and epigenetic mechanisms (Gavery & Roberts, 2017). Epigenetic acts as a hub that enables an organism to cope with the environmental changes that need fast response giving enough time for the genetic adaptation to happen (Banta & Richards, 2018; Russell Bonduriansky & Day, 2009; Geoghegan & Spencer, 2013; Klironomos et al., 2013; Kronholm & Collins, 2016; Torda et al., 2017).

Several studies have reported the association of DNA methylation to complex traits of human disease such as Asthma, Diabetes and Alzheimer (Edris et al., 2019; K. Guo et al., 2020; van den Hove et al., 2020). It has been well documented in an increasing number of studies, showing that DNA methylation could integrate environmental changes to cope with the new context. These DNA methylation changes may affect the transcription process (gene expression) which lead to phenotypic variation (C. L. Richards et al., 2017; Xu et al., 2019).

When subjecting oysters to heat-stress treatment, a positive correlation between the DNA methylation in the gene bodies and gene expression was determined. This finding suggests a divergence in the phenotype that is facilitated by the DNA methylation (Xinxing Wang et al., 2021). Abiotic factors could also induce a global change in DNA methylation in the Pacific oyster such as the salinity and diuron exposure (Rondon et al., 2017; Xin Zhang et al., 2017). Biotic factor as the microbial environment have also been shown to impact gene expression and DNA methylation profile of immune related genes (Fallet et al., 2022).

Overall, there is some evidence illustrating that epigenetic can be an important factor in resistance gain against the POMS.

1.7 Objective of the PhD

In summer 2008, a new OsHV-1 variant the OsHV-1 µVar was responsible for a major mass mortality event that have affected one year old *C. gigas* all over French coasts when seawater temperatures were about 17 °C. The acronym POMS (Pacific Oyster Mortality Syndrome) was used to describe these mass mortalities that is due to a polymicrobial and multifactorial disease that recently it become panzootic.

Research efforts have identified that resistance of *C. gigas* to POMS is associated with early antiviral response to the viral infection. Depending of the studies, heritability values for oyster resistance were shown to range between 12% to 63% and displayed a significant additive (epi)genetic component. While the data about the involvement of epigenetic in POMS resistance are still scarce, some studies focusing on the oyster's genetic determinants were published.

Genome-wide association studies (GWAS) have already been used to identify SNPs significantly associated to oyster resistance but only few genes with an evasive function have been identified. In addition, GWAS study implemented on naturally occurring populations of oyster is still missing to address this question in an ecological context.

Similar to GWAS, epigenome-wide association studies (EWAS) have been widely used in model organism to identify DNA methylation patterns associated with a particular phenotype (Figure 1.12). Many studies showed that the epigenotype should be considered as a factor associated to phenotypic variation (see review Gavery & Roberts, 2017). It is therefore rational and necessary to consider the genotypes and epigenotypes as a material on which the natural selection can act to shape a phenotype that could be transmitted (at least in part) to the offspring. Unfortunately, there is lack of studies assessing the role of epigenetic mechanism in oyster resistance to POMS.

Here we propose a framework to study simultaneously the potential role of genetic and epigenetic in shaping a phenotype by using the *C. gigas*/POMS model at the natural population level.

In this thesis, we hypothesize, that genotype and epigenotype of oyster can play a role in the resistance to POMS and we propose to test this through several operational objectives:

1- Developing an exome capture approach to obtain genetic (SNP) and epigenetic (DNA methylation) information.
2- Sampling natural oyster population with contrasted exposure to POMS and from two geographic scales
3- Phenotype all the sampled oyster through an experimental infection.
4- Identify genetic and epigenetic signatures associated to POMS resistance through GWAS/EWAS.
5- Identify the biological pathways affected by these markers.
6- Weight the genetic and epigenetic effect on phenotypic variation.
7- Disentangle the genetic and epigenetic role in the explanation of the phenotype.

In the second chapter of thesis, I will explain further the protocol used for the exome capture that integrate a bisulfite conversion technique that allow to capture simultaneously genetic (SNP) and epigenetic (DNA methylation) information. This approach needed bench and bioinformatics optimization but finally enabled us to deeply characterize at the population level the genetic (SNPs) and epigenetic (DNA methylation) information.

In the third chapter, I will present the sampling and phenotyping, the GWAS/EWAS analysis and the statistical approaches used to weight and disentangle the genetic and epigenetic effect. This chapter correspond to an article that will be submitted to "**Science of total environment**" journal.

In the fourth Chapter, I will discuss the involvement of the results produce for the fields of Oyster/POMS and Genetic and epigenetic.

Finally, in the last and fifth chapter, I will draw the general conclusion of this work.

Chapter 2

Exome capture

From the Bench to Bioinformatics optimization

Chapter 2 : Exome capture: From the bench to bioinformatics optimization

Context and objectives

Since 2008, with the emergence of a new variant (OsHV-1 µVar), massive oyster mortality events have been reported. These events affected one year old *C. gigas* all over French coasts when seawater temperatures were about 17 °C (Bédier et al., 2009). The acronym POMS (Pacific Oyster Mortality Syndrome) was used to describe these mass mortalities. POMS is polymicrobial and multifactorial disease and recently it become panzootic (for review see Bruno Petton et al., 2021). Understanding the mechanisms behind the resistance and the susceptibility of oysters appeared therefore of fundamental importance.

Recently, using a holistic approach the POMS disease was deciphered (de Lorgeril et al., 2018). This study was based on biparental families with contrasted phenotypes, experimental infection mimicking the natural route and developing a comprehensive molecular analyses of host responses with the characterization through time the microbiota structure and pathogen. Thanks to this last study using the transcriptomics analysis, the authors found that resistant families display strong and early antiviral response to POMS that limit the viral replication (de Lorgeril et al., 2018). Interestingly, the transcriptomic determinant seems to be already but only partly present under basal condition (de Lorgeril et al., 2020). However the phenotypic determinants behind this basal and early response are not yet known.

In this context the work of this thesis aimed at studying two different mechanisms governing phenotypic expression, the genome and the epigenome. This would enable to further understand this complex disease by identifying the genetic (SNPs) and epigenetic (DNA methylation variations) markers of resistant to POMS. Since our aim to work at the population level and to bring information with a functional interest, we had to find a method enabling to deal with each of this characteristic. Analysis of hundreds of individuals are required to perform large scale omics population studies and it seriously impacts the cost of the experiment. In this sense, cost effective reduced representation approaches such as epiGBS, RRBS are generally used for population studies but these necessary approaches led to lower

resolution than whole genome approaches. A previous epiGBS study performed on the fresh water snail *Biomphalaria glabrata* reduced the study of cytosine methylation to 1% of all the CpG (Luviano et al., 2021). This low amount of covered CpG would not have been suitable in our study since this would have reduced the identification of strong selection signatures. To circumvent these limitations we developed a whole exome capture experiment that enabled to characterize most of the exonic regions of the oyster genome and therefore a significant part of the genome and epigenome. The exonic regions are the functional unit of the genome that encode proteins and were DNA methylation changes occur.

Therefore, SeqCap Epi Enrichment System (Wendt et al., 2018) protocol was used. This protocol combines the exome capture along a bisulfite sequencing method, which allows us to have the genetic information along the epigenetic information within exonic regions.

This second chapter is dedicated to the implementation, evaluation and optimization of the exome capture method "SeqCap Epi Enrichment System" (from the bench to the bioinformatics) on Pacific oyster samples collected during my thesis. Pacific oyster is a non-model organism and it was the first time where this method was used. In this chapter, we will focus mostly on the steps that needed to be optimized. I present this chapter as journal article style (not for publishing), starting by giving an introduction for methods to study the DNA methylation, then I further explain the optimized steps, I present the results and end by a discussion about these results.

1. Introduction

DNA methylation is a modification induced by the addition of a methyl group to the 5th carbon of a Cytosine to form 5-methyl- cytosine (5mC). It is one of the several epigenetic mechanisms with an important role in regulation of gene expression, development, response to stress and environmental changes (P. A. Jones, 2012; Z. D. Smith & Meissner, 2013; Tirnaz & Batley, 2019; Verhoeven et al., 2016). Several methods with various principles have been used to study the DNA methylation at different resolution, sensitivity and cost (Cazaly et al., 2019; Fallet et al., 2020; Kurdyukov & Bullock, 2016). These methods are able to determine the organism's pattern of methylation and quantify the DNA methylation at the global or site-specific level. These methods are clustered into three main groups (Pajares et al., 2021). Including the anti-5mC antibodies techniques (MEDIP), methyl sensitive restriction assays (MSAP), and Whole genome bisulfite sequencing approach (WGBS).

The gold standard method is based on the bisulfite treatment of genomic DNA (Kurdyukov & Bullock, 2016). When treating the DNA with sodium bisulfite it converts only the unmethylated cytosine to uracil while the methylated cytosine remains unchanged. Thus, allowing for identification of the DNA methylation at base resolution. Bisulfite sequencing (BS-Seq) has high resolution and sensitivity. However, the cost of BS-Seq remains very high due to its sequencing efforts (especially for large genome as *C. gigas*; the need for high read depth for accurate estimation of DNA methylation). One way to reduce the costs is to combine it with other techniques. For example, method like reduced representation bisulfite sequencing (RRBS), which can be used for enrichment of CpG regions within genome by implementing restriction enzymes that recognize and cuts the CGs sequences (Meissner et al., 2005). Furthermore, exome capture methods would be an interesting solution for reducing the cost of sequencing and focus sequencing on regions of interest (Hodges et al., 2007; Wendt et al., 2018).

This last method is also called target capture, which offers a mean to focus the depth of coverage toward the targeted regions (for example exons) rather than studying the whole genome. The principle of this enrichment method is to fragment the genomic DNA to small fragments of 200 bp. Fragmented genomic DNA is then amplified and targeted DNA sequences (exons) are captured using oligonucleotide complementary probes. These probes are usually biotinylated, allowing them to be attached to streptavidin-coated magnetic beads thus the

targeted DNA/probe/bead complex can be recovered using a magnet rag. This approach relies on prior knowledge about the reference genome of the organism studied to design the desired set of probes. Enrichment will then allow the elimination of off-target regions (in our case most of introns and intergenic regions that do not carry coding information that contain regulatory roles) and maximize the sequencing depth of the regions of interest. One of the very important gain brought by this approach relies on the possibility to significantly increase the number of samples studied in the same time, thus, allowing to perform the study at the population level.

The exome capture technology can be integrated with BS-seq with at least using two workflows; i) capture then convert and ii) convert then capture (implemented by ROCHE) (Wendt et al., 2018). The first workflow comes with the advantage of preparing fewer number of probes for capturing the region of interest. Additionally, it allows for higher capture specificity. However, it comes with several disadvantages; first, small amount of molecular DNA is recovered for sequencing, because the libraries are first captured then bisulfite treated and amplified. The Bisulfite treatment is a very harsh process that can damage up to 90% of the DNA fragments. Altogether, results in a highly redundant and more readily biased dataset, which can affect the accuracy of DNA methylation calling (Wendt et al., 2018). In the second workflow, these problems are avoided, by adding a pre-capture amplification step before the capture, which help to diminish the diversity loses without affecting the DNA methylation accuracy calling. However, this improvement comes with a huge challenge in probe design.

The SeqCap Epi Enrichment System, uses the second workflow and has overcome these last challenges by exploiting a design algorithm that takes into account five conditions of methylation per Cytosine (0%, 25%, 50%, 75% and 100% of methylation) and it takes into the account all the possible condition of methylation at cytosine nucleotides (Figure 2.1) (Wendt et al., 2018). This protocol has been successfully implemented in both plant and mammals (Allum et al., 2015; Q. Li et al., 2014, 2015).

We decided to implement this protocol in Pacific oyster for two main reasons. First, like for other invertebrates the DNA methylation in oyster occurs preferentially in exons and introns (Rondon et al., 2017). Second, with this approach we combine the power of reduced representation in terms of number of samples sequenced and the power of WGBS in the terms of functional information in the coding regions. Additionally, as in the WGBS, the Illumina

sequencing of SeqCap Epi Enrichment system libraries generate sequencing reads that contains the genetic data (SNPs) in addition to the DNA methylation information (Lea et al., 2017). However, the BS-seq data should be carefully handled when calling the genetic data, because the typical SNP-calling software can confuse between real C to T SNPs and the one produced from bisulfite conversion. It is feasible to obtain the accurate genetic information (SNPs), with aid of several developed packages that can overcome this challenge (Barturen et al., 2013).

In summary, I used the SeqCap Epi Enrichment System (Wendt et al., 2018) for my thesis for its ability to produce in functional sequences (e.g. exon) usable data for the study of genetic and epigenetic information on a high number of samples at a reasonable cost. To set up this experiment in oysters several steps were optimized before all the phenotyped oysters (Figure 2.2). This optimization steps include the bench work and bioinformatics pipeline optimizations.

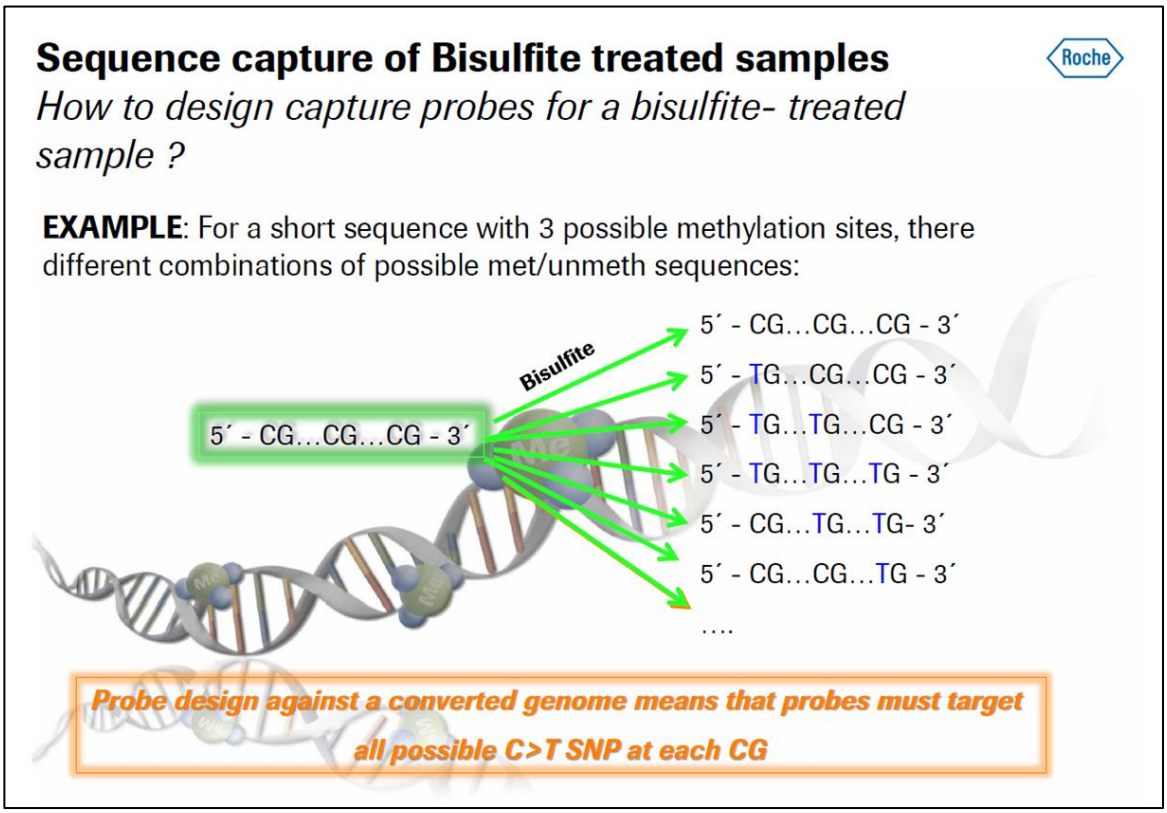


Figure 2.1: The design of the probes, takes into the consideration all the conditions for each cytosine.

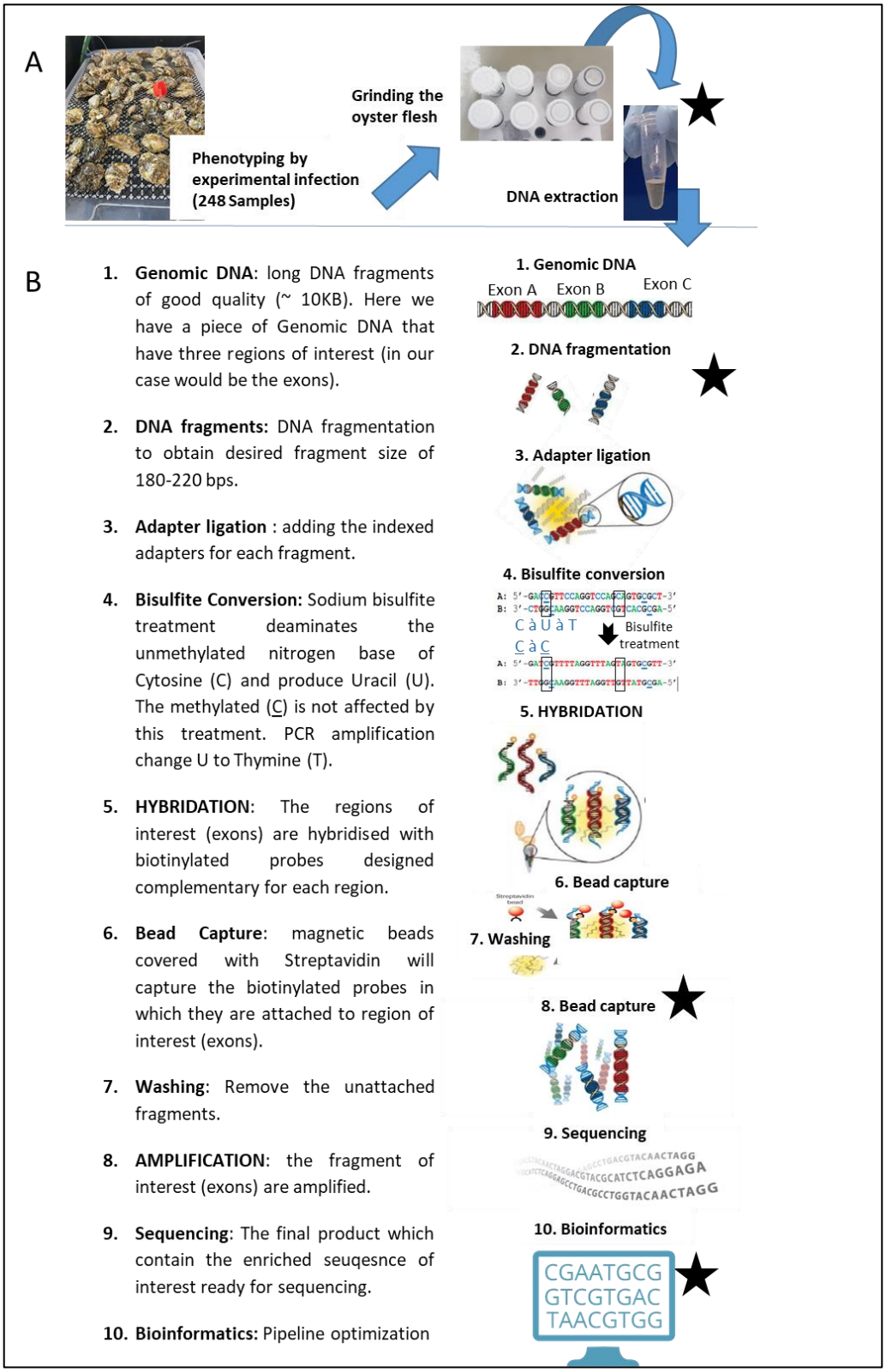


Figure 2.2 Workflow to obtain the sequence data.

A) Sample preparation. B) Adapted figure from ROCHE showing the process of exome capture of the genetic and epigenetic information using "SeqCap Epi Enrichment System" protocol). Star sign indicate the steps needed to be optimised.

2. Material and Methods:

2.1 Probe design

In order to study oyster resistance at the level of the genome and the epigenome we decided to use an exome capture approach. This enable to capture most of the epigenetic information (DNA methylation occurs mainly in gene body in invertebrate) and around 5% of the genetic information (coding regions). With the availability of the *Crassostrea gigas* reference genome (G. Zhang et al., 2012) it is possible to develop probes complementary to exonic sequence, therefore allowing to capture the oyster exome within the whole genome, this was called the primary target. As a first filter the repetitive regions were excluded from this exome to avoid capture bias, this set was called the capture target. Finally, to optimize exome sequencing coverage 100 bp were added to each exon ends and lead to the production of genomic intervals called capture target with padding. The design and probes synthesis were done in collaboration with Roche through the SeqCap epi developer probes kit (Figure 2.3). The probes were developed to capture exons within the 28,027 genes annotated with 2012 genome assembly (G. Zhang et al., 2012).

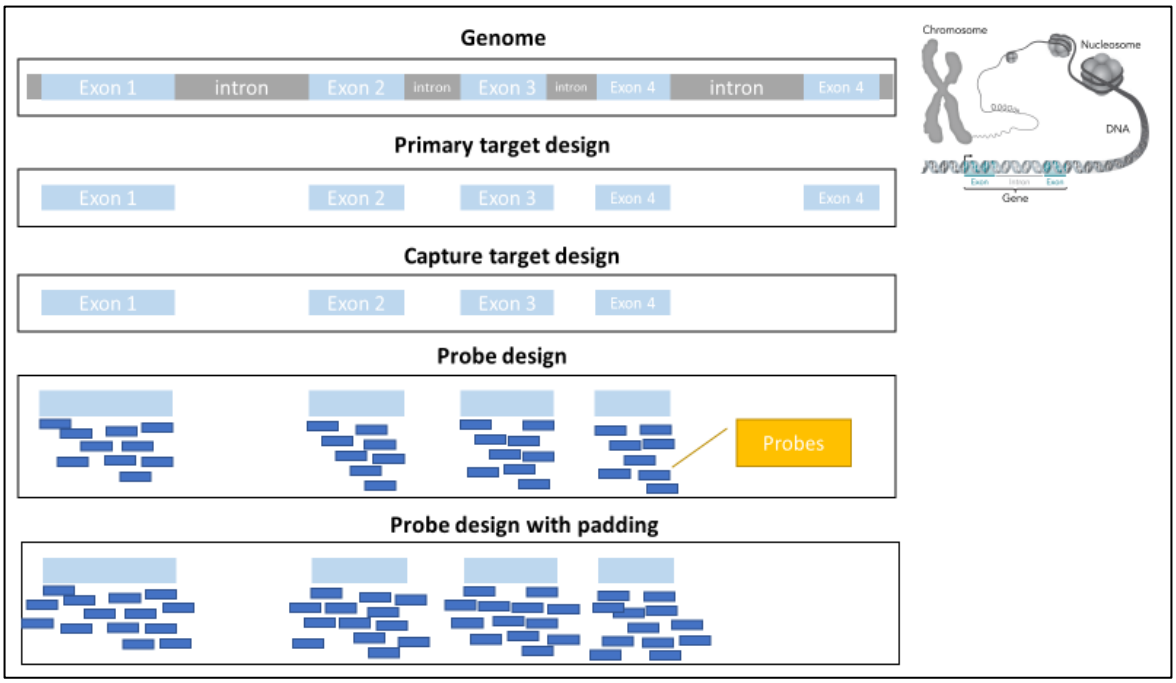


Figure 2.3: Synthesis of probes.

A graphical representation of how probes are designed for capturing the exon regions along with adding the padding sequence in order to capture the exon extremities.

2.2 Sample preparation [phenotyping and DNA extraction]

Once all oysters have been phenotyped (resistant vs susceptible phenotypes) through the experimental infection (see the chapter 3 [118-119] for the details of the experiment if needed), oyster flesh were frozen in liquid nitrogen. The oyster flesh was grounded in liquid nitrogen and the powder produce was kept at -80 °C until DNA extraction. At this step, DNA extraction was slightly optimized to obtain high quality DNA. This procedure and its optimization was done with the NucleoSpin® Tissue kit following manufacturer instructions as a starting step (MACHEREY-NAGEL GmbH & Co. KG). DNA quantity and purity were checked with a Nanodrop One (Thermo Scientific). The quality was checked by 0.8% agarose gel electrophoresis. The extracted DNA was stored at -20°C until use for exome capture. The final aim of this step was to obtain at least 1 µg of good quality DNA characterized by fragments of ~10 Kb and ratio A260/A280 ranging from 1.7-2.0.

2.3 Seq Cap Epi Enrichment System protocol: steps and optimization

2.3.1 DNA fragmentation [producing fragmented DNA]

In order to obtain 200 bp fragments, the sonicator Covaris Focused Ultrasonicator (Covaris, Brighton, UK) was used and different sonication parameters tested. To do this, 1 µg of oyster high quality DNA and 5.8 µL (165 pg) of unmethylated gDNA from the lambda phage (used for bisulfite conversion efficiency) were fragmented in a final volume of 53 µL of Elution Buffer (10mM Tris, 0.1 mM EDTA, pH8). For the fragmentation the microTUBE AFA Fiber Pre-Slit Snap-Cap were used. Fragmentation efficiency was analysed using the Fragment Analyzer (Agilent Technologies France) and the data was processed with the Prosize 3.0 software (Fragment Analyzer software).

2.3.2 End Repair and A-Tailing [End repaired and tailing library]

The DNA fragmentation step does not produce blunt-ended fragments. Therefore, a step called End-Repair is necessary to ensure that each molecule is free of overhangs and contains 5' phosphate and 3' hydroxyl groups. In addition, for Illumina libraries construction it is necessary to incorporate a deoxyadenosine 5’-monophosphate (dAMP) on the 3' end of the DNA fragments (A-tailing). DNA fragments with a deoxyadenosine at their ends bind more

efficiently to adapters possessing complementary deoxythymines. For end repair and A tailling we have added 10 µL of End Repair/A-Tailing mix [consisting of 7 µL of KAPA End Repair & A-Tailing Buffer and 3 µL of KAPA End Repair & A-Tailing enzyme (Roche NimbleGen, Mannheim, Germany)] to 50 µL of fragmented DNA. The reaction was incubated in thermal cycler (Mastercycler Ep Gradient; Eppendorf) for 30 minutes at 20°C (End-repair) followed by 30 minutes at 65°C (A-tailing).

2.3.3 Ligation of adapters to DNA fragments [adapter ligated library]

The adapters used in this protocol (SeqCap Library Indexed Adapter [Roche]) are DNA sequences composed of illumina universal PCR/sequencing primers and index. These indexes are sample-specific DNA sequences that allow sequencing of multiple samples simultaneously (by pooling multiple samples in one tube). For this step 5 µL of SeqCap Library Indexed Adapter are added to tube from previous step (End repaired and tailing library containing 60 µL). Then 45 µL of Master Mix of Adapter ligation (containing 10 µL of KAPA DNA ligase, 30 µL of KAPA ligation buffer and 5 µL of PCR-grade water) are added and incubated at 20°C for 15 minutes in a thermal cycler. This step is followed by a wash to remove the adapters that are not ligated to the DNA using the AMPure XP magnetic beads (Beckman, Roissy, France). These magnetic beads have an outer layer coated with carboxyl groups allowing the reversible binding of DNA (that has a negatively charged phosphate backbone) to the latter depending on the concentration of salts and polyethylene glycol (PEG). This step is performed by adding 88 µL of AMPure XP beads then incubated at room temperature (RT) for 5 minutes. After the incubation, the tubes were placed on the DynaMag magnetic holder (Fisher Scientific, Illkirch, France). The supernatant containing unbound reagents and adapters is discarded. Then the DNA was washed twice with 200 µL of 80% ethanol. The DNA bound to beads was then dried and eluted with 53 µL of elution buffer (10mM Tris, 0.1 mM EDTA, pH8).

2.3.4 Double-sided size selection [Size selected library]

This step is needed for selecting only the fragments with an average insert size of 250 bp plus the 75 bp of adapters on each side of the fragments [total of 350 bp fragments]. Thus, removing the DNA fragments shorter than 250 bp or higher than 450 bp. This was done using AMPure XP beads. The rational is the same as previously, by varying the concentrations of salts and PEG, the AMPure XP beads have the ability to preferentially bind to certain ranges

of DNA fragment size .In order to remove the fragments larger than 450 bp, high quantity of PEG (and therefore of beads) is needed for a preferential attaching to the larger fragments. This was achieved by adding 35 µL of AMPure XP beads to the adapter ligation reaction that contains 50 µL. Larger than 450 bp fragments will be attached to beads, while the shorter fragments remain unattached. The tubes containing the solution are placed on the magnetic rag and the supernatant (fragments less than 450 bp) are recovered. A second step was required to remove fragments shorter than 250 bp. To do this, 10 µL AMPure XP beads are added to recovered supernatant (allowing preferential binding of fragments above 250 bp). The supernatant is removed and the beads with the 250-450 bp fragments were retained, washed and the DNA was eluted in 23 µL of elution buffer (10mM Tris, 0.1 mM EDTA, pH8).

2.3.5 Bisulfite conversion [bisulfite converted library]

The double size selected libraries are then subjected to DNA sodium bisulfite treatment. During this step, the sodium bisulfite treatments will deaminates the unmethylated Cytosine (C) and produce Uracil (U). However, Methylated Cs remain unaffected by this treatment (Frommer et al., 1992). This step is achieved using the EZ DNA Methylation-Lightning kit (Zymo Research, CA).

The conversion was carried out in a dark environment because the reagent used is sensitive to light. We add 130 µL of conversion reagent (Lightning Conversion reagent) to the 20 µL obtained from the double size selection. Since coming incubation will be in a thermal cycler, we split the 150 µL reaction into two PCR tubes. The reaction is then incubated for 8 minutes at 98°C (for DNA denaturation) followed by 60 minutes at 54°C (for bisulfite conversion). Then 600 µL of M-Binding buffer was added to Zymo-spin IC column that is placed in a 1.5 mL collection tube. The column containing the 150 µL of bisulfite converted library is centrifuged for 30 seconds at 12,000g. In order to stop the bisulfite conversion process, the reaction is washed with 200 µL of L- Desulphonation Buffer incubated at RT for 20 minutes. The reaction is again centrifuged for 30 seconds at 12,000g and washed twice by adding 100 µL of M-Wash-Buffer with a centrifugation of 30 seconds at 12,000g. Finally, DNA was eluted with 21.5 µL of PCR-grade water.

2.3.6 Pre-capture PCR and washing [pre-capture library]

After the bisulfite conversion, the methylated Cytosine remain unchanged, while unmethylated Cytosine is converted to Uracil. Then during the PCR amplification, the (U) is replaced by Thymine (T) in the amplified sequences. This results in non-complementary strands (T/G polymorphism).

During this PCR a master mix is prepared by adding 25 µL of KAPA HiFi HotStart Uracil + ready mix, 2 µL of PCR grade water and 3 µL of Pre LM-PCR oligos 1 & 2,5 µM (LM1). Then 30 µL (the total of the mix) is added to the tube from the previous step (bisulfite converted library containing 20 µL). The PCR amplification was run with the following parameters:

Step 1: 2 minutes at 95°C (long DNA denaturation)

X12 {
- Step 2: 30 seconds at 98°C (DNA denaturation)
- Step 3: 30 seconds at 60°C (primer hybridisation)
- Step 4: 4 minutes at 72°C (Elongation)
}

Step 5: 10 minutes at 72°C (Termination)

Step 6: Stored at 4°C

After the PCR amplification the libraries are washed. To do so, 90 µL of Ampure XP beads are added to the amplified libraries, washed twice by adding 180 µL of ETOH 80%. Once dry, 52 µL of PCR grade water are added for the elution. Finally quantity and quality of the converted library was estimated using the NanodropOne (Fisher Scientific) and the Fragment analyzer (Agilent) respectively.

2.3.7 Hybridization of the SeqCap Epi libraries [Hybridised library]

During this step the capture of the target sequences (exons) is achieved using the specifically designed probes (hybridization). However, the design of probes is costly, therefore, several samples can be pooled together. The maximum number of samples in one pool was subjected to optimization. The final aims is to obtain a maximum level of multiplexing but in the same a good equilibrium between samples in order to obtain homogenous sequencing between samples. Several pools containing different numbers of samples were therefore tested (one, two, three, six, eight and 10 samples). The samples were pooled, by adding 1 µg of DNA (for

each pool) in a tube containing 10 µL of Bisulfite Capture Enhancer buffer (SeqCap Epi Accessory Kit, Roche). In addition, 1 µL of HE Universal Oligos and a total of 1 µL of adaptor specific Indexed He Oligos and 1 µL Hybridization Enhancing (HE) Oligos. The Hybridization Enhancing (HE) Oligos are sequences complementary to the adapters that hybridize to the single stranded fragments to prevent the re-hybridizing to each other during probe/library hybridization step. Regardless of the number of samples in a pool, 1 µL of Indexed Oligos HE was added into the tube containing the DNA. Once the DNA and oligos were in the tube, the whole pool was dried using the Eppendorf centrifugal Vacuum concentrator 5301 (Sigma-Aldrich, St Quentin-Fallavier, France) at 60°C for approximately 20 minutes. The pellet was then suspended in 7.5 µL of 2X SC Hybridization Buffer and 3 µL of SC Hybridization Component A (SeqCap Hybridization and wash Kit, Roche). Samples were vortexed and centrifuged, then put in heat block for 10 minutes for denaturation step at 95°C. Samples were transferred to tubes containing 4.5 µL of biotinylated single stranded SeqCap Epi probe pool (Roche) complementary to the regions of interest. This mixture was then incubated at 47°C (lid at 57°C) for up to 45 hours.

2.3.8 DNA capture by streptavidin-coated beads [captured library]

After the 45 hours of hybridization, the complex probes/complementary fragment of interest/streptavidin-coated magnetic beads were recovered. First, 100 µL of SeqCap capture beads are washed twice by adding 150 µL Bead Wash buffer 1X using magnetic rag and by removing the supernatant. Then before the capture beads are dried, the hybridized library are added to the tubes containing the washed capture beads and incubated at 47°C for 45 minutes. Using the magnetic rag, the hybridised library would attach to the magnetic streptavidin beads. The hybridized library is washed several times as shown in the (Table 2.1) and the vortexed and placed on the magnetic rag and the supernatant removed. Finally, the cleaned hybridised library is eluted into 50 µL of PCR grade water.

Table 2.1: washing steps for hybridized library.

Washing buffer	Volume (µL)	Number of wash	Vortex time	Incubation time
Wash Buffer I (@47°C)	**100**	**One**	**10 seconds**	**NA**
Stringent Wash Buffer (47°C)	**200**	**Two**	**NA**	**5 minutes (each was)**
Wash Buffer I (@RT)	**200**	One	90 seconds	NA
Wash Buffer II (@RT)	**200**	One	60 seconds	NA
Wash Buffer III (@RT)	**200**	One	30 seconds	NA

2.3.9 Post-capture PCR and washing [Post-capture library]

This PCR aims to increase the amount of captured libraries. During this PCR a master mix is prepared by adding 50 µL of KAPA HIFI HotStart Ready Mix and 10 µL of Post-LM-PCR Oligos 1&2. Then the captured library was divided into two PCR tubes each containing 20 µL, where, we added 30 µL (the total of the mix) to each tube. The PCR amplification was run with the following parameters:

Step 1: 45 seconds at 98°C (long denaturation)

X14* {
- Step 2: 15 seconds at 98°C (Denaturation)
- Step 3: 30 seconds at 60°C (Primer hybridisation)
- Step 4: 30 minutes at 72°C (Elongation)

Step 5: 1 minute at 72°C (Termination)

Step 6: Stored at 4°C

* Note, the protocol recommended 16 cycles, however, this was reduced to 14 cycles

Once the PCR amplification finished, we pooled again the two PCR tubes of each library into 1.5 mL tube. Then 180 µL of AMPure XP beads were added and washed twice with 250 µL of 80% ethanol. Finally, quantity and quality of captured library was estimated using the NanodropOne (Fisher Scientific) and the Fragment analyzer (Agilent) respectively.

2.4 Illumina sequencing

First a small-scale sequencing (Illumina NextSeq 550 system Paired-End 2x75) was performed for a pool containing three samples. The goal of this sequencing was to check the success of the capture. Next, further sequencing was performed to determine the maximum number of samples that can be pooled together with a homogenous sequencing output. For this, multiplexing of six, eight and 10 samples were multiplexed in three different pools and sequenced. Once the number of samples in each pool was optimized, all the libraries were sequenced with Illumina NextSeq 550 system and Illumina NovaSeq S1 6000 system (PE 2 x 150 bp and PE 2x 100) at an expected coverage of 30X per sample. For both sequencing platform 25% Phix genome was added to each lane to increase the nucleotide diversity.

2.5 Sequencing data analysis

The sequenced samples were analysed and carefully checked to evaluate the success of the SeqCap Epi Enrichment System. First analysis were performed using a GALAXY instance locally installed at IHPE (https://bioinfo.univ-perp.fr/). The quality of the sequencing data was evaluated by checking different parameters according to the supplier's recommendations (ROCHE; Table 2.2).

After Illumina sequencing, the quality of raw reads was checked with FastQC V0.5.3 (Comprehensive QC). Adapter trimming and quality filtering was achieved with TrimGalore V0.4.0 (with quality score threshold of 26 and Maximum allowed error rate of 0.1). To check the effectiveness of the cleaning a second quality analysis was performed for trimmed reads with FastQC. Next the reads were aligned to reference genome 2012 (G. Zhang et al., 2012) using Bismark Mapper Galaxy V0.20.0 (default parameters). Then the duplicate reads were removed using Bismark Deduplicate Galaxy V0.20.0 (default parameters). The number of reads aligned to the exome (on target reads) and the coverage information were calculated using bedtools Intersect intervals Galaxy V2.27.1. Additionally, we looked at other metrics produced by PICARDTOOLS (Linux V2.21.1). Those include coverage depth metrics such as *Fold_Enrichment* and *Fold_80_Penalty* both calculated by CollectHybridSelectionMetrics command. Insert size distribution information were generated using PicardCollectInsertSeizeMetrics. The number of cytosine and cytosine methylation were generated with Bismark mapper (report output). Finally, bisulfite conversion efficiency was

estimated by Bismark Galaxy V0.20.0 by aligning the trimmed reads to the phage lambda genome.

Table 2.2 : Criteria used for evaluating the sequencing results from exome capture.

#	Parameters	Parameter meaning
1	Genome_size	size of the reference genome in bp
2	Primary_target_size	size of the primary targets in bp (regions of sequence coverage is desired)
3	Capture_target_size_padding	size of the capture targets plus padding in bp (covered by capture probes)
4	Median_insert_size	median size of the captured library in bp
5	Mean_insert_size	mean size of the captured library in bp
6	Input_reads	Number of reads (R1+R2) in the input fastq files
7	High_quality_reads	Number of reads (R1+R2) in the fastq files after trimming
8	Pct_hq_reads	Percent high quality reads after trimming
9	Total_mapped_reads	Number of reads Mapped to genome
10	Pct_mapped_reads	Percent mapped reads
11	duplicate_removed_reads	Number of reads after duplicate removal
12	Overall_duplicate_rate	Duplicate rate
13	On_target_reads	Number of reads mapped on target
14	Pct_on_target_reads	Percent of reads mapped on target
15	Mean_depth_of_coverage	Mean depth of coverage
16	Median_depth_of_coverage	Median depth of coverage
17	Pct_>1x	Fraction of primary target bases covered by 1 or more reads
18	Pct_>10x	Fraction of primary target bases covered by 10 or more reads
19	Pct_>20x	Fraction of primary target bases covered by 20 or more reads
20	Pct_>50x	Fraction of primary target bases covered by 50 or more reads
21	Pct_>100x	Fraction of primary target bases covered by 100 or more reads
22	Fold_Enrichment	Fold enrichment calculated by Picard CalculateHsMetrics
23	Fold_80_Penalty	Fold-80 penalty calculated by Picard CalculateHsMetrics
24	Total_c_in_cpg_context	number of C's and C's converted to T's in capture target regions
25	Total_c_in_cpg_methylated	number of C's in CpG context that were methylated
26	Pct_c_in_cpg_methylated	Percent of C's in CpG context that were methylated
27	Total_c_pos_methylated	number of C positions in capture region that had one or more C's methylated
28	Pct_c_pos_methylated	Percent of C positions in capture region that had one or more C's methylated
29	Lambda_cs	Number of C's in lambda control region
30	conversion_efficiency	Percent of C's that were converted to T's in lambda control region

2.6 Bioinformatics pipeline optimization

After the optimization of the exome capture, we were interested in optimizing the bioinformatics pipeline. The idea was to select the best aligner, which can align the higher number of reads, produce the better coverage and finally enable the higher number of SNP and DNA methylation calling. Additionally, to rank the different aligners we compare different metrics (presented in the Table 2.2).

Three different aligners were tested (Figure 2.4) to align the filtered and trimmed reads to the reference genome of *Crassostrea gigas* (G. Zhang et al., 2012). The entire reference genome or masked (for non exonic sequences) reference genome was used.

The three aligners had several steps to follow in order to obtain a bam file optimized for SNPs and DNA methylation calling (PCR duplicate free, properly paired mapped reads, and clipped for overlapping paired reads). Once the clean bam file were produced, we looked at different basic mapping metrics (PICARD-TOOLS V2.21.1; PicardCollectInsertSeizeMetrics) and enrichment metrics (Fold enrichment and fold 80 penalty; by PICARD-TOOLS; CalculateHsMetrics).

Next, SNP and DNA methylation calling were obtained by MethylExtract V1.9. Then, the number of SNPs and DNA methylation sites produced by each aligner were counted and compared between each aligner.

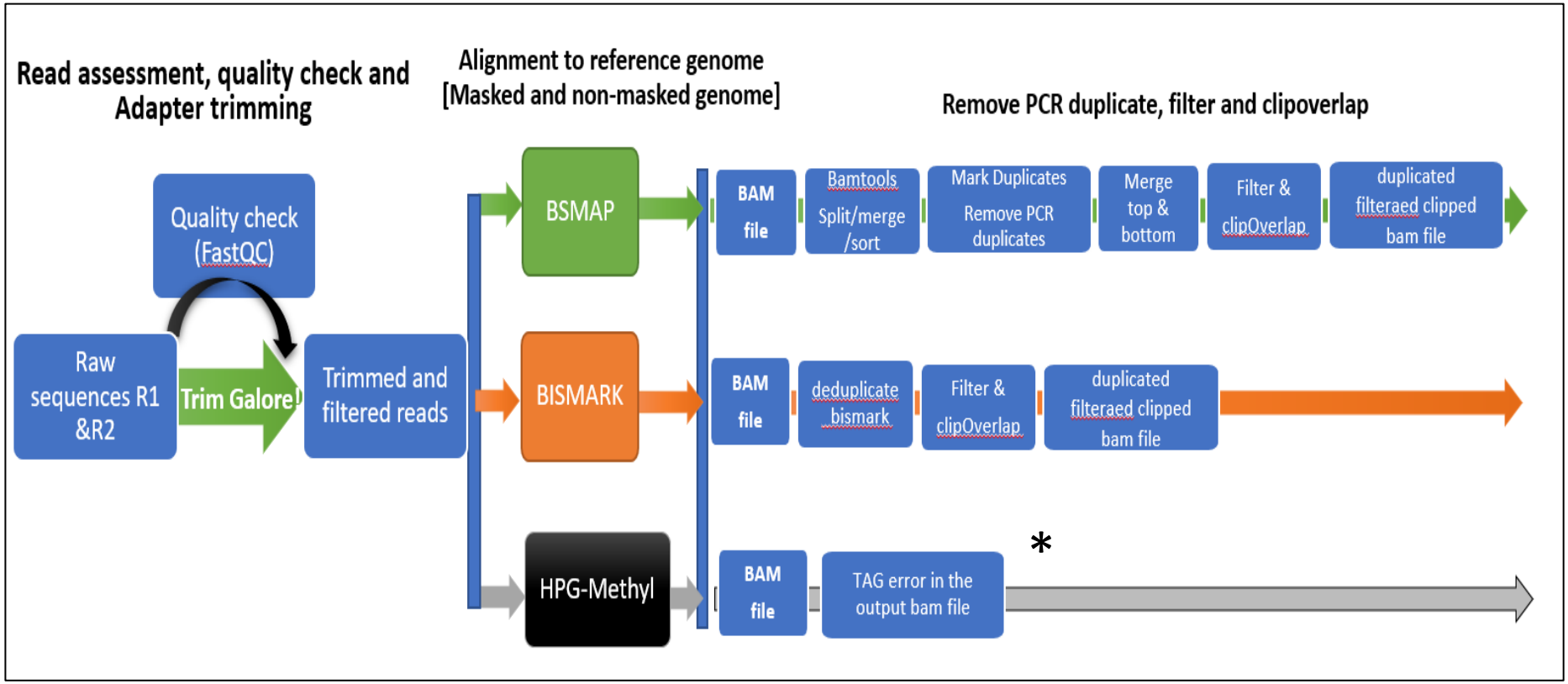


Figure 2.4: Optimization of bioinformatics pipeline.

The HPG-methyl aligner *, we could not proceed with bam file, due to error in the tag of the bam file, so it was excluded.

3. Results of bench work and bioinformatics optimization:

3.1 DNA extraction optimization

The extraction of DNA from oysters was optimised in order to obtain high quality DNA characterized by fragments larger than 10,000 bp at a sufficient concentration (1 µg) and with ratio A260/A280 ranging from 1.7-2.0.

As a first step the quantity of powdered oyster we need to use as starting material was optimized. Tests using either one or two microspoons of powder (20 0r 40 mg) were performed. The concentration of DNA extracted was highly variable, for two microspoons it varied between 77.2 ng/µL to 239.2 ng/µL and one microspoon produce a concentration varies from 31.7 to 208.8 ng/µL (Table 2.3) both with ratio A260/A280 close to 1.8. We conclude that the purity and quantity of the DNA is acceptable in both cases.

Table 2.3 : Quantification of extracted DNA by Nanodrop as a function of the amount of powder used.

sample	Amount of powder (in microspoon)	[DNA] (ng/µL)	A260/A280
C12-2-135	2	118,8	1,81
F37-1-260		189,9	1,81
I1-2-325		28,4	1,79
E36-6-52		77,2	1,82
J39-1-238		239,2	1,83
N37-7-119		187,6	1,82
G11-6-38	1	71,6	1,81
J18-1-256		31,7	1,82
N6-1-93		146,2	1,84
J40-1-232		166,9	1,82
I1-2-325		208,8	1,82
E36-6-52		82	1,83

Secondly, in order to analyse the approximate size of the extracted DNA, a gel electrophoresis was performed on 100 ng of DNA. The gel shows a wide range of DNA quality between samples (Figure 2.5) interpreted as a problem during the elution process.

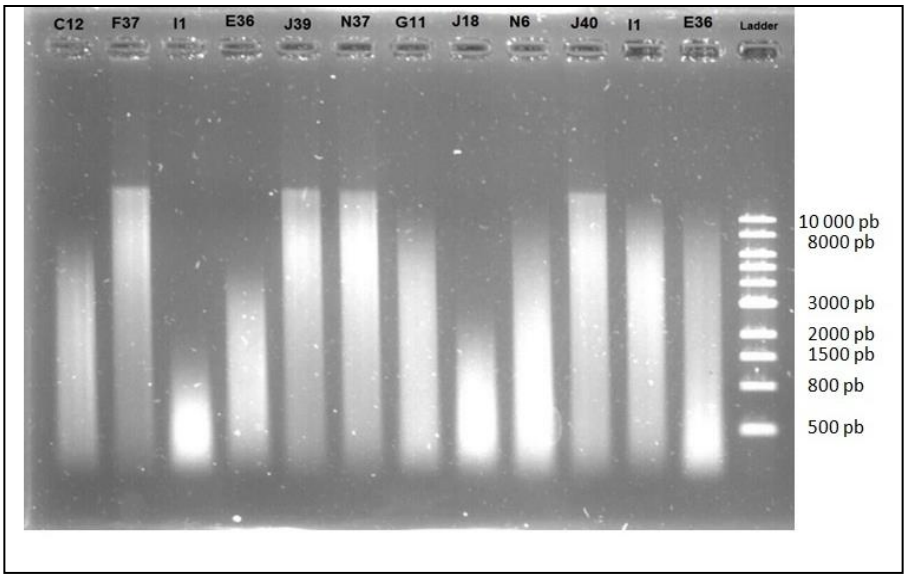


Figure 2.5 : Gel electrophoresis of DNA extracted from 12 oysters.
Using 0.8% agarose gel.

In order to improve the quality of the extracted DNA, further tests were performed: firstly, a lower amount of oyster powder was used (half microspoon), secondly the lysis time was increased from four hours to overnight. Finally the elution buffer was preheated to 70 °C.

As before, high variability in DNA concentration were obtained, but interestingly the overall quantity has increased (ranging from 344 ng/µL to 987 ng/µL and 101 to 784 ng/µL using one microspoon and half microspoon of oyster powder (20 and 10 mg), respectively (Table 2.4). Nevertheless, these results show that half microspoon (10 mg) of oyster is sufficient to obtain high concentration of DNA.

Regarding the quality of the DNA, the ratios obtain were still good (A260/A280 ratio close to 1.8). Gel electrophoresis showed no differences in DNA quality between the different lysis duration (4h and overnight). Finally the gel showed a main band above the 10KB size of the half microspoon set up (Figure 2.6).

Following these last results, the DNA extractions were performed with 1/2 microspoon of oyster powder as starting material, lysis for 4h or overnight and finally the elution buffer were heated at 70°C.

Table 2.4 : Quantification of extracted DNA by Nanodrop as a function of the amount of powder used.

Samples	Amount of powder (in microspoon)	Duration of lysis	[DNA] (ng/µL)	A260/A280
D40	1	4h	479	1,83
		overnight	987	1,80
	0.5	4h	784	1,78
		overnight	450	1,83
M25	1	4h	344	1,83
		overnight	369	1,83
	0.5	4h	101	1,80
		overnight	184	1,82

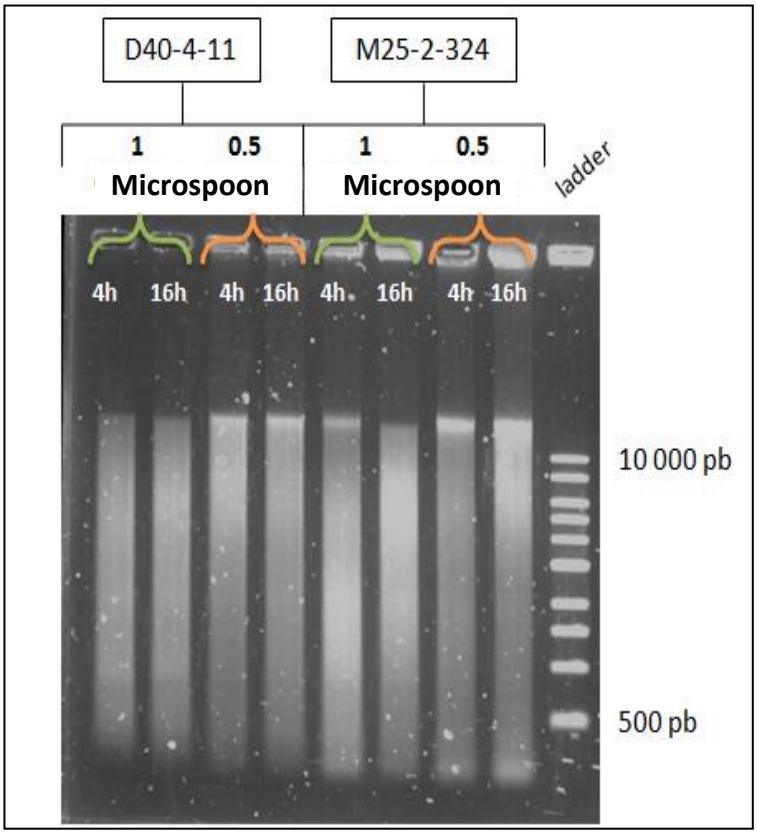


Figure 2.6 : Gel electrophoresis of DNA extracted from two oysters.

Either using one microspoon or half (with lysis for four hours or overnight (16 hours) in 0.8% agarose gel.

3.2 Optimisation of DNA fragmentation

To optimize this step we tested different sonication times with a starting step as recommended by Covaris (the sonication manufacture company) for the microTUBE AFA Fiber Pre-Slit Snap-Cap (Table 2.5). To avoid the formation of too many droplets during sonication a short spin was done in the middle of the sonication period.

Table 2.5: Sonication parameters applied in order to obtain desired fragments size of 200 bp.

parameters		**Duration**
Peak Incident Power (W)	175	(**A**) 85-(**B**) 90-(**C**) 95 seconds
Duty Factor (%)	10	
Cycle per Burst	200	
STOP Centrifugation	10 seconds	
Peak Incident Power (W)	175	**(A**) 85-**(B**) 90-(**C**) 95 seconds
Duty Factor	10	
Cycle per Burst	200	

Theoretically, as the sonication time decreases, the fragment size should be larger. However, the results showed that no correlation between treatment time and fragment size were obtained. In addition, we observed a strong variability in fragment size produced from the different samples that have undergone the same treatment (with an exception for condition C; Figure 2.7). Overall, the fragments size were less than 200 bp, further tests were therefore needed with the consideration of reducing the duration of the sonication.

The tests D, E, F, G and H were carried out with the sonication same parameters but different durations of sonication (Table 2.6). The result obtained shows a slight decrease in the size of the fragments when the processing time is decreased for sample N6-1-93 (Figure 2.8). On the other hand, it is difficult to show such a correlation for sample J39-1-238 where the expected size is still not reached. Surprisingly, for sample J39-1-238 in some tests we did not observe any fragment peak illustrating the difficulty to reproduce this sonication.

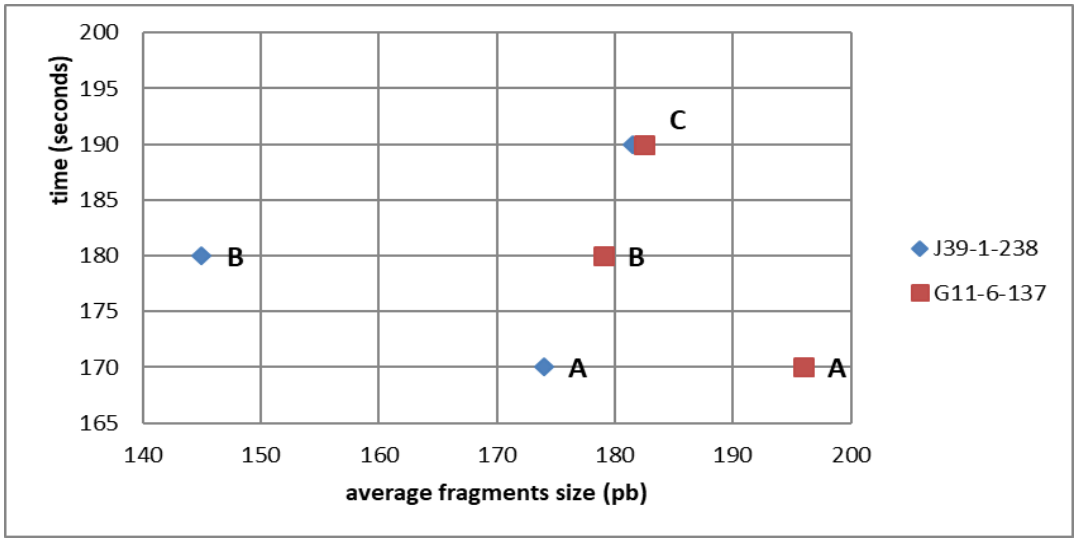


Figure 2.7 Average Size of DNA fragments obtained with different sonication time settings.

Table 2.6: Sonication parameters applied in order to obtain the desired fragments size of 200 bp. In green is the recommended time and parameters used for tube if 55 µL.

parameters		**Duration**
Peak Incident Power (W)	175 - 75 (H')	(**D**) 80-(**E**) 70-(**F**) 60-(**G**) 50- (**H**) 45
Duty Factor (%)	10 – 25 (H')	(H') 45
Cycle per Burst	200 - 1000 (H')	Seconds
STOP Centrifugation	10 seconds	
Peak Incident Power (W)	175 - 75 (H')	(**D**) 80-(**E**) 70-(**F**) 60-(**G**) 50 - (**H**)
Duty Factor	10 - 25(H')	45
Cycle per Burst	200 - 1000 (H')	(**H'**) 45 Seconds

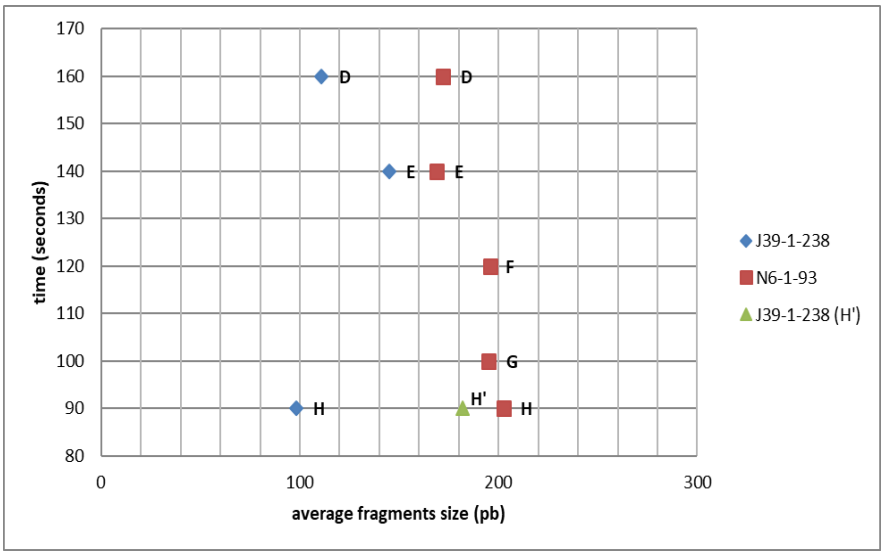


Figure 2.8: Average Size of DNA fragments obtained with different sonication time settings.

However, the H' test (45 seconds) seems to give similar results to the H test (45 seconds). Yet, the sizes are always less than or equal to 200 bp. Knowing that the method is not very reproducible, reducing the processing time would allow us to have fragments with a homogeneous distribution between 180 and 220 bp. Therefore, a final test was carried out with a duration of 80 second (35 seconds – 10 seconds centrifuge – 35 seconds) on six samples (Table 2.7).

Table 2.7: Sonication parameters applied in order to obtain desired fragments size of 200 bp.

parameters		**Duration**
Peak Incident Power (W)	175	35 Seconds
Duty Factor (%)	10	
Cycle per Burst	200	
STOP Centrifugation	10 seconds	
Peak Incident Power (W)	175	35 Seconds
Duty Factor	10	
Cycle per Burst	200	

The average fragment size obtained with this last sonication test (value in red in Figure 2.9) is close to 200 bp but still with variability between samples (Figure 2.9). However, fragment analyser results shows that these last parameters are the most suitable to obtain the desired fragment size. In addition it is important to keep in mind that size selection further homogenize the insert size that will be sequenced.

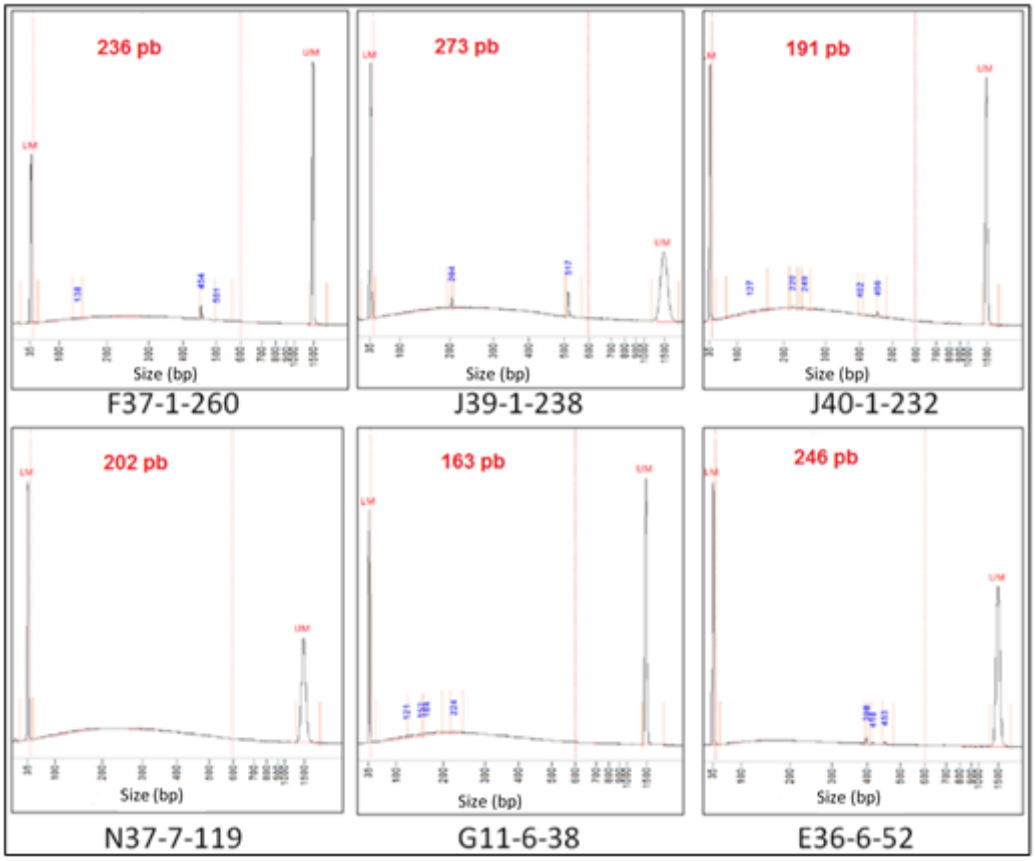


Figure 2.9 : Gel electrophoresis of DNA fragments after sonication.
Showing the average fragment size (number in red).

3.3 DNA fragments size and concentration after size selection with AMPure XP beads

After the DNA fragmentation, End Repair and A-Tailing, Ligation of adapters, Double-sided size selection, Bisulfite-Conversion, Pre-capture PCR amplification and clean up were done. At all these step a control of the DNA concentration and it's the quality of the size distribution obtained where needed. In the last step of Pre- capture PCR amplification and cleaning, the DNA concentration were quite homogeneous between samples and ranged from 38.4 ng/µL to 56.3 ng/µL (Table 2.8). These concentrations were above the expected threshold of 20 ng/µL required to perform the next steps of the protocol. Since we started with an equal amount of 1 µg of DNA for all samples, and since we obtained a close range of DNA concentration for all samples at all steps we can conclude that the protocol enable good reproducibility.

It is worth to mention, that sample E36-6-52 has a slightly lower amount of DNA. This was due to a calculation error leading to a quantity of starting material of 336 ng of DNA instead of 1 µg. Interestingly, we saw that even with 3 time lower DNA quantity (lower than 1 µg) the library construction was feasible.

The next checkpoint is to verify that there has been an enrichment of fragments ranging. In terms of size distribution, the expected results were between 250 and 450 bp (the fragments [~ 250 bp] plus the adapters [150 bp]. The results obtained showed a peak with an average size around 350 bp (Figure 2.10). This confirm the enrichment for fragments in the expected range with the elimination of fragments larger than 450 bp and smaller than 250 bp.

Table 2.8: Quantification of extracted DNA by Nanodrop after double size selection

Samples	DNA (ng/µL)	A260/A280
F37-1-260	54,9	2,04
J39-1-238	45,8	1,89
J40-1-2332	56,3	1,95
N37-7-199	50,3	1,94
G11-6-38	56,0	2,0
E36-6-52	38,4	1,93

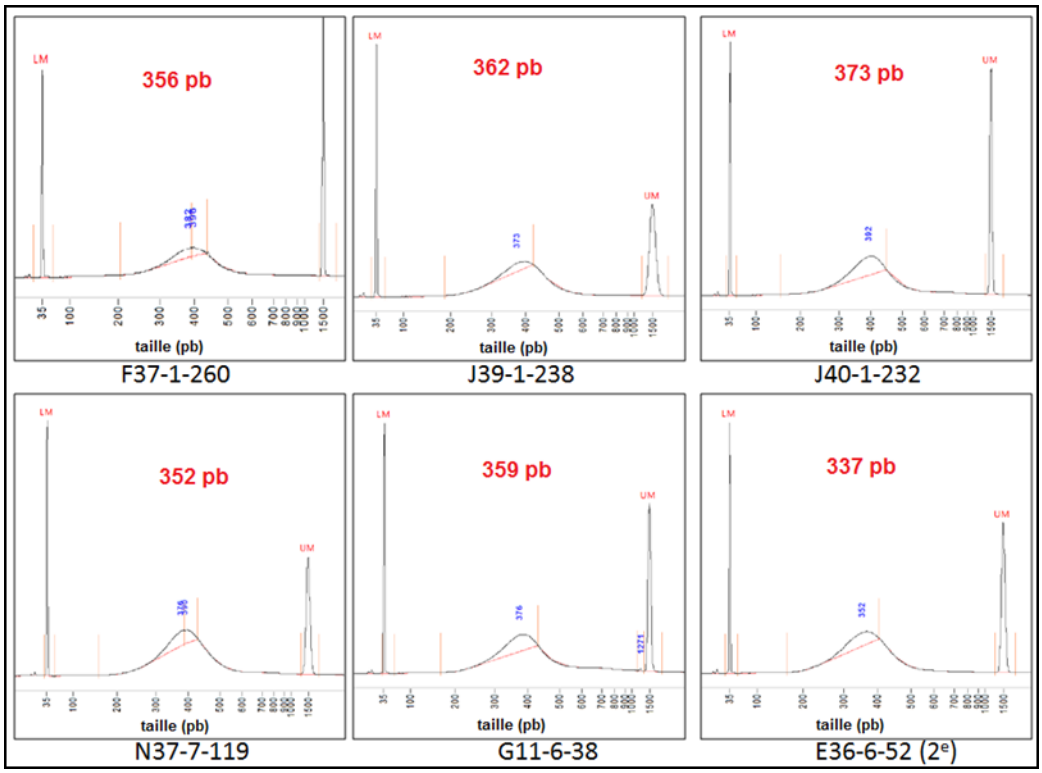


Figure 2.10: Gel electrophoresis of DNA fragments after double size selection. Showing the average fragment size (number in red).

3.4 DNA fragments size and concentration after capture

We first run the protocol on six DNA samples of oyster distributed in three pools, of one, two or three samples in each pool, respectively. The last checkpoint is to look if we have still sufficient DNA concentration and quality (size of the fragments as expected around 350 bp) after the capture step. After the capture the DNA concentrations we recovered were nine times higher than the expected threshold of 10 ng/µL (Table 2.9). This results highlight that some non-specific capture can occur but most probably that the number of PCR cycles done can be reduced. These hypotheses will be verified by the sequencing and bioinformatics analysis. Furthermore, the fragment analyser results showed a peak around 390 bp, higher and finer than previously (Figure 2.11). This clearly demonstrates the enrichment achieved by the capture of exons. Additionally, the good reproducibility between the three pools.

Table 2.9: Quantification of extracted DNA by Nanodrop after capture

Pools	DNA (ng/µL)	A260/A280
Pool A	92,9	1,90
Pool B	94,2	1,90
Pool C	88,0	1,90

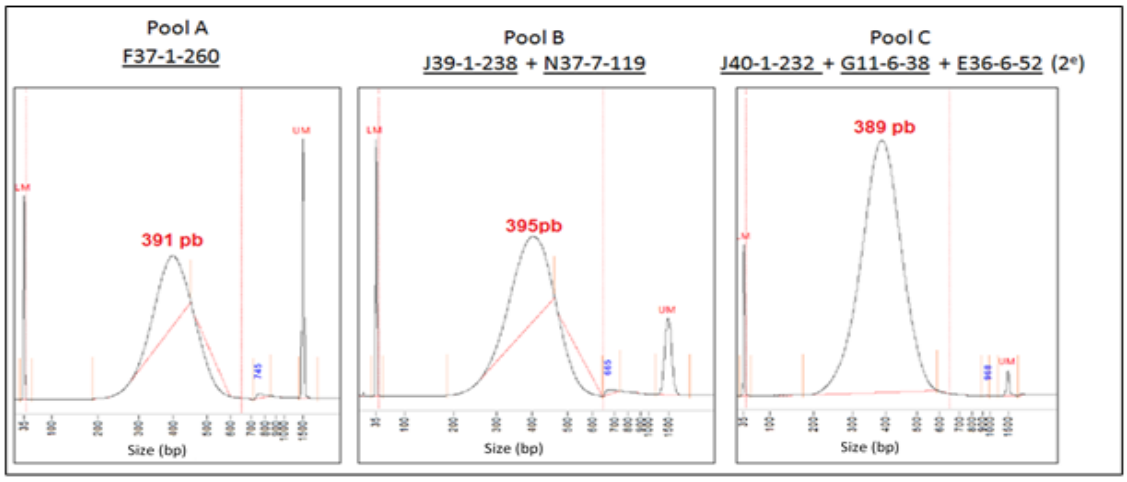


Figure 2.11 : Gel electrophoresis of DNA fragments after capture and pooling. Showing the average fragment size (number in red).

3.6 Illumina sequencing results

3.6.1 Read quality

In the first sequencing test performed for evaluating the exome capture success, a total of, 1 µg of DNA from pool C [containing three samples] was sequenced using Illumina NextSeq 550 system Paired-End 2x75. After sequencing, the qualities of sequenced samples were checked to evaluate the protocol success (strengths and weaknesses) and improve it. Therefore, we checked several points as criteria (Table 2.2) of success of the capture by comparing our results to a reference results from manufacturer (ROCHE) for validating the protocol.

First, we looked at the number of reads yielded per sample. The sequencing returned in average 30 million reads per sample (Table 2.10). The reads given for all the samples were quite close to each other. The quality analysis showed a very good quality for the forward strands reads (R1; sense) with a phred score of 35 on average (0.032% average error in base identification). A decrease in the quality for the bases at the end of the reads was present but usual (Figure 2.12A) and often due to a "phasing" problem in Illumina sequencing. These errors occur with a low probability but over time they accumulate and increasingly pollute the signal sent for calling the nucleotide (Schirmer et al., 2015). On the other side, the quality for the reverse strand (R2; anti-sense), is much lower than R1. This phenomenon of unequal quality between the two strands could rise from the fact that R1 is first sequenced and then followed by R2; therefore clusters grow with time and the sequencer may have difficulty to call base when the analysis is overloaded. Taking in to account the FastQC results, the sequences were cleaned up according to parameters that seemed to be the most appropriate to preserve a maximum of information while eliminating sequences of very poor quality. The parameters used were: quality score threshold of 26 and Maximum allowed error rate of 0.1. The trimming parameters allowed to produce much cleaner reads in both strands (R1 and R2; Figure 2.12B) and with very low percentage of reads removed (Table 2.10; between 1-4% of total reds).

After cleaning, the reads were aligned to the *Crassostrea gigas* genome using Bismark mapper Galaxy V0.20.0. On average 47% of the cleaned reads were aligned to the *C. gigas* genome (Table 2.10). The aligned reads were further filtered by removing the duplicate reads (6-7%) for the three samples (Table 2.10. In order to reduce the duplication rate, the number of PCR

cycles needed to be reduced. During the library preparation, a total of 16 PCR cycles were used. It was therefore reduced to 14 cycles for the next experiments.

Table 2.10 : Main Sequencing, Enrichment and methylation criteria for evaluating the success of SeqCap Epi Enrichment System recommend by manufacturer ROCHE.

#	Parameters	J40	G11	E36	Reference
1	**Genome_size (bp)**	559,000,000			3,101,853,241
2	**Primary_target_size (bp)**	38,546,016			80,321,781
3	**Capture_target_size_padding (bp)**	86,251,471			80,321,781
4	**Median_insert_size**	224	204	197	196
5	**Mean_insert_size**	225	206	199	196
6	**Input_reads**	30,284,216	29,588,762	39,211,926	92,192,812
7	**High_quality_reads**	29,271,190	28,596,070	39,211,256	55,797,974
8	**Pct_hq_reads**	96.7	96.6	100.0	60.5
9	**Total_mapped_reads**	13,633,842	13,475,380	18,241,854	52,950,456
10	**Pct_mapped_reads**	46.6	47.1	46.5	94.9
11	**duplicate_removed_reads**	12,775,678	12,637,012	16,850,302	47,934,528
12	**Overall_duplicate_rate**	6.3	6.2	7.6	5.8
13	**On_target_reads**	8,259,873	8,469,094	15,014,160	35,124,052
14	**Pct_on_target_reads**	64.7	67.0	89.1	73.3
15	**Mean_depth_of_coverage**	13.1	13.5	21.4	40.2
16	**Median_depth_of_coverage**	10.0	10.5	16.0	32.0
17	**Pct_>1x**	86.5	86.7	94.5	99.9
18	**Pct_>10x**	50.4	51.4	72.0	97.2
19	**Pct_>20x**	22.6	23.7	45.8	77.2
20	**Pct_>50x**	1.3	1.4	5.9	26.7
21	**Pct_>100x**	0.2	0.3	1.1	3.5
22	**Fold_Enrichment**	7.6	7.9	7.6	26.3
23	**Fold_80_Penalty**	6.0	6.1	4.3	2.2
24	**Total_c_in_cpg_context**	19,857,155	25,256,034	35,251,194	96,641,750
25	**Total_c_in_cpg_methylated**	4,905,577	5,128,902	6,883,170	30,971,463
26	**Pct_c_in_cpg_methylated**	19.8	20.3	19.5	32.1
27	**Total_c_pos_methylated**	198,366	205,345	276,859	1,786,848
28	**Pct_c_pos_methylated**	38.7	38.7	41.6	57.3
29	**Lambda_cs**	395,683	293,593	1,077,581	836,102
30	**conversion_efficiency**	99.6	99.6	99.6	99.7

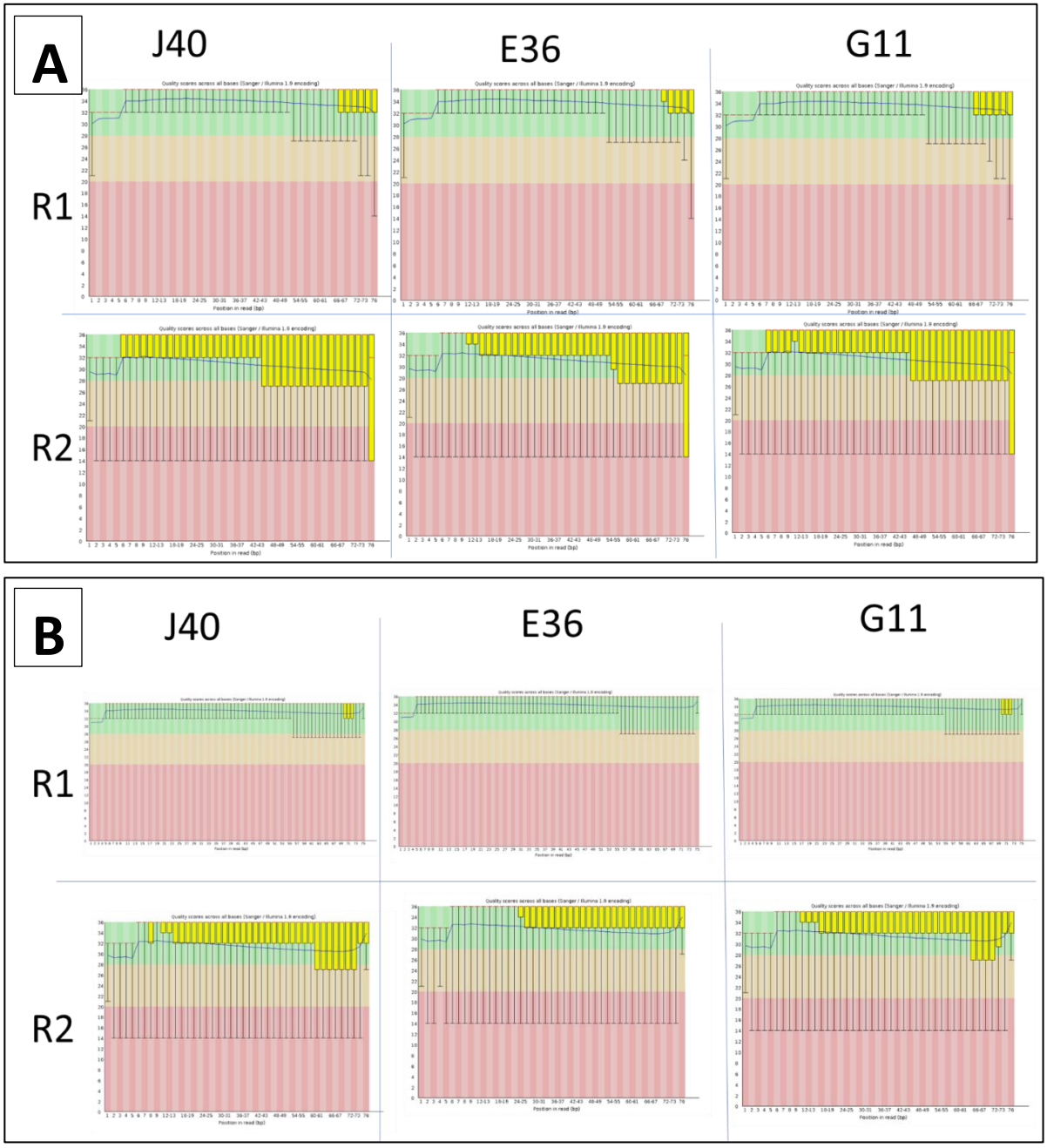


Figure 2.12 : FastQC result showing per base quality score of all reads in forward strands (R1) and reverse strands (R2).

A) Before trimming. B) After trimming.

3.6.2 Enrichment metrics

Amongst the aligned and filtered reads, some probably correspond to non-target regions (outside of exome) and others fall within the exome (on target reads). The percentage of on-target-reads were highly variable between the different samples (64.7% for G11, 67% for J40 and 89.1% for E36; Table 2.10). The sequencing depth is an important measure to determine whether the on-target-reads sequenced were aligned homogeneously on the exome. The analysis showed that mean of sequencing depth of the exome varies between 13X and 21X (Table 2.10), which is close to what we expected (15X). However, the mean depth does not provide a good representation of the homogeneity of this coverage between all the exons. Therefore, we checked the coverage for each exon and then divide it into 20 quantile to see the distribution of the coverage on these quantiles. The quantile results shows that for samples J40 and G11, 70% (14/20 quantiles) of the exome has a coverage of less than 15X, compared with only 45% (9/20 quantiles) for sample E36 (Figure 2.13). The averages close to 15X can be explained by the fact that some regions seem to have a coverage above 15X (over sequencing) and others, on the contrary, have a very little or no coverage. The last quantile of the exome is between 3 and 6 times more covered than what it is intended at the beginning. Moreover, the standard deviations of these 20th quantile show a very wide dispersion around the average, which means that some regions are largely over-represented.

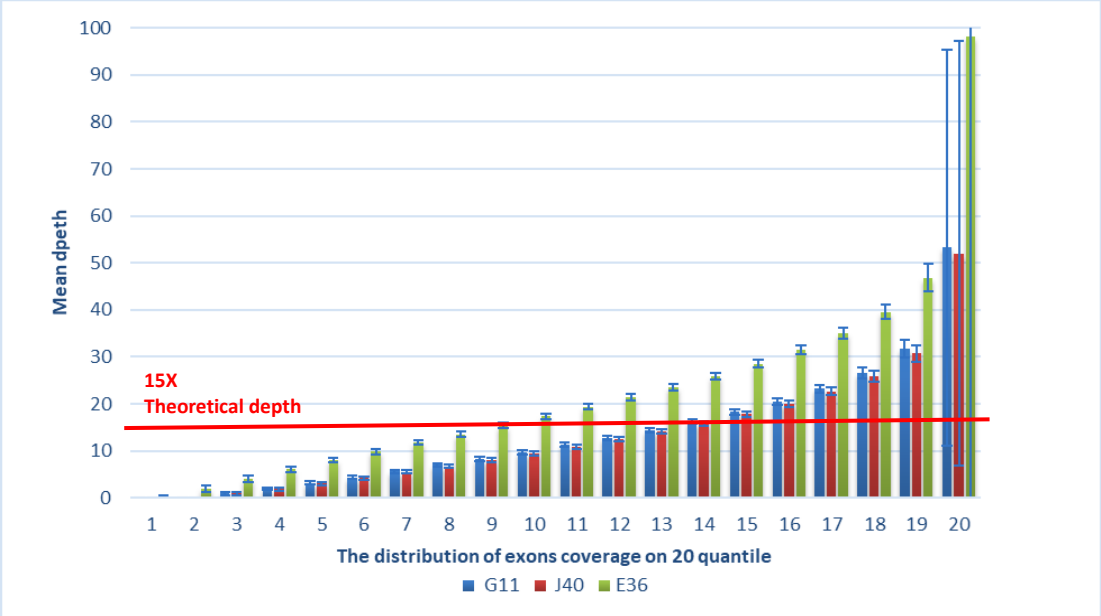


Figure 2.13: Average exome coverage of sequenced reads distributed in 20 quantiles

Additionally, these results were also observed by looking at the *Fold-80-penality* metric, which also measures the homogeneity of the exome coverage. For J40 and G11 the "fold-80-penality" is close to six (Table 2.10). A value of six means that the sample need to be sequenced 6 times more to reach a minimum coverage of 15X for all exons. However, the sample E36 has lower value close to 4, which goes with the fact that this sample was more deeply sequenced due to some bias in the capture or the initial pooling. Another interesting parameter, is the fold enrichment providing information on the number of fold the targeted sequences have been enriched by comparison to the whole genome.

We then further looked at the percentage of exons uncovered and highly covered (over 150X). For sample J40 and G11, about 42 % of exons had a coverage below 8X, among which 23 % are uncovered (Table 2.11). For sample E36, these value were twice lower, which is probably explained by the higher sequencing depth of this sample. When comparing the coverage in the three test samples, only 293 exons (0.09 %) display a coverage below 8x. Concerning the highly covered exon between 0.1-0.5 % of exons displayed a coverage above 150X (Table 2.11). At the level of the three test samples this correspond to only 37 exons (0.01%). Finally, from the 193,263 exons present in the genome, 58% for J40 and G11 samples and 77% for E36 sample were covered at a level between 8 X and150 X (Table 2.11). It is worth to mention, that coverage between 8 and 150 X is a range that is used for calling SNPs and DNA methylation.

Table 2.11: Summary of the coverage depth for the all exons that were captured.

Samples	J40	G11	E36
Total exons	193,263	193,263	193,263
Exons < 8X (=**0X)**	81,722 **(18,933**)	79,751 (**18,966**)	43,283 (**7,189**)
Exons < 8X (=**0X)** %	42.3 (**23.2**)	41.3 (**23.8**)	22.4 (**16.6**)
Exons >= 8X - <= 150X	111,720	113,665	149,292
Exons >= 8X - <= 150X %	57.8	58.8	77.2
Exons > 150X	180	207	1,048
Exons > 150X %	0.1	0.1	0.5

In addition to these analyses other quality check were done, such as the bisulfite conversion performance and the methylation status of each sample. Bisulfite conversion efficiency was quantify using the spike-in with done with the lambda phage DNA (which is completely unmethylated) added to each sample during DNA fragmentation. Therefore, to evaluate the conversion efficiency, the trimmed reads were aligned to the lambda phage genome. This analysis showed that 99.60% of the cytosine were converted which correspond to the expected results (Table 2.10). Finally, we looked at the percentage of methylated Cytosines over all the methylated Cs. The results showed that about 20% of the CpGs were methylated which is consistent with other studies (Xiaotong Wang et al., 2014). Further, we only looked at methylated Cytosines in the exome, and we found that about 40% of the exome CpGs are methylated, in accordance with previously published data (Xiaotong Wang et al., 2014).

Overall, from these results we conclude that the exome capture provide good performance and we further optimized our approach by focusing on the number of samples we would multiplex in a single capture reaction.

3.7 Optimization of maximum number of samples in one pool

To determine the maximum number of samples that can be pooled together, the exome capture was run again with 24 samples which correspond to the number of adapters we had. Three pools were prepared, with either six, eight or 10 samples in each pool, respectively.

DNA quantity and quality were analysed with the Fragment Analyzer and the Nanodrop, respectively. The results obtained during the Benchwork were similar to those previously reported, therefore validating the efficiency of the protocol. Later, the three pools were sequenced and we compared the number of reads produced per samples within each pool.

The comparison showed that pool with three, six and eight samples give a very close number of reads per samples within each pool (Figure 2.14), while the pool of 10 displayed the strongest variation. Therefore, we decided to continue with the pooling of eight samples.

In conclusion, the “SeqCap Epi Enrichment System” protocol was optimised for pooling eight samples together and prepares each time three pools together. In total we prepared and sequenced 248 samples.

However during the second analysis of sequencing data, the sequencing yielded a lower number of reads compared to the first sequencing. This observation highlighted a crucial point of the Illumina sequencing technique used. The data obtained were very low quality. This could be due to the fact that Illumina sequencing requires a balanced base composition (recommendation by Illumina). In the first sequencing we did not have this problem as our samples took only 25% of the flow cell. Thus, the nucleotide diversity was maintained because our libraries were simultaneously sequenced with other base balanced libraries in the same flow cell. While during the second sequencing our libraries took 100% of the flow cell. This problem results from the fact that bisulfite conversion lead to conversion of unmethylated C to T and therefore reducing the nucleotide diversity. In order to avoid such problem in future sequencing, we added 25% of Phix genome to increase the nucleotide diversity or we shared our illumine lane with other project with balanced nucleotide diversity.

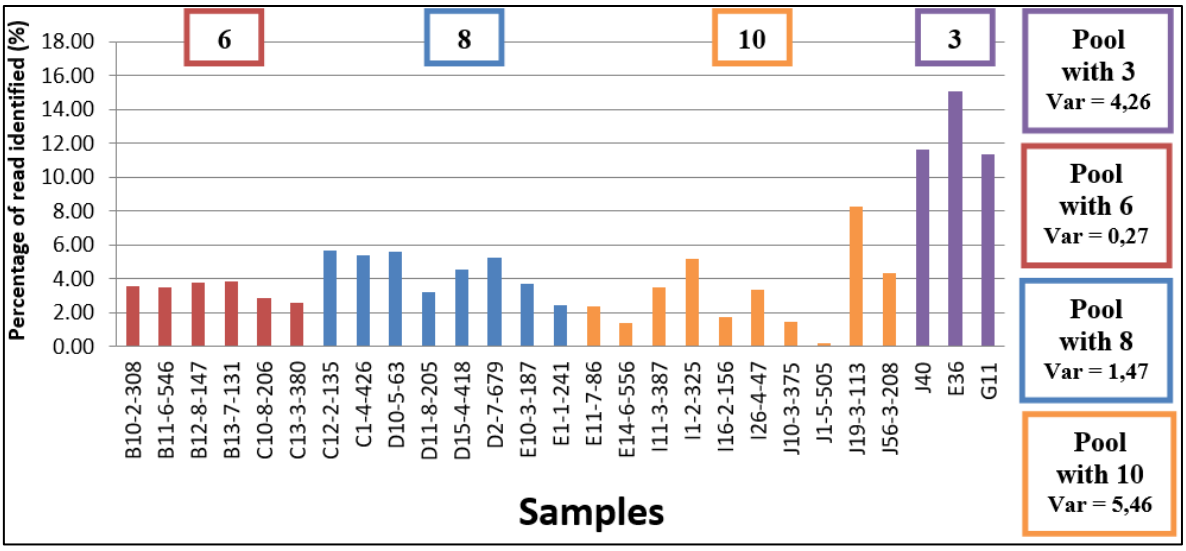


Figure 2.14: Percentage of reads identified in each pool as a number of reads within each flow cell being sequenced (Var: variance)

3.8 Bioinformatics pipeline optimization

The bioinformatics pipeline for the calling of SNP and DNA methylation was optimized with the sequencing data (Illumina NextSeq 550 system; PE 2 x 75) obtained during the Exome capture optimization. We selected six samples that have been sequenced from different pools (three, six or eight samples in a pool). These samples had different number of reads ranging from 8 million to 40 million reads. The goal of the optimization was mostly to select the best aligner (BSMAP, BISMARK or HPG-methyl with masked or non-masked reference genome; Figure 2.4). Unfortunately due to absence of a particular tag in the BAM file needed to separate the Crick and the Watson strands has eliminated HPG-methyl from this comparison. Results showed higher mapping rate for BSMAP on the non-masked genome (Figure 2.15A). After the mapping, the bam file was filtered to only select the properly mapped and the overlapping ends of each paired reads were clipped. Higher number of filtered reads was also obtained for BSMAP non-masked treatment than others (Figure 2.15B). From this filtered BAM file, we calculated the rate of on target read and again, BSMAP used on the non-masked genome display the best metrics (Figure 2.15C).

In a second approach we focus on coverage results for each of the treatment. The homogeneity between exon coverage was evaluated by the *Fold_80_Penalty* metric. This

metric confirms the non-homogeneous sequencing depth obtained (Figure 2.15D) but, showed that BSMAP aligner on the non-masked genome provide the best results (Figure 2.15E).

Third, we called the SNP and DNA methylation using MethylExtract tool (Barturen et al., 2013). We compared the output results for each treatment. The comparison shows that BSMAP and the non-masked treatment provide the higher number of SNPs and CGs comparing to the other aligners (Figure 2.16).

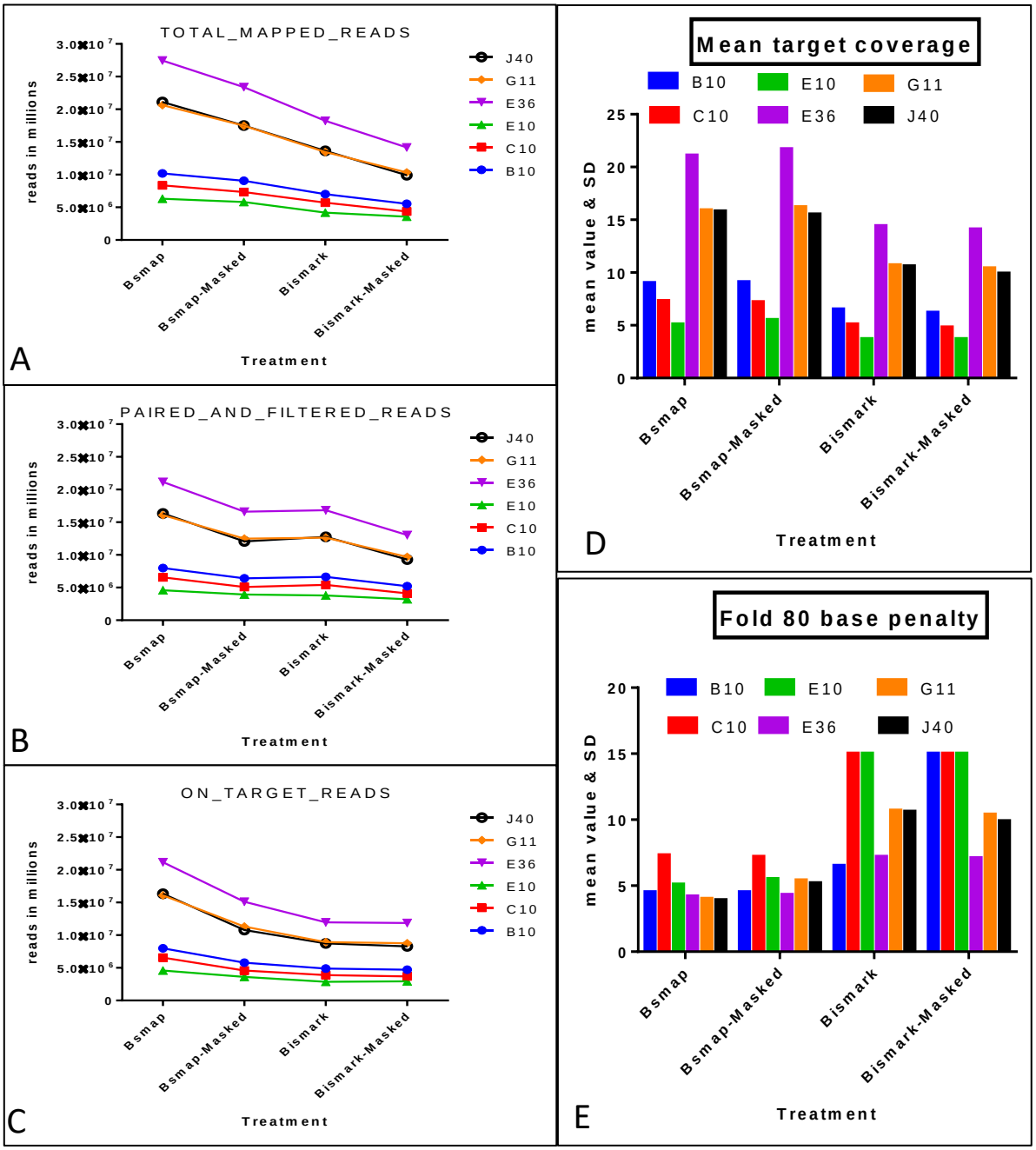


Figure 2.15: Mapping and coverage results.

A) Number of reads (R1+R2) mapped to reference genome. B) Number of reads (R1+R2) after removing the deduplication, filtering for properly mapped and clipped fir overlapping sequence. C) Number of reads (R1+R2) on-target region. D & E) Mean target coverage and Fold_80_Penalty metrics calculated from Picard-tools; Hybrid Selection (hs)

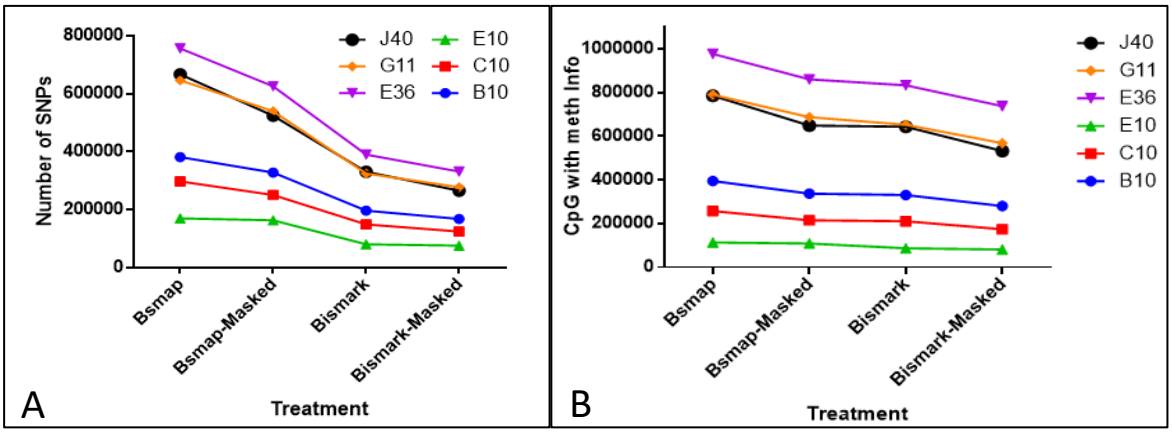


Figure 2.16: A) SNP and B) DNA methylation calling results obtained from

The quality of the methylation calling was evaluated and compared by the calculation of the gene body methylation rate (GBMR) for each aligner. This was then compared to GBMR obtained from the gold standard BS-Seq method used in previously published work. Similar means of GBMR were obtained for all the treatment (Figure 2.17). In addition the distribution of methylation rate among 20 quantiles was also calculated (Figure 2.18). All the results obtained in the same range of what was previously published for *Crassostrea gigas* (Rondon et al., 2017; Xiaotong Wang et al., 2014).

In conclusion, the mapper BSMAP used with a non-masked genome was the best solution for our data. This BSMAP solution was therefore included in a Nextflow pipeline to run autonomously (a fast and scalable way) all the bioinformatics steps from the cleaning of the raw reads to the SNP and DNA methylation calling.

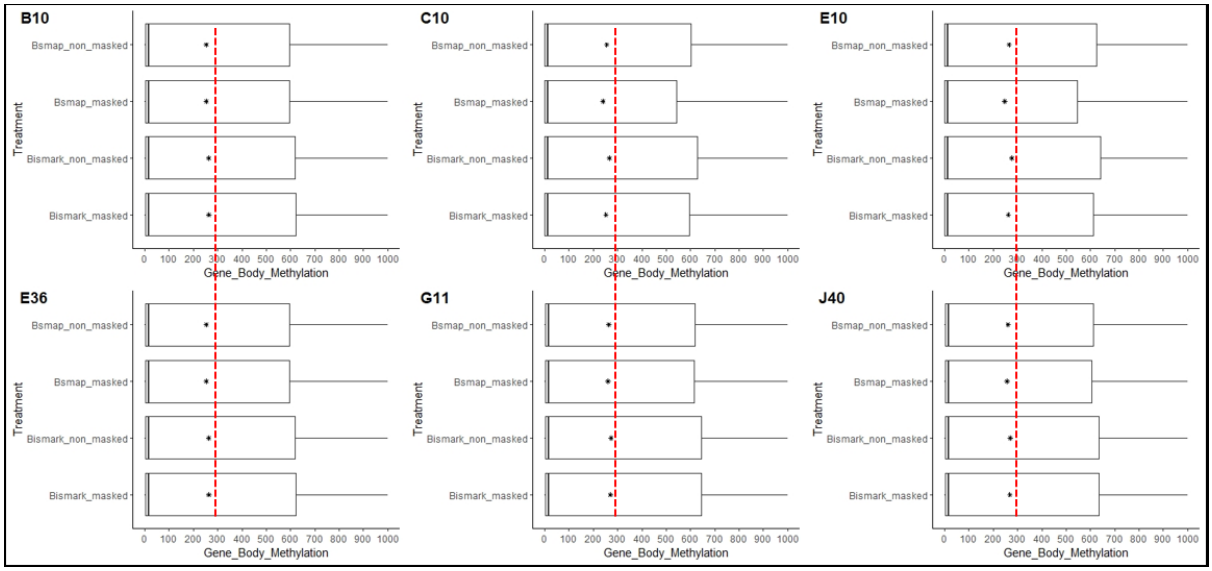


Figure 2.17: Mean gene body methylation.

For the different treatment (BSMAP masked and BSMAP non-masked, BISMARK masked, BISMARK non-masked).

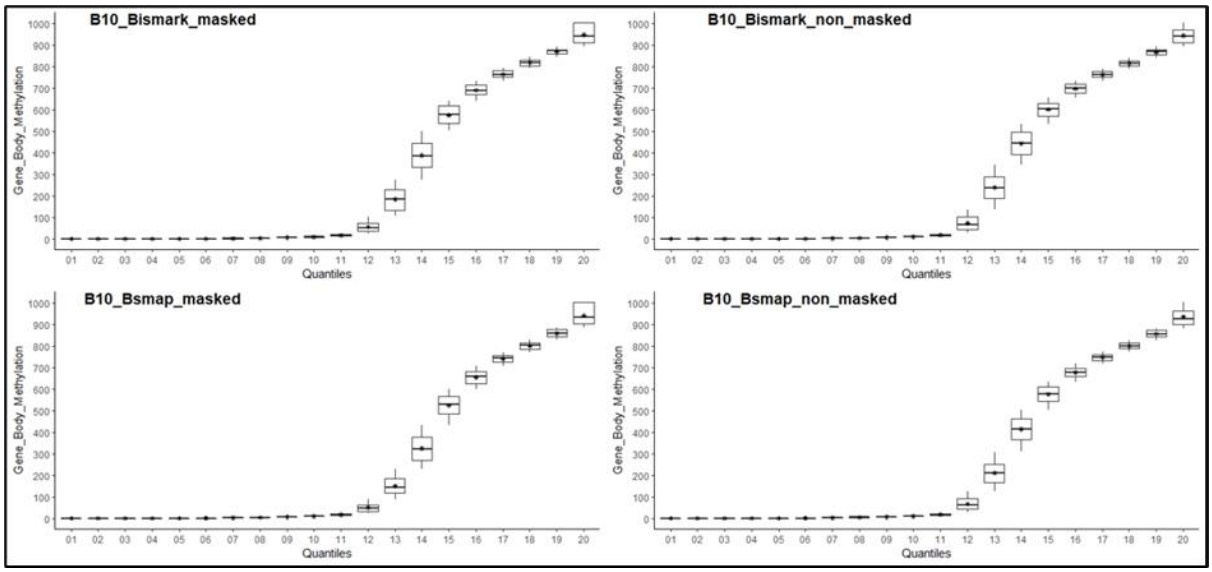


Figure 2.18: Gene body methylation distribution in 20 quantiles.

For the different treatment (BSMAP masked and BSMAP non-masked, BISMARK masked, BISMARK non-masked).

4. Discussion

The aim of this chapter was to present the set up and optimization we did to apply to the "SeqCap Epi Enrichment System" protocol and associated bioinformatics pipeline for our scientific aims. We chose this method because it allows us, i) to capture the region of functional interest exons) and ii) to characterize in the same time and at a reasonable cost the genetic and epigenetic information. This method is quite interesting for our studied model, Pacific oyster. DNA methylation in Pacific oyster is mostly distributed in exonic and intronic parts of genes (de Mendoza et al., 2020; Riviere et al., 2017; Xiaotong Wang et al., 2014). Consequently, by using this method, we were able to find an interesting trade-off between the quantity of genetic and epigenetic information obtained, the quality of this information and the cost per individual. However, as this methods was used for the first time in Pacific oyster the protocol needed to be optimized. These optimizations were done from the DNA extraction to the selection of the best aligner to finally obtain SNPs and DNA methylation information. Thus, we were later able to study the genetic and epigenetic determinants of oyster resistance to POMS. Finally the optimization performed showed that this protocol allowed us to capture in the same time the genetic and epigenetic information of 8 individuals over more than 65 % of all the exons of oyster. Further improvement would be possible and are discussed below.

The main point that would be enhanced concern the homogeneity of the capture between exonic sequences. The main problem encountered was the non-homogeneous coverage obtained and identified by the metric *Fold_80_Penalty*. For a sequencing experiment, this metric indicates that in order to reach the average coverage for 80 % of the exons, the libraries should be sequenced several times more (between six for low depth sequencing to 4 for higher depth sequencing in our case). Interestingly, this value was higher in samples with lower number of reads. The reasons behind this were further interpreted. First, the results showed that some regions of the exome had very high coverage while others had very low coverage (below 8X). This could arise from the fact that probe design was made on the base of the genome assembly published in 2012. This genome is known to be imperfect and it is reasonable to think that regions with high coverage could correspond to repeated regions that have not been well annotated and/or assembled. Consequently, these regions were not discarded during the probe design. Consequently, these repeated regions could affect the

performance of the capture by competing with other regions during capture and lead to capture bias (Roche personal recommendation). In the new published chromosome level genome of the Pacific oyster a higher number of repeats were annotated by comparison to 2012 genome (Peñaloza et al., 2021). A future optimization would be the remove of these sequence from the probe panel in order to reduce this over capture bias.

Another possible reason for this high *Fold_80_Penalty* metric could be that some regions had zero coverage since no sequence were captured (around 20 % of exons had zero coverage). This could result from the fact that the probe design was imperfect. This imperfection or absence of a sufficient complementarity can be due again to the quality the genome assembly. Alternatively this absence or bad complementarity between the probe and the target would come from genetic divergence between the individual used for the probe design (the oyster was from china) and our samples. This divergence could be due to a phenomenon of rapid evolution of gene sequences, or to presence absence of gene.

If the bench part of the exome capture procedure could be further enhanced it is also the case of the bioinformatics. Reads mapping to reference genome is a very important step, especially with BS-seq data. We carefully analysed and test different aligner to select the best (BSMAP in our case). However, the alignment of our test samples showed that only about 46 % of the reads were mapped to the reference genome. This represents a quite low percentage of mapping, although this results is similar or even higher to what is classically seen in other studies (Olson & Roberts, 2014; Xinxing Wang et al., 2021). Usually these unmapped reads are explained by the Bisulfite conversion and too strong differences between the read and the reference. However this unmapped reads could also hold other useful biological information and may reveal source of potential “contamination”. Recently, a paper proposed several strategies to aid in the analysis of this kind of data (Laine et al., 2019). As an example, these unmapped reads could project fruitful source of undiscovered symbiont or parasite of the host we study. Or it could contain information about the sequences of genes that have not been sequenced previously. Additionally, the unmapped reads could result from the differences between the reference genome and the reads coming from oysters that could be genetically distant. Interestingly, when mapping the same reads to two different genomes of oyster (one published in 2012 (G. Zhang et al., 2012) and one recently published in 2021 (Peñaloza et al., 2021)) there were difference in percentage of reads mapped, with more reads mapped to the

newest reference genome (Louis Boismorand, personal communication). This last genome come from a European oyster while the former genome come from a Chinese oyster. However we cannot exclude that such difference can also be due to the huge difference of assembly quality between these two reference genome.

In conclusion, the “Seq Cap Epi Enrichment System” protocol allowed us to capture the genetic and epigenetic information of more than 65 % of exons. We saw that some regions were under sequenced, others oversequenced but the overall results were sufficient to provide information never provided before; the deep characterization of the genetic and epigenetic variation into the functional part of the genome for hundreds of individuals at a reasonable cost. However, a further enhancement of the probe design would be achieved thanks to the new genome and the identification of problematic sequences from our experience.

Chapter 3

Determinants and relative weight of genetic and epigenetic variation in the resistance of the oyster *Crassostrea gigas* to the Pacific Oyster Mortality Syndrome

Chapter 3 : Determinants and relative weight of genetic and epigenetic variation in the resistance of the oyster *Crassostrea gigas* to the Pacific Oyster Mortality Syndrome

3.1 Context and objective:

In summer 2008, major mass mortality events affected spats (around one year old) of *C. gigas* all over French coasts when seawater temperatures were about 17 °C (Bédier et al., 2009). These mortalities were parallel with the appearance of a new variant of herpes-like virus named OSHV-1 µVar. The acronym POMS (Pacific Oyster Mortality Syndrome) was used to describe these mass mortalities, which now has become panzootic.

As we saw in the introduction chapter (section 1.4) research efforts have enabled to better understand the POMS, which is considered as a polymicrobial and multifactorial disease. de Lorgeril et al. (2018) study deciphered the mechanisms that underlie the complex pathosystem affecting the juvenile oysters. They showed that the presence of the virus (OsHV-1 µVar) is the primary step for the onset of infection. In the susceptible oyster, an intense replication of the virus is needed for disease development. However, resistant oyster develops a strong and fast antiviral response by inducing genes involved in antiviral pathway (de Lorgeril et al., 2018).

At the molecular phenotype level, a transcriptomic study performed on biparental families displaying contrasted susceptibility to the POMS revealed that the early induction of the antiviral response is a hallmark of oyster resistance (de Lorgeril et al., 2018). This result was further confirmed by the identification of putative transcriptomic signatures specifically expressed in naïve (e.g. never exposed to POMS) individuals of resistant families (de Lorgeril et al., 2020). Interestingly, early life exposure to a diversified non-pathogenic microbial environments (e.g. immune shaping) was shown to alter the immune transcriptome of *C. gigas* juvenile in a way that it significantly increases the resistance to the POMS of the family studied (Fallet et al., 2022). A close transcriptomic phenomenon with the same phenotypic outcome was also identified in response to an immune priming induced by the injection of Poly I:C (Lafont et al., 2020). These transcriptomic results highlight that the resistance to

POMS is a complex and plastic trait, supported by several genes, subjected to environmental influences, but with an inheritance component.

Moreover, a significant additive genetic component has been identified with evidence of microbiota and epigenetic been involved in resistance (Azéma et al., 2017; Clerissi et al., 2020; de Lorgeril et al., 2018; Dégremont, Garcia, et al., 2015; Gutierrez et al., 2018; Lafont et al., 2020). Genome-wide association studies (GWAS) have identified only few genes associated with the survival, but their exact role have not been further studied (Gutierrez et al., 2018). However, there is no GWAS study implemented on naturally occurring population of oyster. Additionally, there is lack of studies in assessing the role of epigenetic mechanism (DNA methylation) in POMS disease.

Unfortunately, approaches integrating these two components (genetic and epigenetic) remain rare. Here we propose a new framework to study simultaneously the potential role of genetic and epigenetic in shaping a phenotype. Within this context, the objectives of this thesis is to identify genetic and epigenetic signatures of oyster resistance to POMS and to quantify the relative weight of both mechanisms in the phenotypic expression of the resistance. These objectives were addressed by sampling natural oyster populations exposed to different environment and by phenotyping them with an experimental infection that mimic the natural route of infection.

Initially, we decided to answer these objectives within two contexts:

i) At small geographic scale, for this we used the six populations from the bay of Brest. This scale includes two populations from farming areas (high densities of oysters and presence of the virus OsHV-1) and four populations from non-farming areas (low densities of oysters and an undetectable virus).

ii) A wide geographical scale, expanding to three more sites of oyster productions (Marennes-Oléron Bay, Arcachon and Thau Lagoon). From each site, we sampled at least one population of oyster from farming and one from non-farming areas.

3.1.1 Sampling strategy developed:

Samples of wild oyster populations from four different regions of France were collected (bay of Brest, Marennes-Oléron Bay, Arcachon and Thau Lagoon). The sampling was done accordingly to the following dichotomic status: (1) in what called an oyster “farming areas” with high densities of oysters and presence of POMS, (2) in what called an oyster “non-farming areas” with low densities of oysters and no POMS diseases. To bring ecological replication, these two status (farming and non-farming) were present at each location. The sampling scheme was structured over two spatial scales; a small spatial scale centred in the bay of Brest (2 farming and 4 non-farming sites), and a large spatial scale including Marène Olerons, Arcachon and Thau lagoon (each location one farming, one nom-farming site; Figure 3.1). On each site, when possible, 60 individuals were sampled (Table 3.1). Our aim was to sample oysters that were already subjected to a POMS season and that are still in the age windows of susceptibility. As results oyster of 12-16 month old were targeted and sampled in October 2018. In the site of Thau and Vidourle the age of the sampled oysters were significantly lower since these locations were subjected to two seasons of POMS per year, one in spring and one in early fall. The oysters that have recruited in the summer 2018 would have been subjected to the POMS event of the early fall before sampling in October. For the population of Brest 1-6, Agnese, Arcachon and Royan the age was known. For the populations of Mimizan and Chaucre their age were unknown and we have targeted small individuals during the sampling. In total 730 individuals from 13 different natural occurring populations (Table 3.1).

After sampling, all individuals were brought to the IFREMER facilities in Palavas and were acclimatized for 14 days. This acclimatization period was used to gradually acclimatize all oysters from different environment to the laboratory condition used during the phenotyping step. During this acclimatization, the oysters from each site were maintained in 50L tanks and were separated from each other to avoid microflora exchange. Oysters were fed twice a day (early morning and late afternoon), with 3 mL of algae (Shellfish Diet 1800 ® Instant Algae). Seawater was renewed continuously with filtered Mediterranean Sea water at a rate of 50L/h. Water tanks were continuously UV-filtered. All along the acclimatization step the temperature was gradually increase from 13°C to 21°C.

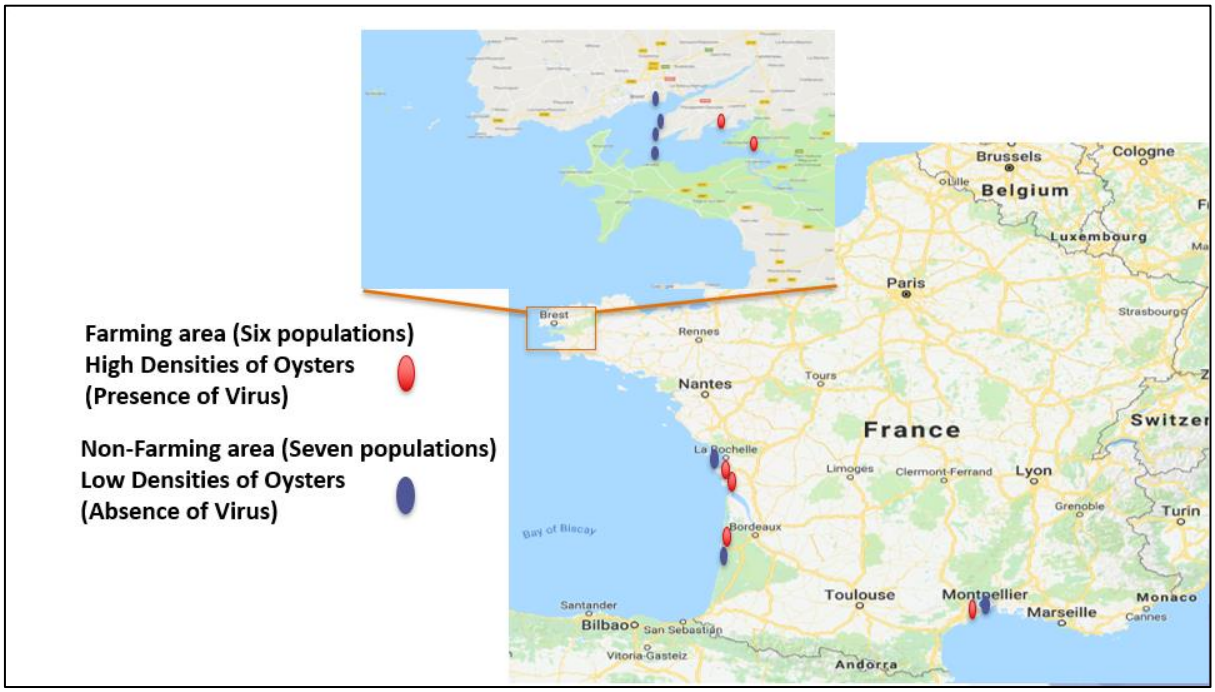


Figure 3.1 : The sampling strategy of natural population of Pacific oyster.

In total 13 population from farming and non-farming areas within three locations in French coast. (Blue for non-farming and red for farming areas).

Table 3.1 : Geographic coordinates for the oyster populations sampled.

Location	Population	Area	Latitude	Longitude	Number of samples
Bay of Brest	Brest 1	Non-Farming	48.379364	-4.446286	61
	Brest 2	Non-Farming	48.341789	-4.441086	59
	Brest 3	Non-Farming	48.322392	-4.454078	61
	Brest 4	Non-Farming	48.296575	-4.451778	56
	Brest 5	Farming	48.32815	-4.321947	60
	Brest 6	Farming	48.34695	-4.338986	59
Bay of Marennes-Oléron	Chaucre	Non-Farming	45.97994	-1.400258	60
	Royan	Farming	45.61677	-1.038838	57
	Agnese	Farming	45.80127	-1.146805	60
Bay of Arcachon	Mimizan	Non-Farming	44.21088	-1.295416	59
	Arcachon	Farming	44.68028	-1.142622	62
Thau lagoon	Thau	Farming	43.39202	3.577774	58
	Vidourle	Non-Farming	43.55615	4.101541	18
	Total				**730**

3.1.2 Results of Phenotyping:

To characterize the resistance or susceptibility of each sampled oyster an experimental infection was performed using a cohabitation approach and a randomized complete block design (Fig. 3.2). This approach starts with the injection of OsHV-1 suspension into donor oysters that will develop the disease and will transmit it through the natural infectious route to oysters of interest (recipient oysters; Figure 3.2). Twenty-four hours after the beginning of the cohabitation, the health status of each recipient oysters, moribund *vs.* alive (*e.g.* susceptible *vs.* resistant phenotypes) was monitored every two hours for 15 days (no mortalities occurred after day 14). An oyster was classified as moribund when it cannot close its valves after 30 seconds of emersion. This checking enabled to sample the susceptible oysters before death (moribund status) to avoid DNA degradation. The resistant oysters were those that were still alive at the end of the experiment.

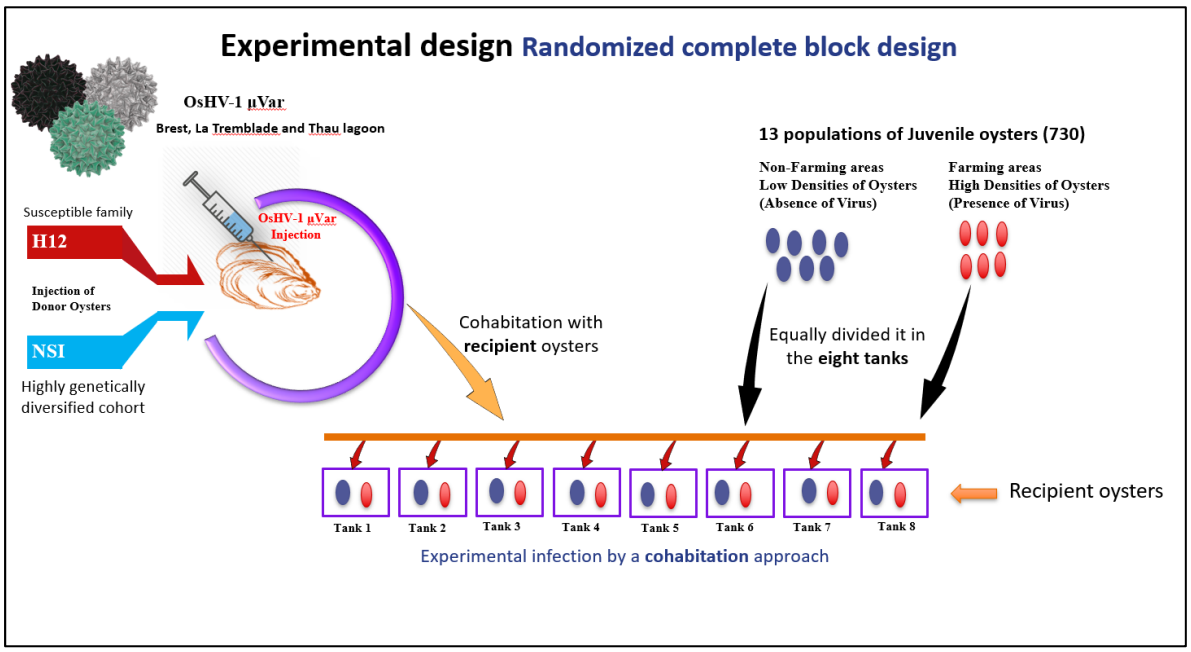


Figure 3.2 : The experimental design of randomized complete block design.

After acclimatization, the 13 populations of oyster where equality divided into eight replicates, then cohabitated with the donner oysters that have been injected with the OsHV-1 µVar.

Once the experiment finished (15 days), of the 730 individuals, 199 had died and were classified as susceptible. The 531 remaining oysters were classified as resistant phenotype. The mortality onset varies from population to population with strong contrasts between them. Some populations exhibit no mortality as populations from farming area Royan and Agnese (100% survival), others had a low survival rate as populations from non-farming area Brest 1 and Brest 3 (30% and 33% survival; Figure 3.3). Hazard ratio analysis statistically confirm this result (*p*value < 4.0376e-52; Figure 3.4) and showed that oysters from areas suffering annual POMS event (farming area) displayed a significantly lower risk of mortality than the other populations (non-farming area). Overall, the populations from farming area (Brest 5, Brest 6, Arcachon, Royan and Agnese) had a significantly higher probability of survival than populations from non-farming area (Brest 1, Brest 2, Brest 3, Brest 4, and Vidourle).

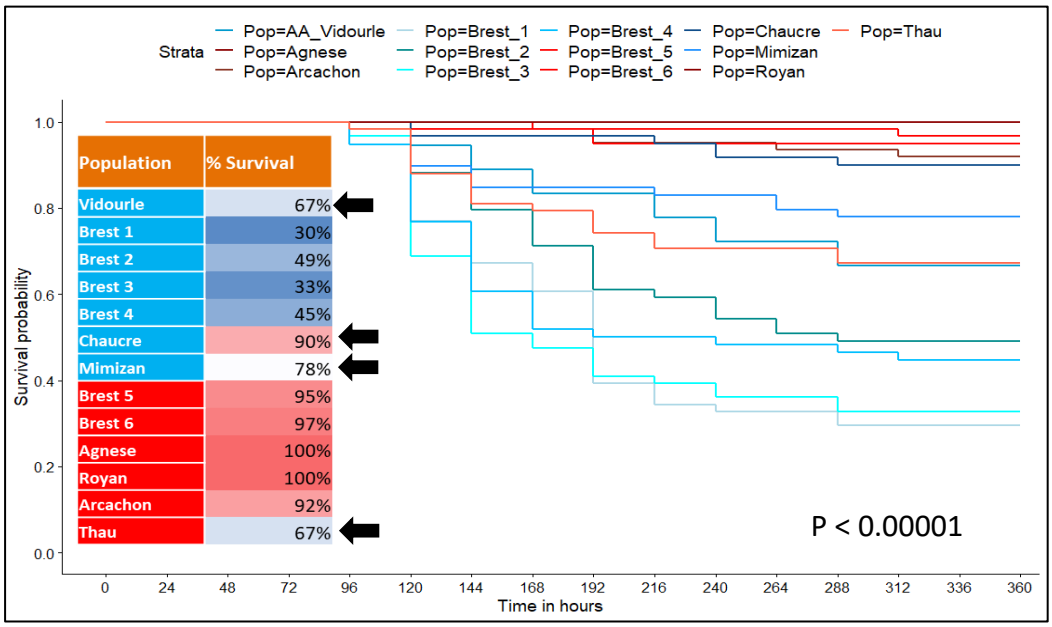


Figure 3.3 : Kaplan-Meier Survival Analysis for the 13 population.

Each line indicates one population; red colour gradients lines represent populations from farming areas, while blue colour gradients lines represent population from non-farming areas time is in hours. In the table attached within the figure shows the survival rates of each population at the end of the experimental infection. The black arrow are the populations with unexpected results.

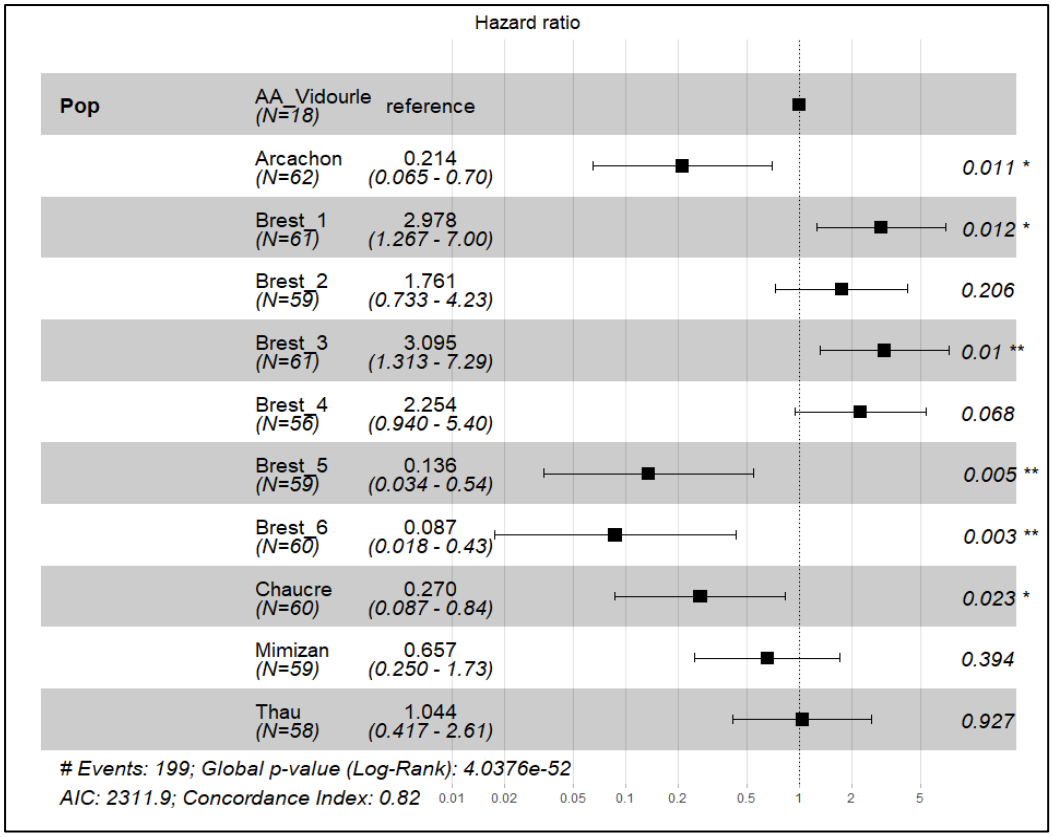


Figure 3.4: Forest plot showing the relative risk of death.

The results are shown only for 11 populations, without the populations Royan and Agnese, because no mortality was observed and the parameter did not converged.

3.1.3 Expected results, obtained results and subsequent change of our strategy

Based on our study design we expected to see strong mortalities in oysters from the non-farming populations and low to no mortalities for those from the farming populations. If the general pattern of mortality follow our hypothesis, unexpected results were also obtained (Figure 3.3; Black arrow); i) a high survival rate for the populations from Chaucre and Mimizan (non-farming population) and ii) the identical susceptibility for the Mediterranean sites, Thau (farming) and Vidourle (non-farming). The age of oysters sampled from the Mimizan and Chaucre sites were unknown, more probably older than two years. On the other side, the oysters sampled from Thau and Vidourle populations were of known age but no POMS event was recorded in summer 2018 in Thau Lagoon because of the harmful algal bloom that occurred in this site. For the populations where the age was uncertain it would be possible to perform a sclerochronology approach for determining the age of each oyster but the unavailability of time lead us to focus on populations where we were sure about the age for genotyping and epigenotyping. This last decision was strengthen by the lower efficiency by comparison to our expectation of the exome capture. The needs for a higher sequencing depth has increased the cost.

Therefore, we decided to focus on the population from the small spatial scale only (six population of Bay of Brest). These populations were those completing all the requirement to produce a qualitative dataset for answering the overall thesis hypothesis and objectives. These results are presented in the following paper that will be submitted to "**Science of total environment**" journal.

3.2 Article resume

In this article, the genetic variation (SNPs) and one component of the epigenetic variation information (DNA methylation at CG context; CpGs) were jointly obtained by optimization of a whole exome capture approach. Genome and Epigenome Wide Association Studies (GWAS and EWAS) were used to identify signature of oyster resistance to POMS. Correlation, MethQTL and variance partition methods were used to quantify the relative involvement of genetic and epigenetic variations in phenotypic expression.

Overall, the work carried out during this paper has enabled us to show: 1) that natural oyster populations differentially exposed to the emerging disease named Pacific Oyster Mortality Syndrome (POMS) display signatures of selections both in their genome (SNP) and in their epigenome (DNA methylation). 2) These signatures are localized in different genes but most of them belong to the same immune related biological processes. 3) The genetic and epigenetic variations are partly correlated and the former was associated in the explanation of a large fraction of the second. 4) The epigenetic variations significantly associated to oyster resistance were independent from the genetic variation and explained a higher part of the phenotypic variation (17.3 and 26.1 % for epigenetic compared to genetic 13.1 and 14.1 %).

These results confirmed that host population facing infectious diseases emergence could rely on genetic and epigenetic variation to rapidly adapt to emerging diseases.

Determinants and relative weight of genetic and epigenetic variation in the resistance of the oyster *Crassostrea gigas* to the Pacific Oyster Mortality Syndrome

Authors

Janan Gawra [1], Julien de Lorgeril [2,3], Yannick Gueguen [2,4], Martin Laporte [5], Fabrice Roux [6], Mathilde Saccas [2], Jean-Michel Escoubas [2], Alejandro Valdivieso [2], Caroline Montagnani [2], Delphine Destoumieux-Garzón [2], Franck Lagarde [4], Marc A. Leroy [2], Philippe Haffner [2], Jean Baptiste Lamy [7], Bruno Petton [8], Céline Cosseau [1], Benjamin Morga [7], Lionel Dégremont [7], Guillaume Mitta [1,9], Christoph Grunau [1]; Jeremie Vidal-Dupiol [2]*

[1] IHPE, Univ Perpignan Via Domitia, CNRS, IFREMER, Univ Montpellier, Perpignan, France

[2] IHPE, Univ Montpellier, CNRS, IFREMER, Univ Perpignan Via Domitia, Montpellier, France

[3] Ifremer, IRD, Université de la Nouvelle-Calédonie, Université de La Réunion, ENTROPIE, Nouméa, Nouvelle- Calédonie, France

[4] MARBEC, Univ Montpellier, CNRS, Ifremer, IRD, Sète, France

[5] Division de l'expertise sur la faune Aquatique, Ministère des Forêts, de la Faune et des Parcs (MFFP), 880 chemin Sainte-Foy, G1S 4X4, Québec, Québec**,** Canada

[6] Laboratoire des Interactions Plantes-Microbes-Environnement, Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement, CNRS, Université de Toulouse, Castanet-Tolosan, France

[7] Ifremer, SG2M-LGPMM, Laboratoire de Genetique et Pathologie des Mollusques Marins, La Tremblade, France

[8] LEMAR UMR 6539, UBO/CNRS/IRD/Ifremer, 11 presqu'île du vivier, 29840 Argenton-en-Landunvez, France

[9] Univ Polynesie Francaise, ILM, IRD, Ifremer, F-98719 Tahiti, French Polynesia, France

* Correspondence: Jeremie.Vidal.Dupiol@ifremer.fr

Abstract

The emergence of pathogens are largely influenced by diverse global changes. The Pacific oyster, *Crassostrea gigas*, is the most exploited oyster species in the world. Since 2008, mass mortality events of juvenile oysters caused by the Pacific Oyster Mortality Syndrome (POMS) have threatened the oyster aquaculture industry. Resistance of *C. gigas* to POMS has demonstrated genetic bases. More recently, it was shown to rely on early transcriptomic response to the viral infection. While data about the involvement of epigenetics in POMS resistance are still scarce.

Here we simultaneously quantified the relative weight and identified the genetic/epigenetic determinants of resistant/susceptible phenotypes from natural oyster populations. A total of 214,263 Single nucleotide polymorphisms (SNPs) and 635,201 polymorphic DNA methylation sites (CpGs) for 102 susceptible and 118 resistant oysters were obtained by sequencing the whole exome capture of bisulfite-converted libraries. We showed that wild oyster populations display signatures of selection in their genome and epigenome to POMS. These signatures were localized in different genes but a high number of these genes belong to immune-related pathways.

While our study confirms the essential role played by the DNA sequence it also shows that other mechanisms (e.g. epigenetic) can interplay with this sequence to encode a resistant phenotype and participate in the expression of resistance. On one side, these results confirm that more holistic approaches of the resistance of host population must be envisioned to have access to most of the mechanisms at stake. On the other side it also demonstrates that epigenetic assisted selection would assist the breeding industry without effect on the DNA sequence.

Introduction:

The emergence of marine and terrestrial pathogens is largely influenced by diverse global changes (Harvell et al., 2002a), such as habitat fragmentation, climate change, pollution, over exploitation, local biodiversity impoverishment or transfer of living organisms (Aguirre & Tabor, 2008). Considering marine diseases, some epizootics significantly disturbed ecosystems, resulted in the extinction of host species (Aguirre & Tabor, 2008), and induced strong economic losses when affecting host species of economic interest (Colleen A. Burge et al., 2014). Understanding by which mechanisms host populations can rapidly adapt to emerging infectious disease pressures appears therefore crucial to propose innovative and eco-friendly management practices.

Host-pathogen interactions are usually characterized by strong reciprocal selective pressures that both partners impose to each other. The case of emerging diseases represents an opportunity to study selective evolutionary processes in action in natural populations, in particular in the case of highly prevalent diseases and massive selection must be engaged by host population which suggest the involvement of all the mechanisms involved in the production of phenotypic variation (Martin et al., 2021) that may be and therefore an important role of genetic and epigenetic variation mechanisms (Danchin, 2013).

Cultivated marine species are often subjected to severe infectious diseases outbreak (Barbosa Solomieu et al., 2015), which makes them interesting models for the study of rapid adaptation to emerging pathogens. This is especially the case when species are cultured in the natural environment closely related to the natural life cycle of wild organisms and without any possibility of control measures. These models can therefore be envisioned as a real time real world evolution experiment of hosts submitted to pathogen pressure. This experiment will also benefit from the extensive resources usually available for bred species and their pathogen (e.g. reference genome, experimental facilities and procedure, general ecological ad physiological knowledge etc.).

In this context, the host-pathogen interaction leading to the Pacific Oyster Mortality Syndrome (POMS) displays all the characteristics needed to address the question of rapid adaptation to emerging pathogen. *Crassostrea gigas*, the cupped oyster is bred worldwide and was in 2019 the first cultivated mollusc species in the world with 6.1 million of tons produced (FAO, 2021). In France, *C. gigas* plays an important socio-economic and

environmental role. Imported from Japan and introduced in the late 60s for its breeding properties and high adaptability, *C. gigas* has successfully replaced in farms the endemic species *Ostrea edulis* that was decimated by two successive infectious diseases outbreak (Buestel et al., 2009; Pernet et al., 2016). Because *C. gigas* has perfectly adapted itself to its new environment, it extensively colonized French coasts and has developed extensive wild populations co-occurring with bred ones (Lapègue et al., 2006). Recurrent summer mortalities of spat and juveniles have been observed in farming areas over the years. Herpes-like virus was associated with these mortalities (Renault et al., 1994, 2000). However, in 2008, a significant increase of this phenomena occurred specifically affecting less than one-year old spats and inducing increased mortalities ranging from 40 to 100% (Bédier et al., 2009). These massive mortalities were associated to the emergence of a new variant of the Ostreid Herpes virus 1 (OsHV-1), the OsHV-1 micro Variant (OsHV-1 µVar; Segarra *et al.*, 2010). This annual disease became panzootic and is now called POMS (EFSA, 2015).

POMS is a polymicrobial disease that is influenced by a series of factor including temperature (Elodie Fleury et al., 2020; Pernet et al., 2012; Renault, Bouquet, et al., 2014), oyster age (Dégremont, 2013b), physiological status (Pernet et al., 2019) and genetic background (Dégremont, 2011) (for review see (Bruno Petton et al., 2021b)). Recent progresses were made in the understanding of successive events leading to oyster death related to POMS. It is initiated by an infection of OsHV-1 µVar that causes a strong and rapid immune-compromised state of the host (de Lorgeril et al., 2018). This primary infection is followed by the colonization of oyster's tissue by opportunistic bacteria that lead to a lethal bacteraemia (de Lorgeril et al., 2018).

The genetic determinism of oyster resistance has been investigated, heritability values between 12% to 63% were estimated and significant additive (epi)genetic components were identified (Azéma et al., 2017; Camara et al., 2017; Dégremont, Garcia, et al., 2015; Dégremont, Lamy, et al., 2015; Gutierrez et al., 2018, 2020). While data about the involvement of epigenetics in POMS resistance are still scarce (but see (Fallet et al., 2022)), some studies focusing on the oyster's genetic determinants were published (Azéma et al., 2017; Dégremont, Garcia, et al., 2015; Gutierrez et al., 2018). A genome-wide association study (GWAS) recently confirmed that resistance to OsHV-1 is polygenic in nature and has identified a significant QTL affecting oyster resistance in the linkage group 6 (Gutierrez et al., 2018). In addition, a transcriptomic study performed on bi-parental families displaying

contrasted susceptibility to the POMS revealed that the early induction of the antiviral response is a hallmark of oyster resistance (de Lorgeril et al., 2018). This result was further corroborated by the identification of putative transcriptomic signatures specifically expressed in naïve (e.g. never exposed to POMS) individuals of resistant families (de Lorgeril et al., 2020). Interestingly, early life exposure to a diversified non-pathogenic microbial environments (e.g. immune shaping) was shown to modify the immune gene expression of *C. gigas* juvenile in a way that it significantly increases the resistance of the family studied (Fallet et al., 2022). A close transcriptomic phenomenon with the same phenotypic outcome was also identified in response to an immune priming induced by the injection of Poly I:C (Lafont et al., 2020). These transcriptomic results highlight that the resistance to POMS is a complex and plastic trait, supported by several genes, subjected to an inheritance component under environmental influence. All these characteristics suggest that ongoing adaptation to POMS should be considered as a dynamic biological system which includes both genetic (*i.e*. DNA sequence) and non-genetic (*i.e.* epigenetic) components (Cosseau et al., 2017).

In this study, we aimed at identifying genetic and epigenetic signatures under massive selection induced by POMS. To do so, we searched for genomic and epigenomic signatures of differences among natural oyster populations exposed to contrasted infectious pressures. Genome and Epigenome Wide Association Studies were used to identify signatures of oyster resistance to POMS. We used a whole exome capture approach to jointly study the genetic variation (Single Nucleotide Polymorphisms; SNPs) and one component of the epigenetic variation (DNA methylation in CG context; hereafter CpGs). Subsequent correlation, MethQTL and variance partition methods allowed us to quantify the relative contribution of genetic and epigenetic variations underlying adaptation to POMS. These variations occurred mostly within the genes involved in the immune response but they are essentially carried by different genes.

Material and Methods:

Sampling strategy:

Wild juvenile oysters of the species *Crassostrea gigas* were collected in the Bay of Brest (France). In total, a maximum of 60 individuals from six natural populations were collected (n=356; Fig. 1A; Supplementary file 1A). While two populations were located in what called an oyster "farming areas" (defined as high densities of oysters and annual event of POMS), the four other populations were located in what called an oyster "non-farming areas" (defined as low densities of oysters and absence of POMS event). Juvenile oysters recruited in summer 2017 were collected in October 2018 after the 2018 POMS event. This sampling design enabled to collect individuals exposed to contrasted environments: non-farming areas (no POMS) and farming areas (POMS) expected to contain a high proportion of susceptible (Fig. 1A; population B1 to B4) or resistant (Fig. 1A; population B5 and B6) oysters, respectively. All individuals were then brought to IFREMER facilities in Palavas-les-Flots (Montpellier, France) where they were acclimatized in a 45L tanks for 14 days. In each tank, seawater temperature was maintained at 21°C, continuously UVC-filtered (BIO-UV) and renewed (30%/h). During this acclimatization period, all populations were separated from each other (separate tanks for each population) and were fed *ad libitum* using Shellfish Diet® 1800 Feeds (Reed Mariculture Inc.).

Experimental infection:

In order to qualify each oyster as resistant or susceptible, we performed an experimental infection mimicking a POMS event. For this purpose, we used a randomized complete block design composed of eight tanks (replicates) of 45 litres each. Each tank was placed in a water bath where the temperature was maintained at 21°C using a chiller/heater apparatus (AQUAVIE ICE 3000). In each tank, a water pump (Aquarium System, Maxijet 1000 L/h) and air bubbling produced water motion and maintained the O_2 level at saturation.

To mimic the POMS, a cohabitation protocol was used as previously describe (D Schikorski et al., 2011). This approach starts with the injection of OsHV-1 suspension into donor oysters that will develop the disease and will transmit it through the natural infectious route to oysters of interest (recipient oysters; Fig. 1B). The ratio between donor and recipient oyster was 1/1. The donor oyster population was composed of 50% of the susceptible H12 family (Azéma et al., 2017) and 50% of a genetically diversified standardised oyster spats

Ifremer (Petton *et al.*, 2013). They were infected by the injection of 200 µL of OsHV-1 suspension (6.00E+7 OsHV-1 genomic units). The viral suspensions were an equimolar mix of viral suspensions extracted from infected oysters collected from three localities (the Rade de Brest, La Tremblade and Thau lagoon). These viral suspensions were prepared as previously described (D Schikorski et al., 2011).

Immediately after OsHV-1 injection into donors, recipient and donor oysters were equally distributed in each of the eight experimental tanks. The disease progression was monitored by checking the health status of donor oysters twice a day during the first week of the experiment. Twenty-four hours after the beginning of the cohabitation, the health status of each recipient oysters, moribund *vs.* alive (*e.g.* susceptible *vs.* resistant phenotypes) was monitored every two hours for 15 days (no mortalities occurred after day 14). An oyster was classified as moribund when it cannot close its valves after 30 seconds of emersion. This checking enabled to sample the susceptible oysters before death (moribund status) to avoid DNA degradation. Dead donor oysters were removed in the course of the experiment, then after 192 hours all remaining donor oysters were removed. The resistant oysters were those that were still alive at the end of the experiment. The flesh of susceptible and resistant oysters was removed from the shell and was immediately snap-frozen in liquid nitrogen and stored at -80 °C until DNA extraction.

Viral load quantification (OsHV-1)

During the first week of the experiment, 1mL of seawater from each tank was sampled daily for viral load quantification. The OsHV-1 DNA was extracted from 200 µL of water using the QIAmp DNA mini Kit and following manufacturer instructions (QIAGEN). Quantitative PCR was performed with 5µL of DNA accordingly to a previously published protocol (Webb et al., 2007).

DNA extraction

Oyster flesh was ground in liquid nitrogen using 50 mL stainless steel bowls and 20-mm-diameter grinding balls. The vibrational frequency used was 30 oscillations per second for a total grinding time of 30 seconds (Retsch MM 400 mill). The resulting powder was used for DNA extraction using the NucleoSpin® Tissue kit following manufacturer instructions (MACHEREY-NAGEL GmbH & Co. KG). Elution buffer was pre-heated (70 °C) for higher DNA quality according to manufacturer instructions. DNA quantity and purity were checked with a

Nanodrop One spectrophotometer (Thermo Scientific). DNA quality was checked by 0.8% agarose gel electrophoresis. The extracted DNA was stored at -20°C until the next use.

Exome capture and Illumina sequencing:

The *C. gigas* exome was captured using the SeqCap Epi Enrichment System protocol (Roche Sequencing Solutions, Inc.; Wendt et al., 2018). In order to capture the region of interest (exons), probes complementary to the whole exonic regions (annotated with reference genome V9 of *Crassostrea gigas*; Zhang et al., 2012) were developed. To ensure optimal coverage of the 5' and 3' ends of each exon, probes were designed to cover the 100 base pairs (bp) upstream and downstream to each exon starts/ends. The genomic regions covered by a probe are provided in Supplementary file 2. Probe design and synthesis were developed by Roche Company.

Exome capture of bisulfite converted libraries were done according to manufacturer instructions (Wendt et al., 2018; check Supplementary file 3 for complete protocol). Briefly, genomic DNA fragmentation was performed on one microgram of oyster DNA in addition to phage lambda DNA as a spike-in control for bisulfite conversion efficiency (GenBank Accession NC_001416). DNA Fragmentation was achieved by sonication with the Covaris S220 apparatus (Covaris, Inc.) using the following custom parameters (Peak Incidence Power: 175, Duty factor: 10, Cycle / Burst: 200, Duration: 70 seconds) to produce fragments of 200 bps in average. After end repair and A-tailing, methylated indexed adapters were ligated to each end of the fragmented DNA. Then 20 µL of cleaned DNA fragments (Ampure beads procedure Beckman Coulter, Inc.) were subjected to sodium bisulfite conversion using the EZ DNA Methylation-Lightning Kit following manufacturer instructions (Zymo Research, CA). After a pre-amplification of the bisulfite-converted library, each eight samples were pooled (equimolar) and subjected to exome capture through their mix with probes complementary to exonic sequences and attached to biotinylated beads. The capture reaction was done at 47°C for 45 h in a thermal cycler (Mastercycler Ep Gradient; Ependorph). After cleaning and elution a final post-capture PCR amplification of 14 cycles was performed. The PCR parameters were: Step 1: 45 seconds at 98°C (long denaturation); Step 2: 15 seconds at 98°C (Denaturation); Step 3: 30 seconds at 60°C (Primer hybridisation). Captured bisulfite-converted Libraries were sequenced either using an Illumina NextSeq 550 system (PE 2 x 150 bp) or an Illumina NovaSeq S1 6000 system (PE 2x 100 bp). Sequencing was design to reach

30x sequencing depth per sample. For both sequencing platforms, 25% of the Phix genome was added to the multiplexed libraries to increase nucleotide diversity and optimized sequencing quality.

SNPs and DNA methylation calling

After Illumina sequencing, the quality of the raw reads were checked with FastQC (Comprehensive QC; v0.53; Andrews, 2010). Adapter trimming and quality filtering were done with TrimGalore (v0.4.0; Krueger, 2015). To estimate bisulfite conversion efficiency, we first aligned the filtered and trimmed reads to the phage lambda genome using BSMAP (v2.90; Xi and Li, 2009), then *methratio.py* function from BSmapz (v1.1.3; Zynda, 2018) was used to estimate the methylation ratio for each bam file produced previously. BSMAP (v2.90; Xi and Li, 2009) was used to align filtered and trimmed reads to reference genome V9 of *Crassostrea gigas* (Zhang et al., 2012). Before SNPs and DNA methylation calling, the BAM files were sorted and duplicate removed following different steps (supplementary file 4 Fig. S1): 1) the reads were split in four sets (top strands [++ and +-]; bottom strands [-+ and --]) using 'split' option from the BamTools (v1.0.14; Barnett et al., 2011); 2) Top strands (++ and +-) and bottom strands (-+ and --) were merged to produce 2 set of reads, a top and a bottom bam using the 'merge' function from BamTools; 3) Top and bottom strands were sorted using 'sort' function from SAMTOOLS (v1.9; Li *et al.*, 2009); 4) PCR duplicate were removed with 'MarkDuplicates' Picard (v2.21.1; http://broadinstitute.github.io/picard/); 5) top and bottom read sets were merged back using 'merge' option from BamTools; 6) overlapping read pairs were clipped using 'clipOverlap' BAMUTIL (v1.0.14; Jun *et al.*, 2015). The scripts used are provided in Supplementary file 5.

To maximize the accuracy of SNPs calling a combination of two caller, FreeBayes dedicated to SNP calling form population data (v1.3.1; Garrison and Marth, 2012) and MethylExtract dedicated to SNP calling from bisulfite converted sequences (v1.9; Barturen *et al.*, 2013) were used. Firstly, FreeBayes was used to call all the SNPs present in the dataset (including those due to the bisulfite conversion; parameters: --use-best-n-alleles=2, --use-mapping-quality, --no-partial-observations, --min-repeat-entropy 1). Secondly, MethylExtract was used to call SNP that were not due to the bisulfite conversion (C/T SNP; parameters: minQ=20, minDepthSNV=8, methNonCpGs=0.9, maxStrandBias=0.7, varFraction=0.1,

maxPval=0.05). Finally, only the SNP identified by both callers were kept and used for GWAS analysis (Supplementary file 5)

DNA Methylation calling in the CG context (hereafter CpGs) was performed using MethylExtract (same parameters as mentioned above). All BED files containing the CpGs (reporting the methylation level ranging from 0 to 1) were combined and used as input for EWAS analyses (Supplementary file 5).

GWAS and EWAS Quality control (QC)

According to the best practices for GWAS (Marees et al., 2018), the following filtering criteria were applied under the PLINK environment (v1.9; Chang *et al.*, 2015) : 1) SNPs supported by a coverage of 8x to 150x were kept; 2) SNPs and individuals with a level of missing data above 5% were discarded; 3) SNPs with a minor allele frequency (MAF) below 0.05 were discarded; 4) SNPs displaying a significant deviation from Hardy–Weinberg equilibrium (HWE) in resistant (HWE $P < 1 \times 10e{-}06$) and susceptible oysters (HWE $P < 1 \times 10e{-}10$) were excluded; 5) individuals with ±3 SD (standard deviations) of samples mean heterozygosity rate were discarded; 6) closely related individuals were excluded (if present) to remove cryptic relatedness.

For EWAS analyses, the following quality controls (QC) were performed under the *R* environment (v4.1.0): 1) CpGs supported by a coverage of 8x to 150x were kept; 2) CpGs and individuals with a level of missing data above 5% were discarded; 3) only the individuals that have passed the above genotyping QC were kept.

For both datasets, the absence of genetic and epigenetic structure between the six populations was checked using an analysis of multivariate homogeneity of group dispersions with the '*betadisper*' function from the VEGAN (v2.5-7) R package (Oksanen et al., 2020). Hierarchical clustering analysis (Euclidian method) and permutational multivariate analysis of variance (PERMANOVA) were performed using the '*adonis*' function from the VEGAN package (Oksanen et al., 2020).

Statistical analyses

Phenotyping

Differences of survival of oysters between the six populations were investigated by a Kaplan Meyer approach with the '*survfit*' and *ggsurvplot* function of the SURVIVAL (v3.2-11; https://cran.r-project.org/web/packages/survival/index.html) and SURVMINER (v0.4.9; https://cran.r-project.org/web/packages/survminer/index.html) R packages, respectively. Cox proportional hazard model was run using the '*coxph*' function from the SURVIVAL package in R (v3.2-11) and was plotted by '*ggforest*' function from SURVMINER package in R (v0.4.9). Results were considered significant below the 5% error level.

Genome / Epigenome wide association studies (GWAS/EWAS) analysis

Two phenotypic traits were considered for GWAS/EWAS analysis, either a binary trait corresponding to susceptible *vs* resistance or a semi-quantitative trait corresponding to the survival time (expressed in hours) of an individual after its exposure to the OsHV-1 virus.

GWA mapping was performed by associating SNPs to the binary trait (using a chi-square allelic test with 1 degree of freedom) and the semi-quantitative trait (using an asymptotic version of usual Student's t test) under the PLINK environment (v1.9; Chang *et al.*, 2015). EWA mapping was performed by associating DNA methylation variation at each CpGs with the binary and semi-quantitative traits (linear regression t.test) using '*cpg.assoc*' function from CPGASSOC R package (v2.60; Barfield et al., 2012). For both GWA and EWA analyses, the significant level of association was defined with a false discovery rate (FDR) below 0.05. GWA/EWA mapping results were visualized using Quantile-Quantile plots and Manhattan plots produced with the R package QQMAN (v0.1.8; Turner, 2018). Because a new reference genome assembled at the chromosomal level was recently released (Peñaloza et al., 2021), homemade scripts were used to locate SNPs and CpGs on this new genome (Supplementary file 6).

Gene annotation and enrichment analysis

To identify the candidate genes with suggestive/significant SNPs/CpGs from GWAS and EWAS, we first located SNPs and CpGs in the individual CGI annotation of *C. gigas* genome v9 assembly (G. Zhang et al., 2012). Then, to identify the functional annotation, we intersected these genes with previously performed functional annotation (de Lorgeril *et al.*,

2018; Supplementary file 7). The sequence of genes related to antiviral and immune response were further assessed by InterProScan (P. Jones et al., 2014) for identification of the conserved domains of each individual gene by searching in protein database.

To test whether genes displaying SNPs or CpGs significantly associated to susceptible or resistant oyster belong to specific biological processes, a Rank-based Gene Ontology Analysis with Adaptive Clustering was performed (RBGOA R package; script can be found at https://github.com/z0on/GO_MWU; Wright et al., 2015). The continuous measure of significance used was a signed –log(*p*value). The following parameters were used for the adaptive clustering: largest=0.4; smallest=10; clusterCutHeight=0.25. A biological process category was considered enriched under an FDR of 0.05. REVIGO (http://revigo.irb.hr; Supek et al., 2011) was used to visualise significant categories containing at least one gene displaying a SNP or a DMP significantly associated with the phenotype of resistance or susceptibility.

Genetic and epigenetic correlation and association

Correlative (Mantel test) and association (methylation quantitative trait loci; MethQTL) approaches between both types of variation were adopted to investigate the relationships between genetic and epigenetic variation. The Mantel test based on the correlation coefficient of Spearman, was applied to estimate the correlation between the genetic and epigenetic matrices of dissimilarity. The '*mantel*' function from the VEGAN (v2.5-7) R package was used to estimate the correlation. The association between SNPs and CpGs levels were identified using a linear regression implemented in the R package GEM (v 0.99.4; Pan *et al.*, 2016) according to the following '*Gmodel*' : lm (G ~ M + covariate), where G is the genetic matrix; M is the methylation level matrix and covariate is the phenotypic trait. This model was run with either the binary trait or the semi-quantitative trait to identify DNA methylation level of each CpG best explained by a SNP (methQTL).

Genetic and Epigenetic variation partition

To estimate the relative contribution of genetic and epigenetic variation to phenotypic variation, we used a method developed by Rougeux *et al.* (2019) and applied in Crotti *et al.* (2021). Briefly, genetic and epigenetic variance were surrogated by producing principal components analyses (PCA) on the same datasets that were used for GWA/EWA mapping

analyses, using the '*prcomp'* function under the R v4.1.0 environment. Then, using a forward selection method '*ordistep'* function from the VEGAN (v2.5-7) R package (Oksanen *et al.,* 2020), the best models explaining variance for the binary and semi-quantitative traits were separately obtained with genetic and epigenetic principal components (PC), resulting in four independent models (2 phenotypic traits X 2 genomic/epigenomic PC). The selected PC for genetic and epigenetic models of each phenotypic trait were retrieved and analysed in a partitioning analysis using '*varpart'* function from the VEGAN (v2.5-7) R package (R scripts in the Supplementary file 8). Variation partitioning is a method of using coefficient of determination to fraction the variation of a response variable into four explanatory variables (Borcard et al., 1992). Two of them correspond to the fractions of variance exclusively explained by one of the two explanatory matrices (e.g. genetic or epigenetic), one corresponds to the fraction of variance shared by the two explanatory matrices (e.g. genetic and epigenetic) and the last one corresponds to the fraction of the variance non-explained by the model.

Results

Experimental infection and phenotyping of oyster populations submitted to different selective pressures

To characterize the resistance or susceptibility of each sampled oyster, an experimental infection was performed using a cohabitation approach and a randomized complete block design (Fig. 1A-B).

The first mortality in donor families was observed 24 hours-post injection (hpi). At 192 hpi, the survival rate dropped to 13.5% and 49.5% for the susceptible H12 and NSI donor families, respectively (Supplementary file 4 Figure S2). Quantification of the OsHV-1 µVar load in seawater showed that the viral excretion from the donor oysters reached a plateau 24 hpi with an average number of 1,755 genome copy per µL (±429.4 SD) and peaked at 72 hpi with 7,185 genome copy per µL (±1,855.7 SD). No significant differences of the viral load was detected between the eight replicate tanks (Kruskal-Wallis *p*value=0.2373; (Supplementary file 4 Figure S3). Mortalities in recipient oysters started 72h after the beginning of the cohabitation with donor oysters and were massive between 96h and 168h (Fig. 1C). No significant differences in the rate of mortality was detected between the eight replicate tanks (Log-rank test; *p*value=0.61; Supplementary file 4 Figure S4). The mortality rates, kinetics of mortalities and viral load into the seawater were consistent with previous studies (de Lorgeril et al., 2018; D Schikorski et al., 2011).

Oysters from the two 'farming area' populations displayed a significantly lower risk of mortality compared to the four 'non-farming area' populations (log-rank test $P < 0.0001$, hazard ratio analysis statistically: $P < 2.2141e\text{-}27$; Supplementary file 4 Figure S5). While the two populations from farming areas, *i.e.* facing annual POMS events and from which susceptible individuals are regularly eliminated, contained almost 100% of resistant oysters (94.9 % and 96.7 %; Fig. 1C), the four populations that were that were confronted to a low POMS challenge displayed a low resistance level varying between 29.5 % and 44.6 % (Fig. 1C).

In total, 150 oysters died (42%) while 206 remained alive (58%). Therefore, 150 oysters were considered as susceptible and 206 as resistant. This phenotype was characterized either as a binary trait with a "0" and "1" corresponding to susceptible and resistant individuals, or as a semi-quantitative trait corresponding to the survival time (expressed in hours) of an

individual after its exposure to the OsHV-1 virus (i.e. the whole duration of the experiment for the resistant oysters; Supplementary file 1B).

Taken together, these results showed that the experimental infection successfully discriminates oyster phenotypes. These phenotyping results were then combined into matrices using the binary or the semi-quantitative methods and used as input for association analyses.

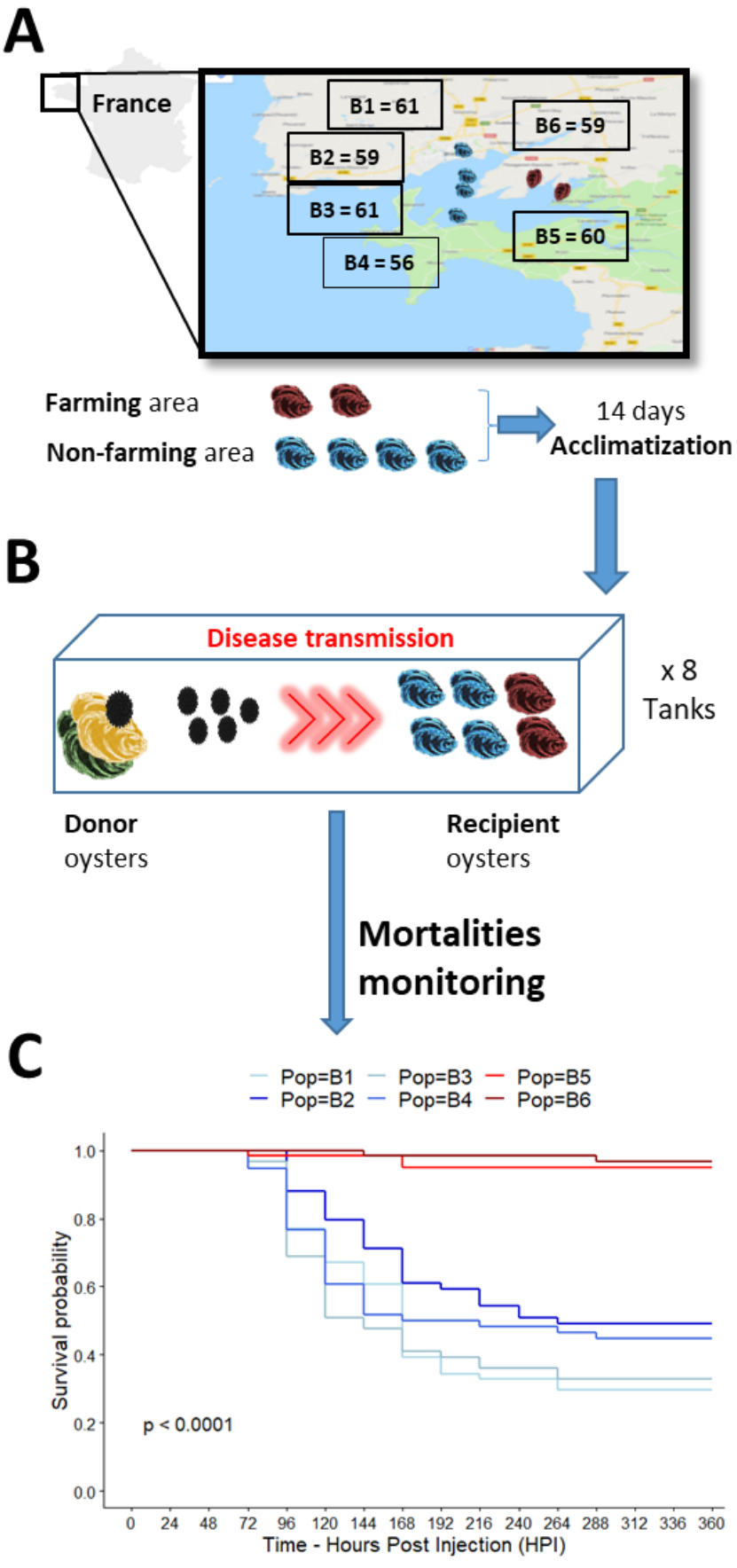

A
France
B1 = 61
B2 = 59
B3 = 61
B4 = 56
B5 = 60
B6 = 59
Farming area
Non-farming area
14 days
Acclimatization
B
Disease transmission
x 8
Tanks
Donor
oysters
Recipient
oysters
Mortalities
monitoring
C
Pop=B1
Pop=B2
Pop=B3
Pop=B4
Pop=B5
Pop=B6
Survival probability
1.0
0.8
0.6
0.4
0.2
0.0
p < 0.0001
0
24
48
72
96
120
144
168
192
216
240
264
288
312
336
360
Time - Hours Post Injection (HPI)

Fig. 1: Experimental infection successfully mimic POMS events.

A) 14 month old *Crassostrea gigas* were sampled in non-farming (No POMS, B1-B4 with blue oyster colour) and farming (annual POMS; B5-B6 with red oyster colour) areas in bay of Brest. In total six populations (356 oysters) were sampled and brought to laboratory facility and acclimatized for 14 days before the experiment. B) Experimental infection was performed accordingly to an eight block Randomized complete block design. Two donor oysters' families (H12 and NSI; yellow and green oysters) have been injected with an OsHV-1 viral suspensions and placed in cohabitation with the six populations of recipient oysters (from 7 to 8 oysters per population and per tank). The health status of each recipient oysters were monitored every two hours for 15 days. Kaplan–Meier survival curves through time. Bleu lines represent populations from non-farming area and red lines populations from farming areas.

Oyster genome and epigenome were deeply characterized by exome-capture

In order to characterize genetic (SNPs) and epigenetic (CpGs) variation in both resistant and susceptible oysters, we performed an exome-capture experiment of bisulfite converted DNA. In total, the exome of 130 resistant and 116 susceptible oysters was captured and sequenced. On average, sequencing yielded in the production of 0.5 – 60 million paired reads per sample (average of 26 million +/- 1 million SD; Supplementary file 1C). Six samples displaying less than 7.8 million of paired reads were discarded from subsequent analysis. On average, 60.2% (+/- 2.7% SD) of the reads were uniquely mapped to the *C. gigas* reference genome (Zhang et al., 2012). Sodium bisulfite conversion was estimated by aligning the filtered and trimmed reads to the phage lambda genome. The efficiency ranged from 99.4%-99.6% (Supplementary file 1C).

SNPs and methylation calling resulted in the identification of 5,110,093 SNPs and 3,449,600 CpGs for the 240 samples analysed. After applying filtering criteria for GWA and EWA mapping analysis, 102 susceptible and 118 resistant oysters characterized by 214,263 SNPs and 635,201 CpGs were kept for subsequent analysis. It is worth to mention, that out of 9,978,551 CpGs dinucleotides in the oyster genome, we capture DNA methylation information at 6.3% of these CpGs. The mean levels of CpG methylation percentage were very similar between all the samples (ranged from 24% to 27%) and similar to what have been previously reported at the exon level (Xiaotong Wang et al., 2014). Additionally, there were no significant difference in the mean between the populations (Supplementary file 4 Figure S6). These results provide a deep characterization of the genetic and epigenetic variation needed for GWAS and EWAS analysis and the understanding of oyster resistance molecular determinants.

Based on PCA and hierarchical cluster analysis, no strong signatures of population structure was detected at the genomic and epigenomic levels (Supplementary file 4 Figure S7-S10). A PERMANOVA analysis allowed to estimate that a very low percentage of the genetic (R^2 = 2.3%, P = 0.091) and epigenetic variance (R^2 = 2.4%, P < 0.001) was explained by differences between the six populations.

Oyster resistance to POMS is associated with genetic variation in antiviral pathways

Visualization of the quantile-quantile plot of p-values resulting from GWA mapping analysis suggested an almost null distribution of p-values, which is in line with the absence of significant effects of population structure detected among the six populations (Fig. 2C & D). GWA mapping revealed one SNP significantly associated with the binary trait (resistant/susceptible) (scaffold1832_479264; A > T, *P* = 5.53E-08; Fig. 2A; Supplementary file 9A) and one SNP with the semi-quantitative trait (time to death in hours) (scaffold364_478394; C > T; *P* = 1.13E-07; Fig. 2B; Supplementary file 9B). While the SNP associated with the binary trait was mapped on chromosome 6 in a gene encoding the SUMO-activating enzyme subunit 2 (CGI_10018487), the SNP associated with the semi-quantitative trait was mapped on chromosome 4 in a gene of unknown function (CGI_10022698). Given this low number of significant SNPs identified and the polygenic nature of POMS resistance (de Lorgeril et al., 2020), we have extended our analysis to the SNPs with a p-value below the value 0.0005, which led to the identification of 113 and 112 SNPs associated with the binary and semi-quantitative traits, respectively. Among these SNPs, 39 were common between the two traits whereas 74 SNPs and 73 SNPs were specific to the binary and semi-quantitative traits, respectively. In total, 186 non-redundant SNPs were associated with resistance, with 111 SNPs located in exons (58 synonymous and 53 non-synonymous), 65 SNPs in introns and 10 SNPs slightly upstream or downstream of annotated genes (Supplementary file 9C). The 186 SNPs were located in 155 genes, with 37 genes being common between the two traits and 58 and 60 being specific to the binary and semi-quantitative traits, respectively (Supplementary file 9D). All these top SNPs were mostly located on chromosomes 6, 7, 10, 3 and 1 (Supplementary file 9C).

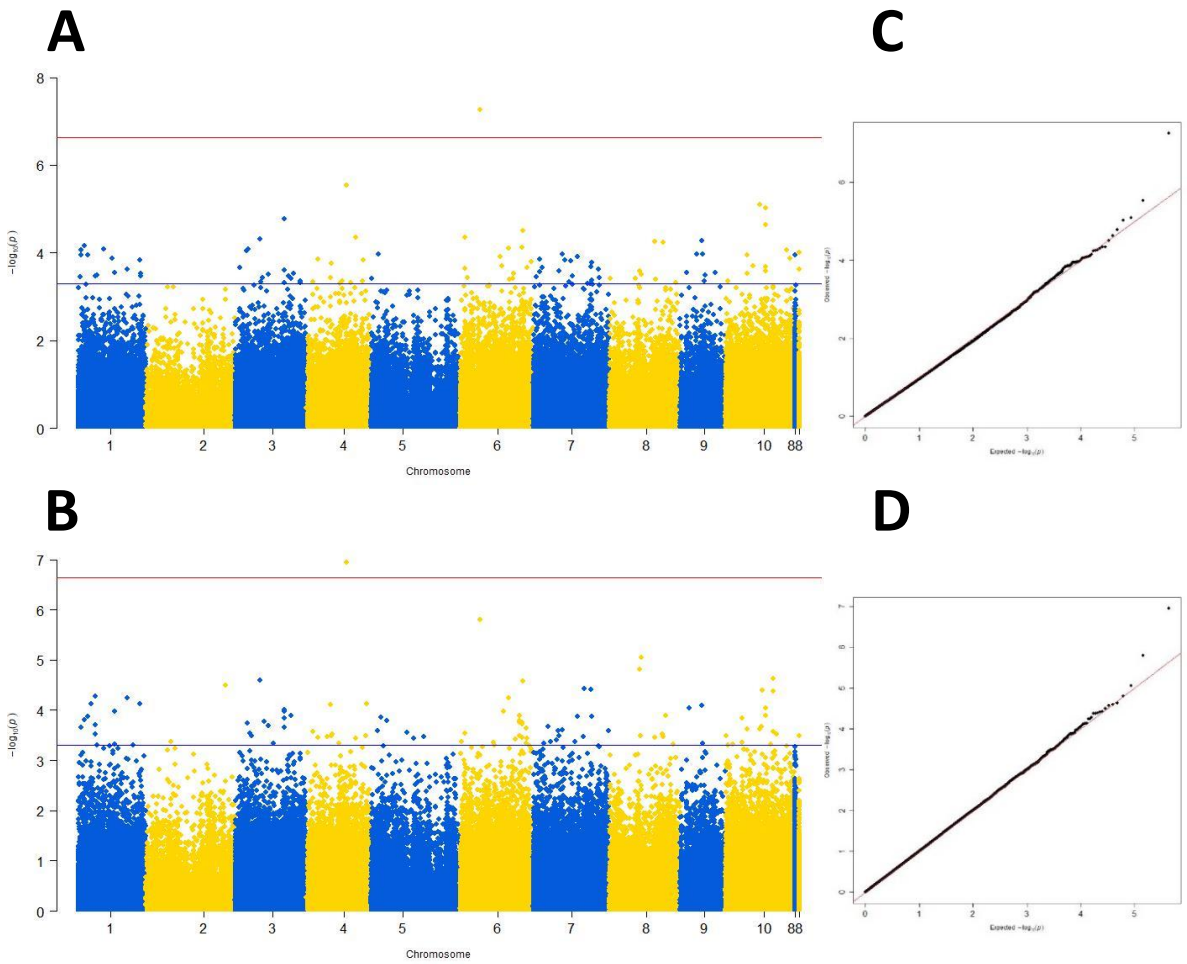


Fig. 2: Manhattan plots and quantile-quantile (QQ) plots for the GWAS.

A and C) Association to the binary trait; **B and D**) association to the semi-quantitative trait. Red line represents the threshold for a FDR<0.05 (significant SNPs) and the blue line the threshold for a *p*value<0.0005 (suggestive threshold). The y-axis shows the negative log10 (p-value), while the x-axis shows the genomic map positions of each the SNP (each dot is a SNP) on the 10 chromosomes of *C. gigas* genome. SNPs with unknown chromosomal location or located in the v9 version of the oyster genome only were grouped on chromosomes 88 and 99, respectively.
Note: chromosome 99 is the yellow color block after the blue 88 block.

In order to identify the enriched biological processes associated with the 186 top SNPs, we performed a gene ontology (GO) term enrichment analysis with the RBGOA package. RBGOA revealed as significant enrichment in biological processes related to immune processes (Supplementary file 10), such as Toll signalling pathway (GO:0008063), cell surface receptor signalling pathway (GO:0007166), response to virus (GO:0009615), response to bacterium (GO:0009617), immune system process (GO:0002376), response to external stimulus (GO:0009605), G-protein coupled receptor signalling pathway (GO:0007186) and response to stimulus (GO:0050896) (Fig. 3). Other functions related to metabolic processes, translation processes, cell cycle and cell structure were also enriched (Fig. 3).

Amongst the significantly enriched biological processes to immunity, we identified genes known to be actors or regulators of the JAK/STAT pathway (e.g. PRMT5, AIMP1, UBA2, and DCST1), the STING/RLRs pathway (e.g. TRIM33, TRAF3), the TLR/NF-KB pathway (e.g. MIB2, MyD88, PRGP, TBK1), the RNAi pathway (Dicer) and pathogen recognition (e.g. C1q, DSCAM, MR) (Fig. 4A).

Taken together, This GWAS analysis enabled the identification of several SNPs significantly or suggestively associated to oyster resistance/susceptibility to POMS. The biological processes and genes affected by these SNPs concerned key antiviral and immune pathways which biologically validated and strengthened the results of this analysis.

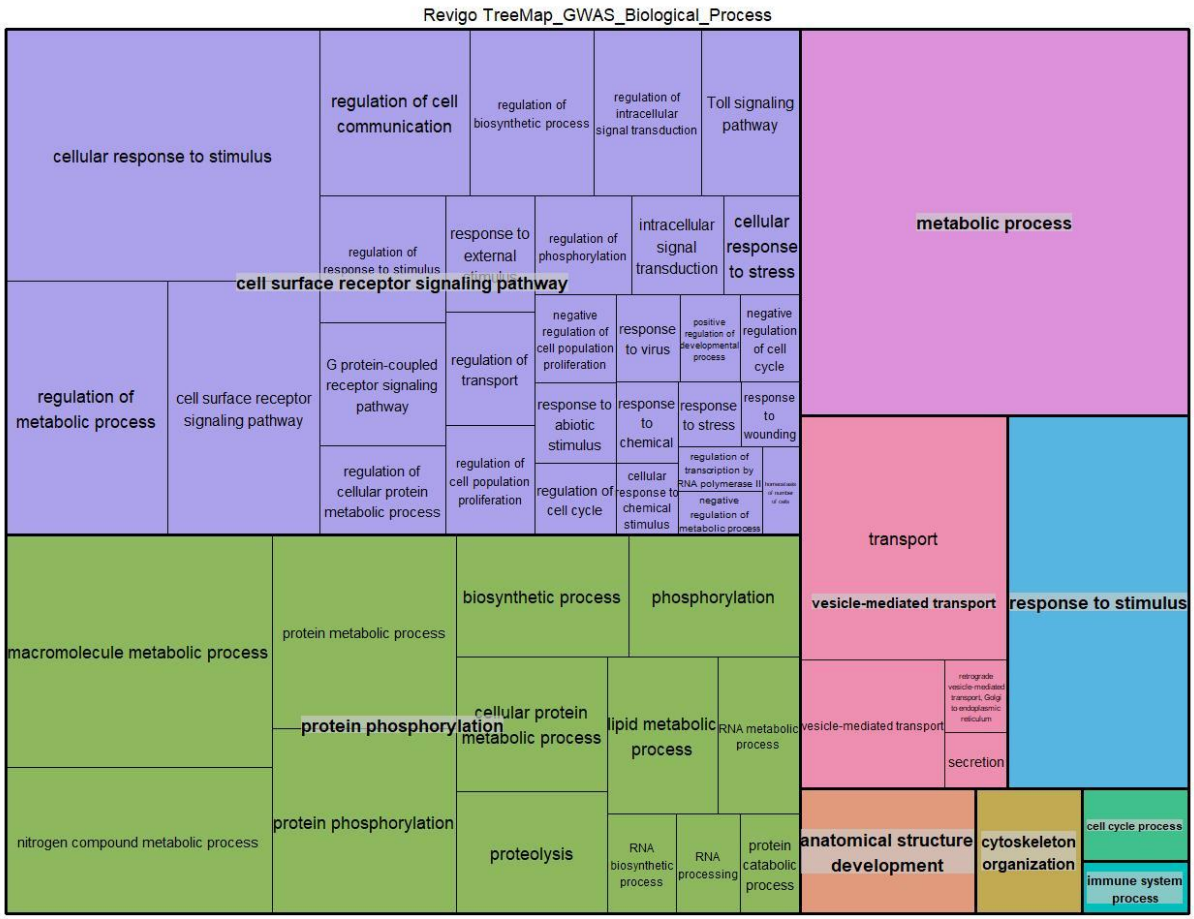


Fig. 3: Oyster resistance to POMS is associated with genetic variation in immune pathways.

GO term of the biological process root enriched (RBGOA) from the set of genes displaying a suggestive or a significant SNPs associated either to the binary or continuous phenotype. RBGOA results were summarized using Revigo treemap. Rectangles size depends on the adjusted *p*value from RBGOA analysis**.** Not all the terms are reported due to space restrictions (see Additional file10 for details).

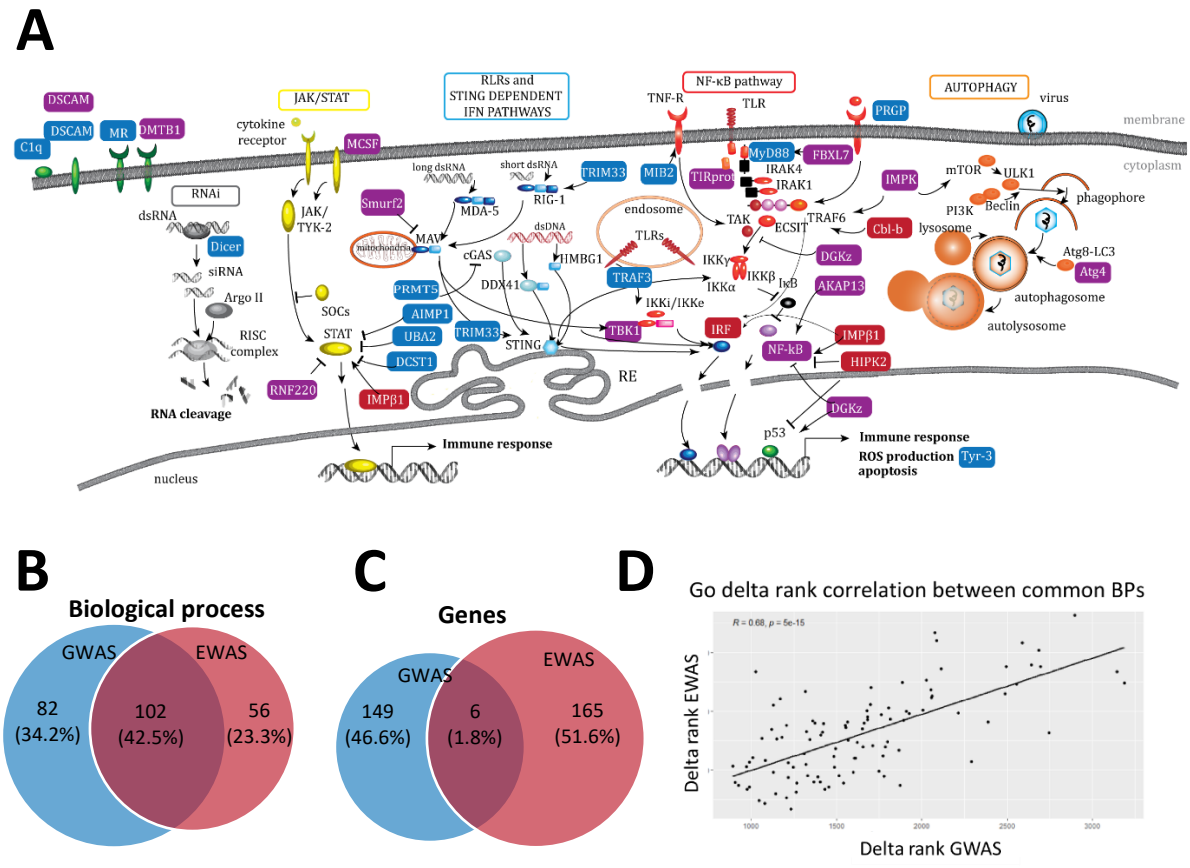


Fig. 4: Genes involved in innate immune pathways display genetic and/ epigenetic variation.

A) Genes of innate immune signalling pathways displaying genetic variation (blue rectangle), epigenetic variation (red rectangle) or a mix of genetic and epigenetic variation (SNP plus CpG or MethQTL, violet rectangle). B and C): Venn diagram illustrating biological processes (B) or genes (C) displaying specific and/or shared genetic and epigenetic variation. D) Correlation between the delta ranks of GO terms significantly enriched from the GWAS and EWAS. Figure A was adapted from Green, et al. 2015.

Oyster resistance to POMS is associated with differential methylation of immune genes

EWA mapping revealed 240 and 226 CpGs significantly associated with the binary and semi-quantitative traits, respectively (Fig. 5A & B; Supplementary file 11A-B). Among the CpGs, 161 were common between the two traits, whereas 79 and 65 CpGs were specific to the binary and the semi-quantitative traits, respectively. Among the 305 non-redundant CpGs identified, 23 CpGs were hypermethylated and 282 CpGs hypomethylated in resistant oysters compared to the susceptible ones. While 292 CpGs were located in exons, nine were located in introns and four in the upstream or downstream region of a gene (Supplementary file 11C). In total, 171 genes displayed at least one CpGs, with 99 genes being common to the two traits and 41 and 31 genes being specific to the binary or semi-quantitative traits, respectively (Supplementary file 11D). Significantly associated CpGs were mainly located on chromosomes 10, 7, 6, and 4 (Supplementary file 11C). Quantile-Quantile plot suggested a departure from a uniform distribution of p-values (Fig. 5C and D), which may result from the weak population structure observed at the epigenetic level. To test the robustness of the significant CpGs identified by CPGASSOC, we estimated for each CpGs the correlation coefficient of Spearman between CpGs and the two phenotypic traits. Spearman's *rho* values were highly correlated with results from EWA mapping (r^2=0.75 when considering all CpGs; r^2=0.81 when only considering the significant CpGs identified by EWA mapping using binary trait), which suggest the identification of true positives by EWA mapping.

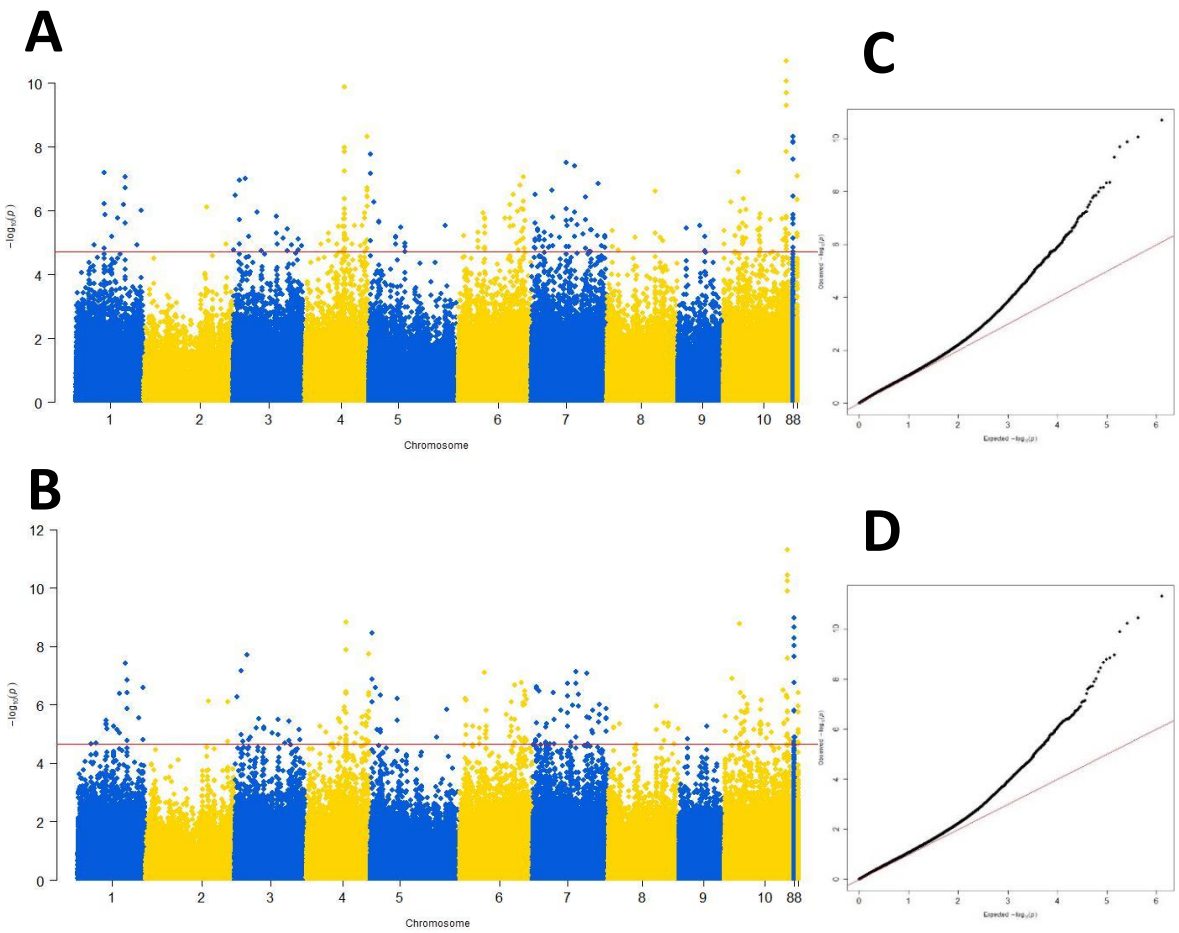


Fig. 5: Manhattan plots and quantile-quantile (QQ) plots of the EWAS.

A and C) Association to the binary trait; **B and D)** association to the semi-quantitative trait. Red line represent the threshold for a FDR < 0.05 (significant CpGs). The y-axis shows the negative log10 (p-value), while the x-axis shows the genomic map positions of each the CpG (each dot is a CpG) on the 10 chromosomes of *C. gigas* genome. CpGs with unknown chromosomal location or located in the v9 version of the oyster genome only were grouped on chromosomes 88 and 99, respectively.
Note: chromosome 99 is the yellow color block after the blue 88 block.

Based on the 305 top CpGs, RBGOA also revealed as significant enrichment in biological processes related to immunity (Supplementary file 12), such as Toll signalling pathway (GO:0008063); cell surface receptor signalling pathway (GO:0007166), response to stimulus (GO:0050896), regulation of autophagy (GO:0010506) and negative regulation of response to stimulus (GO:0048585) (Fig. 6). A significant enrichment of biological processes linked to the metabolism, the cell cycle and tissue structuration was also detected (Fig. 6).

Similar to GWA mapping results, we identified immune genes known to be actors or regulators of the JAK/STAT pathway (e.g. MCSF, RNF220, IMPβ1), the STING/RLRs pathway (e.g. Smurf2 and TBK1), the TLR/NF-KB pathway (e.g. TIRprot, FBXL7, IMPK, Cb1-b, DGKz, AKAP13, AIMP1, IMPβ1, HIPK2, TBK1, NF-kB, IRF), recognition (e.g. DMTB1) and the autophagy pathway (IMPK, ATG4) (Fig. 4A).

EWAS analysis enabled the identification of 305 CpGs in resistant compared to susceptible oysters. As for the GWAS, molecular pathways and genes affected by these CpGs concerned key antiviral and immune related functions which biologically validated and strengthened the results of this analysis.

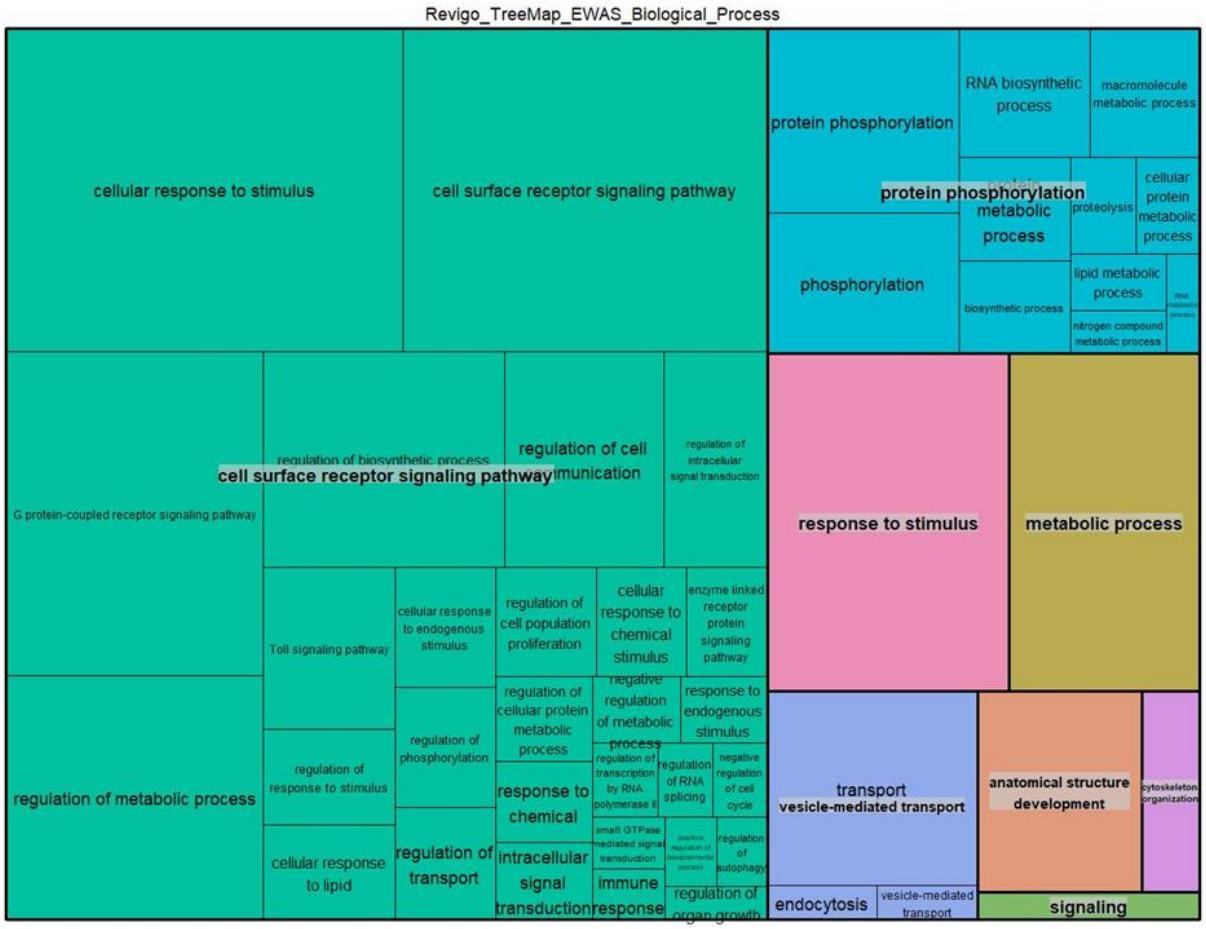


Fig. 6: Oyster resistance to POMS is associated to epigenetic variation in immune pathways.

GO term of the biological process root enriched (RBGOA) from the set of genes displaying a suggestive or a significant CpG associated either to the binary or continuous phenotype. RBGOA results were summarized using Revigo treemap. Rectangles size depends on the adjusted *p*value from RBGOA analysis. Not all the terms are reported due to space restrictions (see Additional file12 for details).

Genetic and epigenetic selection occurred on the same biological processes but on different genes

We compared the genetic and epigenetic mechanisms of resistance identified by GWA and EWA mapping by considering both enriched biological processes and the underlying candidate genes. At the biological process level, a total of 240 GO terms were enriched, 82 being specific to genetic variation, 56 to epigenetic variation and 102 common to genetic and epigenetic variation (Fig. 4B; Supplementary file 13A). Correlation between the delta rank of the GO terms significantly enriched both in GWA and EWA (Fig. 4D) was significantly positive (Pearson correlation coefficient: R=0.68, $P < 0.01$). At the gene level, 320 genes displayed one SNP or CpG associated with resistance traits, with 149 genes being specific to genetic variation and 165 specific to epigenetic variation. Only six genes displayed both genetic and epigenetic variation (Fig. 4C; Supplementary file 13B). From these set of six genes, TBK1 is known to be a major activator of antiviral pathways as the NF-KB and IRF3/7 pathway (Fig. 4A).

These results suggest that selection acting on the genetic and epigenetic information occurred on the same biological functions, in particular on innate immune processes, but not on the same genes.

Genetic and epigenetic information are not independent but epigenetic variation explains more phenotypic variation

To quantify the relative contribution of genetic and epigenetic variation to phenotypic variation, we first tested the presence of relationship between the matrix of pairwise genetic and the matrix of pairwise epigenetic distance among 220 individuals. A significant but weak correlation was detected between the two matrices of distance (Mantel statistic r = 0.089*, P* = 0.0184). Interestingly, this correlation between genetic and epigenetic distances was almost three times higher with genetic distances calculated with synonymous SNPs (Mantel statistic r = 0.3287, *p*value < 1e-04) than with genetic distances calculated with non-synonymous SNPs (r = 0.1142, *p*value = 0.0064). This result illustrates that epigenetic variation is probably not independent from the genetic variation especially in a synonymous context.

As a second approach, MethQTL analysis was performed to identify SNPs best explaining the CpGs methylation level. From the 214,263 SNPs and 635,201 CpGs, 5,151,194 and 5,152,611 SNP-CpG pairs (MethQTLs; FDR < 0.05) were identified when using as a covariate the binary and the semi-quantitative trait, respectively. When we removed redundancy, 160,325 (binary trait) and 160,220 (semi-quantitative trait) SNPs were associated with the methylation level of 557,703 (binary trait) and 557,850 (semi-quantitative phenotype) CpGs. From significant 240 CpGs identified by binary trait EWA mapping, 126 CpGs were significantly associated to 207 SNPs. (Table 1; Supplementary file 14A). With the semi-quantitative trait, out of the 226 CpGs associated with POMS, 111 were significantly associated to 198 SNPs (Table 1; Supplementary file 14B). When considering the intersect between significant SNP-CpG pairs, top SNPs identified by GWA and significant CpGs identified by EWA, only three and eight SNPs associated with 18 and 15 CpGs methylation level were identified for the binary and semi-quantitative traits, respectively (Table 1). These 18 and 15 CpGs were located in four and seven genes for the binary and semi-quantitative traits, respectively (Supplementary file 14C-D). Among these genes, only TBK1 was linked to immunity. In this unique case, TBK1 was displaying a SNP (identified by the GWAS) was associated *in trans* with the methylation level of a CpG (identified by the EWAS) that included in a gene encoding a Transcription terminator factor 2 (TRF2). This MethQTL concern both the binary and the semi-quantitative traits. These results highlight that most of the methylation

level of CpGs were associated in *cis* or *trans* by a SNP. However, this signal is weakened when only the genetic and epigenetic variation significantly associated to the trait is considered.

Finally, a variance partition analysis (RDA) showed that genetic and epigenetic variation jointly explained the highest percentage of phenotypic variation, with 33.5 % and 34.2 % for the binary and semi-quantitative traits (Fig. 7 A-B). When taken individually, epigenetic variation (binary trait = 26.1 % and semi-quantitative trait = 17.3 %; Fig. 7 A-B) explained a higher proportion of phenotypic variation than genetic variation (binary trait = 13.1 % and semi-quantitative trait = 14.1 %; Fig. 7 A-B). Finally, 27.3 % and 34.4 % of phenotypic variation was not explained either by epigenetic and/or genetic variation for the binary and semi-quantitative traits, respectively (Fig. 7 A-B).

The obtained results highlights that genetic and epigenetic information are partially correlated and that genetic was associated to significant part of the epigenetic at the methylation level. However, these results also showed that this two components of the inheritance system can display independent signals associated to a change in phenotypic frequency.

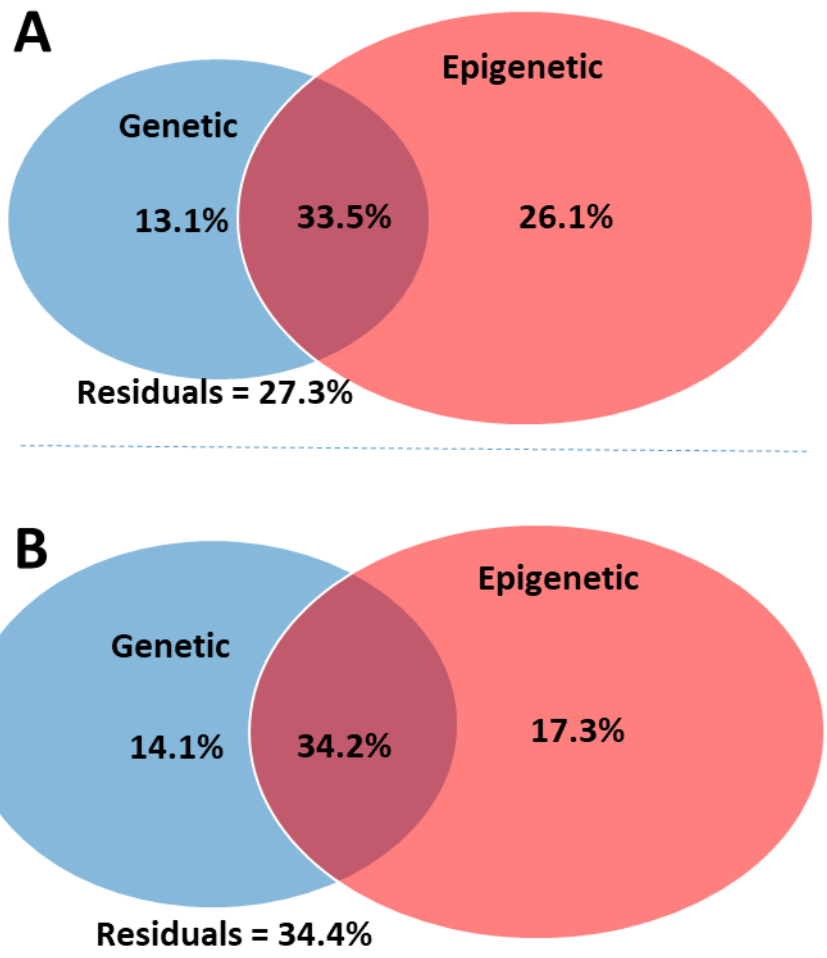


Fig. 7: Variation in oyster survival is explained only by genetic variation, epigenetic variation and their interaction.

RDA performed to disentangle the portion of phenotypic variation explained by the genetic variation (blue), the epigenetic variation (red) or their interaction. (A) Analysis using the binary phenotype and B) using the Semi-continuous phenotype.

Table 1: MethQTL association analysis

	Covariate	
	Binary trait	**Semi-quantitative trait**
Total number of SNPs	214,263	214,263
Total number of CpGs	635,201	635,201
Significant SNP-CpG pairs (MethQTL)	5,151,194	5,152,611
Number of non-redundant SNPs involved in a MethQTL	160,325	160,220
Number of non-redundant CpGs involved in a MethQTL	557,703	557,850
MethQTL associated with a CpG identified by the EWAS #	207	198
MethQTL identified by the GWAS (suggestive threshold) †	3	8
Number of CpGs identified by the EWAS and associated by a MethQTL *	126	111
MethQTL involving a CpG and a SNP identified by the EWAS and GWAS ‡	18	15

How many SNPs are associated to a significant CpGs associated to POMS (EWAS)

† How many GWAS suggestive SNPs are associated to a significant CpGs associated to POMS (EWAS)

* How many CpGs associated to POMS (EWAS) are associated to a MethQTL

‡ How many CpGs associated to POMS (EWAS) are associated to a GWAS suggestive SNPs

Discussion

In the present study we showed that wild oyster populations differentially exposed to the emerging disease named Pacific Oyster Mortality Syndrome (POMS) display signatures of selections both in their genome (SNP) and epigenome (CpGs). A high number of these SNPs and CpGs were located in genes encoding immune relating functions although genetic and epigenetic signatures occurred in different loci. Although, the genetic and epigenetic variations are partly correlated and the former was associated with a large fraction of the second, we also showed that a part of the epigenetic variation significantly associated to oyster resistance was independent and explained a higher part of the phenotypic variation. These results confirmed that host population facing pathogen emergence can rely on genetic and epigenetic variation to rapidly adapt to emerging diseases.

One of the most challenging issue of this study was to capture a sufficient portion of the genome and epigenome to identify signatures of selection in wild oyster populations. Analysis of hundreds of individuals are required to perform large scale omics population studies and it seriously impacts the cost of the experiment. In this sense, cost effective reduced representation approaches such as epiGBS (Van Gurp et al., 2016), RRBS (Gu et al., 2011)or epiRAD (Schield et al., 2016) are generally used for population studies but these necessary approaches led to lower resolution than whole genome approaches. A previous epiGBS study performed on the fresh water snail *Biomphalaria glabrata* reduced the study of cytosine methylation to 1% of all the CpGs (Luviano et al., 2021). This low amount of covered CpG would not have been suitable in our study since this would have reduced the identification of strong selection signatures. This problem of detection also increased for species with short linkage disequilibrium (Lowry et al., 2017). To circumvent these limitations we developed a whole exome capture experiment that enabled us to characterized most of the exonic sequences of the oyster genome with the advantage of covering a significant part of DNA methylation information since methylated CpGs is essentially restricted to gene-body in mollusc species (Xiaotong Wang et al., 2014). Interestingly, the DNA methylation for 6% of the total CpGs were captured in our study, which represent six fold more with what epiGBS had obtained in the *B. glabrata* (Luviano et al., 2021). The Exome capture approach was successfully used to identify genetic and epigenetic diseases in human (Precone et al., 2015) and was more recently applied to non-model organisms (Bitter et al., 2019; Heer et al., 2018).

Another interesting advantage of this method relies in the localization of the genetic and epigenetic variation within coding sequences which brings information at the functional level. In regards to the processes of adaption to infectious diseases it is probable that signatures of selection occurred in the gene body or the regulatory region rather than in the intergenic region (Hoban et al., 2016).

The whole exome capture, sequencing and downstream analysis performed in this study showed that oyster's signature of resistance to POMS were found associated both to the genetic and epigenetic information. At the genetic level, 2 SNPs were significantly associated to oyster resistance and when this threshold was increased to consider suggestive SNPs this value reached 186. At the epigenetic level 305 CpGs were differentially methylated between the resistant and the susceptible oysters. Enrichment analysis performed with the set of genes showing genetic and/or epigenetic variation showed a strong enrichment of biological processes linked to immunity (Fig. 3 and 6). Emblematic genes of these immune-related process were affected by genetic and/or epigenetic variation (Fig. 4A). This span over the JAK/STAT pathway with genes such as UBA2 and RNF220; RLR/STING with genes such as TRIM33, TBK1 and IRF; NF-KB with genes such as TIRprot, NF-KB and MyD88; RNAi with the gene DICER; autophagy with the gene ATG4 and several pathogen recognition receptors such as DSCAM, Mannose Receptors, C1q and PRGP (Green et al., 2015). Accordingly, previous work have reported the polygenic nature of POMS resistance (Gutierrez et al., 2018, 2020; Sauvage et al., 2010). At the phenotypic level, transcriptomic and proteomic studies have demonstrated the key role played by an early antiviral response (de Lorgeril et al., 2018; Leprêtre et al., 2021). During experimental infections, resistant families were shown to express as soon as 6 hours post contact genes involved in the RLR/STING, JAK/STAT, apoptosis and autophagy pathways (de Lorgeril et al., 2018; Leprêtre et al., 2021). Interestingly, a transcriptomic comparison characterizing gene expression signature of resistance under non-infectious condition has shown a strong enrichment of biological processes linked to the innate immune response. (de Lorgeril et al., 2020). In this study, authors have identified only one gene (encoding Toll-like receptor (TLR13) protein) significantly over expressed in the all three resistant families but many immune genes were retrieved in shared between at least two families. Based on these results, they conclude that while TLR13 should be essential the over expression of the anti-viral response seem to be as important than the over expression

of a single gene (de Lorgeril et al., 2020). This last conclusion seems to be in agreement with our results showing signature of genetic and epigenetic selection disseminated in several immune genes and pathways. Whether this feature is a characteristic of a multi genic resistance based on the overall immune capabilities of an individual or a signature of a recent and rapid adaptation will need to be further studied.

At the phenotypic level, the POMS resistance phenotype has been shown to rely on antiviral gene expression, either constitutively expressed in naturally occurring resistant families (de Lorgeril et al., 2020) or environmentally induced in immune primed oysters (Fallet et al., 2022; Lafont et al., 2020). In all these studies, a substantial amount of immune related genes are key players of the POMS resistance phenotype displayed by the different resistant oyster families, but few expressed immune genes are common to the different families. This emphasizes the polygenic response of the POMS resistant phenotype and underlies its immune network nature.

The question about the role of epigenetic variation in the generation of phenotypic variation and subsequent transgenerational adaptation is still hotly debated (C. L. Richards & Pigliucci, 2020). While an increasing number of studies showed a link between epigenetic change and phenotypic change the extent to which this epigenetic change is under a genetic determination is still unclear (Husby, 2022). This statement raises the question about the relative weight of genetic and epigenetic information in the expression of adaptive phenotype. A first level of answer was obtained by taking advantages of evolutionary experiments in controlled condition performed on different model and non-model organisms (for example see (Liew et al., 2018; Ryu et al., 2018; Schmid et al., 2018)), and in our opinion these first results show that epigenetic information can solely encode adaptive phenotypic variation without the sequence variation. However, the relative contribution of both genetic and epigenetic for adaptive phenotypic variation has not been addressed in wild population so far. In our study, we took advantage of a natural differential environmental pressure related to the viral POMS disease events occurring in the field (farming area *vs.* non farming area) to disentangle the contribution of genetic and epigenetic components for differential resistant phenotype observed in wild oysters. GWAS and EWAS applied in wild oyster populations displaying contrasted resistant phenotypes enabled us to identify signatures of selection both in the genome and the epigenome. However, since such analyses were done independently

from each other the questions of the independence of the epigenetic variation from the genetic variation was still present.

As a first approach to disentangle the effect of each phenotypic determinant we first tested for the presence of a correlation between the genetic and the epigenetic distance and we showed that the epigenetic variation is not entirely independent from the genetic variation. Such a conclusion was already provided but also its opposite despite identical ecological context (Fargeot et al., 2021; Foust et al., 2016). In these studies both groups have proposed that this opposite pattern may reflect a species-dependent effect. The main difference between these two studies and ours reside in the density of genetic and epigenetic markers which is an order of magnitude higher within our datasets. An alternative but non-exclusive hypothesis would be that in certain condition the detection of such correlation need a high density of information to provide a sufficient statistical power.

In a second step dedicated to the disentangling of the genetic and epigenetic variations and their relative effect on the phenotypic expression we developed MethQTL and variance partition analysis. MethQTL analysis enables to identify the SNP(s) that are associated with the methylation level of a CpG(s). The number of identified MethQTL was surprisingly high and highlights the strong interlinking present among these two information. Many SNPs were associated to the methylation level of many CpGs. In some cases, a single SNP associated to a single CpG but in most cases the interlinking is much more intricate with a single SNP that associated with several CpGs; or several SNPs associated with the same CpG. However, independent SNPs and independent CpGs were also identified which confirm that a part of the epigenetic information is independent from the DNA sequence. Strong determination of methylation patterns by genetic variation were demonstrated several time and associated to major phenotypic change (Gibbs et al., 2010; Höglund et al., 2020). However the absence of MethQTL for a given methylation pattern linked to a phenotypic change was also reported previously (Cortijo et al., 2014; Rathod et al., 2020). Interestingly, when our GWAS, EWAS and MethQTL result are combined with a focus on the set of genes with a clear role in host/pathogen interaction (e.g. immune genes displaying a significant SNP or CpG identified by the WAS) pure genetic effect, pure epigenetic effect and a mix of both are present. This partly independence and partly interlinking of each mechanisms was further confirmed by the variance partition analysis that shows that phenotypic variation of oyster resistance to POMS

is partly explained by the genetic variance (independent SNPs), by the epigenetic variance (independent CpGs) and by both (MethQTL). These associations corroborate what has been previously published and suggest that under a given selective pressure DNA methylation pattern that are dependent or independent of the DNA sequence can be selected. The question whether these independent methylation patterns are environmentally induced or the results of a random phenomenon similar to standing genetic variation need to be further explored, but recent works on oyster and POMS interaction can bring some interesting information about this question.

Oyster resistance to POMS have already been demonstrated to be environmentally sensitive (Bruno Petton et al., 2021b). Links with molecular mechanisms and a potential role of an environmentally induced epigenetic modification are currently available for two kinds of phenomenon, immune priming (Lafont et al., 2017, 2020) and immune shaping (Fallet et al., 2022). The former can be broadly defined as increased protection to a pathogen following previous exposure to a pathogen or an immune elicitor. The second consists of the modulation of the immune capabilities of an individual by an interaction during its early life with microorganisms. In the case of immune priming in oysters, injections of Poly(I:C), a viral mimic, into susceptible oysters prior to POMS infection led to a resistant phenotype associated with and increased viral protection that can reach 100% (Lafont et al., 2017). This phenotypic inversion was characterized at the transcriptomic level by a strong antiviral response that impaired OsHV-1 replication and POMS disease development (Lafont et al., 2020). Although epigenetic changes associated with this immune priming in oyster was not yet demonstrated, recent studies on innate immune memory in mammals and plants strongly point towards epigenome remodelling as a potent driver of these mechanisms (Netea et al., 2016; Thellier & Lüttge, 2013). In the case of immune shaping, as for the immune priming, the exposure to a non-pathogenic but rich micro-flora during early life, has enhanced the immune capabilities of the oysters. The phenotypes obtained are less contrasted since the resistance level of a susceptible family was increased by 9 to 13% but was characterized by significant differences of the transcriptomic response of several antiviral gene families (Fallet et al., 2022). Interestingly, it was also shown with this system that the exposure to this rich micro-flora has modified the epigenome (Fallet et al., 2022). Finally it was demonstrated that both the enhanced phenotype and the epigenetic modifications were transmitted for at least two

generations without any new exposure to a rich non-pathogenic micro-flora at the F2 (Fallet et al., 2022). This study developed under controlled condition and under a controlled genetic background highlights that the environment can induce heritable epigenetic modifications in oysters that subsequently lead to a higher resistance to the POMS. In the present study, the epigenetic modifications observed could in part reflect such immune priming/shaping mechanisms related to a more resistant phenotype. Oysters from non-farming areas are cultivated in the vicinity of farming area that could expose them to very low quantity of OsHV-1, insufficient to induce disease and/or to a rich microbial flora found in farming area that would induce immune priming/shaping. Alternatively, late recruitment (when temperature decreased below 16°C) in farming areas would also enable such exposures resulting in the induction of immune priming or immune shaping. In conclusion it is clearly possible that such kind of exposure can be the environmental triggers of the independent epigenetic signatures of resistance we have identified in our study.

The present work showed that in response to the recent emergence of a viral variant inducing a strong selective pressure, host populations were selected both at the genetic and epigenetic level. While our study confirms the essential role played by the DNA sequence it also shows that other mechanisms can interplay with this sequence to encode a resistant phenotype; but they can also be independent from this sequence and participate to the expression of resistance. On one side, these results confirm that more holistic approaches of the resistance of host population must be envisioned to have access to most of the mechanisms at stake. On the other side it also demonstrates that epigenetic assisted selection would be a way to assist breeding industry without effect on the DNA sequence.

Acknowledgments

Authors are grateful to the staff of the Ifremer stations of Argenton, Bouin and Palavas-sur-Mer for technical support in animal sampling and experimentation. We are grateful to Dr. Clémentine Vitte and Dr. Benoit Pujol for fruitful discussions. Data used in this work were (partly) produced through the GenSeq technical facilities. We thank the Bio-Environment platform (University of Perpignan Via Domitia) and Jean-François Allienne for support in sequencing. The present study was supported by the ANR project DECIPHER (ANR-14-CE19-0023), by the FEAMP project GESTINNOV (grant n°PFEA470020FA1000007), by the EU funded project VIVALDI (H2020 program, n°678589) and by Ifremer (grant politique de site GEM). This study is set within the framework of the "Laboratoires d'Excellences (LABEX)" TULIP (ANR-10-LABX-41). We thank the SEBIMER bioinformatics platform from Ifremer for all the bioinformatics support.

References

Aguirre AA, Tabor GM (2008). Global factors driving emerging infectious diseases: Impact on wildlife populations. Ann N Y Acad Sci 1149: 1–3.

Andrews S (2010). FastQC a Quality Control Tool for High Throughput Sequence Data [Online]. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Azéma P, Lamy JB, Boudry P, Renault T, Travers MA, Dégremont L (2017). Genetic parameters of resistance to Vibrio aestuarianus, and OsHV-1 infections in the Pacific oyster, *Crassostrea gigas*, at three different life stages. Genet Sel Evol 49: 1–16.

Barbosa Solomieu V, Renault T, Travers MA (2015). Mass mortality in bivalves and the intricate case of the Pacific oyster, *Crassostrea gigas*. J Invertebr Pathol 131: 2–10.

Barfield RT, Kilaru V, Smith AK, Conneely KN (2012). CpGassoc: An R function for analysis of DNA methylation microarray data. Bioinformatics 28: 1280–1281.

Barturen G, Rueda A, Oliver JL, Hackenberg M (2013). MethylExtract: High-Quality methylation maps and SNV calling from whole genome bisulfite sequencing data. F1000Research 2: 217.

Bédier E, D'Amico F, Annezo J-P, Auby I, Barret J, Bouget J-F, et al. (2009). Observatoire national Conchylicole. Rapport 2009.

Bitter MC, Kapsenberg L, Gattuso JP, Pfister CA (2019). Standing genetic variation fuels rapid adaptation to ocean acidification. Nat Commun 10: 1–10.

Borcard D, Legendre P, Drapeau P (1992). Partialling out the Spatial Component of Ecological Variation. Ecology 73: 1045–1055.

Buestel D, Ropert M, Prou J, Goulletquer P (2009). History, status, and future of oyster culture in France. J Shellfish Res 28: 813–820.

Burge CA, Mark Eakin C, Friedman CS, Froelich B, Hershberger PK, Hofmann EE, et al. (2014). Climate change influences on marine infectious diseases: Implications for management and society. Ann Rev Mar Sci 6: 249–277.

Camara MD, Yen S, Kaspar HF, Kesarcodi-Watson A, King N, Jeffs AG, 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.

Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. Gigascience 4: 1–16.

Clerissi C, de Lorgeril J, Petton B, Lucasson A, Escoubas JM, Gueguen Y, et al. (2020). Microbiota Composition and Evenness Predict Survival Rate of Oysters Confronted to Pacific Oyster Mortality Syndrome. Front Microbiol 11: 1–11.

Cortijo S, Wardenaar R, Colomé-Tatché M, Gilly A, Etcheverry M, Labadie K, et al. (2014). Mapping the epigenetic basis of complex traits. Science (80-) 343: 1145–1148.

Cosseau C, Wolkenhauer O, Padalino G, Geyer KK, Hoffmann KF, Grunau C (2017). (Epi)genetic Inheritance in Schistosoma mansoni: A Systems Approach. Trends Parasitol 33: 285–294.

Crotti M, Yohannes E, Winfield IJ, Lyle AA, Adams CE, Elmer KR, et al. (2021). Rapid adaptation through genomic and epigenomic responses following translocations in an endangered salmonid. Evol Appl 00: 1–20.

Danchin É (2013). Avatars of information: Towards an inclusive evolutionary synthesis. Trends Ecol Evol 28: 351–358.

Dégremont L (2011). Evidence of herpesvirus (OsHV-1) resistance in juvenile *Crassostrea gigas* selected for high resistance to the summer mortality phenomenon. Aquaculture 317: 94–98.

Dégremont L (2013). Size and genotype affect resistance to mortality caused by OsHV-1 in *Crassostrea gigas*. Aquaculture 416–417: 129–134.

Dégremont L, Garcia C, Allen SK (2015). Genetic improvement for disease resistance in oysters: A review. J Invertebr Pathol 131: 226–241.

Dégremont L, Lamy JB, Pépin JF, Travers MA, Renault T (2015). New insight for the genetic evaluation of resistance to ostreid herpesvirus infection, a worldwide disease, in *Crassostrea gigas*. PLoS One 10: 1–12.

EFSA PAHW (2015). Oyster mortality. FASEB J 13: 4u22–4n/au22.

Fallet M, Montagnani C, Petton B, Ird UBOC, Umr L, Chaparro C, et al. (2022). Early life microbial exposures shape the *Crassostrea gigas* immune system for lifelong and intergenerational disease protection. BMC microbiome.

FAO (2021). Fisheries and Aquaculture Information and Statistics Branch - 07/09/2021 - http://www.fao.org/figis/servlet/TabSelector.

Fargeot L, Loot G, Prunier JG, Rey O, Veyssière C, Blanchet S (2021). Patterns of epigenetic diversity in two sympatric fish species: Genetic vs. environmental determinants. Genes (Basel) 12: 1–18.

Fleury E, Barbier P, Petton B, Normand J, Thomas Y, Pouvreau S, et al. (2020). Latitudinal drivers of oyster mortality: deciphering host, pathogen and environmental risk factors. Sci Rep 10.

Foust CM, Preite V, Schrey AW, Alvarez M, Robertson MH, Verhoeven KJF, et al. (2016). Genetic and epigenetic differences associated with environmental gradients in replicate populations of two salt marsh perennials. Mol Ecol 25: 1639–1652.

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

Gibbs JR, van der Brug MP, Hernandez DG, Traynor BJ, Nalls MA, Lai SL, et al. (2010). Abundant quantitative trait loci exist for DNA methylation and gene expression in Human Brain. PLoS Genet 6: 29.

Green TJ, Raftos D, Speck P, Montagnani C (2015). Antiviral immunity in marine molluscs. J Gen Virol 96: 2471–2482.

Gu H, Smith ZD, Bock C, Boyle P, Gnirke A, Meissner A (2011). Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. Nat Protoc 6: 468–481.

Van Gurp TP, Wagemaker NCAM, Wouters B, Vergeer P, Ouborg JNJ, Verhoeven KJF (2016). EpiGBS: Reference-free reduced representation bisulfite sequencing. Nat Methods 13: 322–324.

Gutierrez AP, Bean TP, Hooper C, 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, Genet 8: 1273–1280.

Gutierrez AP, Symonds J, King N, Steiner K, Bean TP, Houston RD (2020). Potential of genomic selection for improvement of resistance to ostreid herpesvirus in Pacific oyster (*Crassostrea gigas*). Anim Genet 51: 249–257.

Harvell CD, Mitchell CE, Ward JR, Altizer S, Dobson AP, Ostfeld RS, et al. (2002). Climate warming and disease risks for terrestrial and marine biota. Science (80-) 296: 2158–2162.

Heer K, Ullrich KK, Hiss M, Liepelt S, Schulze Brüning R, Zhou J, et al. (2018). Detection of somatic epigenetic variation in Norway spruce via targeted bisulfite sequencing. Ecol Evol 8: 9672–9682.

Hoban S, Kelley JL, Lotterhos KE, Antolin MF, Bradburd G, Lowry DB, et al. (2016). Finding the genomic basis of local adaptation: Pitfalls, practical solutions, and future directions. Am Nat 188: 379–397.

Höglund A, Henriksen R, Fogelholm J, Churcher AM, Guerrero-Bosagna CM, Martinez-Barrio A, et al. (2020). The methylation landscape and its role in domestication and gene regulation in the chicken. Nat Ecol Evol 4: 1713–1724.

Husby A (2022). Wild epigenetics: Insights from epigenetic studies on natural populations. Proc R Soc B Biol Sci 289: 20211633.

Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, et al. (2014). InterProScan 5: Genome-scale protein function classification. Bioinformatics 30: 1236–1240.

Jun G, Wing MK, Abecasis GR, Kang HM (2015). An efficient and scalable analysis framework for variant extraction and refinement from population-scale DNA sequence data. Genome Res 25: 918–925.

Krueger F (2015). Trim Galore!. [http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/].

Lafont M, Petton B, Vergnes A, Pauletto M, Segarra A, Gourbal B, et al. (2017). Long-lasting antiviral innate immune priming in the Lophotrochozoan Pacific oyster, *Crassostrea gigas*. Sci Rep 7: 1–14.

Lafont M, Vergnes A, Vidal-Dupiol J, De Lorgeril J, Gueguen Y, Haffner P, et al. (2020). A sustained immune response supports long-term antiviral immune priming in the pacific oyster, *Crassostrea gigas*. MBio 11.

Lapègue S, Boudry P, Goulletquer P (2006). Pacific cupped oyster -*Crassostrea gigas*.in GENINPACT-Evaluation of genetic impact of aquaculture activies on native population. A European network, WP1 workshop Genetics of domestication, breeding and enhancement of performance of fish and shellfish.

Leprêtre M, Faury N, Segarra A, Claverol S, Degremont L, Palos-Ladeiro M, et al. (2021).

Comparative Proteomics of Ostreid Herpesvirus 1 and Pacific Oyster Interactions With Two Families Exhibiting Contrasted Susceptibility to Viral Infection. Front Immunol 11: 1–16.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. (2009). The Sequence Alignment / Map format and SAMtools. Bioinformatics 25: 2078–2079.

Liew YJ, Zoccola D, Li Y, Tambutté E, Venn AA, Michell CT, et al. (2018). Epigenome-associated phenotypic acclimatization to ocean acidification in a reef-building coral. Sci Adv 4: eaar8028.

de Lorgeril J, Lucasson A, Petton B, Toulza E, Montagnani C, Clerissi C, et al. (2018). Immune-suppression by OsHV-1 viral infection causes fatal bacteraemia in Pacific oysters. Nat Commun 9: 4215.

de Lorgeril J, Petton B, Lucasson A, Perez V, Stenger PL, Dégremont L, et al. (2020). Differential basal expression of immune genes confers *Crassostrea gigas* resistance to Pacific oyster mortality syndrome. BMC Genomics 21: 1–14.

Lowry DB, Hoban S, Kelley JL, Lotterhos KE, Reed LK, Antolin MF, et al. (2017). Breaking RAD: an evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. Mol Ecol Resour 17: 142–152.

Luviano N, Lopez M, Gawehns F, Chaparro C, Arimondo PB, Ivanovic S, et al. (2021). The methylome of Biomphalaria glabrata and other mollusks: enduring modification of epigenetic landscape and phenotypic traits by a new DNA methylation inhibitor. Epigenetics and Chromatin 14: 1–25.

Marees AT, de Kluiver H, Stringer S, Vorspan F, Curis E, Marie-Claire C, et al. (2018). A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. Int J Methods Psychiatr Res 27.

Martin LB, Hanson HE, Hauber ME, Ghalambor CK (2021). Genes, Environments, and Phenotypic Plasticity in Immunology. Trends Immunol 42: 198–208.

Netea MG, Joosten LAB, Latz E, Mills KHG, Natoli G, Stunnenberg HG, et al. (2016). Trained immunity: A program of innate immune memory in health and disease. Science (80-) 352: 427.

Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, Dan M, et al. (2020). vegan: Community Ecology Package. R package version 2.5-7. https://CRAN.R-project.org/package=vegan.

Pan H, Holbrook JD, Karnani N, Kwoh CK (2016). Gene, Environment and Methylation (GEM): A tool suite to efficiently navigate large scale epigenome wide association studies and integrate genotype and interaction between genotype and environment. BMC Bioinformatics 17: 1–8.

Peñaloza C, Gutierrez AP, Eöry L, Wang S, Guo X, Archibald AL, et al. (2021). A chromosome-level genome assembly for the Pacific oyster *Crassostrea gigas*. Gigascience 10: 1–9.

Pernet F, Barret J, Le Gall P, Corporeau C, Dégremont L, Lagarde F, et al. (2012). Mass mortalities of Pacific oysters *Crassostrea gigas* reflect infectious diseases and vary with

farming practices in the Mediterranean Thau lagoon, France. Aquac Environ Interact 2: 215–237.

Pernet F, Lupo C, Bacher C, Whittington RJ (2016). Infectious diseases in oyster aquaculture require a new integrated approach. Philos Trans R Soc B Biol Sci 371.

Pernet F, Tamayo D, Fuhrmann M, Petton B (2019). Deciphering the effect of food availability, growth and host condition on disease susceptibility in a marine invertebrate. J Exp Biol 222: jeb210534.

Petton B, Destoumieux-Garzón D, Pernet F, Toulza E, de Lorgeril J, Degremont L, et al. (2021). The Pacific Oyster Mortality Syndrome, a Polymicrobial and Multifactorial Disease: State of Knowledge and Future Directions. Front Immunol 12.

Petton B, Pernet F, Robert R, Boudry P (2013). Temperature influence on pathogen transmission and subsequent mortalities in juvenile pacific oysters *Crassostrea gigas*. Aquac Environ Interact 3: 257–273.

Precone V, Del Monaco V, Esposito MV, De Palma FDE, Ruocco A, Salvatore F, et al. (2015). Cracking the Code of Human Diseases Using Next-Generation Sequencing: Applications, Challenges, and Perspectives. Biomed Res Int 2015.

Rathod A, Duan J, Zhang H, Holloway JW, Ewart S, Arshad SH, et al. (2020). Interweaving Between Genetic and Epigenetic Studies on Childhood Asthma. Epigenetics Insights 13.

Renault T, Bouquet AL, Maurice J., Lupo C, Blachier P (2014). Ostreid herpesvirus 1 infection among Pacific oyster (*Crassostrea gigas*) Spat: Relevance of water temperature to virus replication and circulation prior to the onset of mortality. Appl Environ Microbiol 80: 5419–5426.

Renault T, Le Deuff RM, Chollet B, Cochennec N, Gerard A (2000). Concomitant herpes-like virus infections in hatchery-reared larvae and nursery-cultured spat *Crassostrea gigas* and *Ostrea edulis*. Dis Aquat Organ 42: 173–183.

Renault T, Le Deuff RM, Cochennec N, Maffart P (1994). Herpesvirus associated with mortalities among Pacific oyster. *Crassostrea gigas*, in France comparative study. Rev Méd Vét 145: 735–742.

Richards CL, Pigliucci M (2020). Epigenetic Inheritance. A Decade into the Extended Evolutionary Synthesis. Paradigmi XXXVIII: 463–494.

Rougeux C, Laporte M, Gagnaire PA, Bernatchez L (2019). The role of genomic vs. epigenomic variation in shaping patterns of convergent transcriptomic variation across continents in a young species complex. bioRxiv: 1–27.

Ryu T, Veilleux HD, Donelson JM, Munday PL, Ravasi T (2018). The epigenetic landscape of transgenerational acclimation to ocean warming. Nat Clim Chang 8: 504–509.

Sauvage C, Boudry P, De Koning DJ, Haley CS, Heurtebise S, Lapègue S (2010). QTL for resistance to summer mortality and OsHV-1 load in the Pacific oyster (*Crassostrea gigas*). Anim Genet 41: 390–399.

Schield DR, Walsh MR, Card DC, Andrew AL, Adams RH, Castoe TA (2016). EpiRADseq: Scalable analysis of genomewide patterns of methylation using next-generation sequencing.

Methods Ecol Evol 7: 60–69.

Schikorski D, Faury N, Pepin F, Saulnier D, Tourbiez D, Renault T (2011). Experimental ostreid herpesvirus 1 infection of the Pacific oyster *Crassostrea gigas*: Kinetics of virus DNA detection by q-PCR in seawater and in oyster samples. Virus Res 155: 28–34.

Schmid MW, Heichinger C, Coman Schmid D, Guthörl D, Gagliardini V, Bruggmann R, et al. (2018). Contribution of epigenetic variation to adaptation in Arabidopsis. Nat Commun 9.

Segarra A, Pépin JF, Arzul I, Morga B, Faury N, Renault T (2010). Detection and description of a particular Ostreid herpesvirus 1 genotype associated with massive mortality outbreaks of Pacific oysters, *Crassostrea gigas*, in France in 2008. Virus Res 153: 92–99.

Supek F, Bošnjak M, Škunca N, Šmuc T (2011). Revigo summarizes and visualizes long lists of gene ontology terms. PLoS One 6: 21800.

Thellier M, Lüttge U (2013). Plant memory: A tentative model. Plant Biol 15: 1–12.

Turner SD (2018). qqman : an R package for visualizing GWAS results using Q-Q and manhattan plots. 3: 2–3.

Wang X, Li Q, Lian J, Li L, Jin L, Cai H, et al. (2014). Genome-wide and single-base resolution DNA methylomes of the Pacific oyster *Crassostrea gigas* provide insight into the evolution of invertebrate CpG methylation. BMC Genomics 15.

Webb SC, Fidler A, Renault T (2007). Primers for PCR-based detection of ostreid herpes virus-1 (OsHV-1): Application in a survey of New Zealand molluscs. Aquaculture 272: 126–139.

Wendt J, Rosenbaum H, Richmond TA, Jeddeloh JA, Burgess DL (2018). Targeted bisulfite sequencing using the SeqCap epi enrichment system. In: Methods in Molecular Biology, Vol 1708, pp 383–405.

Wright RM, Aglyamova G V., Meyer E, Matz M V. (2015). Gene expression associated with white syndromes in a reef building coral, Acropora hyacinthus. BMC Genomics 16: 1–12.

Xi Y, Li W (2009). BSMAP: Whole genome bisulfite sequence MAPping program. BMC Bioinformatics 10: 1–9.

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.

Zynda G (2018). 'BSMAPz.' GitHub Repository. https://github.com/zyndagj/BSMAPz.

Supplementary files:

Supplementary file 1A

Supplementary file 1-A: Geographic coordinate of the oyster populations sampled and phenotyped in this study

Population	Area	Latitude	Longitude	Number of samples
B1	Non-Farming	48.379364	-4.446286	61
B2	Non-Farming	48.341789	-4.441086	59
B3	Non-Farming	48.322392	-4.454078	61
B4	Non-Farming	48.296575	-4.451778	56
B5	Farming	48.32815	-4.321947	59
B6	Farming	48.34695	-4.338986	60
Total				356

Supplementary file 1B

Due to the large size of the file, the file can be reached through the link below.

https://drive.google.com/drive/folders/114V0oY5ySOUtpgUb_E97zhZtr3speIMf?usp=sharing

Supplementary file 1C

Due to the large size of the file, the file can be reached through the link below.

https://drive.google.com/drive/folders/114V0oY5ySOUtpgUb_E97zhZtr3speIMf?usp=sharing

Supplementary file 2

Due to the large size of the file, the file can be reached through the link below.

https://drive.google.com/drive/folders/114V0oY5ySOUtpgUb_E97zhZtr3speIMf?usp=sharing

Supplementary file 3

Seq Cap Epi Enrichment System

SeqCap Epi Library (with Hyper Prep plus version 1.0, 1µg) : Workshop protocol

1. Preparation of Bisulfite Conversion Control (CC)

1. Briefly spin down the tube containing the Bisulfite-Conversion Control (**CC**), from the SeqCap Epi Accessory Kit. Add **1 ml** of **PCR-grade water** (from the same kit) directly to the tube containing the Bisulfite-Conversion Control (**CC**).
2. Vortex to mix and briefly spin down.
3. Aliquot **990 µl** of PCR-grade water into a **new** 1.5 ml tube.
4. Add **10 µl** of the diluted Bisulfite-Conversion Control from point 1.1 to the tube containing 990 µl of PCR-grade water (from point 1.3).
5. Vortex to mix and briefly spin down.
6. The Bisulfite-Conversion Control is now ready for use as detailed in the following step. It should be stored later on at -15 to -25°C.

2. Fragmentation of the DNA (with a Covaris shearing instrument)

1. Add **5.8 µl** of the diluted Bisulfite-Conversion Control (from point 1.6) to **1 µg** of the gDNA sample of interest in a **new** 1.5 ml tube.
2. Adjust the volume of the combined gDNA sample and Bisulfite-Conversion Control to a total volume of 53 **µl** using 1 x TE (low EDTA).
3. Vortex for 3 seconds to mix and briefly spin down.
4. Transfer 53 µl of the input DNA to a microTUBE AFA fiber **screw**-cap (**50 µl tubes**, # 500096) in case of Covaris **M220**, and to a **snap**-cap (**50 µl tubes**, #520045) for another Covaris model.

5. Set your Covaris instrument to get an average size range of 180 – 220 bp. The following settings have been successfully used, but it could be that you have to optimize them for your concrete device (discuss this point with your trainer) :

M220 (with 50 µl screw microtubes) :

Parameters suggested for a M220	Setting
Peak Incident Power (W)	50
Duty Factor	20%
Cycles per Burst	200
Temperature	20°C
Duration (seconds)	220

S220 (with 50 µl snap microtubes):

Parameters suggested for a S220	Setting
Peak Incident Power (W)	175
Duty Factor	10%
Cycles per Burst	200
Temperature	7°C
Mode	Frequency Sweeping
Duration (seconds)	100

6. Spin down briefly the tube each 30 seconds during the sonication process, as well as at the end.

7. Transfer the fragmented DNA to a 200 µl PCR tube (you should be able to transfer at least 52 µl). 50 µl are required for the End Repair step.

8. Run 1 µl in a Bioanalyzer DNA High-Sensitivity chip in duplicate. A successfully fragmented gDNA should look like figure below, with an average fragment size between 180 and 220 bp.

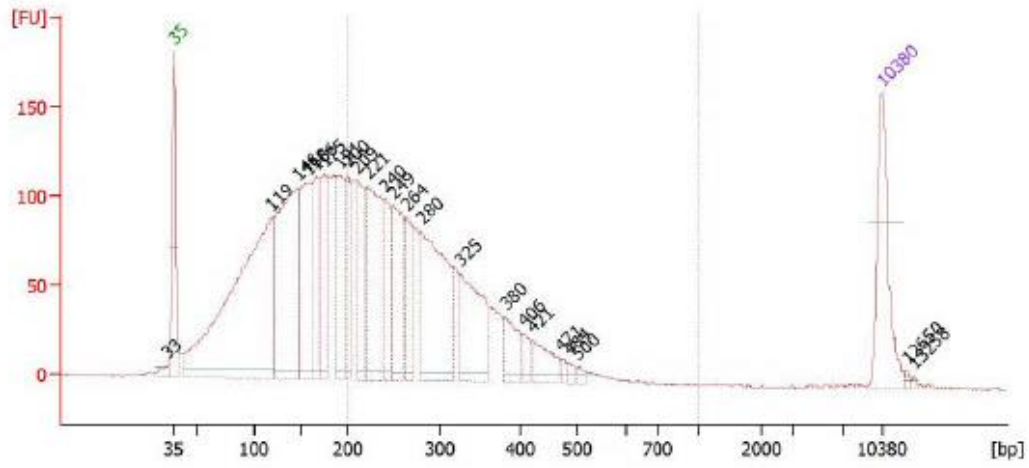


Figure 1 : Example of a proper fragmented DNA in a Bioanalyzer High-Sensitivity chip

3. End Repair and A-Tailing

1. Prepare the End Repair and A-Tailing **Master Mix** as follows:

Reagent	Quantity
KAPA End Repair & A-Tailing Buffer	7 µl
KAPA End Repair & A-Tailing Enzyme Mix	3 µl
Total	**10 µl**

2. Assemble each End Repair and A-Tailing reaction as follows:

Reagent	Quantity
Fragmented, double-stranded DNA	50 µl
End Repair & A-Tailing Master Mix (from 2.1)	10 µl
Total	**60 µl**

3. Mix the End Repair and A-Tailing reaction thoroughly.

4. Place **on ice** and immediately proceed to next step.

5. Perform the End Repair and A-Tailing incubation in a thermocycler using the following program with heated lid:
   - Step 1: 30 minutes @ 20°C
   - Step 2: 30 minutes @ 65°C
   - Step 3: Hold @ 4°C

6. In the meantime, keep the AMPure XP beads at room temperature (to be used in point 3.8) and prepare fresh 80% Ethanol.

7. Following the 60 minute incubation, proceed **immediately** to the next step.

---

4. Adapter Ligation

1. Prepare the Adapter Ligation Master Mix as follows:

Ligation Master Mix	Quantity
PCR-grade water	5 µl
KAPA Ligation Buffer	30 µl
KAPA DNA Ligase	10 µl
Total	**45 µl**

2. Add the selected indexed adapter to each sample:

Reagent	Quantity
Tube containing the End Repair and A-Tailing mix (from 2.7)	60 µl
SeqCap Library **Indexed** Adapter	5 µl
Total	**65 µl**

3. Ensure that you record the index used for each sample.

4. To each tube that contains 65 µl End Repair and A-Tailing mix/DNA/adapter add 45 µl of the Ligation Master Mix, resulting in a total volume of 110 µl :

Reagent	Quantity
Sample with index adapter (from 3.2)	65 µl
Adapter Ligation Master Mix	45 µl
Total	**110 µl**

5. Mix the Ligation Reaction thoroughly.

6. Perform the Adapter Ligation in a thermocycler using the following program:

- 15 minutes @ 20°C

7. Following the incubation, proceed **immediately** to the next step.

8. To each Ligation Reaction product add thoroughly resuspended, **room temperature** AMPure XP beads as indicated here:

First Post-Ligation Clean Up	Per Individual Sample Library
Ligation Reaction product	110 µl
AMPure XP Reagent	88 µl
Total	**198 µl**

9. Mix the Ligation Reaction product and AMPure XP beads thoroughly.

10. Incubate the sample at room temperature for 5 **minutes** to allow the DNA to bind to the beads.

11. Place the sample on a magnetic particle collector to capture the beads. Incubate until the liquid is clear.

12. Carefully remove and discard the supernatant.

13. Keeping the samples on the magnetic particle collector, add **200 µl** of freshly-prepared 80% ethanol.

14. Incubate the sample at room temperature for ≥30 seconds.

15. Carefully remove and discard the ethanol.

16. Keeping the sample on the magnetic particle collector, add **200 µl** of freshly-prepared 80% ethanol.

17. Incubate the sample at room temperature for ≥30 seconds.

18. Carefully remove and discard the ethanol. Try to remove all residual ethanol without disturbing the beads.

19. Allow the beads to dry at room temperature, sufficiently for all the ethanol to evaporate. **Overdrying** the beads may result in dramatic library loss.

20. Remove the sample from the magnetic particle collector.

21. Thoroughly resuspend the beads in 53 µl of elution buffer (10 mM Tris-HCl, pH 8.0).

22. Incubate the sample at room temperature for 2 minutes to allow the DNA to elute off the beads.

23. Place the sample on a magnetic particle collector to capture the beads. Incubate until the liquid is clear.

24. Transfer **50 µl** supernatant to a fresh tube/well.

25. Proceed **immediately** to next point (double-sided size selection).

5. Double-sided Size Selection

1. To each sample from point 3.24, add 35 µl well resuspended, room temperature AMPure XP beads for a total volume of 85 µl.

Double-sided Size Selection	Per Individual Sample Library
Resuspended DNA (from point 3.24)	50 µl
AMPure XP beads	35 µl
Total	**85 µl**

2. Mix thoroughly.
3. Incubate the sample at room temperature for 5 **minutes** to allow library fragments larger than ~450 bp to bind to the beads.
4. Place the samples on a magnetic particle collector to capture the beads. Incubate until the liquid is clear.
5. Carefully transfer **80 µl** of the supernatant containing library fragments smaller than ~450 bp to a **new** tube. **It is critical to not transfer any beads with the supernatant.**
6. Discard the old tube with the beads carrying library fragments larger than ~450 bp.
7. To the 80 µl supernatant add **10 µl** of well mixed, **room temperature** AMPure XP beads.
8. Thoroughly mix by pipetting up and down multiple times.
9. Incubate the tube at room temperature for 5 **minutes** to allow library fragments larger than ~250 bp to bind to the beads.
10. Place the tube on a magnet to capture the beads. Incubate until the liquid is clear.
11. Carefully remove and discard most of the supernatant.
12. Keeping the tube on the magnet, add **200 µl** of 80% ethanol.
13. Incubate the tube at room temperature for **30 seconds**.
14. Carefully remove and discard the ethanol.
15. Repeat steps 4.12 to 4.14 **once**.

16. Spin briefly the tube, place it again on the magnet, and remove all residual ethanol without disturbing the beads.

17. Allow the beads to dry at **room temperature**, sufficiently for all the ethanol to evaporate. **Overdrying** the beads may result in dramatic yield loss (discuss this point with your trainer). Remove the tube from the magnet and proceed then **immediately** with the following point.

18. Thoroughly resuspend the beads in **23 µl** of elution buffer (10 mM Tris-HCl, pH 8.0) or PCR-grade water (ask your trainer). Incubate the tube at **room temperature** for 2 minutes to allow the DNA to elute off the beads.

19. Place the tube on a magnet to capture the beads. Incubate until the liquid is clear.

20. Transfer the **clear supernatant** to a new labelled tube. 20 µl will be required for the next step (bisulfite-conversion of the sample library).

6. Bisulfite-Conversion of the Sample Library

1. Take the Zymo Research "**EZ DNA Methylation-Lightning Kit**".

2. Add **24 ml** of **EtOH** to the **M-Wash Buffer** and mix the first time you use the kit.

3. Transfer **20 µl** of the DNA sample from point 4.20 to a **new** tube.

4. Add **130 µl** of **Lightning Conversion Reagent** to the previous 20 µl of the DNA sample (it is light sensitive, and it should be handled fast).

5. Mix and then centrifuge briefly to ensure there are no droplets in the cap or sides of the tube.

6. Split the 150 µl in **two** 0.2 ml tubes (**2 x 75 µl**).

7. Place the two PCR tubes in a thermocycler and perform the following steps :

   - Step 1: 8 minutes @ 98°C
   - Step 2: 60 minutes @ 54°C
   - Step 3: storage up to 20 hours @ 4°C (optional step)

   STOP **For a safe stopping point, leave the bisulfite converted DNA at 4°C in the thermocycler for up to 20 hours.**

8. Add 600 µl of **M-Binding Buffer** to a **Zymo-Spin IC Column** and place the column into a provided **collection tube**.

9. Load the samples from the two tubes (2 x 75 µl) into the **Zymo-Spin IC Column** containing the **M-Binding Buffer**. Close the cap and mix by inverting the column several times.

10. Centrifuge at full speed (> 10,000 x *g*) for 30 seconds. Discard the flow-through.

11. Add 100 µl of **M-Wash Buffer** to the column (be sure that EtOH has been previously added to the wash buffer). Centrifuge at full speed for 30 seconds.

12. Add 200 µl of **L-Desulphonation Buffer** to the column and let stand at room temperature for **20 minutes** (not shorter).

13. After 20 minutes incubation, centrifuge at full speed for 30 seconds.

14. Add 200 µl of **M-Wash Buffer** to the column. Centrifuge at full speed for 30 seconds.

15. Repeat the wash step with another 200 µl of **M-Wash Buffer** to the column. Centrifuge at full speed for 30 seconds.

16. Place the column into a **new** 1.5 ml microcentrifuge tube and add **21.5 µl of PCR-grade water** directly to the column matrix. Wait for 1 minute. Centrifuge for 30 seconds at full speed to elute the DNA.

STOP **For a safe stopping point, leave the purified bisulfite converted DNA at -20°C**

7. Pre-Captured LM-PCR

1. To resuspend the "Pre-LM-PCR Oligos 1 & 2 (LP1)" (from the SeqCap Adapter Kit), spin the lyophilized primers briefly. Please note that both primers are contained within a single tube.

2. Add 550 µl PCR-grade water to the tube labeled "**Pre-LM-PCR Oligo 1 & 2 (LP1)**" from the SeqCap Adapter Kit A or B. Vortex briefly and spin down. The resuspended oligo tube should be stored later on at -15 to -25 °C.

3. Prepare the LM-PCR Master Mix on ice according to the following table :

Pre-Captured LM-PCR Master Mix	Number of Reaction	Quantity for one Reaction	Master Mix Recipe
KAPA HiFi HotStart **Uracil** + Ready Mix (2x)		25 µl	
PCR grade water		2 µl	
Pre LM-PCR Oligos 1 & 2, 5 µM		3 µl	
Total		30 µl	

4. Pipette 30µl of the LM-PCR master mix into a 0.2ml tube. Add **20 µl** of sample library from step 5.16 (or PCR grade water for negative control). Mix well by pipetting up and down five times. Do not vortex.

Reagent	Quantity
Sample Library (from step 7.16)	20 µl
Pre-Captured LM-PCR Master Mix	30 µl
Total	**50 µl**

5. Amplify samples in thermocycler using the following Pre-Capture LM-PCR program.
   - Step 1: 2 minutes @ 95°C
   - Step 2: 30 seconds @ 98°C
   - Step 3: 30 seconds @ 60°C
   - Step 4: 4 minutes @ 72°C
   - Step 5: Go back to Step 2-4 and repeat 11 times (12 cycles in total) (ask your trainer)
   - Step 6: 10 minute @ 72°C
   - Step 7: Hold @ 4°C

6. Allow the **AMPure XP Beads** (from the SeqCap Pure Capture Bead Kit) to warm to room temperature while the PCR is running. Prepare fresh 80% ethanol (400 µl per PCR reaction are required).

8. Clean up Amplified Sample Library using DNA Purification Beads

Remark: a **1.8 x** AMPure XP Beads ratio is here used (ask your trainer)

1. Be sure that the AMPure XP Beads are warmed to room temperature for at least **30 minutes** before use.
2. Transfer each amplified sample library into a separate 1.5 ml microcentrifuge tube (**low-binding tubes** are suggested).
3. Vortex the beads for 10 seconds before use to ensure a homogenous mixture of beads.
4. Add **90 µl** (or 1.8x volume) AMPure XP Beads to the **50 µl** amplified sample library.
5. Vortex briefly and incubate at room temperature for **15 minutes** to allow the DNA to bind the beads.
6. Place the tube containing the bead bound DNA on a magnetic particle concentrator and allow the solution to clear.
7. Once clear, remove and discard the supernatant being careful not to disturb the bead pellet.
8. Add 200 µl freshly prepared 80% ethanol to the tube containing the beads plus DNA, leaving the tube in the magnetic particle concentrator during this step.
9. Incubate at room temperature for 30 seconds.

10. Remove and discard the supernatant.
11. Repeat steps 7.8 to 7.10 for a total of **two washes** with 80% ethanol.
12. Spin briefly the tube, place it again on the magnet, and remove all residual ethanol without disturbing the beads.
13. Allow the beads to dry at room temperature with the tube lid open until dry. Over drying of the beads can result in yield loss.
14. Remove the tube from the magnetic particle concentrator and resuspend the DNA using 52 **µl** of **PCR-grade water** (it is critical that the amplified bisulfite-converted sample library is eluted with PCR grade water and not buffer EB or 1x TE). Pipette up and down ten times to mix to ensure that all of the beads are resuspended.
15. Incubate at room temperature for 2 minutes.
16. Place the tube back on the magnetic particle concentrator and allow the solution to clear.
17. Transfer 50 µl of the supernatant (that now contains the amplified sample library) to a new 1.5 ml tube.
18. Measure A_{260}/A_{280} on a NanoDrop spectrophotometer to determine the concentration. The sample library yield should be > 1 µg, with a A_{260}/A_{280} between 1.7 and 2.0.

Sample	Concentration (ng/µl)
Sample 1	_____ng/µl
Sample 2	_____ng/µl
Sample 3	_____ng/µl
Sample 4	_____ng/µl

19. Run 1µl pre-capture LM-PCR product on a Bioanalyzer DNA 1000 chip. A successfully constructed library should look like figure below, with an average fragment size between 150 and 500 bp. A sharp peak may be visible below 150 bp. The presence of this material will not interfere with the capture process.

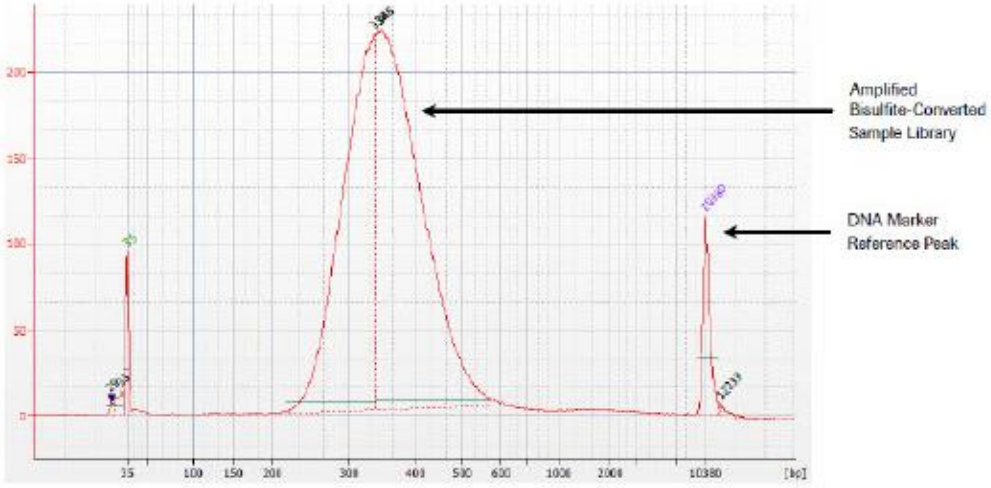


Figure 2 : Example of a constructed library in a Bioanalyzer DNA 1000 (Pre-Capture LM-PCR)

9. Hybridizing the Sample and SeqCap Epi libraries

1. Turn on heat block to 95°C and let it equilibrate to the set temperature.
2. Spin the lyophilized SeqCap HE Universal and required SeqCap HE Index oligo tubes briefly.
3. Add 120 µl PCR-grade water to SeqCap HE Universal Oligo tube (1,000 µM final concentration). Vortex 5 seconds and spin down.
4. Add 10 µl PCR-grade water to each required SeqCap HE Index Oligo tube (1,000 µM final concentration). Vortex 5 seconds and spin down.
5. To prevent damage of the Hybridization Enhancing (HE) oligos due to multiple freeze/thaw cycles, the resuspended oligos can be aliquoted into smaller volumes to minimize the number of freeze/thaw cycles. Store the aliquots at -15 to -25°C.
6. Add 10 µl of **Bisulfite Capture Enhancer**, contained in the SeqCap Epi Accessory Kit, to a new 1.5 ml tube.
7. Add 1 µg of the amplified bisulfite-converted DNA sample (from point 7.17) to the 1.5 ml tube (from point 10.6).
8. Add 1 µl of SeqCap HE Universal Oligo and 1 µl of the appropriate SeqCap HE Index Oligo to the amplified bisulfite-converted DNA sample plus Bisulfite Capture Enhancer.
9. Close the tube's lid and make a hole in the top of the tube's cap with an 18 – 20 gauge or thinner needle, to suppress contamination in the DNA vacuum concentrator.
10. Dry the Multiplex DNA Sample Library Pool / Bisulfite Capture Enhancer / Hybridization Enhancing Oligos in a DNA vacuum concentrator at high heat (+60°C).
11. Once the sample is dry, **cover the hole** with a sticker or small piece of laboratory tape.
12. Add the following reagents to each sample:

Reagent	Volume for One Reaction
2X SC Hybridization Buffer (vial 5)	7.5µl
SC Hybridization Component A (vial 6)	3µl

13. Vortex the sample for 10 seconds and centrifuge at maximum speed for 10 seconds.
14. Place each sample in a 95°C heat block for 10 minutes.
15. During the denaturation step, remove the appropriate number of 4.5 µl SeqCap Epi Library aliquots (one per sample library) from the -15 to -25°C freezer and leave them chilled on ice.
16. After the 10 minutes denaturation step, centrifuge samples at maximum speed for 10 seconds.
17. Transfer the sample to the aliquot of SeqCap Epi Library (in a 0.2 ml tube).
18. Vortex for 3 seconds and centrifuge at maximum speed for 10 seconds.
19. Incubate in a thermocycler at 47°C (with heated lid turned on at 57°C). The hybridization time should be 64 – 72 hours. For training proposals, the time can be reduced (ask your trainer about this point).

Washing and Recovering Captured Multiplex DNA Sample

Allow the Capture Beads to warm to room temperature (at least **30 minutes** prior to use).

10.Prepare Sequence Capture Wash Buffers

1. Dilute 10X SC Wash Buffers (I, II, and III), 10X Stringent Wash Buffer and 2.5X Bead Wash Buffer with PCR grade water to create 1X working solutions. For **one** capture you will need the following volumes:

Buffer	Amount of Concentrated Buffer	Amount of PCR- Grade Water	Total Volume of 1X Buffer	Temperature
Stringent Wash Buffer	40 µl	360 µl	400 µl	47°C
1x Wash Buffer I	10 µl	90 µl	100 µl	47°C
	20 µl	180 µl	200 µl	RT
1x Wash Buffer II	20 µl	180 µl	200 µl	RT
1x Wash Buffer III	20 µl	180 µl	200 µl	RT
1x Bead Wash Buffer	200 µl	300 µl	500 µl	RT

2. Place the working solutions at the appropriate temperature.

11. Preparing Capture Beads

1. Allow the Capture Beads to warm to room temperature for **30 minutes** prior to use.
2. Vortex the beads for 15 seconds.
3. Aliquot **100 µl** of beads for each capture into a single 1.5 ml tube (low-binding tubes are recommended). Enough beads for six captures can be prepared in a single tube.
4. Place the tube in a DynaMag-2 device. When the liquid becomes clear (should take less than 5 minutes), remove the supernatant, being careful to leave all of the beads in the tube (any remaining traces of liquid will be removed with subsequent wash steps).
5. Add **200 µl** of the 1x Bead Wash Buffer from Step 9.1 to the beads (400µl in case of two captures, 600 µl in case of three captures, …).
6. Remove the tube from the DynaMag-2 device and vortex for 10 seconds.

7. Place the tube back into the DynaMag-2 device and, once clear, remove the supernatant.
8. Repeat steps 10.5-10.7 for a total of **two washes**.
9. Resuspend the beads in 100µl 1x Bead Wash Buffer (200 µl in case of two captures, …).
10. Transfer the **100 µl** of resuspended beads into a **new 0.2 ml** tube.
11. Place the tube against the magnet from the DynaMag-2 device to bind the beads.
12. With the tube on the DynaMag-2 device, remove and discard the supernatant.
13. Proceed to point 11.1 as quickly as possible. Do not allow the Capture Beads to dry out.

12.Bind DNA to the Capture Beads

1. Transfer the hybridization samples to the Capture Beads prepared in Step 10.12.
2. Mix by pipetting up and down 10 times.
3. Place the tubes containing the beads and the hybridized samples at 47°C for 45 minutes.
4. Vortex for 3 seconds at 15 minute intervals to ensure the beads remain in suspension.

13.Washing

1. Wash the beads using the table below as a guide:

Wash Buffer	Wash Volume	Mixture Time	Time in Water Bath	Temperature
1x Wash Buffer I	100 µl	Vortex for 10 seconds	-	47°C
1x Stringent Wash Buffer	200 µl	Pipette up and down 10 times	5 minutes	47°C
1x Stringent Wash Buffer	200 µl	Pipette up and down 10 times	5 minutes	47°C
1x Wash Buffer I	200 µl	Vortex for 2 minutes	-	RT
1x Wash Buffer II	200 µl	Vortex for 1 minute	-	RT
1x Wash Buffer III	200 µl	Vortex for 30 seconds	-	RT

2. After the 45 minutes, add 100 µl of 1 x Wash Buffer I heated to 47°C to the 15 µl of Capture Beads plus bound DNA.
3. Mix by vortexing 10 seconds and transfer the entire content of each 0.2 ml tube to a new 1.5 ml tube. Place the tube in the DynaMag-2 device to bind the beads and remove the Wash Buffer I.

4. Remove the tubes from the DynaMag-2 and add 200 µl of 47°C Stringent Wash Buffer to the beads.
5. Pipette up and down 10 times to mix. The temperature should not drop much below 47°C.
6. Incubate in the 47°C water bath for 5 minutes.
7. Place the tube back in the DynaMag-2 device to bind the beads and remove the Stringent Wash Buffer.
8. Remove the tubes from the DynaMag-2 device and add fresh 200 µl of 47°C Stringent Wash Buffer to the beads.
9. Pipette up and down 10 times to mix. The temperature should not drop much below 47°C.
10. Incubate in the 47°C water bath for 5 minutes.
11. Place the tube back in the DynaMag-2 device to bind the beads and remove the Stringent Wash Buffer.
12. Remove the tubes from the DynaMag-2 and add 200 µl of room temperature Wash Buffer I.
13. Mix by vortexing for 2 minutes.
14. Place the tube back in the DynaMag-2 device to bind the beads and remove the Wash Buffer I.
15. Remove the tubes from the DynaMag-2 and add 200 µl of room temperature Wash Buffer II.
16. Mix by vortexing for 1 minute.
17. Place the tube back in the DynaMag-2 device to bind the beads and remove the Wash Buffer II.
18. Remove the tubes from the DynaMag-2 and add 200 µl of room temperature Wash Buffer III.
19. Mix by vortexing for 30 seconds.
20. Place the tube back in the DynaMag-2 device to bind the beads and remove the Wash Buffer III.
21. Remove the tubes from the DynaMag-2 and add 50 µl **PCR-grade water** to the tube of bead-bound captured sample.

14. Post-Captured LM-PCR

1. Spin the lyophilized oligos, contained in the SeqCap Epi Accessory Kit, briefly to allow the contents to pellet at the bottom of the tube. Please note that both oligos are contained within a single tube.
2. Add **480 µl** PCR-grade water to the tube labeled "Post-LM-PCR Oligos 1 & 2" from the SeqCap Epi Accessory Kit.
3. Briefly vortex the primers plus PCR-grade water and spin down the resuspended oligo tube. It should be stored at -15 to -25°C after use.
4. Prepare the Post-Capture LM-PCR Master Mix in a 1,5 ml tube as follows :

Post-Captured LM-PCR Master Mix	Number of Reaction	Quantity for Two Reactions (for one captured DNA sample or negative control)	Master Mix Recipe
KAPA HiFi HotStart Ready Mix		50 µl	
Post-LM-PCR Oligos 1 & 2, 5 µM		10 µl	
Total		**60 µl**	

5. Vortex the bead-bound captured DNA from Step 12.21 to ensure homogenous mixture of beads.
6. Aliquot the following reagents into 2 **PCR tubes** per capture. Mix well by pipetting up and down five times. Do not vortex.

Reagent	Quantity
Bead-bound Captured gDNA from Step 10.21	20 µl
LM-PCR Master Mix from Step 11.4	30 µl
Total	**50 µl**

4. Amplify samples in thermocycler using the following Post-Capture LM-PCR program :

- Step 1: 45 seconds @ 98°C
- Step 2: 15 seconds @ 98°C
- Step 3: 30 seconds @ 60°C
- Step 4: 30 seconds @ 72°C
- Step 5: Go back to Step 2-4 and repeat 15 times (for a total of 16 cycles)
- Step 6: 1 minute @ 72°C
- Step 7: Hold @ 4°C

5. Allow the **AMPure XP Beads** to warm to room temperature while the PCR is running. Prepare fresh 80% ethanol (400 µl per PCR reaction are required).

15.Clean up Amplified Sample Library using DNA Purification Beads

Remark: a **1.8 x** AMPure XP Beads ratio is here used (ask your trainer)

1. Be sure that the AMPure XP Beads are warmed to room temperature for at least **30 minutes** before use.
2. **Pool** the like amplified captured Multiplex DNA Sample Libraries into a separate 1.5 ml microcentrifuge tube (low-binding tubes are suggested).
3. Vortex the beads for 10 seconds before use to ensure a homogenous mixture of beads.
4. Add **180 µl** (or 1.8x volume) AMPure XP Beads to the **100 µl** pooled amplified captured Multiplex DNA Sample Library.
5. Vortex briefly and incubate at room temperature for **15 minutes** to allow the DNA to bind the beads.
6. Place the tube containing the bead bound DNA on a magnetic particle concentrator and allow the solution to clear.
7. Once clear, remove and discard the supernatant being careful not to disturb the beads.
8. Add 200 µl freshly prepared 80% ethanol to the tube containing the beads plus DNA, leaving the tube in the magnetic particle concentrator during this step.
9. Incubate at room temperature for 30 seconds.
10. Remove and discard the 80% ethanol.
11. Repeat steps 14.8 to 14.10 for a total of **two washes** with 80% ethanol.
12. Spin briefly the tube, place it again in the magnet, and remove all residual ethanol without disturbing the beads.
13. Allow the beads to dry at room temperature with the tube lid open. Over drying of the beads can result in yield loss.
14. Remove the tube from the magnetic particle concentrator and resuspend the DNA using **52 µl** of **PCR-grade water**. Pipette up and down ten times to mix to ensure that all of the beads are resuspended.
15. Incubate at room temperature for 2 minutes.
16. Place the tube back on the magnetic particle concentrator and allow the solution to clear.

17 Transfer 50 µl of the supernatant (that now contains the amplified sample library) to a new 1.5 ml tube.

18 Measure A_{260}/A_{280} on a NanoDrop spectrophotometer to determine the concentration. The sample library yield should be ≥ 500 ng, with a A_{260}/A_{280} between 1.7 and 2.0.

Sample	Concentration (ng/µl)
Sample 1	_____ng/µl
Sample 2	_____ng/µl
Sample 3	_____ng/µl
Sample 4	_____ng/µl

19 Run 1µl post-capture LM-PCR product on a Bioanalyzer DNA 1000 chip. A successfully constructed library should look like Figure 3 below, with an average fragment size between 150 and 500 bp.

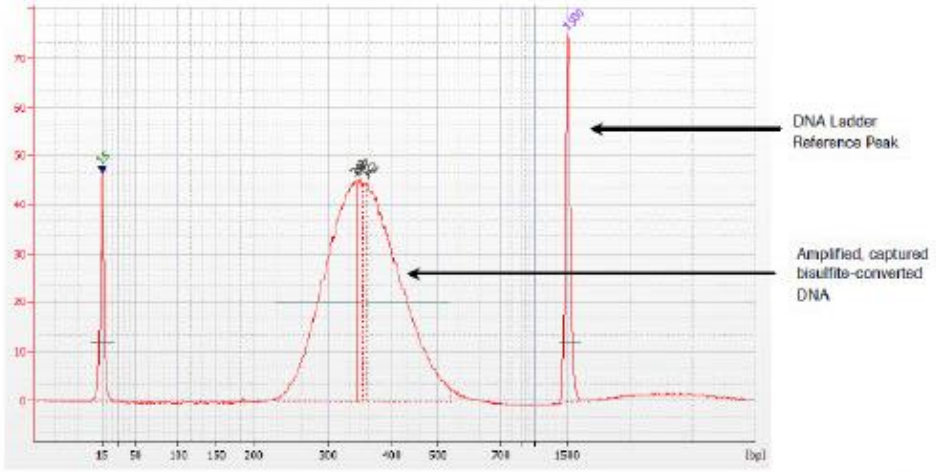


Figure 3 : Example of successfully amplified, captured bisulfite-converted DNA analyzed using an Agilent Bioanalyzer DNA 1000 chip.

12/16

NimbleGen

Published by
Roche NimbleGen, Inc.
504 S. Rosa Rd
Madison, WI 53719 USA

www.sequencing.roche.com/

Supplementary file 4

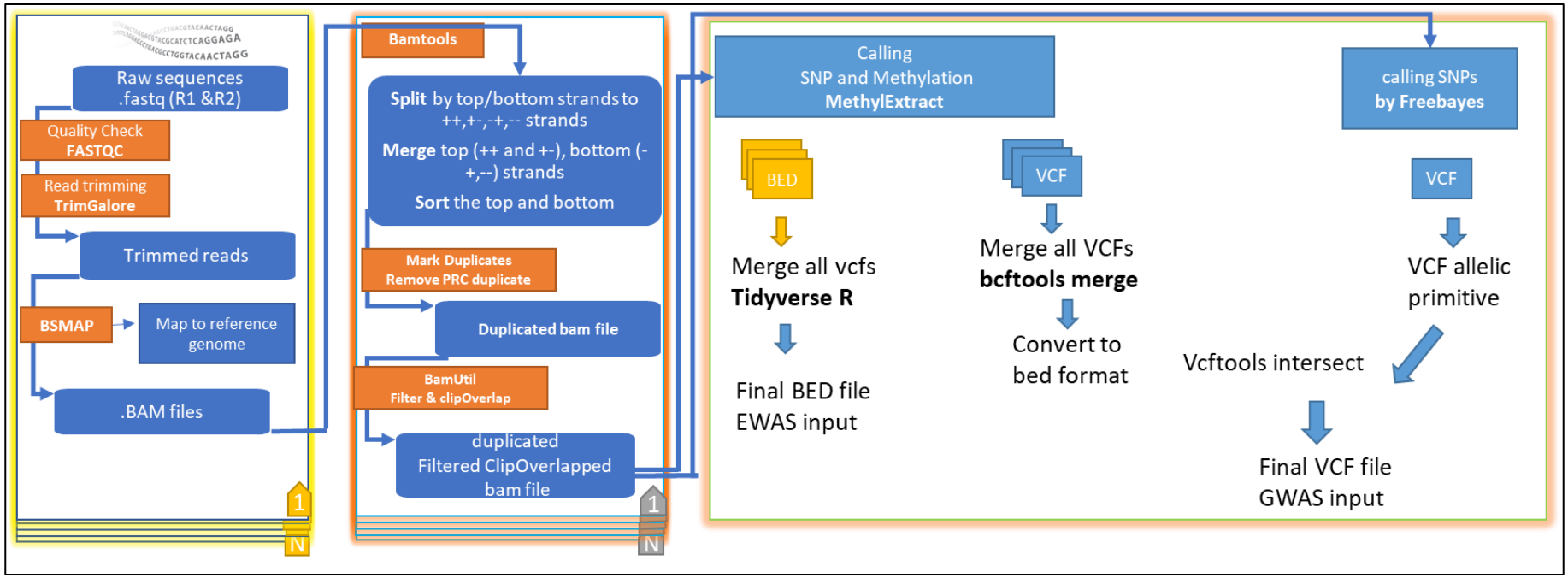


Figure S1: Bioinformatics pipeline for SNP and DNA methylation calling.

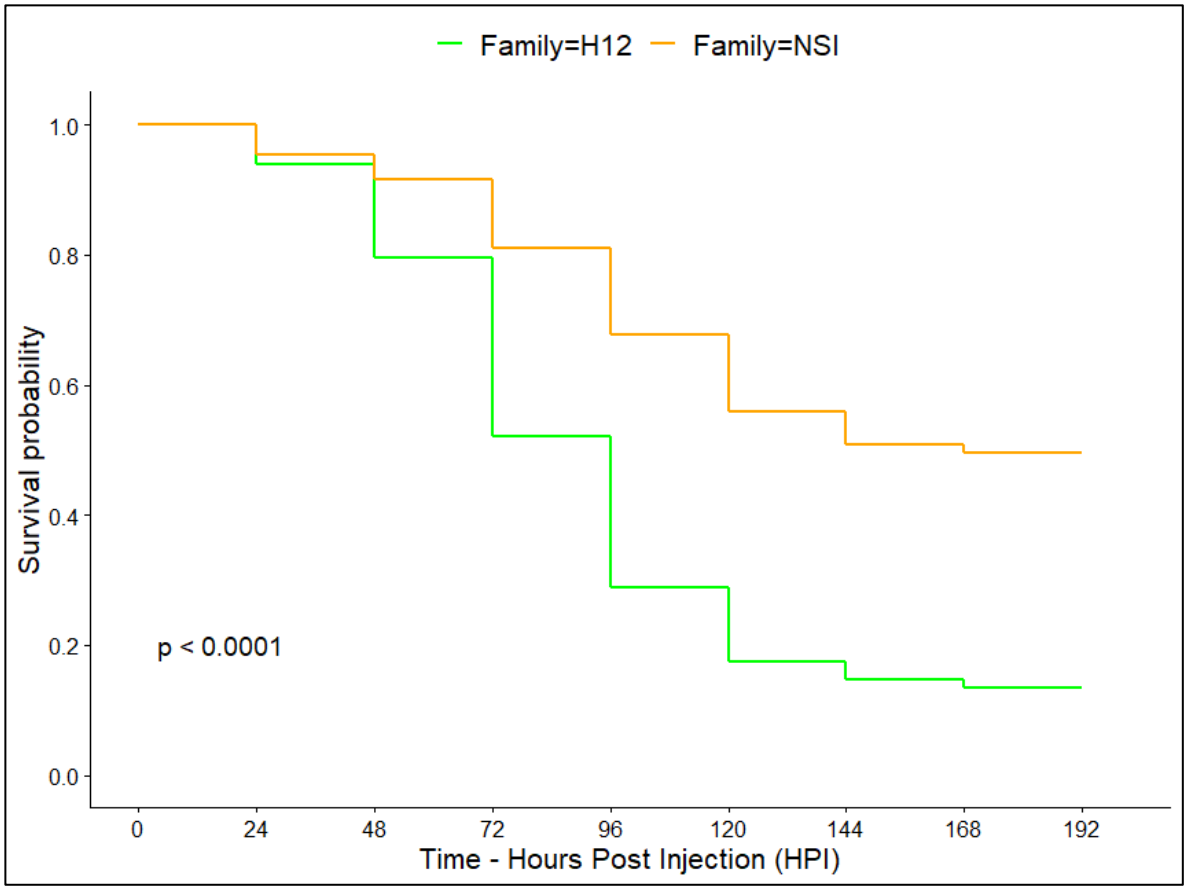


Figure S2: Kaplan-Meier survival curve for the OsHV-1 µVar injected donor families. In green colour, the very susceptible family (H12 families) and in orange the highly genetically diversified cohort (NSI family).

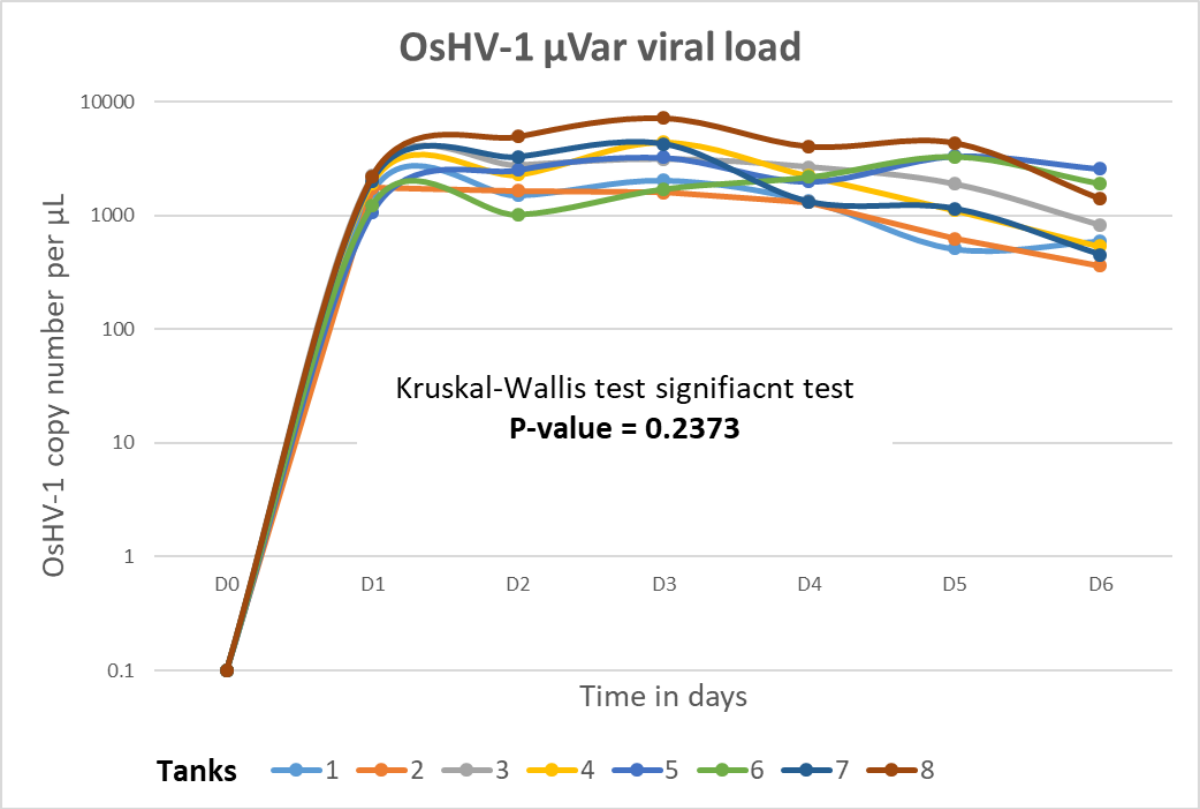


Figure S3: The OsHV-1 *µVar* viral load in eight replicate tanks.
The Y-axis is the viral load on the X-axis is the time in days (D0 to D6; D=day), where the D0 is the beginning of the infection. In the middle is the Kruskal-Wallis test showing no significant differences between eight tanks

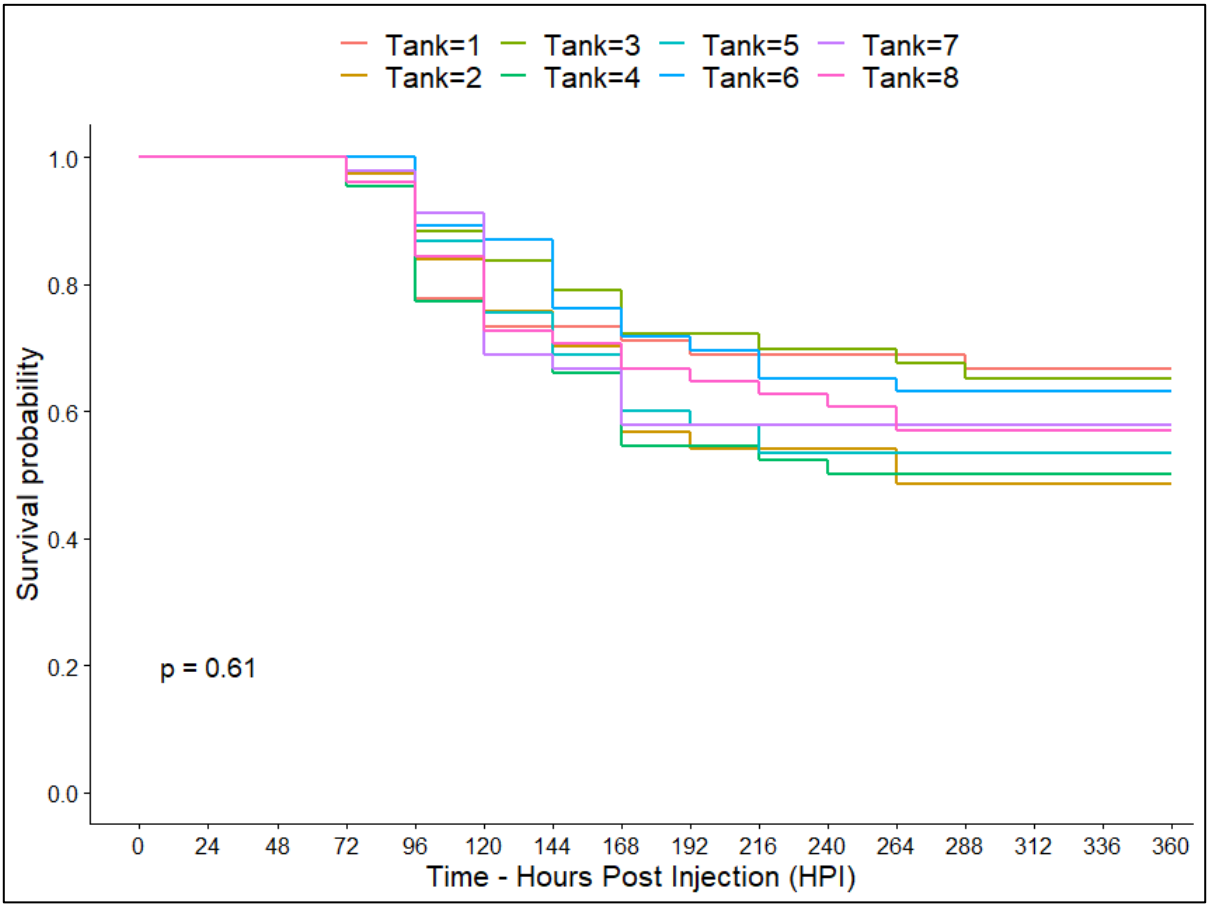


Figure S4: Kaplan-Meier survival curve in all eight replicate tanks.

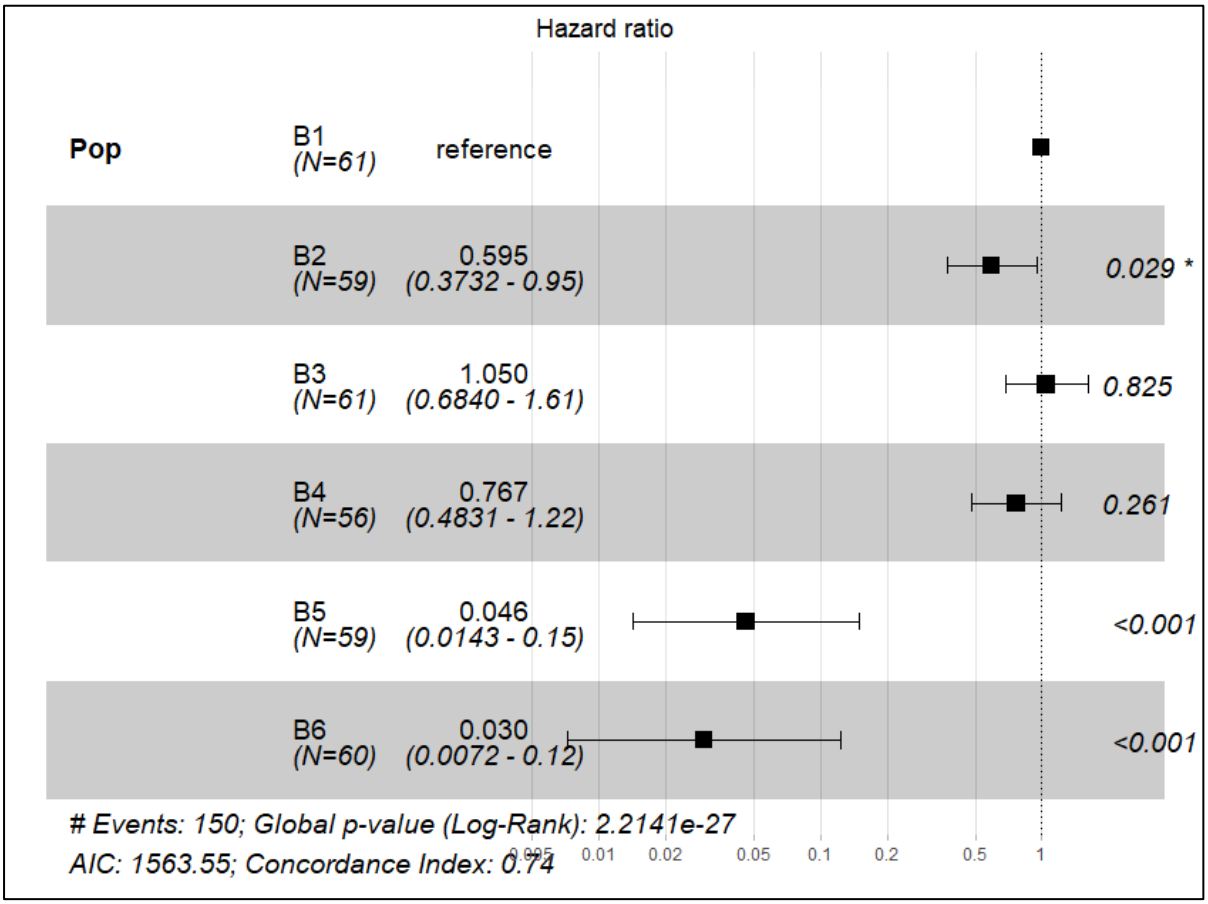


Figure S5: Forest plot showing the relative risk of death for all six populations.

The non-farming populations are B1-B4 and the farming populations are B5 and B6.

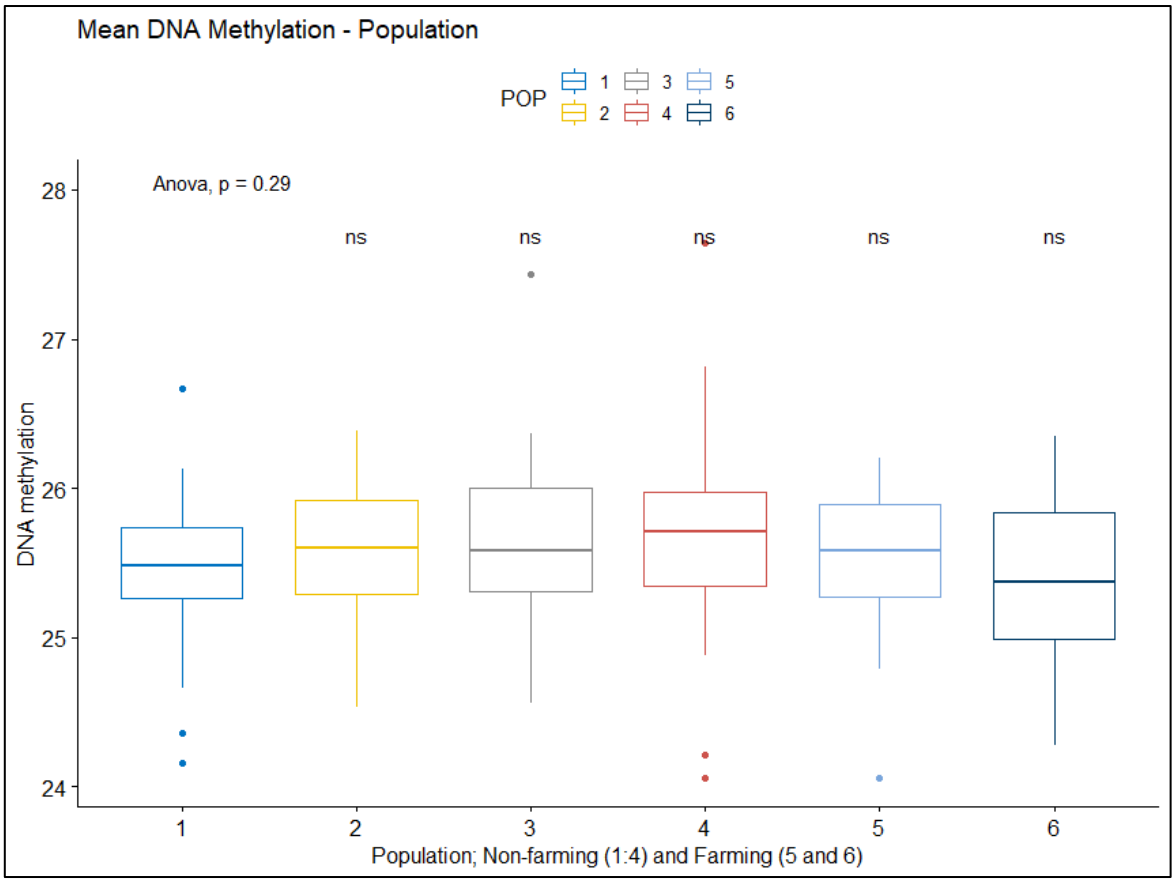


Figure S6: Mean DNA methylation level.
Boxplot showing mean methylation for each population (1-4 non-farming populations; 5-6 farming populations). The overall anova test showing no difference between groups. Additionally the t.test comparing population 1 to all the other population show no significant differences

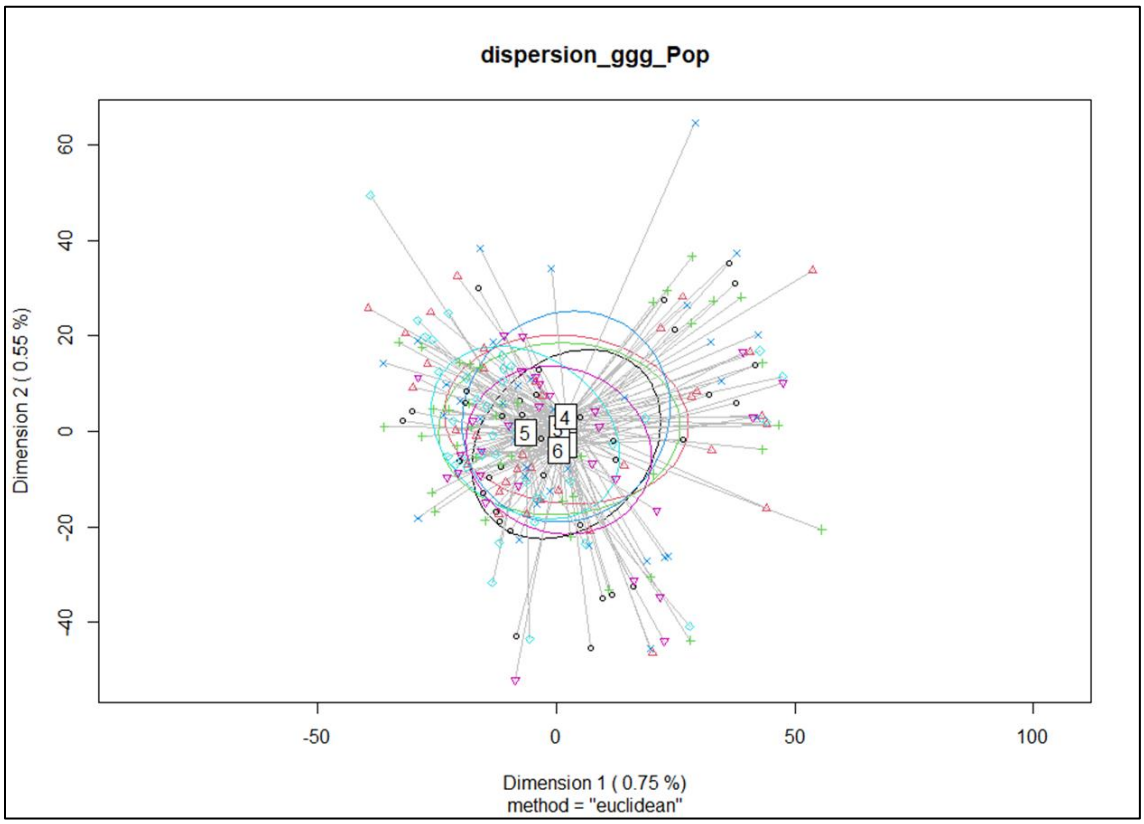


Figure S7: Analysis of multivariate homogeneity of group dispersions (variances) between the six population (1-4= non-farming, 5-6 = farming) for genetic data. The X-axis and Y-axis showing the first and second dimensions (Dim1 and Dim2) respectively, which both represent the highest amount of variance. Between the brackets is the percentages of variation explained by each dimensions.

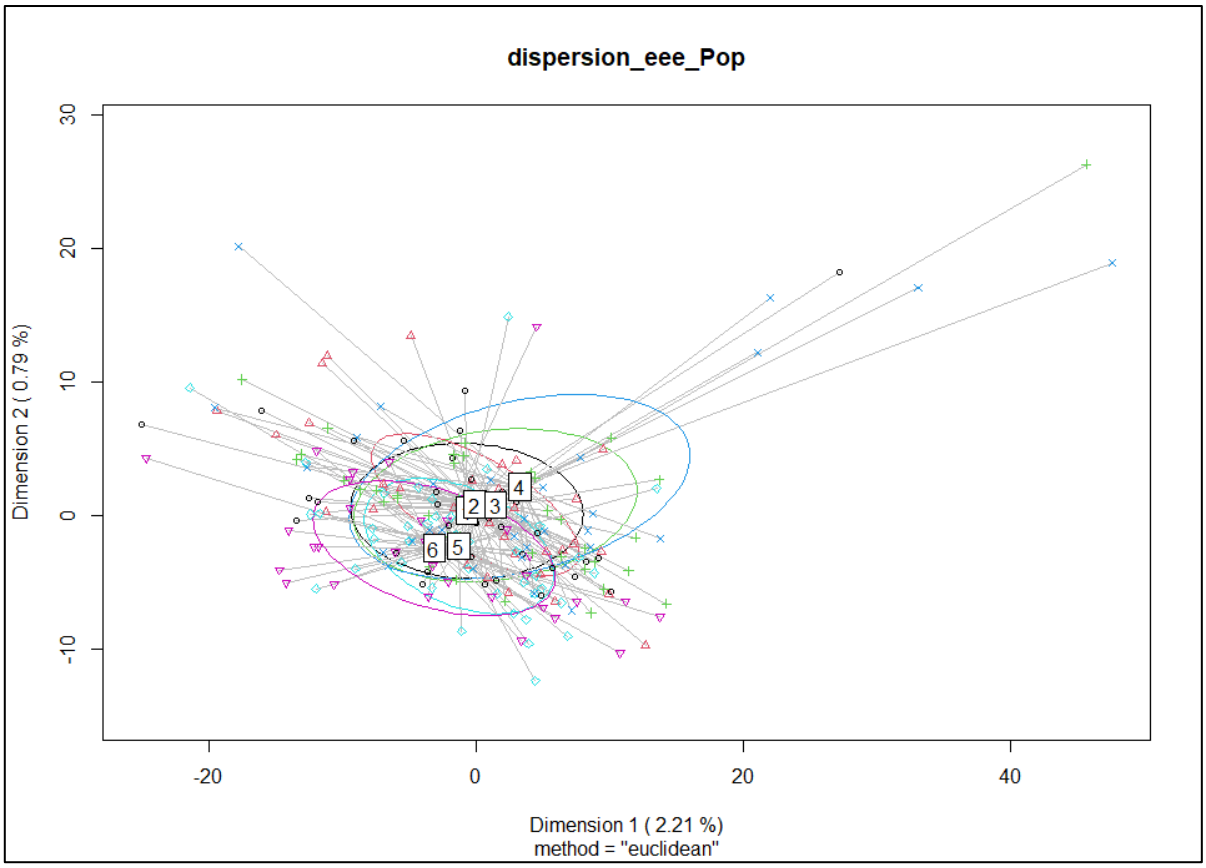


Figure S8: Analysis of multivariate homogeneity of group dispersions (variances) between the six population (1-4= non-farming, 5-6 = farming) for epigenetic data.
The X-axis and Y-axis showing the first and second dimensions (Dim1 and Dim2) respectively, which both represent the highest amount of variance. Between the brackets is the percentages of variation explained by each dimensions.

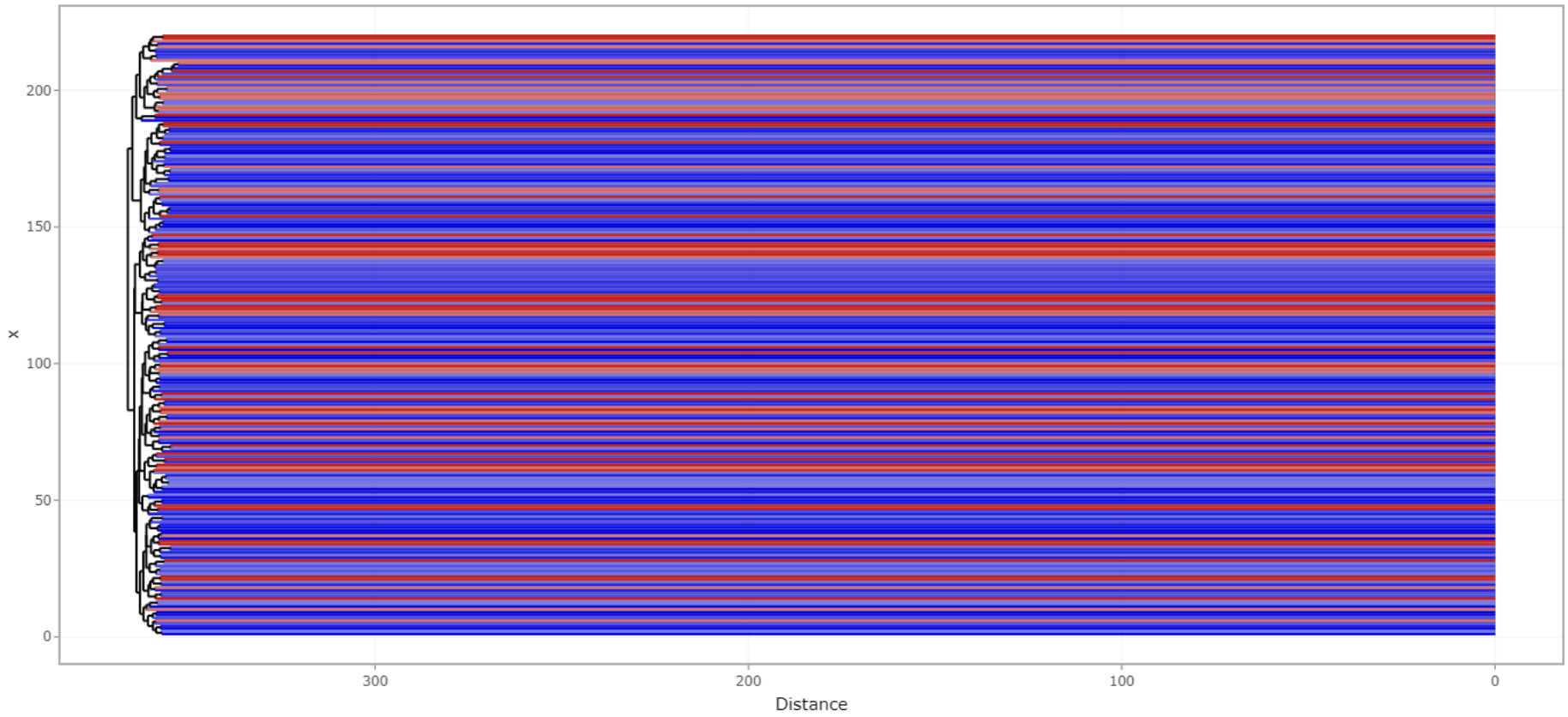


Figure S9: Hierarchical cluster analysis for the genetic data.
The blue gradient colour are the non-farming populations; the red colour gradient are the farming population

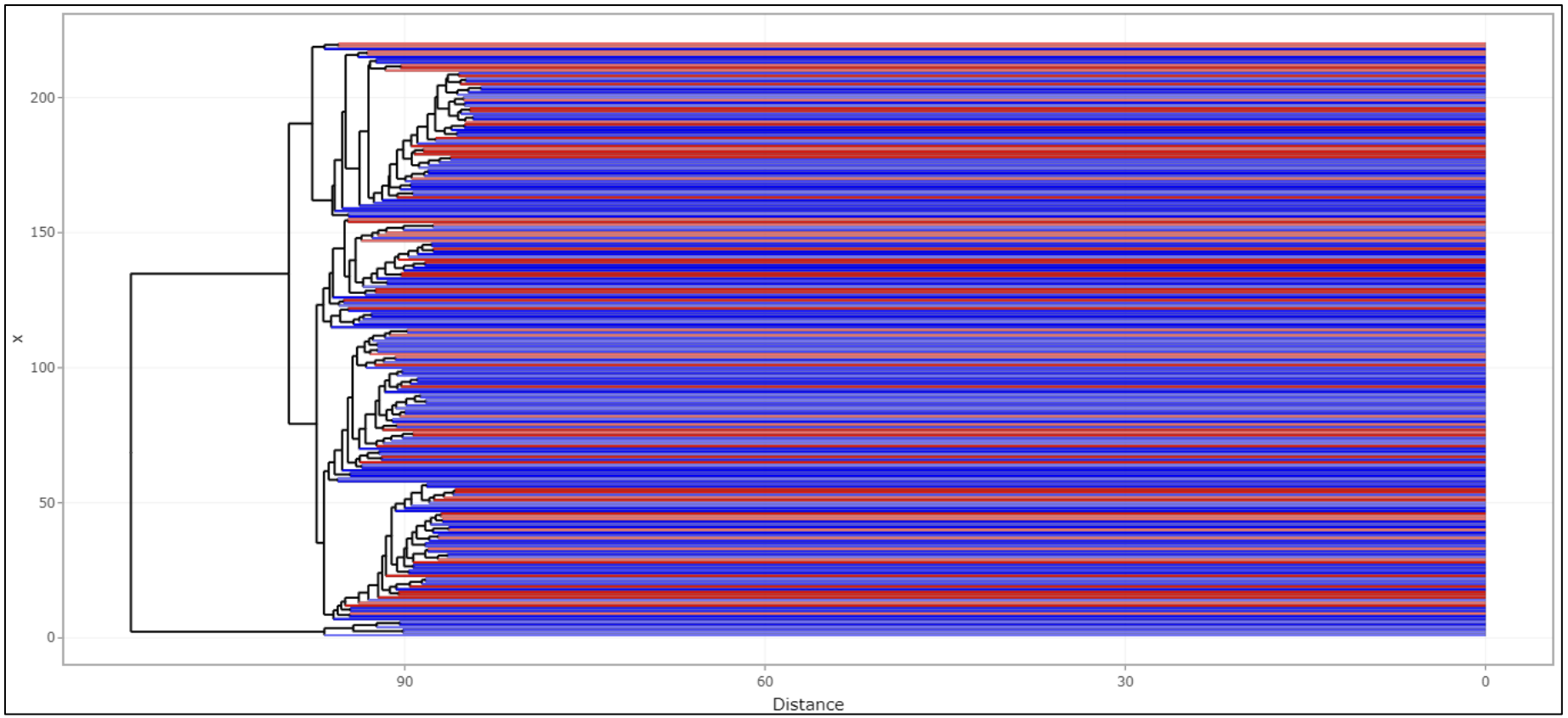


Figure S10: Hierarchical cluster analysis for the epigenetic data.
The blue gradient colour are the non-farming populations; the red colour gradient are the farming population

Supplementary file 5

SNP and DNA methylation Calling

Before the first step - trimming the data to remove the adapter and quality check.

The parameters for trimming: For 150 bp, 2x Paired End reads [remove the last 50]

trim_galore --paired --illumina --quality ${params.quality} --three_prime_clip_R1 50 --three_prime_clip_R2 50

While for trimming reads with 100 bp 2x Paired End reads

trim_galore --paired --illumina --quality ${params.quality} --clip_R1 1 --clip_R2 1

Then to do the Mapping - filtering and SNP - Methylation calling

The pipeline for this is ready. On gitlab Ifremer

https://gitlab.ifremer.fr/bioinfo/nf-core-gem.git

Note is better to select only samples that have a closely number of reads otherwise it will lead to many missing data in the VCF (SNP) and Bed (DNA methylation) files

Step 1 - Map reads to *reference* genome using bsmap

bsmap -r ${params.bsmap_repeat} -n ${params.bsmap_mapstrand} -s ${params.bsmap_seedsize} -p ${task.cpus} -d ${params.genome} -a ${name}_R1_val_1.fq -b ${name}_R2_val_2.fq -o ${name}.sam &> bsmap-${name}.log 2>&1

picard -Xms512m -Xmx${task.memory.toGiga()}g AddOrReplaceReadGroups RGID=${name} RGLB=${name} RGPL=illumina RGSM=${name} RGPU=@A00902:117:HKKNJDRXX:2 VALIDATION_STRINGENCY=LENIENT I=${name}.sam O=${name}.bam &> picard-${name}.log 2>&1

STEP 2 - Split, merge and sort mapped reads using bamtools

bamtools split -tag ${params.bamtools_tag} -in ${bam} &> bamtools-${name}.log 2>&1

bamtools merge -in ${name}.TAG_${params.bamtools_tag}_++.bam -in ${name}.TAG_${params.bamtools_tag}_+-.bam -out ${name}_top_merged.bam &>> bamtools-${name}.log 2>&1

bamtools merge -in ${name}.TAG_${params.bamtools_tag}_-+.bam -in ${name}.TAG_${params.bamtools_tag}_--.bam -out ${name}_bottom_merged.bam &>> bamtools-${name}.log 2>&1

samtools sort ${name}_top_merged.bam > ${name}_top_merged_sorted.bam 2> samtools-${name}.log

samtools sort ${name}_bottom_merged.bam > ${name}_bottom_merged_sorted.bam 2>> samtools-${name}.log

STEP 3 - Mark duplicates with picard tools - remove duplicates

picard -Xms512m -Xmx${task.memory.toGiga()}g -Djava.io.tmpdir=./picard MarkDuplicates \

VALIDATION_STRINGENCY=SILENT \

INPUT=${topbam} \

OUTPUT=${name}_top_rm_dupl.bam \

METRICS_FILE=${name}_top_rm_dupl_metrics.txt \

ASSUME_SORTED=TRUE \

REMOVE_DUPLICATES=TRUE \

CREATE_INDEX=TRUE &> ${name}_top_picard.log 2>&1

picard -Xms512m -Xmx${task.memory.toGiga()}g -Djava.io.tmpdir=./picard MarkDuplicates \

VALIDATION_STRINGENCY=SILENT \

INPUT=${bottombam} \

OUTPUT=${name}_bottom_rm_dupl.bam \

METRICS_FILE=${name}_bottom_rm_dupl_metrics.txt \

ASSUME_SORTED=TRUE \

REMOVE_DUPLICATES=TRUE \

CREATE_INDEX=TRUE &> ${name}_bottom_picard.log 2>&1

STEP 4 - Merge reads with bamtools

bamtools merge -in ${topbam_rm_dupl} -in ${bottombam_rm_dupl} -out ${name}_bsmap_non_masked_rm-dupl.bam &> bamtools-${name}.log 2>&1

STEP 5 - Filter merged reads with bamtools

bamtools filter \

-isMapped true \

-isPaired true \

-isProperPair true \

-forceCompression \

-in ${merged_bam} \

-out ${name}_filtered.bam &> bamtools-${name}.log 2>&1

STEP 6 - Filter clipoverlap with bamutils and index bam file

*bam clipOverlap *

*--stats *

*--in ${filtered_bam} *

--out ${name}_clipped.bam &> bamtutils-${name}.log 2>&1

STEP 7 - Methylation maps and SNP calling with MethylExtract

MethylExtract.pl p=${task.cpus} seq=${params.genome} inDir=. outDir=. minDepthMeth=${params.methylextract_mindepthmeth} minDepthSNV=${params.methylextract_mindepthsnv} context=ALL wigOut=Y bedOut=Y flagW=${params.methylextract_flagw} flagC=${params.methylextract_flagc} &> ${name}_methylextract.log

Then IS MERGE THE BED FILES AND VCF FILES

Two problems here to deal with. That maybe come from the not well having a homogenous number of reads, if a closely number of reads have selected maybe these problem probably won't show up.

First: merge the vcf files (Single Nucleotide Polymorphisms; SNPs containing file)

Because with the *METHYLEXTRACT* package the SNP calling is done based on single sample producing single vcf file reporting only the SNPs. When merging many VCF files from different sample would lead to many missing data. Simply because one sample or many samples would have a homozygote genotype for reference allele and others would have a heterozygote or homozygote genotype for alternative allele.

How to tackle this issue?

First, we use FreeBayes to obtain a single VCF file for all the samples.

First, we use FreeBayes to obtain a single VCF file.

```
#PBS -q omp
#PBS -l walltime=120:00:00
#PBS -l mem=115g
#PBS -l ncpus=56
```

```
## Manage script history
```

```
INPUT_DIR=/home/datawork-ihpe/gem/06_clipped-bam-files
FreeBayes_TOOLS=". /appli/bioinfo/freebayes/latest/env.sh"
#GENOME=/home1/datawork/jgawra/GWAS_EWAS_TEST/vcfcd $INPUT_DIR
$FreeBayes_TOOLS
##########################################
## Shell variables ##
##########################################
#INPUT_DIR=/home1/datawork/jgawra/GWAS_EWAS_TEST
GENOME=/home1/datawork/jgawra/GWAS_EWAS_TEST/vcf/oyster.v9.fa
OUTPUT_DIR=/home1/scratch/jgawra/GWAS_EWAS_TEST
##########################################
## prepapre input file ##
##########################################
 ls -d "${INPUT_DIR}/"*"_clipped.bam" > "${INPUT_DIR}/SAMPLES_clipped_bam.txt"
##########################################
## Freebays variant calliing ##
##########################################
echo "create a list of bam files... "`cat "${INPUT_DIR}/SAMPLES_clipped_bam.txt"`
echo "variant calling..."
freebayes-parallel <(fasta_generate_regions.py "${GENOME}.fai" 10000) 56 -p 2 -f $GENOME
--use-best-n-alleles 2 --use-mapping-quality --min-coverage 8 --no-partial-observations --min-
repeat-entropy 1 -L ${INPUT_DIR}/SAMPLES_clipped_bam.txt >&
${OUTPUT_DIR}/freebayes_248_Brest_samples_s
econd_try.vcf 2> ${OUTPUT_DIR}/freebayes_248_Brest_samples_second_try.vcf.log
```

Then I need to check if the vcf file is ok. Then is to change the vcf from the haplotype to single SNP type by using the below script

#!/usr/bin/env bash

#PBS -q omp

#PBS -l walltime=250:00:00

#PBS -l mem=115g

#PBS -l ncpus=56

DATA=/home/datawork-ihpe-gem-nos

BCFTOOLS_TOOLS=". /appli/bioinfo/vcflib/1.0.0_rc1/env.sh"

cd $DATA

$BCFTOOLS_TOOLS

vcfallelicprimitives -kg freebayes_248_Brest_samples_second_try_remove-bad-lines.vcf > freebayes_248_Brest_samples_second_try_remove_bad_lines_vcfallelicprimitives.vcf

FreeBayes is not able to differentiate a real SNP from an SNP produced by bisulfite treatment.

Therefore, MethylExtract was used to call real SNPs in separate VCF files for each sample independently. This is done in STEP 7 already [page 4].

Then, we used BCFTOOLS to merge all the VCF files into a single VCF file and convert it to a bed file format (which contains the real SNP genomic location).

#!/usr/bin/env bash

#PBS -q sequentiel

#PBS -l walltime=00:30:00

#PBS -l mem=5g

DATA=/home1/datawork/jgawra/GWAS_EWAS_TEST/vcf/vcf-sub

BCFTOOLS_TOOLS=". /appli/bioinfo/bcftools/latest/env.sh"

cd $DATA

$BCFTOOLS_TOOLS

First is to bgzip the vcf file to index it.

*for file in *_sub.vcf*

do

bgzip -c $file

done

Then to index it.

*#for file in *.vcf.gz ; do bcftools index -c $file ; done*

Finally to merge it.

*#bcftools merge --force-samples *vcf.gz -Oz -o Merged.vcf.gz*

Then is to make a bed file like to of all the vcf position (the real SNPs) that to be used in the next step

Finally, we intersected it with a VCF file produced by FreeBayes using vcfintersect from VCFTOOLS (version 0.1.16) and obtained the final VCF file that was will be used for GWAS analyses.

#!/usr/bin/env bash

#PBS -q omp

#PBS -l walltime=50:00:00

#PBS -l mem=50g

#PBS -l ncpus=28

DATA=/home/datawork-ihpe-gem-nos

BED_DATA=/home/datawork-ihpe/gem/06_clipped-bam-files

#BCFTOOLS_TOOLS=". /appli/bioinfo/vcflib/1.0.0_rc1/env.sh"

VCFTOOLS=". /appli/bioinfo/vcftools/latest/env.sh"

cd $DATA

$VCFTOOLS

Vcftools

--vcf freebayes_248_Brest_samples_second_try_remove_bad_lines_vcfallelicprimitives.vcf

--bed ${BED_DATA}/vcf2bed_jb.bed

--out freebayes_248_samples_vcfallelicprimitive_with_region_methylextract.vcf

--temp $SCRATCH –recode

This file is ready for GWAS analysis

####### This is a vcf file that would be used for GWAS analysis.

Second: How to deal with Bed file

First I used R to merge all the bed files (produced by METHYEXTRACT in step 7).

#first I need to prepare the bed files

```
#!/usr/bin/env bash
#PBS -q omp
#PBS -l walltime=48:00:00
#PBS -l select=1:ncpus=28:mem=115g
DATA=/home/datawork-ihpe/gem/09_methylextract_results/
cd $DATA
for i in *.bed
do
 sed '1d' "$i" > ${i%.bed}_temp1.bed
done
for i in *_temp1.bed
do
 awk '{print $1,($3 - 1),($5/10)}' "$i" > ${i}_temp2.bed
done

for i in *_temp2.bed
do
 sed  -i '1i chrom pos  methratio' "$i"
done

for i in *_temp2.bed
do
 sed -e 's/  */\t/g' "$i" > ${i%CG_temp1.bed_temp2.bed}CG_2.bed
done
```

Then is to merge all the bed files

```
Merge <- tibble(chrom = "C13972",
                pos = 83,
                methratio_ValueUseless = 0)

for (file in list.files(path=".", pattern="*CG_2.bed")     ) {
  filename <- file
  Merge <- full_join(Merge,read_tsv(file) %>%
                rename(!!filename := methratio) %>%
                as_tibble() )
}
Merged3.bed <- Merge %>% select(-methratio_ValueUseless)
```

This file is ready for EWAS analysis

Supplementary file 6

Locating the SNPs and CpGs in the new released Genome of *Crassostrea gigas*

Locating the SNP passing the PLINK quality control to the NEW Roslin GENOME (with chromosome information). This is for visualization purpose, so we can have a Manhattan plot with the ten chromosome.

Preparing a FASTA file

The final plink output file (binary fileset; that was used for GWAS association study; containing 214,263 SNPs) were mapped to the new genome CGA using the vcfprimer from

First the final plink binary fileset were converted to vcf file.

Then for each SNP, a 100 bps were added and finally producing a fasta file that each line is a SNP with 100 bps following.

Script for making a FASTA file by vcfprimer v1.0.0. on Linux.

#!/usr/bin/env bash

#PBS -q omp

#PBS -l walltime=40:00:00

#PBS -l mem=20g

#PBS -l ncpus=8

##call the vcflib tool

. /appli/bioinfo/vcflib/1.0.0_rc1/env.sh

data location

DATA=/home/datawork-ihpe-gem-nos/CGA-genome

cd $DATA

vcfprimers /home/datawork-ihpe-gem-nos/filtered_Depth_vcf_8-150/plink.vcf -f oyster.v9.fa -l 100 > reads-100bp_plink_214k.fasta

Align the fasta file to new genome

Then the fasta file is used to align it on the new genome using the BOWTIE2 v2.3.5

B.1- First we make an index for the genome

#!/usr/bin/env bash

#PBS -q omp

#PBS -l mem=50G

#PBS -l ncpus=28

#PBS -l walltime=10:00:00

BANK_DIR=/home/datawork-ihpe-gem-nos/CGA-genome

BANK_FILE_NAME=/home/datawork-ihpe-gem-nos/CGA-genome/GCA_902806645.1_cgigas_uk_roslin_v1_genomic.fna

 # Genome directory

INDEX_NAME=/home/datawork-ihpe-gem-nos/CGA-genome/GCA_902806645.1_cgigas_uk_roslin_v1_genomic

 # Lancement de Bowtie

bowtie2_cpus=$((${NCPUS}-2))

. /appli/bioinfo/bowtie2/2.3.5/env.sh

cd ${BANK_DIR}

bowtie2-build ${BANK_FILE_NAME} ${INDEX_NAME} -p ${bowtie2_cpus} >& $BANK_DIR/mkbowtie.log 2>&1

B.2- Align the fasta to the new genome and produce a SAM file

#!/usr/bin/env bash

#PBS -q omp

#PBS -l mem=50G

#PBS -l ncpus=28

#PBS -l walltime=10:00:00

BANK_DIR=/home/datawork-ihpe-gem-nos/CGA-genome

#BANK_FILE_NAME=/home/datawork-ihpe-gem-nos/CGA-genome/GCA_902806645.1_cgigas_uk_roslin_v1_genomic.fna

 # Genome directory

INDEX_NAME=/home/datawork-ihpe-gem-nos/CGA-genome/GCA_902806645.1_cgigas_uk_roslin_v1_genomic

 # launching the BOWTIE2

bowtie2_cpus=$((${NCPUS}-2))

. /appli/bioinfo/bowtie2/2.3.5/env.sh

cd ${BANK_DIR}

bowtie2 -x ${INDEX_NAME} -f ${BANK_DIR}/reads-100bp_plink_414k.fasta -S reads-50bp_plink_214.fasta.sam

Convert the SAM file to Bed file

Then convert the SAM file were converted to bed using Linux cat, grep and sed and awk tools

cat reads-100bp_plink_214.fasta.sam | grep -v "@HD" | grep -v "@SQ" |grep -v "@PG" | awk '{print $3"\t"$1"\t"($4+100)"\t"$5}' | sed 's/_LEFT//g' > LR_reads-100bp_plink_214K.fasta.sam.bed

The output would be a SNP, and the position of the SNP in the chromosome and its coordinates.

Intersecting with the GWAS output file to be used for Manhattan plot

Then this bed file were intersected with the GWAS association output using the tidyverse package by left_join function in R.

library(tidyverse)

setwd("E:/GEM/PhD-Thesis/Lab Methodology/GWAS_EWAS_ANALYSIS/plink/Final_GWAS")

read the file with the SNPs and their coordinates in the new genome

df <- read.delim("E:/GEM/PhD-Thesis/Lab Methodology/GWAS_EWAS_ANALYSIS/plink/Final_GWAS/LR_reads-100bp_plink_214K.fasta.sam.bed", header=FALSE)

load the GWAS association file to intersect it with the bed file to located each SNP in the new genome

assoc_results <- read.csv("E:/GEM/PhD-Thesis/Lab Methodology/GWAS_EWAS_ANALYSIS/plink/Final_GWAS/assoc_results.assoc", sep="")

assoc_results_pq.qassoc <- read.csv("E:/GEM/PhD-Thesis/Lab Methodology/GWAS_EWAS_ANALYSIS/plink/Final_GWAS/assoc_results-pq.qassoc", sep="")

colnames(df)[1] <- "CHR"

colnames(df)[2] <- "SNP"

colnames(df)[3] <- "POS"

colnames(df)[4] <- "Q"

df2<- cbind(df, read.table(text = as.character(df$CHR), sep = '_'))

df2$SNP <- paste(df2$V2,df2$V3, sep="_")

```
df3 <- df2[c(5,1,2,3)]
library(dplyr)
df4 <- df3 %>%
  group_by(SNP) %>%
  filter(POS==max(POS))
df5 =df4[!duplicated(df4$SNP), ]

logistic_adjusted_merged <- left_join(df5, assoc_results , by=c("SNP"))
logistic_adjusted_merged <- left_join(df5, assoc_results_pq.qassoc , by=c("SNP"))

colnames(logistic_adjusted_merged)[1] <- "CHR"

logistic_adjusted_merged$CHR [grepl("CADCXH*", logistic_adjusted_merged$CHR)] <- "88"
logistic_adjusted_merged$CHR [grepl("LR761634.1", logistic_adjusted_merged$CHR)] <-"1"
logistic_adjusted_merged$CHR [grepl("LR761635.1", logistic_adjusted_merged$CHR)] <-"2"
logistic_adjusted_merged$CHR [grepl("LR761636.1", logistic_adjusted_merged$CHR)] <-"3"
logistic_adjusted_merged$CHR [grepl("LR761637.1", logistic_adjusted_merged$CHR)] <-"4"
logistic_adjusted_merged$CHR [grepl("LR761638.1", logistic_adjusted_merged$CHR)] <-"5"
logistic_adjusted_merged$CHR [grepl("LR761639.1", logistic_adjusted_merged$CHR)] <-"6"
logistic_adjusted_merged$CHR [grepl("LR761640.1", logistic_adjusted_merged$CHR)] <-"7"
logistic_adjusted_merged$CHR [grepl("LR761641.1", logistic_adjusted_merged$CHR)] <-"8"
logistic_adjusted_merged$CHR [grepl("LR761642.1", logistic_adjusted_merged$CHR)] <-"9"
logistic_adjusted_merged$CHR [grepl("LR761643.1", logistic_adjusted_merged$CHR)] <-"10"
##
logistic_adjusted_merged$CHR <- gsub("\\*", "99", logistic_adjusted_merged$CHR)
###write.table(logistic_adjusted_merged, file = "ADD_logistic_merged.txt", sep = "\t", quote
= FALSE, row.names =F)
logistic_adjusted_merged$CHR<-as.numeric(logistic_adjusted_merged$CHR)
logistic_adjusted_merged$POS<-as.numeric(logistic_adjusted_merged$POS)
logistic_adjusted_merged$P<-as.numeric(logistic_adjusted_merged$P)
colnames(logistic_adjusted_merged)[2] <- "chr_position_dir"
```

colnames(logistic_adjusted_merged)[5] <- "chr_v9"

write.table(logistic_adjusted_merged, file ="assoc_results_merged.txt", sep = "\t", row.names = FALSE)

write.table(logistic_adjusted_merged, file ="assoc_results_pq.qassoc_merged.txt", sep = "\t", row.names = FALSE)

Plotting the Manhattan plot and QQplot

library(qqman)

Manhattan plot

jpeg(filename = "2Manhattan_plot_gwas_binary.jpg", width = 1200, height = 550)

manhattan(x = assoc_results_merged, chr = "CHR", bp = "POS", p = "P", genomewideline = -log10(0.05/214318), suggestiveline = -log10(0.0005), col = c("blue", "red"))

dev.off()

jpeg(filename = "2Manhattan_plot_gwas_coninous.jpg", width = 1200, height = 550)

manhattan(x = assoc_results_pq.qassoc_merged, chr = "CHR", bp = "POS", p = "P", genomewideline = -log10(0.05/214318), suggestiveline = -log10(0.0005), col = c("blue", "red"))

dev.off()

####qqplot

jpeg(filename = "qq_plot_gwas_binary.jpg")

qq(assoc_results_merged$P)

dev.off()

jpeg(filename = "qq_plot_gwas_continous.jpg")

qq(assoc_results_pq.qassoc_merged$P)

dev.off()

Intersect the CpG to the new Genome

First prepare bed file

First we prepare a bed file. It is a CpG and its start position and end position for a CpG.

CpG_List = results_95_220_max [c(1)]

df2<- cbind(CpG_List, read.table(text = as.character(CpG_List$TargetID), sep = "_"))

CpG_List = df2[c(2,3)]

write.table(CpG_List, file ="CpG_List-635K.txt", sep = "\t",row.names = F, quote = F,col.names = F)

Prepare a fasta file

To do this, a getfasta function from bedtools was used to get a fasta file.

awk '{print $1"\t"$2"\t"($2+49)"\t"(49-$2)}' CpG_List-635K.txt > tmp1

sed 's/-//g' tmp1 > tmp2

awk '{print $1"\t"$4"\t"$3}' tmp2 > tmp3

bedtools getfasta -fi oyster.v9.fa -bed tmp3 -name > tmp99

Convert the Fasta to bed file

cat 2getfasta_CpG_filtered_bed.fasta.sam | grep -v "@HD" | grep -v "@SQ" |grep -v "@PG" | awk '{print $3"\t"$1"\t"($4+49)"\t"$5}' | sed 's/_LEFT//g' > 2getfasta_CpG_filtered_bed.fasta.sam.bed

Intersecting the Bed with EWAS association output

get2<- read.table("E:/GEM/PhD-Thesis/Lab Methodology/GWAS_EWAS_ANALYSIS/cpgassoc/Final_EWAS/2getfasta_CpG_filtered_bed.fasta.sam.bed", quote="\"", comment.char="")

get = get2 [c(1,2,3)]

colnames(get)[1] <- "CHR"

colnames(get)[2] <- "CHR_Pos_Start_End"

colnames(get)[3] <- "MAPINFO"

get$CHR <- as.character(get$CHR)

get$CHR [grepl("CADCXH*", get$CHR)] <- "88"

get$CHR [grepl("LR761634.1", get$CHR)] <-"1"

get$CHR [grepl("LR761635.1", get$CHR)] <-"2"

get$CHR [grepl("LR761636.1", get$CHR)] <-"3"

get$CHR [grepl("LR761637.1", get$CHR)] <-"4"

get$CHR [grepl("LR761638.1", get$CHR)] <-"5"

get$CHR [grepl("LR761639.1", get$CHR)] <-"6"

get$CHR [grepl("LR761640.1", get$CHR)] <-"7"

```
get$CHR [grepl("LR761641.1", get$CHR)] <-"8"
get$CHR [grepl("LR761642.1", get$CHR)] <-"9"
get$CHR [grepl("LR761643.1", get$CHR)] <-"10"

get$CHR <- as.numeric(get$CHR)
#newthing
get$CHR_Pos_Start_End <- gsub (":", "-", get$CHR_Pos_Start_End)

df2<- cbind(get, read.table(text = as.character(get$CHR_Pos_Start_End), sep = "-"))
df2$TargetID = paste(df2$V1, df2$V2+49,sep="_")
# colnames(df2)[4] <- "TargetID"
# colnames(df2)[5] <- "Pos"
#
# df2$TargetID <- gsub (":", "_", df2$TargetID)
library(tidyverse)
colnames(results_95_220_max)[1] <- "TargetID"
colnames(results_cpg_death2)[1] <- "TargetID"

merge_binary <- left_join(results_cpg_death2, df2, by="TargetID")
merge_continous<- left_join(results_95_220_max, df2, by="TargetID")

merge_binary$CHR[is.na(merge_binary$CHR)] <- 99
merge_binary$MAPINFO[is.na(merge_binary$MAPINFO)] <- 100

merge_continous$CHR[is.na(merge_continous$CHR)] <- 99
merge_continous$MAPINFO[is.na(merge_continous$MAPINFO)] <- 100

write.table(merge_binary, file ="merge_binary_Ewas.txt", sep = "\t", row.names = FALSE)
write.table(merge_continous, file ="merge_continous_Ewas.txt", sep = "\t", row.names = FALSE)
```

```
merge_continous_Ewas <- read.delim("C:/GEM_THESE/Final_EWAS/merge_continous_Ewas_CpG_GENE2.txt")

merge_binary_Ewas <- read.delim("C:/GEM_THESE/Final_EWAS/merge_Binary_Ewas_CpG_GENE2.txt")

colnames(merge_continous_Ewas)[1] <- "SNP"

colnames(merge_binary_Ewas)[1] <- "SNP"

summary(merge_binary_Ewas)

merge_binary_Ewas$P.value[is.na(merge_binary_Ewas$P.value)] <- 1

merge_continous_Ewas$P.value[is.na(merge_continous_Ewas$P.value)] <- 1

library(qqman)

#### plotting by qqman package

jpeg(filename = "Manhattan_plot_ewas_binary.jpg", width = 1200, height = 550 )

manhattan(x = merge_binary_Ewas, chr = "CHR", bp = "MAPINFO", p = "P.value", genomewideline = -log10(1.901340e-05),suggestiveline = FALSE,  col = c("blue", "red"))

dev.off()

jpeg(filename = "Manhattan_plot_ewas_continous.jpg", width = 1200, height = 550 )

manhattan(x = merge_continous_Ewas, chr = "CHR", bp = "MAPINFO", p = "P.value", genomewideline = -log10(2.147985e-05),suggestiveline = FALSE,  col = c("blue", "red"))

dev.off()


jpeg(filename = "qq_plot_ewas_binary.jpg")

qq(merge_binary_Ewas$P.value)

dev.off()

jpeg(filename = "qq_plot_ewas_continous.jpg")

qq(merge_continous_Ewas$P.value)

dev.off()
```

Supplementary file 7

Due to the large size of the file, the file can be reached through the link below.

https://drive.google.com/drive/folders/114V0oY5ySOUtpgUb_E97zhZtr3speIMf?usp=sharing

Supplementary file 8

```
setwd("C:/GEM_THESE/Final_distangle_EWAS_GWAS/Varpart")

library(vegan)

####load the epigenotype file with no NA

# df_E <- read.delim("C:/GEM_THESE/Final_distangle_EWAS_GWAS/epigenotype_220_varpart_input_no_NA.txt")

###### prepare the file for the PCA

data <- df_E

rnames <- data[[1]]# assign labels in column 1 to "rnames"

mat_data <- data.matrix(data[,2:221])  # transform column 2 - end into a matrix

rownames(mat_data) <- rnames

data <- as.matrix(mat_data)

data2 <- data/(100)

EE <- data2

#### transpnse the dataframe

EEE <- t(EE)

m_E <- data.frame (EEE)

####load the genotype file with no NA

# G_df2 <- read.delim("C:/GEM_THESE/Final_distangle_EWAS_GWAS/Final_Variation_partiton/genotype_220_varpart_input_no_NA.txt")

df2_G <- G_df2[, names(df_E)]

data <- df2_G

rnames <- data[[1]]# assign labels in column 1 to "rnames"

mat_data <- data.matrix(data[,2:221])  # transform column 2 - end into a matrix

rownames(mat_data) <- rnames

data <- as.matrix(mat_data)

GG <- data
```

```
GGG <- t(GG)

m_G_1 <- data.frame (GGG)


### make the PCA for epigenetic data

meth.pc=prcomp(m_E)

save(meth.pc, file="meth.pc.Rdata")

load("meth.pc.Rdata")

summary(meth.pc)

meth.bs=meth.pc$x[,1:220]

write.table(meth.bs, file ="prcomp_epigenotype_220_varpart_input.txt", sep = "\t",row.names = T)

### load the Meth.bs

meth.bs <- read.delim("C:/GEM_THESE/Final_distangle_EWAS_GWAS/Varpart/prcomp_epigenotype_220_varpart_input.txt")


### make the PCA for epigenetic data

genet.pc=prcomp(m_G_1)

save(genet.pc, file="genet.pc.Rdata")

load("genet.pc.Rdata")

summary(genet.pc)

genet.bs <- genet.pc$x[,1:220]

write.table(genet.bs, file ="prcomp_genotype_220_varpart_input.txt", sep = "\t", row.names = T)

genet.bs <- read.delim("C:/GEM_THESE/Final_distangle_EWAS_GWAS/Varpart/prcomp_genotype_220_varpart_input.txt")


### load the phenotype

Final_phenotype_binary_220 <- read.delim("C:/GEM_THESE/Final_distangle_EWAS_GWAS/Varpart/Final_phenotype_binary_220.txt")

# df_PP <- Final_phenotype_binary_220[, names(df_E)]
```

```
data <- df_PP

names <- data[[1]]# assign labels in column 1 to "rnames"

mat_data <- data.matrix(data[,2:221])  # transform column 2 - end into a matrix

rownames(mat_data) <- rnames

data <- as.matrix(mat_data)

PP <- data

PPP <- t(PP)

m_P <- data.frame (PPP)

write.table(m_P, file ="phenotype_bin_220_varpart_input.txt", sep = "\t", row.names = T)

m_P <- read.delim("C:/GEM_THESE/Final_distangle_EWAS_GWAS/Varpart/phenotype_bin_220_varpart_input.txt")

Final_phenotype_continous_220 <- read.delim("C:/GEM_THESE/Final_distangle_EWAS_GWAS/Varpart/Final_phenotype_continous_220.txt")

df_PP_con <- Final_phenotype_continous_220[, names(df_E)]

data <- df_PP_con

rnames <- data[[1]]# assign labels in column 1 to "rnames"

mat_data <- data.matrix(data[,2:221])  # transform column 2 - end into a matrix

rownames(mat_data) <- rnames

data <- as.matrix(mat_data)

PP <- data

PPP <- t(PP)

m_P_con <- data.frame (PPP)

write.table(m_P_con, file ="phenotype_con_220_varpart_input.txt", sep = "\t", row.names = T)

m_P_con <- read.delim("C:/GEM_THESE/Final_distangle_EWAS_GWAS/Varpart/phenotype_con_220_varpart_input.txt")

### run the model to select the best genetic axis explain the binary phenotype

mod0=rda(m_P~1)

mod1=rda(m_P~genet.bs[, 1]+genet.bs[, 2]+genet.bs[, 3]+genet.bs[, 4]+genet.bs[, 5]+genet.bs[, 6]+genet.bs[, 7]+genet.bs[, 8]+genet.bs[, 9]+genet.bs[, 10]+genet.bs[, 11]+genet.bs[, 12]+genet.bs[, 13]+genet.bs[, 14]+genet.bs[, 15]+genet.bs[, 16]+genet.bs[,
```

```
17]+genet.bs[, 18]+genet.bs[, 19]+genet.bs[, 20]+genet.bs[, 21]+genet.bs[, 22]+genet.bs[,
23]+genet.bs[, 24]+genet.bs[, 25]+genet.bs[, 26]+genet.bs[, 27]+genet.bs[, 28]+genet.bs[,
29]+genet.bs[, 30]+genet.bs[, 31]+genet.bs[, 32]+genet.bs[, 33]+genet.bs[, 34]+genet.bs[,
35]+genet.bs[, 36]+genet.bs[, 37]+genet.bs[, 38]+genet.bs[, 39]+genet.bs[, 40]+genet.bs[,
41]+genet.bs[, 42]+genet.bs[, 43]+genet.bs[, 44]+genet.bs[, 45]+genet.bs[, 46]+genet.bs[,
47]+genet.bs[, 48]+genet.bs[, 49]+genet.bs[, 50]+genet.bs[, 51]+genet.bs[, 52]+genet.bs[,
53]+genet.bs[, 54]+genet.bs[, 55]+genet.bs[, 56]+genet.bs[, 57]+genet.bs[, 58]+genet.bs[,
59]+genet.bs[, 60]+genet.bs[, 61]+genet.bs[, 62]+genet.bs[, 63]+genet.bs[, 64]+genet.bs[,
65]+genet.bs[, 66]+genet.bs[, 67]+genet.bs[, 68]+genet.bs[, 69]+genet.bs[, 70]+genet.bs[,
71]+genet.bs[, 72]+genet.bs[, 73]+genet.bs[, 74]+genet.bs[, 75]+genet.bs[, 76]+genet.bs[,
77]+genet.bs[, 78]+genet.bs[, 79]+genet.bs[, 80]+genet.bs[, 81]+genet.bs[, 82]+genet.bs[,
83]+genet.bs[, 84]+genet.bs[, 85]+genet.bs[, 86]+genet.bs[, 87]+genet.bs[, 88]+genet.bs[,
89]+genet.bs[, 90]+genet.bs[, 91]+genet.bs[, 92]+genet.bs[, 93]+genet.bs[, 94]+genet.bs[,
95]+genet.bs[, 96]+genet.bs[, 97]+genet.bs[, 98]+genet.bs[, 99]+genet.bs[, 100]+genet.bs[,
101]+genet.bs[,   102]+genet.bs[,   103]+genet.bs[,   104]+genet.bs[,   105]+genet.bs[,
106]+genet.bs[,   107]+genet.bs[,   108]+genet.bs[,   109]+genet.bs[,   110]+genet.bs[,
111]+genet.bs[,   112]+genet.bs[,   113]+genet.bs[,   114]+genet.bs[,   115]+genet.bs[,
116]+genet.bs[,   117]+genet.bs[,   118]+genet.bs[,   119]+genet.bs[,   120]+genet.bs[,
121]+genet.bs[,   122]+genet.bs[,   123]+genet.bs[,   124]+genet.bs[,   125]+genet.bs[,
126]+genet.bs[,   127]+genet.bs[,   128]+genet.bs[,   129]+genet.bs[,   130]+genet.bs[,
131]+genet.bs[,   132]+genet.bs[,   133]+genet.bs[,   134]+genet.bs[,   135]+genet.bs[,
136]+genet.bs[,   137]+genet.bs[,   138]+genet.bs[,   139]+genet.bs[,   140]+genet.bs[,
141]+genet.bs[,   142]+genet.bs[,   143]+genet.bs[,   144]+genet.bs[,   145]+genet.bs[,
146]+genet.bs[,   147]+genet.bs[,   148]+genet.bs[,   149]+genet.bs[,   150]+genet.bs[,
151]+genet.bs[,   152]+genet.bs[,   153]+genet.bs[,   154]+genet.bs[,   155]+genet.bs[,
156]+genet.bs[,   157]+genet.bs[,   158]+genet.bs[,   159]+genet.bs[,   160]+genet.bs[,
161]+genet.bs[,   162]+genet.bs[,   163]+genet.bs[,   164]+genet.bs[,   165]+genet.bs[,
166]+genet.bs[,   167]+genet.bs[,   168]+genet.bs[,   169]+genet.bs[,   170]+genet.bs[,
171]+genet.bs[,   172]+genet.bs[,   173]+genet.bs[,   174]+genet.bs[,   175]+genet.bs[,
176]+genet.bs[,   177]+genet.bs[,   178]+genet.bs[,   179]+genet.bs[,   180]+genet.bs[,
181]+genet.bs[,   182]+genet.bs[,   183]+genet.bs[,   184]+genet.bs[,   185]+genet.bs[,
186]+genet.bs[,   187]+genet.bs[,   188]+genet.bs[,   189]+genet.bs[,   190]+genet.bs[,
191]+genet.bs[,   192]+genet.bs[,   193]+genet.bs[,   194]+genet.bs[,   195]+genet.bs[,
196]+genet.bs[,   197]+genet.bs[,   198]+genet.bs[,   199]+genet.bs[,   200]+genet.bs[,
201]+genet.bs[,   202]+genet.bs[,   203]+genet.bs[,   204]+genet.bs[,   205]+genet.bs[,
206]+genet.bs[,   207]+genet.bs[,   208]+genet.bs[,   209]+genet.bs[,   210]+genet.bs[,
211]+genet.bs[,   212]+genet.bs[,   213]+genet.bs[,   214]+genet.bs[,   215]+genet.bs[,
216]+genet.bs[, 217]+genet.bs[, 218])


###ordistep for binary with all the PCs, here almost most of PCs are significant

gen_bin= ordistep(mod0, mod1, Pin=0.05, permutations=999)

save(gen_bin, file="ordistep_gen_bin.Rdata")

load("ordistep_gen_bin.Rdata")
```

WITH 999 PERM, select the significant axis that been selected by ordistep

GENET=data.frame(cbind(genet.bs[, 151] ,genet.bs[, 126] ,genet.bs[, 28] ,genet.bs[, 42] ,genet.bs[, 178] ,genet.bs[, 35] ,genet.bs[, 102] ,genet.bs[, 166] ,genet.bs[, 149] ,genet.bs[, 209] ,genet.bs[, 135] ,genet.bs[, 95] ,genet.bs[, 6] ,genet.bs[, 156] ,genet.bs[, 73] ,genet.bs[, 158] ,genet.bs[, 23] ,genet.bs[, 133] ,genet.bs[, 21] ,genet.bs[, 199] ,genet.bs[, 183] ,genet.bs[, 39] ,genet.bs[, 5] ,genet.bs[, 116] ,genet.bs[, 177] ,genet.bs[, 186] ,genet.bs[, 86] ,genet.bs[, 88] ,genet.bs[, 90] ,genet.bs[, 34]))

run the model to select the best epigenetic axis explain the binary phenotype

mod2=rda(m_P~meth.bs[, 1]+meth.bs[, 2]+meth.bs[, 3]+meth.bs[, 4]+meth.bs[, 5]+meth.bs[, 6]+meth.bs[, 7]+meth.bs[, 8]+meth.bs[, 9]+meth.bs[, 10]+meth.bs[, 11]+meth.bs[, 12]+meth.bs[, 13]+meth.bs[, 14]+meth.bs[, 15]+meth.bs[, 16]+meth.bs[, 17]+meth.bs[, 18]+meth.bs[, 19]+meth.bs[, 20]+meth.bs[, 21]+meth.bs[, 22]+meth.bs[, 23]+meth.bs[, 24]+meth.bs[, 25]+meth.bs[, 26]+meth.bs[, 27]+meth.bs[, 28]+meth.bs[, 29]+meth.bs[, 30]+meth.bs[, 31]+meth.bs[, 32]+meth.bs[, 33]+meth.bs[, 34]+meth.bs[, 35]+meth.bs[, 36]+meth.bs[, 37]+meth.bs[, 38]+meth.bs[, 39]+meth.bs[, 40]+meth.bs[, 41]+meth.bs[, 42]+meth.bs[, 43]+meth.bs[, 44]+meth.bs[, 45]+meth.bs[, 46]+meth.bs[, 47]+meth.bs[, 48]+meth.bs[, 49]+meth.bs[, 50]+meth.bs[, 51]+meth.bs[, 52]+meth.bs[, 53]+meth.bs[, 54]+meth.bs[, 55]+meth.bs[, 56]+meth.bs[, 57]+meth.bs[, 58]+meth.bs[, 59]+meth.bs[, 60]+meth.bs[, 61]+meth.bs[, 62]+meth.bs[, 63]+meth.bs[, 64]+meth.bs[, 65]+meth.bs[, 66]+meth.bs[, 67]+meth.bs[, 68]+meth.bs[, 69]+meth.bs[, 70]+meth.bs[, 71]+meth.bs[, 72]+meth.bs[, 73]+meth.bs[, 74]+meth.bs[, 75]+meth.bs[, 76]+meth.bs[, 77]+meth.bs[, 78]+meth.bs[, 79]+meth.bs[, 80]+meth.bs[, 81]+meth.bs[, 82]+meth.bs[, 83]+meth.bs[, 84]+meth.bs[, 85]+meth.bs[, 86]+meth.bs[, 87]+meth.bs[, 88]+meth.bs[, 89]+meth.bs[, 90]+meth.bs[, 91]+meth.bs[, 92]+meth.bs[, 93]+meth.bs[, 94]+meth.bs[, 95]+meth.bs[, 96]+meth.bs[, 97]+meth.bs[, 98]+meth.bs[, 99]+meth.bs[, 100]+meth.bs[, 101]+meth.bs[, 102]+meth.bs[, 103]+meth.bs[, 104]+meth.bs[, 105]+meth.bs[, 106]+meth.bs[, 107]+meth.bs[, 108]+meth.bs[, 109]+meth.bs[, 110]+meth.bs[, 111]+meth.bs[, 112]+meth.bs[, 113]+meth.bs[, 114]+meth.bs[, 115]+meth.bs[, 116]+meth.bs[, 117]+meth.bs[, 118]+meth.bs[, 119]+meth.bs[, 120]+meth.bs[, 121]+meth.bs[, 122]+meth.bs[, 123]+meth.bs[, 124]+meth.bs[, 125]+meth.bs[, 126]+meth.bs[, 127]+meth.bs[, 128]+meth.bs[, 129]+meth.bs[, 130]+meth.bs[, 131]+meth.bs[, 132]+meth.bs[, 133]+meth.bs[, 134]+meth.bs[, 135]+meth.bs[, 136]+meth.bs[, 137]+meth.bs[, 138]+meth.bs[, 139]+meth.bs[, 140]+meth.bs[, 141]+meth.bs[, 142]+meth.bs[, 143]+meth.bs[, 144]+meth.bs[, 145]+meth.bs[, 146]+meth.bs[, 147]+meth.bs[, 148]+meth.bs[, 149]+meth.bs[, 150]+meth.bs[, 151]+meth.bs[, 152]+meth.bs[, 153]+meth.bs[, 154]+meth.bs[, 155]+meth.bs[, 156]+meth.bs[, 157]+meth.bs[, 158]+meth.bs[, 159]+meth.bs[, 160]+meth.bs[, 161]+meth.bs[, 162]+meth.bs[, 163]+meth.bs[, 164]+meth.bs[, 165]+meth.bs[, 166]+meth.bs[, 167]+meth.bs[, 168]+meth.bs[, 169]+meth.bs[, 170]+meth.bs[, 171]+meth.bs[, 172]+meth.bs[, 173]+meth.bs[, 174]+meth.bs[, 175]+meth.bs[, 176]+meth.bs[, 177]+meth.bs[, 178]+meth.bs[, 179]+meth.bs[, 180]+meth.bs[, 181]+meth.bs[, 182]+meth.bs[, 183]+meth.bs[, 184]+meth.bs[, 185]+meth.bs[, 186]+meth.bs[, 187]+meth.bs[, 188]+meth.bs[, 189]+meth.bs[, 190]+meth.bs[, 191]+meth.bs[, 192]+meth.bs[, 193]+meth.bs[, 194]+meth.bs[, 195]+meth.bs[, 196]+meth.bs[,

```
197]+meth.bs[,  198]+meth.bs[,  199]+meth.bs[,  200]+meth.bs[,  201]+meth.bs[,
202]+meth.bs[,  203]+meth.bs[,  204]+meth.bs[,  205]+meth.bs[,  206]+meth.bs[,
207]+meth.bs[,  208]+meth.bs[,  209]+meth.bs[,  210]+meth.bs[,  211]+meth.bs[,
212]+meth.bs[,  213]+meth.bs[,  214]+meth.bs[,  215]+meth.bs[,  216]+meth.bs[,
217]+meth.bs[, 218])

###ordistep for binary with all the PCs, here almost most of PCs are significant

meth_bin= ordistep(mod0, mod2, Pin=0.05, permutations=999)

save(meth_bin, file="ordistep_meth_bin.Rdata")

load("ordistep_meth_bin.Rdata")


### WITH 999 PERM, select the significant axis that been selected by ordistep

METH=data.frame(cbind(meth.bs[, 2] ,meth.bs[, 6] ,meth.bs[, 4] ,meth.bs[, 35] ,meth.bs[, 24]
,meth.bs[, 1] ,meth.bs[, 72] ,meth.bs[, 15] ,meth.bs[, 100] ,meth.bs[, 106] ,meth.bs[, 86]
,meth.bs[, 26] ,meth.bs[, 77] ,meth.bs[, 22] ,meth.bs[, 145] ,meth.bs[, 208] ,meth.bs[, 141]
,meth.bs[, 200] ,meth.bs[, 5] ,meth.bs[, 43] ,meth.bs[, 32] ,meth.bs[, 13] ,meth.bs[, 158]
,meth.bs[, 197] ,meth.bs[, 81] ,meth.bs[, 108] ,meth.bs[, 153] ,meth.bs[, 20] ,meth.bs[, 121]
,meth.bs[, 148] ,meth.bs[, 55] ,meth.bs[, 149] ,meth.bs[, 103] ,meth.bs[, 90] ,meth.bs[, 94]
,meth.bs[, 129] ,meth.bs[, 82] ,meth.bs[, 83] ,meth.bs[, 68] ,meth.bs[, 87]))


varpart_bin = varpart(m_P,GENET,METH)

varpart_bin

# Partition of variance in RDA

# Call: varpart(Y = m_P, X = GENET, METH)

# Explanatory tables:

#   X1:  GENET

# X2:  METH

# No. of explanatory tables: 2

# Total variation (SS): 54.709

# Variance: 0.24981

# No. of observations: 220

# Partition table:

#   Df R.squared Adj.R.squared Testable

# [a+b] = X1         30   0.53922      0.46608     TRUE

# [b+c] = X2         40   0.66981      0.59602     TRUE
```

```
# [a+b+c] = X1+X2      70   0.81431      0.72707     TRUE
# Individual fractions
# [a] = X1|X2          30               0.13105     TRUE
# [b]                  0                0.33503     FALSE
# [c] = X2|X1          40               0.26098     TRUE
# [d] = Residuals                       0.27293     FALSE
# ---
#   Use function 'rda' to test significance of fractions of interest
plot(varpart_bin,  digits = 1, Xnames = c('Genetic', 'Epigenetic'), bg = c('Blue', 'red'))

################### With semi-continuous phenotype #######
### run the model to select the best genetic axis explain the semi-quantitative phenotype

modA=rda(m_P_con~1)
modB=rda(m_P_con~genet.bs[,  1]+genet.bs[,  2]+genet.bs[,  3]+genet.bs[,  4]+genet.bs[,
5]+genet.bs[,  6]+genet.bs[,  7]+genet.bs[,  8]+genet.bs[,  9]+genet.bs[,  10]+genet.bs[,
11]+genet.bs[,  12]+genet.bs[,  13]+genet.bs[,  14]+genet.bs[,  15]+genet.bs[,  16]+genet.bs[,
17]+genet.bs[,  18]+genet.bs[,  19]+genet.bs[,  20]+genet.bs[,  21]+genet.bs[,  22]+genet.bs[,
23]+genet.bs[,  24]+genet.bs[,  25]+genet.bs[,  26]+genet.bs[,  27]+genet.bs[,  28]+genet.bs[,
29]+genet.bs[,  30]+genet.bs[,  31]+genet.bs[,  32]+genet.bs[,  33]+genet.bs[,  34]+genet.bs[,
35]+genet.bs[,  36]+genet.bs[,  37]+genet.bs[,  38]+genet.bs[,  39]+genet.bs[,  40]+genet.bs[,
41]+genet.bs[,  42]+genet.bs[,  43]+genet.bs[,  44]+genet.bs[,  45]+genet.bs[,  46]+genet.bs[,
47]+genet.bs[,  48]+genet.bs[,  49]+genet.bs[,  50]+genet.bs[,  51]+genet.bs[,  52]+genet.bs[,
53]+genet.bs[,  54]+genet.bs[,  55]+genet.bs[,  56]+genet.bs[,  57]+genet.bs[,  58]+genet.bs[,
59]+genet.bs[,  60]+genet.bs[,  61]+genet.bs[,  62]+genet.bs[,  63]+genet.bs[,  64]+genet.bs[,
65]+genet.bs[,  66]+genet.bs[,  67]+genet.bs[,  68]+genet.bs[,  69]+genet.bs[,  70]+genet.bs[,
71]+genet.bs[,  72]+genet.bs[,  73]+genet.bs[,  74]+genet.bs[,  75]+genet.bs[,  76]+genet.bs[,
77]+genet.bs[,  78]+genet.bs[,  79]+genet.bs[,  80]+genet.bs[,  81]+genet.bs[,  82]+genet.bs[,
83]+genet.bs[,  84]+genet.bs[,  85]+genet.bs[,  86]+genet.bs[,  87]+genet.bs[,  88]+genet.bs[,
89]+genet.bs[,  90]+genet.bs[,  91]+genet.bs[,  92]+genet.bs[,  93]+genet.bs[,  94]+genet.bs[,
95]+genet.bs[,  96]+genet.bs[,  97]+genet.bs[,  98]+genet.bs[,  99]+genet.bs[,  100]+genet.bs[,
101]+genet.bs[,    102]+genet.bs[,    103]+genet.bs[,    104]+genet.bs[,    105]+genet.bs[,
106]+genet.bs[,    107]+genet.bs[,    108]+genet.bs[,    109]+genet.bs[,    110]+genet.bs[,
111]+genet.bs[,    112]+genet.bs[,    113]+genet.bs[,    114]+genet.bs[,    115]+genet.bs[,
116]+genet.bs[,    117]+genet.bs[,    118]+genet.bs[,    119]+genet.bs[,    120]+genet.bs[,
121]+genet.bs[,    122]+genet.bs[,    123]+genet.bs[,    124]+genet.bs[,    125]+genet.bs[,
126]+genet.bs[,    127]+genet.bs[,    128]+genet.bs[,    129]+genet.bs[,    130]+genet.bs[,
131]+genet.bs[,    132]+genet.bs[,    133]+genet.bs[,    134]+genet.bs[,    135]+genet.bs[,
136]+genet.bs[,    137]+genet.bs[,    138]+genet.bs[,    139]+genet.bs[,    140]+genet.bs[,
```

141]+genet.bs[, 142]+genet.bs[, 143]+genet.bs[, 144]+genet.bs[, 145]+genet.bs[, 146]+genet.bs[, 147]+genet.bs[, 148]+genet.bs[, 149]+genet.bs[, 150]+genet.bs[, 151]+genet.bs[, 152]+genet.bs[, 153]+genet.bs[, 154]+genet.bs[, 155]+genet.bs[, 156]+genet.bs[, 157]+genet.bs[, 158]+genet.bs[, 159]+genet.bs[, 160]+genet.bs[, 161]+genet.bs[, 162]+genet.bs[, 163]+genet.bs[, 164]+genet.bs[, 165]+genet.bs[, 166]+genet.bs[, 167]+genet.bs[, 168]+genet.bs[, 169]+genet.bs[, 170]+genet.bs[, 171]+genet.bs[, 172]+genet.bs[, 173]+genet.bs[, 174]+genet.bs[, 175]+genet.bs[, 176]+genet.bs[, 177]+genet.bs[, 178]+genet.bs[, 179]+genet.bs[, 180]+genet.bs[, 181]+genet.bs[, 182]+genet.bs[, 183]+genet.bs[, 184]+genet.bs[, 185]+genet.bs[, 186]+genet.bs[, 187]+genet.bs[, 188]+genet.bs[, 189]+genet.bs[, 190]+genet.bs[, 191]+genet.bs[, 192]+genet.bs[, 193]+genet.bs[, 194]+genet.bs[, 195]+genet.bs[, 196]+genet.bs[, 197]+genet.bs[, 198]+genet.bs[, 199]+genet.bs[, 200]+genet.bs[, 201]+genet.bs[, 202]+genet.bs[, 203]+genet.bs[, 204]+genet.bs[, 205]+genet.bs[, 206]+genet.bs[, 207]+genet.bs[, 208]+genet.bs[, 209]+genet.bs[, 210]+genet.bs[, 211]+genet.bs[, 212]+genet.bs[, 213]+genet.bs[, 214]+genet.bs[, 215]+genet.bs[, 216]+genet.bs[, 217]+genet.bs[, 218])

###ordistep for binary with all the PCs, here almost most of PCs are significant

genet_con = ordistep(modA, modB, Pin=0.05, permutations=999)

save(genet_con, file="ordistep_genet_con.Rdata")

genet_con$anova

WITH 999 PERM, select the significant axis that been selected by ordistep

GENET2=data.frame(cbind(genet.bs[, 35] ,genet.bs[, 151] ,genet.bs[, 28] ,genet.bs[, 126] ,genet.bs[, 42] ,genet.bs[, 73] ,genet.bs[, 21] ,genet.bs[, 178] ,genet.bs[, 102] ,genet.bs[, 158] ,genet.bs[, 166] ,genet.bs[, 23] ,genet.bs[, 45] ,genet.bs[, 6] ,genet.bs[, 209] ,genet.bs[, 146] ,genet.bs[, 55] ,genet.bs[, 5] ,genet.bs[, 133] ,genet.bs[, 116] ,genet.bs[, 95] ,genet.bs[, 135] ,genet.bs[, 118] ,genet.bs[, 149] ,genet.bs[, 215] ,genet.bs[, 205] ,genet.bs[, 72] ,genet.bs[, 159] ,genet.bs[, 98]))

run the model to select the best epigenetic axis explain the semi-quantitative phenotype

modC=rda(m_P_con~meth.bs[, 1]+meth.bs[, 2]+meth.bs[, 3]+meth.bs[, 4]+meth.bs[, 5]+meth.bs[, 6]+meth.bs[, 7]+meth.bs[, 8]+meth.bs[, 9]+meth.bs[, 10]+meth.bs[, 11]+meth.bs[, 12]+meth.bs[, 13]+meth.bs[, 14]+meth.bs[, 15]+meth.bs[, 16]+meth.bs[, 17]+meth.bs[, 18]+meth.bs[, 19]+meth.bs[, 20]+meth.bs[, 21]+meth.bs[, 22]+meth.bs[, 23]+meth.bs[, 24]+meth.bs[, 25]+meth.bs[, 26]+meth.bs[, 27]+meth.bs[, 28]+meth.bs[, 29]+meth.bs[, 30]+meth.bs[, 31]+meth.bs[, 32]+meth.bs[, 33]+meth.bs[, 34]+meth.bs[, 35]+meth.bs[, 36]+meth.bs[, 37]+meth.bs[, 38]+meth.bs[, 39]+meth.bs[, 40]+meth.bs[, 41]+meth.bs[, 42]+meth.bs[, 43]+meth.bs[, 44]+meth.bs[, 45]+meth.bs[, 46]+meth.bs[, 47]+meth.bs[, 48]+meth.bs[, 49]+meth.bs[, 50]+meth.bs[, 51]+meth.bs[, 52]+meth.bs[, 53]+meth.bs[, 54]+meth.bs[, 55]+meth.bs[, 56]+meth.bs[, 57]+meth.bs[, 58]+meth.bs[, 59]+meth.bs[, 60]+meth.bs[, 61]+meth.bs[, 62]+meth.bs[, 63]+meth.bs[, 64]+meth.bs[, 65]+meth.bs[, 66]+meth.bs[, 67]+meth.bs[, 68]+meth.bs[, 69]+meth.bs[, 70]+meth.bs[,

```
71]+meth.bs[, 72]+meth.bs[, 73]+meth.bs[, 74]+meth.bs[, 75]+meth.bs[, 76]+meth.bs[,
77]+meth.bs[, 78]+meth.bs[, 79]+meth.bs[, 80]+meth.bs[, 81]+meth.bs[, 82]+meth.bs[,
83]+meth.bs[, 84]+meth.bs[, 85]+meth.bs[, 86]+meth.bs[, 87]+meth.bs[, 88]+meth.bs[,
89]+meth.bs[, 90]+meth.bs[, 91]+meth.bs[, 92]+meth.bs[, 93]+meth.bs[, 94]+meth.bs[,
95]+meth.bs[, 96]+meth.bs[, 97]+meth.bs[, 98]+meth.bs[, 99]+meth.bs[, 100]+meth.bs[,
101]+meth.bs[, 102]+meth.bs[, 103]+meth.bs[, 104]+meth.bs[, 105]+meth.bs[,
106]+meth.bs[, 107]+meth.bs[, 108]+meth.bs[, 109]+meth.bs[, 110]+meth.bs[,
111]+meth.bs[, 112]+meth.bs[, 113]+meth.bs[, 114]+meth.bs[, 115]+meth.bs[,
116]+meth.bs[, 117]+meth.bs[, 118]+meth.bs[, 119]+meth.bs[, 120]+meth.bs[,
121]+meth.bs[, 122]+meth.bs[, 123]+meth.bs[, 124]+meth.bs[, 125]+meth.bs[,
126]+meth.bs[, 127]+meth.bs[, 128]+meth.bs[, 129]+meth.bs[, 130]+meth.bs[,
131]+meth.bs[, 132]+meth.bs[, 133]+meth.bs[, 134]+meth.bs[, 135]+meth.bs[,
136]+meth.bs[, 137]+meth.bs[, 138]+meth.bs[, 139]+meth.bs[, 140]+meth.bs[,
141]+meth.bs[, 142]+meth.bs[, 143]+meth.bs[, 144]+meth.bs[, 145]+meth.bs[,
146]+meth.bs[, 147]+meth.bs[, 148]+meth.bs[, 149]+meth.bs[, 150]+meth.bs[,
151]+meth.bs[, 152]+meth.bs[, 153]+meth.bs[, 154]+meth.bs[, 155]+meth.bs[,
156]+meth.bs[, 157]+meth.bs[, 158]+meth.bs[, 159]+meth.bs[, 160]+meth.bs[,
161]+meth.bs[, 162]+meth.bs[, 163]+meth.bs[, 164]+meth.bs[, 165]+meth.bs[,
166]+meth.bs[, 167]+meth.bs[, 168]+meth.bs[, 169]+meth.bs[, 170]+meth.bs[,
171]+meth.bs[, 172]+meth.bs[, 173]+meth.bs[, 174]+meth.bs[, 175]+meth.bs[,
176]+meth.bs[, 177]+meth.bs[, 178]+meth.bs[, 179]+meth.bs[, 180]+meth.bs[,
181]+meth.bs[, 182]+meth.bs[, 183]+meth.bs[, 184]+meth.bs[, 185]+meth.bs[,
186]+meth.bs[, 187]+meth.bs[, 188]+meth.bs[, 189]+meth.bs[, 190]+meth.bs[,
191]+meth.bs[, 192]+meth.bs[, 193]+meth.bs[, 194]+meth.bs[, 195]+meth.bs[,
196]+meth.bs[, 197]+meth.bs[, 198]+meth.bs[, 199]+meth.bs[, 200]+meth.bs[,
201]+meth.bs[, 202]+meth.bs[, 203]+meth.bs[, 204]+meth.bs[, 205]+meth.bs[,
206]+meth.bs[, 207]+meth.bs[, 208]+meth.bs[, 209]+meth.bs[, 210]+meth.bs[,
211]+meth.bs[, 212]+meth.bs[, 213]+meth.bs[, 214]+meth.bs[, 215]+meth.bs[,
216]+meth.bs[, 217]+meth.bs[, 218])

meth_con = ordistep(modA, modC, Pin=0.05, permutations=999)

save(meth_con, file="ordistep_meth_con.Rdata")

### WITH 999 PERM, select the significant axis that been selected by ordistep

METH2=data.frame(cbind(meth.bs[, 4] ,meth.bs[, 2] ,meth.bs[, 1] ,meth.bs[, 6] ,meth.bs[, 24]
,meth.bs[, 35] ,meth.bs[, 15] ,meth.bs[, 5] ,meth.bs[, 100] ,meth.bs[, 106] ,meth.bs[, 22]
,meth.bs[, 86] ,meth.bs[, 145] ,meth.bs[, 32] ,meth.bs[, 72] ,meth.bs[, 141] ,meth.bs[, 153]
,meth.bs[, 26] ,meth.bs[, 108] ,meth.bs[, 177] ,meth.bs[, 30] ,meth.bs[, 139] ,meth.bs[, 66]
,meth.bs[, 202] ,meth.bs[, 103] ,meth.bs[, 77] ,meth.bs[, 112] ,meth.bs[, 82] ,meth.bs[, 119]))

varpart_con = varpart(m_P_con,GENET2,METH2)

varpart_con

# Partition of variance in RDA

# Call: varpart(Y = m_P_con, X = GENET2, METH2)

# Explanatory tables:
```

```
#   X1:  GENET2
# X2:  METH2
# No. of explanatory tables: 2
# Total variation (SS): 2627146
# Variance: 11996
# No. of observations: 220
# Partition table:
#   Df R.squared Adj.R.squared Testable
# [a+b] = X1          29   0.55111      0.48259     TRUE
# [b+c] = X2          29   0.57945      0.51527     TRUE
# [a+b+c] = X1+X2     58   0.74700      0.65586     TRUE
# Individual fractions
# [a] = X1|X2         29              0.14059     TRUE
# [b]                 0               0.34200    FALSE
# [c] = X2|X1         29              0.17327     TRUE
# [d] = Residuals                     0.34414    FALSE
# ---
#   Use function 'rda' to test significance of fractions of interest


plot(varpart_con,  digits = 1, Xnames = c('Genetic', 'Epigenetic'), bg = c('Blue', 'red'))
```

Supplementary file 9

Due to the large size of the file, the file A, B and D can be reached through the link below.

https://drive.google.com/drive/folders/114V0oY5ySOUtpgUb_E97zhZtr3speIMf?usp=sharing

The Supplementary file 9C is below:

TOP associated SNPs identified either by Binary or semi-quantitative or both trait. Also showing their location in the genome (exon, intron or close to a gene). The SNPs in immune related genes (are the once with gene abbreviation column mentioned in the last column).

SNP	Gene_IDS	Gene Annotation or (family or domain containing name)	Approch	Location in the gene	Synonomous	SNP in Immune-related genes
scaffold1832_479264	CGI_10018487	SUMO-activating enzyme subunit 2	Common	Intron	NA	UBA2
scaffold248_39153	CGI_10017214	tyrosinase tyr-3	Semi-quantative	Exon	YES	Tyr-3
scaffold248_39163	CGI_10017214	tyrosinase tyr-3	Semi-quantative	Exon	YES	Tyr-3
scaffold43598_225332	CGI_10011327	E3 ubiquitin- ligase TRIM33	Binary	Intron	NA	TRIM33
scaffold1533_317105	CGI_10012880	Tripartite motif-containing 2	Common	Exon	YES	TRIM
scaffold1315_95359	CGI_10021595	Tripartite motif-containing 3	Binary	Exon	NO	TRIM
scaffold377_119115	CGI_10019401	TNF receptor-associated factor 3-like	Binary	Intron	NA	TRAF3
scaffold117_77443	CGI_10016954	Serine threonine- kinase TBK1	Semi-quantative	Intron	NA	TBK1
scaffold39008_19787	CGI_10003120	arginine N-methyltransferase 5-like	Semi-quantative	Intron	NA	PRMT5
scaffold36490_15835	CGI_10001975	N-acetylmuramoyl-L-alanine amidase	Binary	Intron	NA	PGRP
scaffold870_10131	CGI_10007490	Myeloid differentiation primary response 88	Common	Intron	NA	MyD88
scaffold42674_138357	CGI_10007445	C-type mannose receptor 2-partial	Semi-quantative	Exon	NO	MR
scaffold42674_138366	CGI_10007445	C-type mannose receptor 2-partial	Semi-quantative	Exon	NO	MR
scaffold1352_475640	CGI_10018431	E3 ubiquitin- ligase MIB2	Binary	Exon	YES	MIB2
scaffold987_350721	CGI_10015851	E3 ubiquitin- ligase HERC2-like	Binary	Exon	YES	E3lig
scaffold556_406324	CGI_10016808	(Down syndrome cell adhesion molecule (DSCAM))	Binary	Intron	NA	DSCAM
scaffold556_406344	CGI_10016808	(Down syndrome cell adhesion molecule (DSCAM))	Binary	Intron	NA	DSCAM
scaffold1584_375564	CGI_10015093	endoribonuclease Dicer-like	Binary	Intron	NA	DICER
scaffold707_138248	CGI_10007724	DC-STAMP domain-containing 1-like	Semi-quantative	Exon	NO	DCST1
scaffold1506_8031	CGI_10003367	Neuroendocrine convertase 1	Binary	Exon	YES	C1q
scaffold547_555803	CGI_10018862	aminoacyl tRNA synthase complex-interacting multifunctional 1	Semi-quantative	Exon	NO	AIMP1
C32984_7712	CGI_10000975	Xaa-Pro partial	Semi-quantative	Intron	NA	
C33664_12120	CGI_10001114	hypothetical protein CGI_10001114	Binary	Exon	YES	
C34274_5145	CGI_10001263	Inositol 1,4,5-trisphosphate receptor	Semi-quantative	Exon	YES	

C37024_18634	CGI_10002229	seipin [Orussus abietinus]	Common	Exon	YES	
C37024_18682	CGI_10002229	seipin [Orussus abietinus]	Common	Exon	YES	
scaffold857_6284	CGI_10002412	cyclin-dependent kinase 1-like	Semi-quantative	Intron	NA	
scaffold749_47184	CGI_10002826	mitochondrial import inner membrane translocase subunit Tim22	Common	close to gene	NA	
scaffold341_24460	CGI_10002986	F-box only 42	Common	Intron	NA	
scaffold38922_19952	CGI_10003058	brefeldin A-inhibited guanine nucleotide-exchange 1-like isoform X4	Semi-quantative	Exon	NO	
scaffold39064_3039	CGI_10003156	Acyl- synthetase short-chain family member mitochondrial	Binary	Exon	YES	
scaffold39366_53158	CGI_10003339	organic cation transporter - like	Common	Intron	NA	
scaffold39390_25090	CGI_10003347	SCO-spondin-like	Common	Exon	YES	
scaffold39390_25137	CGI_10003347	SCO-spondin-like	Common	Exon	NO	
scaffold39390_25054	CGI_10003347	SCO-spondin-like	Binary	Exon	YES	
scaffold39390_25184	CGI_10003347	SCO-spondin-like	Binary	Intron	NA	
scaffold1437_30576	CGI_10003665	Anosmin-1	Semi-quantative	Intron	NA	
scaffold409_75099	CGI_10004318	PREDICTED: uncharacterized protein LOC105319668 [*Crassostrea gigas*]	Common	Intron	NA	
scaffold201_38432	CGI_10004417	dedicator of cytokinesis 7-like isoform X4	Semi-quantative	Intron	NA	
scaffold41034_72841	CGI_10004652	Dynein heavy cytoplasmic	Semi-quantative	Exon	NO	
scaffold41296_65578	CGI_10004841	disintegrin and metallo ase domain-containing 12-like isoform X2	Binary	Exon	NO	
scaffold41296_65579	CGI_10004841	disintegrin and metallo ase domain-containing 12-like isoform X2	Binary	Exon	YES	
scaffold41816_122449	CGI_10005652	tectonic-1-like isoform X3	Binary	Intron	NA	
scaffold41824_11146	CGI_10005707	hypothetical protein CGI_10005707	Semi-quantative	Exon	NO	
scaffold41994_119855	CGI_10005879	0	Semi-quantative	Exon	YES	
scaffold41994_119856	CGI_10005879	0	Semi-quantative	Exon	YES	
scaffold41994_119858	CGI_10005879	0	Semi-quantative	Exon	YES	
scaffold1180_96270	CGI_10005985	RWD domain-containing 2B-like [*Crassostrea gigas*]	Semi-quantative	Exon	YES	
scaffold1813_2902	CGI_10005987	SWI SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A 1	Binary	Exon	NO	
scaffold42184_48477	CGI_10006328	Iporin [*Crassostrea gigas*]	Binary	Exon	YES	
scaffold42184_48478	CGI_10006328	Iporin [*Crassostrea gigas*]	Binary	Exon	NO	
scaffold1774_67280	CGI_10006381	Scm-like with four MBT domains 1	Binary	Exon	NO	
scaffold1774_67281	CGI_10006381	Scm-like with four MBT domains 1	Binary	Exon	YES	
scaffold141_26231	CGI_10006577	0	Semi-quantative	Exon	YES	
scaffold215_69859	CGI_10006639	fibrocystin-L-like	Binary	Exon	NO	
scaffold1871_120731	CGI_10007357	E3 ubiquitin- ligase UBR3-like	Semi-quantative	Exon	NO	
scaffold42684_151364	CGI_10007485	hypothetical protein CGI_10007485	Binary	Exon	YES	
scaffold42930_101979	CGI_10008126	hypothetical protein CGI_10008126	Binary	Intron	NA	
scaffold1841_87559	CGI_10008271	ankyrin repeat and KH domain-containing 1-like isoform X2	Binary	Intron	NA	

scaffold665_177785	CGI_10008301	fatty acid-binding heart [Myotis brandtii]	Binary	Intron	NA	
scaffold138_192913	CGI_10008435	cell wall integrity and stress response component 4-like	Common	Exon	NO	
scaffold161_81939	CGI_10008586	mannan-binding lectin serine protease 2-like [*Crassostrea gigas*]	Semi-quantative	Exon	NO	
scaffold500_172897	CGI_10008719	heat shock 70 kDa 12A-like	Common	Exon	NO	
scaffold1247_153770	CGI_10008948	Dimethyladenosine transferase mitochondrial	Semi-quantative	Exon	YES	
scaffold1116_151400	CGI_10009132	Ras and EF-hand domain-containing partial	Common	Exon	NO	
scaffold993_221506	CGI_10009380	Serine threonine-phosphatase 6 catalytic subunit	Common	Intron	NA	
scaffold1503_84500	CGI_10009530	Ankyrin repeat domain-containing 55	Semi-quantative	Exon	NO	
scaffold43240_207293	CGI_10009595	neurobeachin 1	Semi-quantative	Exon	NO	
scaffold43272_87487	CGI_10009666	PREDICTED: uncharacterized protein LOC105333728 [*Crassostrea gigas*]	Semi-quantative	Exon	NO	
scaffold43302_125743	CGI_10009764	BTB POZ domain-containing 17-like isoform X4 [Octopus bimaculoides]	Semi-quantative	Exon	NO	
scaffold323_119555	CGI_10009779	hypothetical protein CGI_10009779	Semi-quantative	close to gene	NA	
scaffold752_73843	CGI_10010040	ATP-dependent zinc metalloprotease YME1L1-like [Biomphalaria glabrata]	Semi-quantative	Exon	YES	
scaffold1751_152761	CGI_10010088	nicotinamide nicotinic acid mononucleotide adenylyltransferase 1-like isoform X1	Common	Exon	YES	
scaffold43452_197656	CGI_10010486	beta-1-syntrophin-like	Semi-quantative	Intron	NA	
scaffold43500_220913	CGI_10010622	phospholipase D1-like isoform X1 [Parasteatoda tepidariorum]	Semi-quantative	Exon	YES	
scaffold43520_262571	CGI_10010725	CREB-regulated transcription coactivator 1-like isoform X4	Semi-quantative	Exon	NO	
scaffold954_26349	CGI_10010737	hypothetical protein CGI_10010737	Common	Exon	YES	
scaffold917_270662	CGI_10011091	Fibroblast growth factor receptor 2	Semi-quantative	Exon	YES	
scaffold1825_5644	CGI_10011136	hypothetical protein CGI_10011136	Binary	Exon	NO	
scaffold1382_191376	CGI_10011175	Collagen alpha-5(VI) chain	Semi-quantative	Intron	NA	
scaffold508_227821	CGI_10011205	sortilin-related receptor-like isoform X1	Binary	Exon	NO	
scaffold1309_55873	CGI_10011558	proteasome subunit beta type-7-like	Binary	Intron	NA	
scaffold1235_76550	CGI_10011831	X-ray radiation resistance-associated 1	Common	Intron	NA	
scaffold421_199035	CGI_10012156	myosin-2 essential light chain-like	Common	Intron	NA	
scaffold1125_265921	CGI_10012342	monocarboxylate transporter 12-like [*Crassostrea gigas*]	Binary	Exon	YES	
scaffold1164_299687	CGI_10013010	ABC transporter G family member 9	Semi-quantative	close to gene	NA	
scaffold1144_159627	CGI_10013961	neuronal acetylcholine receptor subunit alpha-10-like	Binary	Intron	NA	
scaffold631_17116	CGI_10014166	zinc finger 768-like isoform X3 [Lingula anatina]	Common	Exon	YES	
scaffold631_17106	CGI_10014166	zinc finger 768-like isoform X3 [Lingula anatina]	Binary	Exon	YES	
scaffold1053_66069	CGI_10014293	centrosomal of 152 kDa	Semi-quantative	Exon	YES	

scaffold873_45939	CGI_10014415	limbic system-associated membrane -like [*Crassostrea gigas*]	Binary	Intron	NA	
scaffold716_178771	CGI_10014446	kinase C-binding 1-like isoform X1 [*Crassostrea gigas*]	Binary	Exon	NO	
scaffold43956_229870	CGI_10014626	Cubilin	Binary	Exon	NO	
scaffold43986_74669	CGI_10015018	solute carrier family 26 member 6-like [Lingula anatina]	Semi-quantative	Exon	YES	
scaffold165_175329	CGI_10015321	clarin-3-like [Biomphalaria glabrata]	Binary	close to gene	NA	
scaffold1901_58205	CGI_10015545	calcitonin receptor-like isoform X1 [*Crassostrea gigas*]	Binary	Intron	NA	
scaffold479_40550	CGI_10015678	synaptophysin 1 isoform X2	Semi-quantative	Intron	NA	
scaffold324_45440	CGI_10016183	tyrosine phosphatase domain-containing 1	Common	Exon	YES	
scaffold61_65263	CGI_10016339	myosin heavy striated muscle-like	Semi-quantative	close to gene	NA	
scaffold794_140753	CGI_10017142	Transcription elongation factor SPT6	Semi-quantative	Exon	NO	
scaffold789_230713	CGI_10017197	homer homolog 2-like isoform X3	Binary	Exon	YES	
scaffold751_393837	CGI_10017401	serine arginine-rich splicing factor 1B	Semi-quantative	close to gene	NA	
scaffold853_362477	CGI_10017506	notchless homolog 1-like	Binary	Intron	NA	
scaffold853_362479	CGI_10017506	notchless homolog 1-like	Binary	Intron	NA	
scaffold1670_92453	CGI_10017599	ankyrin repeat domain-containing 6-like [Lingula anatina]	Binary	Exon	YES	
scaffold44098_149726	CGI_10017704	probable aminopeptidase NPEPL1	Binary	Exon	YES	
scaffold557_239727	CGI_10017968	ubiquitin carboxyl-terminal hydrolase 19-like	Binary	Intron	NA	
scaffold630_124042	CGI_10018327	Organic cation transporter [*Crassostrea gigas*]	Semi-quantative	Intron	NA	
scaffold163_174660	CGI_10018615	Hypoxia-inducible factor 1-alpha inhibitor	Common	Intron	NA	
scaffold163_174670	CGI_10018615	Hypoxia-inducible factor 1-alpha inhibitor	Binary	Intron	NA	
scaffold547_290607	CGI_10018853	Nucleoside diphosphate kinase [*Crassostrea gigas*]	Semi-quantative	Exon	NO	
scaffold1249_71396	CGI_10019046	laminin subunit gamma-1-like	Binary	Exon	YES	
scaffold1249_348949	CGI_10019064	Cleft lip and palate transmembrane 1	Common	Exon	NO	
scaffold126_377388	CGI_10019196	short-chain collagen C4-like	Common	close to gene	NA	
scaffold1794_282542	CGI_10019584	26S protease regulatory subunit 6B	Binary	Intron	NA	
scaffold1794_353208	CGI_10019588	Eukaryotic translation initiation factor 2-alpha kinase 4	Binary	Exon	NO	
scaffold1512_509767	CGI_10019874	KAT8 regulatory NSL complex subunit 3	Common	Exon	YES	
scaffold140_178153	CGI_10020125	tctex1 domain-containing 1-B-like	Semi-quantative	Exon	YES	
scaffold140_178154	CGI_10020125	tctex1 domain-containing 1-B-like	Semi-quantative	Exon	YES	
scaffold140_199703	CGI_10020127	RNA exonuclease 1 homolog isoform X2	Semi-quantative	Exon	NO	
scaffold535_373565	CGI_10020291	Hemicentin-1	Common	close to gene	NA	
scaffold160_609047	CGI_10020612	hypothetical protein CGI_10020612	Binary	Exon	NO	
scaffold288_304055	CGI_10020664	contactin-like	Semi-quantative	close to gene	NA	
scaffold203_656390	CGI_10021081	fibrillin-1-like [*Crassostrea gigas*]	Binary	Exon	NO	

scaffold203_656391	CGI_10021081	fibrillin-1-like [*Crassostrea gigas*]	Binary	Exon	NO	
scaffold203_656395	CGI_10021081	fibrillin-1-like [*Crassostrea gigas*]	Binary	Exon	YES	
scaffold203_656403	CGI_10021081	fibrillin-1-like [*Crassostrea gigas*]	Binary	Exon	NO	
scaffold610_325703	CGI_10021110	sodium hydrogen exchanger 9-like isoform X3 [*Crassostrea gigas*]	Semi-quantative	Intron	NA	
scaffold157_403653	CGI_10021254	ecdysoneless homolog isoform X1 [Sus scrofa]	Semi-quantative	Exon	NO	
scaffold157_403656	CGI_10021254	ecdysoneless homolog isoform X1 [Sus scrofa]	Semi-quantative	Exon	NO	
scaffold1750_124830	CGI_10021469	cilia- and flagella-associated 54-like	Binary	Exon	NO	
scaffold721_513573	CGI_10022007	4-hydroxyphenylpyruvate dioxygenase-like	Semi-quantative	Intron	NA	
scaffold721_513577	CGI_10022007	4-hydroxyphenylpyruvate dioxygenase-like	Semi-quantative	Intron	NA	
scaffold721_514268	CGI_10022007	4-hydroxyphenylpyruvate dioxygenase-like	Semi-quantative	Intron	NA	
scaffold721_514463	CGI_10022007	4-hydroxyphenylpyruvate dioxygenase-like	Semi-quantative	Exon	NO	
scaffold394_50488	CGI_10022062	serine threonine kinase-like domain-containing STKLD1	Semi-quantative	Intron	NA	
scaffold593_316932	CGI_10022523	0	Binary	Exon	NO	
scaffold364_478394	CGI_10022698	AAEL010828- partial [Aedes aegypti]	Common	Exon	YES	
scaffold365_158645	CGI_10022721	Polycystic kidney disease 1-like 2	Binary	Exon	YES	
scaffold365_158654	CGI_10022721	Polycystic kidney disease 1-like 2	Binary	Exon	YES	
scaffold896_535018	CGI_10022856	eukaryotic translation initiation factor 4 gamma 3-like isoform X1 [*Crassostrea gigas*]	Binary	Intron	NA	
scaffold348_435696	CGI_10022984	myotubularin-related 9	Binary	Intron	NA	
scaffold348_435697	CGI_10022984	myotubularin-related 9	Binary	Intron	NA	
scaffold432_280475	CGI_10023093	PREDICTED: uncharacterized protein LOC105330599 [*Crassostrea gigas*]	Binary	Exon	YES	
scaffold1132_613978	CGI_10023325	ras-associated and pleckstrin homology domains-containing 1-like isoform X1 [*Crassostrea gigas*]	Semi-quantative	Intron	NA	
scaffold383_664299	CGI_10023551	Neuron navigator 2	Semi-quantative	Exon	NO	
scaffold54_367971	CGI_10023676	Dual serine threonine and tyrosine kinase	Common	Intron	NA	
scaffold48_585158	CGI_10023907	inactive tyrosine- kinase 7-like	Semi-quantative	Exon	YES	
scaffold48_677936	CGI_10023915	cAMP-regulated D2 -like	Binary	Exon	YES	
scaffold1219_910726	CGI_10024309	heat shock 75 mitochondrial-like	Common	Exon	NO	
scaffold469_659950	CGI_10024691	Patatin-like phospholipase domain-containing 7	Common	Intron	NA	
scaffold107_570179	CGI_10025062	All-trans-retinol 13,14-reductase	Common	Exon	YES	
scaffold1267_535371	CGI_10025156	JNK-interacting 1	Binary	Exon	YES	
scaffold82_218546	CGI_10025248	PREDICTED: uncharacterized protein LOC105333164	Common	Exon	NO	
scaffold82_290465	CGI_10025253	cholecystokinin receptor-like [*Crassostrea gigas*]	Binary	Exon	YES	
scaffold149_822536	CGI_10025323	ATP-dependent Clp protease proteolytic mitochondrial	Common	Exon	NO	
scaffold226_793063	CGI_10025375	serine threonine- kinase Chk1-like	Binary	Exon	YES	
scaffold370_117367	CGI_10025524	Remodeling and spacing factor 1	Semi-quantative	Exon	YES	

scaffold168_446915	CGI_10025594	advillin-like [Lingula anatina]	Binary	Intron	NA	
scaffold168_849986	CGI_10025614	0	Semi-quantative	Exon	YES	
scaffold1583_870363	CGI_10025848	alpha-N-acetylgalactosamine-specific lectin-like [*Crassostrea gigas*]	Common	close to gene	NA	
scaffold334_960606	CGI_10025969	Uncharacterized protein C3orf59-like protein	Semi-quantative	Exon	YES	
scaffold425_1022500	CGI_10026011	chromodomain-helicase-DNA-binding 4-like isoform X3	Common	Intron	NA	
scaffold1154_759038	CGI_10026186	calpain-B-like isoform X1 [*Crassostrea gigas*]	Semi-quantative	Intron	NA	
scaffold678_201258	CGI_10026340	Telomerase-binding EST1A	Common	Intron	NA	
scaffold678_201171	CGI_10026340	Telomerase-binding EST1A	Semi-quantative	Exon	NO	
scaffold678_201291	CGI_10026340	Telomerase-binding EST1A	Semi-quantative	Intron	NA	
scaffold678_201295	CGI_10026340	Telomerase-binding EST1A	Semi-quantative	Intron	NA	
scaffold678_201352	CGI_10026340	Telomerase-binding EST1A	Semi-quantative	Intron	NA	
scaffold678_377478	CGI_10026355	hypothetical protein CGI_10026355	Semi-quantative	Exon	NO	
scaffold100_688108	CGI_10026448	Dynein heavy chain axonemal	Semi-quantative	Intron	NA	
scaffold100_1110867	CGI_10026464	Polycystic kidney disease 1-like 2	Binary	Exon	YES	
scaffold156_265848	CGI_10026488	Membrane metallo-endopeptidase-like 1	Semi-quantative	Exon	YES	
scaffold142_1138894	CGI_10026790	crossover junction endonuclease EME1-like [Priapulus caudatus]	Semi-quantative	Exon	NO	
scaffold393_461178	CGI_10026893	adenylate kinase isoenzyme 1 isoform X2 [Monodelphis domestica]	Binary	Intron	NA	
scaffold471_44041	CGI_10026998	kinase D-interacting substrate of 220 kDa isoform X4 [Vollenhovia emeryi]	Binary	Exon	NO	
scaffold471_44042	CGI_10026998	kinase D-interacting substrate of 220 kDa isoform X4 [Vollenhovia emeryi]	Binary	Exon	NO	
scaffold3_200765	CGI_10027281	polycystic kidney disease 1-like 1-like	Binary	Intron	NA	
scaffold1179_179587	CGI_10027347	coatomer subunit alpha isoform X2	Common	Exon	NO	
scaffold198_870932	CGI_10027830	hypothetical protein CGI_10027830	Binary	Intron	NA	
scaffold102_1062943	CGI_10028467	ER degradation-enhancing alpha-mannosidase-like 3	Semi-quantative	Intron	NA	
scaffold102_1125734	CGI_10028472	KAT8 regulatory NSL complex subunit 1-like	Common	Intron	NA	
scaffold1009_65957	CGI_10028712	caveolin [Saccoglossus kowalevskii]	Common	Intron	NA	
scaffold22_19417	CGI_10028823	PREDICTED: uncharacterized protein LOC105336120	Semi-quantative	Intron	NA	
scaffold22_385783	CGI_10028849	piezo-type mechanosensitive ion channel component 2-like	Binary	Exon	YES	
scaffold22_385784	CGI_10028849	piezo-type mechanosensitive ion channel component 2-like	Binary	Exon	NO	

Supplementary file 10

Due to the large size of the file, the file can be reached through the link below.

https://drive.google.com/drive/folders/114V0oY5ySOUtpgUb_E97zhZtr3speIMf?usp=sharing

Supplementary file 11

Due to the large size of the file, the file A, B and D can be reached through the link below.

https://drive.google.com/drive/folders/114V0oY5ySOUtpgUb_E97zhZtr3speIMf?usp=sharing

The Supplementary file 11C is below:

CpGs	Approch	Methylation in Resistant	Gene_IDS	Gene Annotation	Location in the gene	
scaffold364_467953	Binary	hyper	CGI_10022697	Transcription termination factor 2	Exon	TTF2
scaffold364_467674	Binary	hyper	CGI_10022697	Transcription termination factor 2	Exon	TTF2
scaffold364_471615	common	hyper	CGI_10022697	Transcription termination factor 2	Exon	TTF2
scaffold364_471528	common	hyper	CGI_10022697	Transcription termination factor 2	Exon	TTF2
scaffold364_471627	common	hyper	CGI_10022697	Transcription termination factor 2	Exon	TTF2
scaffold364_471609	common	hyper	CGI_10022697	Transcription termination factor 2	Exon	TTF2
scaffold364_473230	common	hyper	CGI_10022697	Transcription termination factor 2	Exon	TTF2
scaffold364_473238	common	hyper	CGI_10022697	Transcription termination factor 2	Exon	TTF2
scaffold364_473225	common	hyper	CGI_10022697	Transcription termination factor 2	Exon	TTF2
scaffold364_473193	Binary	hyper	CGI_10022697	Transcription termination factor 2	Exon	TTF2
scaffold39074_60729	common	hypo	CGI_10003134	toll-like receptor 4 (Proteins with epidermal growth factor (EGF) and TIR domains (EGF-TIR))	Exon	TIRprot
scaffold117_79908	Binary	hypo	CGI_10016954	Serine threonine- kinase TBK1	Exon	TBK1
scaffold485_51416	common	hypo	CGI_10024137	E3 ubiquitin- ligase SMURF2	Exon	SMURF2
scaffold485_51405	common	hypo	CGI_10024137	E3 ubiquitin- ligase SMURF2	Exon	SMURF2
scaffold305_13951	common	hypo	CGI_10019110	E3 ubiquitin- ligase RNF220-like isoform X1	Exon	RNF220
scaffold305_13943	common	hypo	CGI_10019110	E3 ubiquitin- ligase RNF220-like isoform X1	Exon	RNF220
scaffold305_13966	Semi-quantitative	hypo	CGI_10019110	E3 ubiquitin- ligase RNF220-like isoform X1	Exon	RNF220
scaffold305_13960	Semi-quantitative	hypo	CGI_10019110	E3 ubiquitin- ligase RNF220-like isoform X1	Exon	RNF220
scaffold237_123498	Binary	hypo	CGI_10021567	NF-kappa B	Exon	NF-kB

scaffold43446_184674	Binary	hyper	CGI_10010404	Macrophage colony-stimulating factor 1 receptor 2	Exon	MCSF
scaffold4_156840	common	hyper	CGI_10021170	Interferon regulatory factor 2	Exon	IRF
scaffold522_402047	Binary	hypo	CGI_10020330	Inositol polyphosphate multikinase	Exon	IPMK
scaffold376_54858	Binary	hypo	CGI_10019491	importin subunit beta-1-like	Exon	IMPβ1
scaffold1185_127091	Semi-quantitative	hypo	CGI_10006263	homeodomain-interacting kinase 2-like isoform X1	Exon	HIPK2
scaffold43726_271694	common	hypo	CGI_10012028	F-box LRR-repeat 7-like isoform X2	Exon	FBXL7
scaffold43726_271521	common	hypo	CGI_10012028	F-box LRR-repeat 7-like isoform X2	Exon	FBXL7
scaffold43726_271721	Binary	hypo	CGI_10012028	F-box LRR-repeat 7-like isoform X2	Exon	FBXL7
scaffold43726_275164	Semi-quantitative	hypo	CGI_10012028	F-box LRR-repeat 7-like isoform X2	Exon	FBXL7
scaffold425_801672	common	hypo	CGI_10025995	E3 ubiquitin- ligase MYLIP	Exon	E3lig
scaffold522_293879	common	hypo	CGI_10020325	tyrosine- phosphatase Lar-like isoform X6	Intron	DSCAM
scaffold211_1161890	common	hypo	CGI_10026985	deleted in malignant brain tumors 1 - like	Exon	DMTB1
scaffold211_1161693	Binary	hypo	CGI_10026985	deleted in malignant brain tumors 1 - like	Exon	DMTB1
scaffold211_1161696	Binary	hypo	CGI_10026985	deleted in malignant brain tumors 1 - like	Exon	DMTB1
scaffold211_1161137	Semi-quantitative	hypo	CGI_10026985	deleted in malignant brain tumors 1 - like	Exon	DMTB1
scaffold42918_99262	common	hypo	CGI_10008117	diacylglycerol kinase zeta-like isoform X10 [*Crassostrea gigas*]	Exon	DGKz
scaffold42918_99184	common	hypo	CGI_10008117	diacylglycerol kinase zeta-like isoform X10 [*Crassostrea gigas*]	Exon	DGKz
scaffold42918_100003	Binary	hypo	CGI_10008117	diacylglycerol kinase zeta-like isoform X10 [*Crassostrea gigas*]	Exon	DGKz
scaffold1815_183378	Binary	hypo	CGI_10010268	E3 ubiquitin- ligase CBL-B-like isoform X1	Exon	CBLB
scaffold40894_73489	common	hypo	CGI_10004539	tilB homolog (cysteine protease ATG4C-like)	Intron	ATG4
scaffold40894_73510	Semi-quantitative	hypo	CGI_10004539	tilB homolog (cysteine protease ATG4C-like)	Exon	ATG4
scaffold1746_93149	Binary	hypo	CGI_10007280	A-kinase anchor 13	Exon	AKAP13
scaffold1746_93142	common	hypo	CGI_10007280	A-kinase anchor 13	Exon	AKAP13
scaffold705_279705	Binary	hypo	NA	#N/A	NA	
scaffold617_159991	common	hypo	NA	#N/A	NA	
scaffold363_348272	Semi-quantitative	hypo	NA	#N/A	NA	
scaffold165_246880	Semi-quantitative	hypo	NA	#N/A	NA	
C35262_15314	common	hypo	CGI_10001446	hypothetical protein CGI_10001446	Exon	
scaffold36278_14345	Binary	hypo	CGI_10001847	hypothetical protein CGI_10001847	Exon	
scaffold37576_3762	common	hypo	CGI_10002419	alpha-tubulin N-acetyltransferase-like isoform X1 [Polistes dominula]	Exon	
scaffold38688_19593	common	hypo	CGI_10002921	tyrosine- kinase SRK2-like isoform X2 [*Crassostrea gigas*]	Exon	
scaffold1842_51345	common	hypo	CGI_10003201	phosphatase 1 regulatory subunit 12A	Exon	

scaffold39716_13084	Semi-quantitative	hypo	CGI_10003560	Arrestin domain-containing 2	Exon	
scaffold1735_60628	common	hypo	CGI_10003744	probable domain-containing histone demethylation 2C isoform X2	Exon	
scaffold1735_60573	common	hypo	CGI_10003744	probable domain-containing histone demethylation 2C isoform X2	Exon	
scaffold1735_60640	common	hypo	CGI_10003744	probable domain-containing histone demethylation 2C isoform X2	Exon	
scaffold1735_60600	common	hypo	CGI_10003744	probable domain-containing histone demethylation 2C isoform X2	Exon	
scaffold1735_60535	common	hypo	CGI_10003744	probable domain-containing histone demethylation 2C isoform X2	Exon	
scaffold1735_60635	Binary	hypo	CGI_10003744	probable domain-containing histone demethylation 2C isoform X2	Exon	
scaffold1735_60543	Binary	hypo	CGI_10003744	probable domain-containing histone demethylation 2C isoform X2	Exon	
scaffold1735_64183	Binary	hypo	CGI_10003744	probable domain-containing histone demethylation 2C isoform X2	Exon	
scaffold40156_67725	common	hypo	CGI_10003886	septin-7-like isoform X1	Exon	
scaffold40156_67738	common	hypo	CGI_10003886	septin-7-like isoform X1	Exon	
scaffold40254_36476	common	hypo	CGI_10004006	dual specificity tyrosine-phosphorylation-regulated kinase 1A isoform X1 [Rattus norvegicus]	Exon	
scaffold467_63261	Binary	hypo	CGI_10004063	transport sec31-like	Exon	
scaffold40456_74583	Binary	hypo	CGI_10004091	polypyrimidine tract-binding 2-like isoform X1 [Biomphalaria glabrata]	Exon	
scaffold40412_59324	common	hypo	CGI_10004113	3-hydroxybutyrate dehydrogenase [Thalassospira xiamenensis]	Exon	
scaffold40412_59299	Semi-quantitative	hypo	CGI_10004113	3-hydroxybutyrate dehydrogenase [Thalassospira xiamenensis]	Exon	
scaffold1654_40130	common	hypo	CGI_10004124	zinc finger 346-like [Octopus bimaculoides]	Exon	
scaffold1747_72751	Binary	hypo	CGI_10004402	BTB POZ domain-containing 19	Exon	
scaffold201_38576	Semi-quantitative	hypo	CGI_10004417	dedicator of cytokinesis 7-like isoform X4	Exon	
scaffold1562_112860	common	hypo	CGI_10005776	leucine-rich repeat-containing 16A-like isoform X3	Exon	
scaffold41890_53247	Semi-quantitative	hypo	CGI_10005796	kinesin KIF26B isoform X1 [*Crassostrea gigas*]	Intron	
scaffold41890_61523	Semi-quantitative	hypo	CGI_10005796	kinesin KIF26B isoform X1 [*Crassostrea gigas*]	Exon	
scaffold42060_34953	Binary	hypo	CGI_10005937	neuralized-like isoform X2	Exon	
scaffold42060_34961	Binary	hypo	CGI_10005937	neuralized-like isoform X2	Exon	
scaffold944_50892	Binary	hypo	CGI_10005966	hypothetical protein CGI_10005966	Exon	
scaffold42096_50602	Binary	hypo	CGI_10006162	zinc finger basonuclin-2-like isoform X2 [*Crassostrea gigas*]	Exon	
scaffold929_101400	Semi-quantitative	hypo	CGI_10006344	histone-lysine N-methyltransferase ASH1L-like isoform X10 [Lingula anatina]	Exon	
scaffold42198_47868	Binary	hypo	CGI_10006391	rho GTPase-activating 12-like isoform X1 [*Crassostrea gigas*]	Exon	
scaffold42198_47639	Binary	hypo	CGI_10006391	rho GTPase-activating 12-like isoform X1 [*Crassostrea gigas*]	Exon	
scaffold42198_47703	Binary	hypo	CGI_10006391	rho GTPase-activating 12-like isoform X1 [*Crassostrea gigas*]	Exon	
scaffold1121_73907	common	hypo	CGI_10006499	FERM domain-containing 4A-like isoform X1	Exon	

scaffold1121_73997	Semi-quantitative	hypo	CGI_10006499	FERM domain-containing 4A-like isoform X1	Exon	
scaffold42422_80488	common	hypo	CGI_10006851	serine-rich adhesin for platelets-like	Exon	
scaffold42466_140022	common	hypo	CGI_10006943	hemicentin-1 isoform X3 [Macaca fascicularis]	Intron	
scaffold42466_156331	Binary	hypo	CGI_10006944	Lysocardiolipin acyltransferase 1	Exon	
scaffold42558_55424	common	hypo	CGI_10007217	tensin-1-like isoform X1 [Bombus terrestris]	Exon	
scaffold42558_55404	Binary	hypo	CGI_10007217	tensin-1-like isoform X1 [Bombus terrestris]	Exon	
scaffold493_108458	common	hypo	CGI_10007248	metastasis suppressor 1-like isoform X6	Exon	
scaffold493_108441	Binary	hypo	CGI_10007248	metastasis suppressor 1-like isoform X6	Exon	
scaffold1373_68600	Semi-quantitative	hypo	CGI_10007907	nuclear hormone receptor HR96-like isoform X3 [Lingula anatina]	Exon	
scaffold42850_53693	Binary	hypo	CGI_10007921	set1 Ash2 histone methyltransferase complex subunit ASH2-like isoform X1 [Biomphalaria glabrata]	Intron	
scaffold42850_50287	common	hypo	CGI_10007922	origin recognition complex subunit 1 isoform X1 [Monodelphis domestica]	Exon	
scaffold42850_50317	common	hypo	CGI_10007922	origin recognition complex subunit 1 isoform X1 [Monodelphis domestica]	Exon	
scaffold42850_50276	common	hypo	CGI_10007922	origin recognition complex subunit 1 isoform X1 [Monodelphis domestica]	Exon	
scaffold42850_50295	common	hypo	CGI_10007922	origin recognition complex subunit 1 isoform X1 [Monodelphis domestica]	Exon	
scaffold42850_50327	common	hypo	CGI_10007922	origin recognition complex subunit 1 isoform X1 [Monodelphis domestica]	Exon	
scaffold42850_50341	common	hypo	CGI_10007922	origin recognition complex subunit 1 isoform X1 [Monodelphis domestica]	Exon	
scaffold42850_52743	common	hypo	CGI_10007922	origin recognition complex subunit 1 isoform X1 [Monodelphis domestica]	Exon	
scaffold42850_53821	common	hypo	CGI_10007922	origin recognition complex subunit 1 isoform X1 [Monodelphis domestica]	Exon	
scaffold42850_53813	common	hypo	CGI_10007922	origin recognition complex subunit 1 isoform X1 [Monodelphis domestica]	Exon	
scaffold42850_53238	Semi-quantitative	hypo	CGI_10007922	origin recognition complex subunit 1 isoform X1 [Monodelphis domestica]	Exon	
scaffold42850_151076	common	hypo	CGI_10007928	MKL myocardin 2	Exon	
scaffold309_32176	common	hypo	CGI_10007935	tyrosine- kinase CSK-like isoform X1 [*Crassostrea gigas*]	Exon	
scaffold309_32178	common	hypo	CGI_10007935	tyrosine- kinase CSK-like isoform X1 [*Crassostrea gigas*]	Exon	
scaffold42904_91418	Binary	hypo	CGI_10008103	Histone-lysine N-methyltransferase MLL4	Exon	
scaffold1578_61158	Binary	hyper	CGI_10008160	ATP synthase subunit mitochondrial	Exon	
scaffold43028_146629	common	hypo	CGI_10008509	breast cancer anti-estrogen resistance 1-like isoform X3	Exon	
scaffold1786_146248	common	hypo	CGI_10008580	Regulator of G- signaling 3 [*Crassostrea gigas*]	Exon	
scaffold1786_146260	Binary	hypo	CGI_10008580	Regulator of G- signaling 3 [*Crassostrea gigas*]	Exon	
scaffold1855_117683	Binary	hypo	CGI_10008869	FAM102A-like isoform X2	Exon	
scaffold1855_119214	common	hypo	CGI_10008869	FAM102A-like isoform X2	Exon	
scaffold1855_119197	common	hypo	CGI_10008869	FAM102A-like isoform X2	Exon	
scaffold635_218340	Binary	hypo	CGI_10008929	tyrosine- kinase yes	Exon	
scaffold208_160522	Binary	hypo	CGI_10010030	LIX1 [Polistes canadensis]	Exon	

scaffold1704_115075	Binary	hypo	CGI_10010169	galactose-3-O-sulfotransferase 2-like isoform X2 [Aplysia californica]	Exon	
scaffold1704_114924	Binary	hypo	CGI_10010169	galactose-3-O-sulfotransferase 2-like isoform X2 [Aplysia californica]	Exon	
scaffold43526_119440	common	hypo	CGI_10010787	frizzled-5-like [*Crassostrea gigas*]	Exon	
scaffold43574_197354	Binary	hypo	CGI_10010880	la-related 1B-like isoform X1 [Aplysia californica]	Exon	
scaffold43574_197341	Semi-quantitative	hypo	CGI_10010880	la-related 1B-like isoform X1 [Aplysia californica]	Exon	
scaffold1895_144890	common	hypo	CGI_10010980	PREDICTED: uncharacterized protein LOC105340519	Exon	
scaffold1895_144832	common	hypo	CGI_10010980	PREDICTED: uncharacterized protein LOC105340519	Exon	
scaffold1895_144903	common	hypo	CGI_10010980	PREDICTED: uncharacterized protein LOC105340519	Exon	
scaffold400_254566	Semi-quantitative	hypo	CGI_10011439	PREDICTED: uncharacterized protein LOC105330553 [*Crassostrea gigas*]	Exon	
scaffold1883_152896	common	hypo	CGI_10011547	stearoyl- desaturase 5-like isoform X1 [Octopus bimaculoides]	Exon	
scaffold43692_222785	common	hypo	CGI_10011819	MAP kinase-interacting serine threonine- kinase 1	Exon	
scaffold43692_222790	common	hypo	CGI_10011819	MAP kinase-interacting serine threonine- kinase 1	Exon	
scaffold122_94552	Semi-quantitative	hypo	CGI_10011867	ataxin-1 isoform X1 [Monodelphis domestica]	Exon	
scaffold1884_209825	common	hypo	CGI_10012017	cGMP-inhibited 3 ,5 -cyclic phosphodiesterase A-like isoform X2	Exon	
scaffold1884_209839	Semi-quantitative	hypo	CGI_10012017	cGMP-inhibited 3 ,5 -cyclic phosphodiesterase A-like isoform X2	Exon	
scaffold1822_216848	Semi-quantitative	hypo	CGI_10012063	26S proteasome non-ATPase regulatory subunit 1-like	Exon	
scaffold43786_74601	Binary	hypo	CGI_10012348	Hemicentin-1	Exon	
scaffold43786_100502	Binary	hypo	CGI_10012348	Hemicentin-1	Exon	
scaffold825_239310	common	hypo	CGI_10012713	lateral signaling target 2 homolog	Exon	
scaffold825_239329	Binary	hypo	CGI_10012713	lateral signaling target 2 homolog	Exon	
scaffold825_239567	Semi-quantitative	hypo	CGI_10012713	lateral signaling target 2 homolog	Exon	
scaffold617_161235	Semi-quantitative	hypo	CGI_10012729	LIM domain and actin-binding 1	Exon	
scaffold617_189928	common	hypo	CGI_10012731	enolase-phosphatase E1-like isoform X4	Exon	
scaffold617_189910	common	hypo	CGI_10012731	enolase-phosphatase E1-like isoform X4	Exon	
scaffold790_53116	common	hypo	CGI_10013164	tropomyosin [*Crassostrea gigas*]	Exon	
scaffold1870_209437	common	hypo	CGI_10013186	Neurogenic locus Notch	Exon	
scaffold888_252528	Binary	hypo	CGI_10013455	[Nematostella vectensis]	Exon	
scaffold1865_226561	common	hypo	CGI_10013544	heparan-alpha-glucosaminide N-acetyltransferase-like isoform X2	Exon	
scaffold43868_15906	Semi-quantitative	hypo	CGI_10013649	kelch 5 isoform X1	Exon	
scaffold786_92224	common	hypo	CGI_10013694	dual specificity phosphatase 16-like isoform X1 [Lingula anatina]	Exon	
scaffold786_92205	common	hypo	CGI_10013694	dual specificity phosphatase 16-like isoform X1 [Lingula anatina]	Exon	

scaffold786_92212	common	hypo	CGI_10013694	dual specificity phosphatase 16-like isoform X1 [Lingula anatina]	Exon	
scaffold976_40875	common	hypo	CGI_10014002	Beta-hexosaminidase subunit alpha	Exon	
scaffold43932_243395	Binary	hypo	CGI_10014124	Multiple epidermal growth factor-like domains 6	Exon	
scaffold43940_133880	common	hypo	CGI_10014153	supervillin-like isoform X4	Exon	
scaffold43940_133902	common	hypo	CGI_10014153	supervillin-like isoform X4	Exon	
scaffold43940_133784	Binary	hypo	CGI_10014153	supervillin-like isoform X4	Exon	
scaffold43940_133867	Binary	hypo	CGI_10014153	supervillin-like isoform X4	Exon	
scaffold43940_166586	common	hypo	CGI_10014156	hypothetical protein CGI_10014156	Exon	
scaffold737_49656	Binary	hypo	CGI_10014304	bromo adjacent homology (BAH) domain-containing [Ixodes scapularis]	Exon	
scaffold659_274707	common	hypo	CGI_10014708	ETS translocation variant 4-like isoform X3	Exon	
scaffold1490_326985	common	hypo	CGI_10014920	LIM domain-containing jub-like	Exon	
scaffold1490_355150	Semi-quantitative	hypo	CGI_10014920	LIM domain-containing jub-like	Exon	
scaffold1584_56857	common	hypo	CGI_10015072	BTB POZ domain-containing 7 isoform X3	Exon	
scaffold1584_59994	common	hypo	CGI_10015072	BTB POZ domain-containing 7 isoform X3	Exon	
scaffold193_111838	Binary	hyper	CGI_10015225	ecto-NOX disulfide-thiol exchanger 2 isoform X1	Exon	
scaffold165_246671	common	hypo	CGI_10015327	Serine threonine- phosphatase 2A regulatory subunit B subunit alpha	Exon	
scaffold165_246844	Semi-quantitative	hypo	CGI_10015327	Serine threonine- phosphatase 2A regulatory subunit B subunit alpha	Exon	
scaffold934_53353	Semi-quantitative	hypo	CGI_10015432	mRNA export factor	Exon	
scaffold705_397028	common	hypo	CGI_10015519	Beta-hexosaminidase subunit alpha	Exon	
scaffold705_397129	Binary	hypo	CGI_10015519	Beta-hexosaminidase subunit alpha	Exon	
scaffold705_397049	common	hypo	CGI_10015519	Beta-hexosaminidase subunit alpha	Exon	
scaffold705_397038	common	hypo	CGI_10015519	Beta-hexosaminidase subunit alpha	Exon	
scaffold705_397016	Semi-quantitative	hypo	CGI_10015519	Beta-hexosaminidase subunit alpha	Exon	
scaffold705_397031	Semi-quantitative	hypo	CGI_10015519	Beta-hexosaminidase subunit alpha	Exon	
scaffold1901_292161	common	hypo	CGI_10015554	rho GTPase-activating 7-like	Exon	
scaffold1901_304667	common	hypo	CGI_10015554	rho GTPase-activating 7-like	Exon	
scaffold1901_304577	common	hypo	CGI_10015554	rho GTPase-activating 7-like	Exon	
scaffold1901_306385	Binary	hypo	CGI_10015554	rho GTPase-activating 7-like	Exon	
scaffold1901_304631	Semi-quantitative	hypo	CGI_10015554	rho GTPase-activating 7-like	Exon	
scaffold562_398198	common	hypo	CGI_10016101	calcium calmodulin-dependent kinase kinase 1-like isoform X1 [*Crassostrea gigas*]	Exon	
scaffold1222_401016	common	hyper	CGI_10016252	ras-responsive element-binding 1-like	Exon	

scaffold1083_75762	Binary	hypo	CGI_10016523	A-kinase anchor 6 [*Crassostrea gigas*]	Exon	
scaffold1834_203169	common	hypo	CGI_10016659	leucine-rich repeat-containing 15-like	Exon	
scaffold556_237671	common	hypo	CGI_10016803	Myosin light chain smooth muscle	Exon	
scaffold556_262294	common	hypo	CGI_10016804	myosin light chain smooth muscle isoform X4 [Monodelphis domestica]	Exon	
scaffold789_46005	Binary	hypo	CGI_10017177	amyloid beta A4 precursor -binding family A member 1-like isoform X1 [*Crassostrea gigas*]	Exon	
scaffold248_180571	common	hypo	CGI_10017224	CCR4-NOT transcription complex subunit 6-like isoform X2	Exon	
scaffold248_180583	common	hypo	CGI_10017224	CCR4-NOT transcription complex subunit 6-like isoform X2	Exon	
scaffold248_180636	common	hypo	CGI_10017224	CCR4-NOT transcription complex subunit 6-like isoform X2	Exon	
scaffold378_198035	Semi-quantitative	hypo	CGI_10017640	SH3 domain-binding glutamic acid-rich	Intron	
scaffold459_287065	common	hypo	CGI_10018176	GH23898 [Drosophila grimshawi]	Exon	
scaffold1737_408222	Semi-quantitative	hypo	CGI_10018250	delta 4	Exon	
scaffold67_52564	common	hypo	CGI_10018263	leucine-rich repeat-containing 24-like [*Crassostrea gigas*]	Exon	
scaffold67_53929	common	hypo	CGI_10018263	leucine-rich repeat-containing 24-like [*Crassostrea gigas*]	Exon	
scaffold67_52652	common	hypo	CGI_10018263	leucine-rich repeat-containing 24-like [*Crassostrea gigas*]	Exon	
scaffold67_393362	Binary	hypo	CGI_10018273	COUP transcription factor 1 isoform X1	Exon	
scaffold67_393456	Semi-quantitative	hypo	CGI_10018273	COUP transcription factor 1 isoform X1	Exon	
scaffold514_360602	Binary	hypo	CGI_10018361	rab3 GTPase-activating catalytic subunit-like	Exon	
scaffold1832_440199	common	hypo	CGI_10018482	Tetratricopeptide repeat 17	Exon	
scaffold472_111208	common	hypo	CGI_10018501	rhotekin-like isoform X1	Exon	
scaffold472_111205	Semi-quantitative	hypo	CGI_10018501	rhotekin-like isoform X1	Exon	
scaffold189_102523	common	hypo	CGI_10018670	Ras association domain-containing 5	Exon	
scaffold576_255259	Binary	hypo	CGI_10018766	unconventional myosin-XVIIIa-like	Exon	
scaffold576_255093	common	hypo	CGI_10018766	unconventional myosin-XVIIIa-like	Exon	
scaffold576_255256	common	hypo	CGI_10018766	unconventional myosin-XVIIIa-like	Exon	
scaffold576_254666	common	hypo	CGI_10018766	unconventional myosin-XVIIIa-like	Exon	
scaffold576_254509	common	hypo	CGI_10018766	unconventional myosin-XVIIIa-like	Exon	
scaffold576_255070	common	hypo	CGI_10018766	unconventional myosin-XVIIIa-like	Exon	
scaffold42_404870	Semi-quantitative	hypo	CGI_10018916	CREB-binding -like	Exon	
scaffold305_377074	Binary	hypo	CGI_10019135	spermine oxidase	Exon	
scaffold305_376622	common	hypo	CGI_10019135	spermine oxidase	Exon	
scaffold305_376629	common	hypo	CGI_10019135	spermine oxidase	Exon	

scaffold305_376540	Semi-quantitative	hypo	CGI_10019135	spermine oxidase	Exon	
scaffold305_376536	Semi-quantitative	hypo	CGI_10019135	spermine oxidase	Exon	
scaffold980_471641	Semi-quantitative	hypo	CGI_10019237	chondroitin sulfate proteoglycan 4-like	Exon	
scaffold1763_99888	Binary	hypo	CGI_10019527	CLEC16A-like isoform X2	Exon	
scaffold1763_178440	Semi-quantitative	hypo	CGI_10019530	ankyrin-2-like isoform X5	Exon	
scaffold1763_341220	common	hypo	CGI_10019541	axin-1-like [Octopus bimaculoides]	Exon	
scaffold1896_37860	Semi-quantitative	hypo	CGI_10019632	FAM81A-like isoform X1 [*Crassostrea gigas*]	Exon	
scaffold1018_310256	Binary	hypo	CGI_10019765	fatty acid desaturase 1-like isoform X1 [Aplysia californica]	Exon	
scaffold1093_15464	Semi-quantitative	hypo	CGI_10019783	membrane-associated guanylate WW and PDZ domain-containing 1-like isoform X1	Exon	
scaffold1093_15450	Semi-quantitative	hypo	CGI_10019783	membrane-associated guanylate WW and PDZ domain-containing 1-like isoform X1	Exon	
scaffold563_119172	common	hypo	CGI_10019822	discoidin domain-containing receptor 2-like isoform X1 [Lingula anatina]	Exon	
scaffold563_142249	Semi-quantitative	hypo	CGI_10019822	discoidin domain-containing receptor 2-like isoform X1 [Lingula anatina]	Exon	
scaffold1788_290510	Semi-quantitative	hypo	CGI_10019922	eyes absent homolog 1-like isoform X1 [Lingula anatina]	Exon	
scaffold140_294630	common	hypo	CGI_10020133	Spermatogenesis-associated 13	Exon	
scaffold140_299245	common	hypo	CGI_10020133	Spermatogenesis-associated 13	Exon	
scaffold1014_181646	common	hypo	CGI_10020265	zinc finger 845-like isoform X2 [*Crassostrea gigas*]	Exon	
scaffold57_160924	common	hypo	CGI_10020544	afadin-like isoform X1	Exon	
scaffold43_610305	Semi-quantitative	hypo	CGI_10021364	Spermatogenesis-associated 1	Exon	
scaffold973_362887	Binary	hypo	CGI_10021552	rap guanine nucleotide exchange factor 1-like isoform X1	Exon	
scaffold1315_223733	Semi-quantitative	hypo	CGI_10021613	Prostaglandin E2 receptor EP4 subtype	Exon	
scaffold1315_306409	common	hypo	CGI_10021617	transducin-like enhancer 4 isoform X1	Exon	
scaffold1315_306418	Semi-quantitative	hypo	CGI_10021617	transducin-like enhancer 4 isoform X1	Exon	
scaffold1032_208914	common	hypo	CGI_10021668	BTB POZ domain-containing 17-like [*Crassostrea gigas*]	Exon	
scaffold1032_208920	common	hypo	CGI_10021668	BTB POZ domain-containing 17-like [*Crassostrea gigas*]	Exon	
scaffold1032_208984	common	hypo	CGI_10021668	BTB POZ domain-containing 17-like [*Crassostrea gigas*]	Exon	
scaffold1710_366330	common	hypo	CGI_10022355	muscle M-line assembly unc-89-like isoform X3	Exon	
scaffold1021_473573	common	hypo	CGI_10022500	biogenesis of lysosome-related organelles complex 1 subunit 3-like isoform X2	Exon	
scaffold364_478491	common	hyper	CGI_10022698	AAEL010828- partial [Aedes aegypti]	Exon	

scaffold364_478392	Binary	hyper	CGI_10022698	AAEL010828- partial [Aedes aegypti]	Exon	
scaffold364_479613	common	hyper	CGI_10022698	AAEL010828- partial [Aedes aegypti]	Exon	
scaffold364_479623	Binary	hyper	CGI_10022698	AAEL010828- partial [Aedes aegypti]	Exon	
scaffold364_479563	Binary	hyper	CGI_10022698	AAEL010828- partial [Aedes aegypti]	Exon	
scaffold364_479550	Binary	hyper	CGI_10022698	AAEL010828- partial [Aedes aegypti]	Exon	
scaffold364_479031	common	hyper	CGI_10022698	AAEL010828- partial [Aedes aegypti]	Exon	
scaffold364_479053	Binary	hyper	CGI_10022698	AAEL010828- partial [Aedes aegypti]	Exon	
scaffold413_532227	Binary	hypo	CGI_10022808	activating transcription factor of chaperone	Exon	
scaffold950_792404	Binary	hypo	CGI_10022937	dual specificity phosphatase CDC14A-like	Exon	
scaffold432_156804	Binary	hypo	CGI_10023084	chitin binding beak 1	Exon	
scaffold432_401827	common	hypo	CGI_10023103	rho guanine nucleotide exchange factor 11-like isoform X24 [*Crassostrea gigas*]	Intron	
scaffold432_401758	Semi-quantitative	hypo	CGI_10023103	rho guanine nucleotide exchange factor 11-like isoform X24 [*Crassostrea gigas*]	Exon	
scaffold602_413737	common	hypo	CGI_10023187	PREDICTED: uncharacterized protein LOC105331822 isoform X4 [*Crassostrea gigas*]	Exon	
scaffold602_413729	common	hypo	CGI_10023187	PREDICTED: uncharacterized protein LOC105331822 isoform X4 [*Crassostrea gigas*]	Exon	
scaffold1132_599160	common	hypo	CGI_10023325	ras-associated and pleckstrin homology domains-containing 1-like isoform X1 [*Crassostrea gigas*]	Exon	
scaffold383_242277	Binary	hypo	CGI_10023529	vang 2	Exon	
scaffold383_650323	Binary	hypo	CGI_10023550	neuron navigator 2-like isoform X6	Exon	
scaffold383_663929	Binary	hypo	CGI_10023551	Neuron navigator 2	Exon	
scaffold192_94565	common	hypo	CGI_10023700	autism susceptibility gene 2 -like isoform X2	Exon	
scaffold192_94620	common	hypo	CGI_10023700	autism susceptibility gene 2 -like isoform X2	Exon	
scaffold192_94572	common	hypo	CGI_10023700	autism susceptibility gene 2 -like isoform X2	Exon	
scaffold192_94529	Semi-quantitative	hypo	CGI_10023700	autism susceptibility gene 2 -like isoform X2	Exon	
scaffold1589_79672	Semi-quantitative	hypo	CGI_10024106	segment polarity dishevelled homolog DVL-3-like isoform X2	Exon	
scaffold1589_79667	Semi-quantitative	hypo	CGI_10024106	segment polarity dishevelled homolog DVL-3-like isoform X2	Exon	
scaffold1589_79646	Semi-quantitative	hypo	CGI_10024106	segment polarity dishevelled homolog DVL-3-like isoform X2	Exon	
scaffold485_672490	common	hypo	CGI_10024187	chondroitin sulfate synthase 1	Exon	
scaffold485_672472	common	hypo	CGI_10024187	chondroitin sulfate synthase 1	Exon	
scaffold485_672360	Semi-quantitative	hypo	CGI_10024187	chondroitin sulfate synthase 1	Exon	
scaffold271_395493	common	hypo	CGI_10024474	zinc finger 142-like	Intron	
scaffold271_395448	common	hypo	CGI_10024474	zinc finger 142-like	Exon	

scaffold271_395333	common	hypo	CGI_1002447 4	zinc finger 142-like	Exon	
scaffold271_395379	Binary	hypo	CGI_1002447 4	zinc finger 142-like	Exon	
scaffold271_395286	common	hypo	CGI_1002447 4	zinc finger 142-like	Exon	
scaffold70_533625	common	hypo	CGI_1002477 6	plexin-A2-like [Limulus polyphemus]	Exon	
scaffold1017_366393	Semi-quantitative	hypo	CGI_1002500 1	calponin homology domain-containing DDB_G0272472-like isoform X1 [*Crassostrea gigas*]	Exon	
scaffold226_1019449	common	hypo	CGI_1002539 7	FAM179B-like isoform X1 [*Crassostrea gigas*]	Intron	
scaffold168_677764	Semi-quantitative	hypo	CGI_1002560 3	ankyrin repeat and SOCS box 12-like [Priapulus caudatus]	Exon	
scaffold168_743638	common	hypo	CGI_1002561 0	eukaryotic elongation factor 2 kinase-like isoform X2	Exon	
scaffold168_743658	common	hypo	CGI_1002561 0	eukaryotic elongation factor 2 kinase-like isoform X2	Exon	
scaffold168_743629	common	hypo	CGI_1002561 0	eukaryotic elongation factor 2 kinase-like isoform X2	Exon	
scaffold168_778029	common	hypo	CGI_1002561 2	ankyrin repeat and fibronectin type-III domain-containing 1-like isoform X1 [*Crassostrea gigas*]	Exon	
scaffold168_777982	common	hypo	CGI_1002561 2	ankyrin repeat and fibronectin type-III domain-containing 1-like isoform X1 [*Crassostrea gigas*]	Exon	
scaffold121_70185	Semi-quantitative	hypo	CGI_1002563 4	tyrosine- phosphatase non-receptor type 4-like isoform X3	Exon	
scaffold121_70234	Semi-quantitative	hypo	CGI_1002563 4	tyrosine- phosphatase non-receptor type 4-like isoform X3	Exon	
scaffold733_685155	common	hypo	CGI_1002588 5	nuclear receptor coactivator 2-like isoform X5 [*Crassostrea gigas*]	Exon	
scaffold204_117424	common	hypo	CGI_1002609 4	disks large-associated 4-like isoform X1 [*Crassostrea gigas*]	Exon	
scaffold204_117376	Semi-quantitative	hypo	CGI_1002609 4	disks large-associated 4-like isoform X1 [*Crassostrea gigas*]	Exon	
scaffold1154_771923	common	hypo	CGI_1002618 6	calpain-B-like isoform X1 [*Crassostrea gigas*]	Exon	
scaffold678_188930	Semi-quantitative	hypo	CGI_1002633 9	cysteine sulfinic Acid Decarboxylase	Exon	
scaffold100_662090	common	hypo	CGI_1002644 7	Cyclic AMP-responsive element-binding 3 2	Exon	
scaffold313_186165	common	hypo	CGI_1002658 7	disks large homolog 5-like isoform X2	Exon	
scaffold313_443123	Semi-quantitative	hypo	CGI_1002660 5	chitinase-3 1	Exon	
scaffold301_1200972	Binary	hypo	CGI_1002675 1	Exoglucanase xynX	Exon	
scaffold142_1065353	Binary	hypo	CGI_1002678 5	winged eye-like isoform X1 [Lingula anatina]	Exon	
scaffold142_1065204	Binary	hypo	CGI_1002678 5	winged eye-like isoform X1 [Lingula anatina]	Exon	
scaffold471_967227	Binary	hypo	CGI_1002703 1	nuclear hormone receptor HR96-like isoform X3 [Lingula anatina]	Exon	
scaffold433_1022606	Semi-quantitative	hypo	CGI_1002710 5	DC-STAMP domain-containing 2-like	Exon	
scaffold1301_195125	common	hypo	CGI_1002770 9	rap guanine nucleotide exchange factor 2-like isoform X14 [*Crassostrea gigas*]	Exon	
scaffold1301_195237	common	hypo	CGI_1002770 9	rap guanine nucleotide exchange factor 2-like isoform X14 [*Crassostrea gigas*]	Exon	
scaffold1301_195135	common	hypo	CGI_1002770 9	rap guanine nucleotide exchange factor 2-like isoform X14 [*Crassostrea gigas*]	Exon	

scaffold1301_1119020	common	hypo	CGI_10027761	AF4 FMR2 family member 4	Exon	
scaffold77_112668	common	hypo	CGI_10027967	phospholipase A-2-activating	Exon	
scaffold77_112673	Semi-quantitative	hypo	CGI_10027967	phospholipase A-2-activating	Exon	
scaffold77_1265639	common	hypo	CGI_10028030	rho GTPase-activating 190-like isoform X1 [Lingula anatina]	Exon	
scaffold419_94810	common	hypo	CGI_10028320	FRAS1-related extracellular matrix 2-like	Exon	
scaffold419_94803	Binary	hypo	CGI_10028320	FRAS1-related extracellular matrix 2-like	Exon	
scaffold419_354029	Semi-quantitative	hypo	CGI_10028340	sprouty homolog 2	Exon	
scaffold150_889014	Semi-quantitative	hypo	CGI_10028630	Serine threonine- phosphatase 4 regulatory subunit 4	Exon	
scaffold150_1471517	common	hypo	CGI_10028686	zinc finger MIZ domain-containing 1-like isoform X1 [*Crassostrea gigas*]	Exon	

Supplementary file 12

Due to the large size of the file, the file can be reached through the link below.

https://drive.google.com/drive/folders/114V0oY5ySOUtpgUb_E97zhZtr3speIMf?usp=sharing

Supplementary file 13

Due to the large size of the file, the file can be reached through the link below.

https://drive.google.com/drive/folders/114V0oY5ySOUtpgUb_E97zhZtr3speIMf?usp=sharing

Supplementary file 14

Due to the large size of the file, the files A and B can be reached through the link below.

https://drive.google.com/drive/folders/114V0oY5ySOUtpgUb_E97zhZtr3speIMf?usp=sharing

The Supplementary file 14C is below

List of Significant CpGs best explained by suggestive SNPs using the binary phenotype as a covariate

CpGs	Location in GENE	Cis or trans	Gene_IDS	Gene Annotation (family or domain containing name)	SNPs	Beta	Stats	pvalue	FDR	Location in GENE	Gene_IDS	Gene Annotation (family or domain containing name)
scaffold364_471615	Exon	Trans	CGI_10022697	Transcription termination factor 2	scaffold117_77443	-0.10270651	-5.15835673	5.61E-07	0.0246272	Intron	CGI_10016954	Serine threonine-kinase TBK1
scaffold193_111838	Exon	Cis	CGI_10015225	ecto-NOX disulfide-thiol exchanger 2 isoform X1	scaffold364_478394	-0.06864863	-5.27387609	3.22E-07	0.0177013	Exon	CGI_10022698	AAEL010828-partial [Aedes aegypti]
scaffold364_467953	Exon	Cis	CGI_10022697	Transcription termination factor 2	scaffold364_478394	-0.08152261	-4.94300803	1.54E-06	0.0444018	Exon	CGI_10022698	AAEL010828-partial [Aedes aegypti]
scaffold364_471528	Exon	Cis	CGI_10022697	Transcription termination factor 2	scaffold364_478394	-0.11638338	-5.68815443	4.12E-08	0.0051074	Exon	CGI_10022698	AAEL010828-partial [Aedes aegypti]
scaffold364_471609	Exon	Cis	CGI_10022697	Transcription termination factor 2	scaffold364_478394	-0.11135953	-5.37081997	2.01E-07	0.0133366	Exon	CGI_10022698	AAEL010828-partial [Aedes aegypti]
scaffold364_471615	Exon	Cis	CGI_10022697	Transcription termination factor 2	scaffold364_478394	-0.11177886	-6.18649859	3.02E-09	0.0010378	Exon	CGI_10022698	AAEL010828-partial [Aedes aegypti]
scaffold364_471627	Exon	Cis	CGI_10022697	Transcription termination factor 2	scaffold364_478394	-0.12072229	-6.61116303	2.92E-10	0.0002194	Exon	CGI_10022698	AAEL010828-partial [Aedes aegypti]
scaffold364_473225	Exon	Cis	CGI_10022697	Transcription termination factor 2	scaffold364_478394	-0.1556275	-5.34047018	2.33E-07	0.0145797	Exon	CGI_10022698	AAEL010828-partial [Aedes aegypti]

scaffold364_473230	Exon	Cis	CGI_10022697	Transcription termination factor 2	scaffold364_478394	-0.13021534	-8.08765822	4.26E-14	9.86E-08	Exon	CGI_10022698	AAEL010828- partial [Aedes aegypti]
scaffold364_473238	Exon	Cis	CGI_10022697	Transcription termination factor 2	scaffold364_478394	-0.13606206	-7.06648238	2.14E-11	2.82E-05	Exon	CGI_10022698	AAEL010828- partial [Aedes aegypti]
scaffold364_478392	Exon	Cis	CGI_10022698	AAEL010828- partial [Aedes aegypti]	scaffold364_478394	-0.24313477	-13.8336023	1.27E-31	1.29E-24	Exon	CGI_10022698	AAEL010828- partial [Aedes aegypti]
scaffold364_478491	Exon	Cis	CGI_10022698	AAEL010828- partial [Aedes aegypti]	scaffold364_478394	-0.22215064	-12.105978	4.09E-26	2.72E-19	Exon	CGI_10022698	AAEL010828- partial [Aedes aegypti]
scaffold364_479031	Exon	Cis	CGI_10022698	AAEL010828- partial [Aedes aegypti]	scaffold364_478394	-0.19509848	-10.1960993	3.39E-20	1.39E-13	Exon	CGI_10022698	AAEL010828- partial [Aedes aegypti]
scaffold364_479053	Exon	Cis	CGI_10022698	AAEL010828- partial [Aedes aegypti]	scaffold364_478394	-0.14597148	-6.35162867	1.23E-09	0.00058626	Exon	CGI_10022698	AAEL010828- partial [Aedes aegypti]
scaffold364_479550	Exon	Cis	CGI_10022698	AAEL010828- partial [Aedes aegypti]	scaffold364_478394	-0.17724841	-6.8317992	8.33E-11	8.56E-05	Exon	CGI_10022698	AAEL010828- partial [Aedes aegypti]
scaffold364_479563	Exon	Cis	CGI_10022698	AAEL010828- partial [Aedes aegypti]	scaffold364_478394	-0.15960649	-5.94996016	1.06E-08	0.0022468	Exon	CGI_10022698	AAEL010828- partial [Aedes aegypti]
scaffold364_479613	Exon	Cis	CGI_10022698	AAEL010828- partial [Aedes aegypti]	scaffold364_478394	-0.16177233	-6.47266041	6.32E-10	0.00037645	Exon	CGI_10022698	AAEL010828- partial [Aedes aegypti]
scaffold364_479623	Exon	Cis	CGI_10022698	AAEL010828- partial [Aedes aegypti]	scaffold364_478394	-0.18990609	-7.3043288	5.24E-12	8.33E-06	Exon	CGI_10022698	AAEL010828- partial [Aedes aegypti]
scaffold522_293879	Intron	Cis	CGI_10020325	tyrosine-phosphatase Lar-like isoform X6	scaffold522_301782	-0.086644	-4.9077293	1.81E-06	0.04872247	Exon	CGI_10020325	tyrosine- phosphatase Lar-like isoform X6

The Supplementary file 14D is below

List of Significant CpGs best explained by suggestive SNPs using the binary phenotype as a covariate

CpGs	Location in GENE	Cis or trans	Gene_IDS	Gene Annotation (family or domain containing name)	SNPs	Beta	Stats	pvalue	FDR	Location in GENE	Gene_IDS	Gene Annotation (family or domain containing name)
scaffold364_471615	Exon	Trans	CGI_10022697	Transcription termination factor 2	scaffold117_77443	-0.101768859	-5.018813289	1.08E-06	0.036232646	Intron	CGI_10016954	Serine threonine- kinase TBK1
scaffold522_293879	Intron	Cis	CGI_10020325	tyrosine- phosphatase Lar-like isoform X6	scaffold522_301782	-0.087151346	-4.946341718	1.51E-06	0.044010311	Exon	CGI_10020325	tyrosine- phosphatase Lar-like isoform X6
scaffold364_478491	Exon	Cis	CGI_10022698	AAEL010828- partial [Aedes aegypti]	scaffold364_478394	-0.222399728	-11.99443912	9.20E-26	5.94E-19	Exon	CGI_10022698	AAEL010828- partial [Aedes aegypti]
scaffold201_38576	Exon	Cis	CGI_10004417	dedicator of cytokinesis 7-like isoform X4	scaffold201_38576	-0.120838852	-10.37982707	9.40E-21	4.04E-14	Exon	CGI_10004417	dedicator of cytokinesis 7-like isoform X4
scaffold364_479031	Exon	Cis	CGI_10022698	AAEL010828- partial [Aedes aegypti]	scaffold364_478394	-0.195984807	-10.15208858	4.60E-20	1.87E-13	Exon	CGI_10022698	AAEL010828- partial [Aedes aegypti]
scaffold364_473230	Exon	Cis	CGI_10022697	Transcription termination factor 2	scaffold364_478394	-0.131989477	-8.046580936	5.52E-14	1.26E-07	Exon	CGI_10022698	AAEL010828- partial [Aedes aegypti]
scaffold364_473238	Exon	Cis	CGI_10022697	Transcription termination factor 2	scaffold364_478394	-0.138094786	-7.066661553	2.13E-11	2.82E-05	Exon	CGI_10022698	AAEL010828- partial [Aedes aegypti]
scaffold364_471627	Exon	Cis	CGI_10022697	Transcription termination factor 2	scaffold364_478394	-0.124938464	-6.725587071	1.53E-10	0.000136372	Exon	CGI_10022698	AAEL010828- partial [Aedes aegypti]
scaffold364_479613	Exon	Cis	CGI_10022698	AAEL010828- partial [Aedes aegypti]	scaffold364_478394	-0.167534502	-6.560326767	3.88E-10	0.00026868	Exon	CGI_10022698	AAEL010828- partial [Aedes aegypti]
scaffold364_471615	Exon	Cis	CGI_10022697	Transcription termination factor 2	scaffold364_478394	-0.113436701	-6.193862348	2.91E-09	0.001011788	Exon	CGI_10022698	AAEL010828- partial [Aedes aegypti]
scaffold364_471528	Exon	Cis	CGI_10022697	Transcription termination factor 2	scaffold364_478394	-0.120153717	-5.772653812	2.68E-08	0.003928608	Exon	CGI_10022698	AAEL010828- partial [Aedes aegypti]

scaffold364_471609	Exon	Cis	CGI_10022697	Transcription termination factor 2	scaffold364_478394	-0.113854019	-5.414198943	1.63E-07	0.01172554	Exon	CGI_10022698	AAEL010828- partial [Aedes aegypti]
scaffold364_473225	Exon	Cis	CGI_10022697	Transcription termination factor 2	scaffold364_478394	-0.157010827	-5.329001423	2.47E-07	0.015070723	Exon	CGI_10022698	AAEL010828- partial [Aedes aegypti]
scaffold121_70185	Exon	Trans	CGI_10025634	tyrosine- phosphatase non-receptor type 4-like isoform X3	scaffold107_575382	-0.082312478	-5.062014792	8.84E-07	0.032196101	Exon	CGI_10025062	All-trans-retinol 13,14-reductase
scaffold1032_208984	Exon	Cis	CGI_10021668	BTB POZ domain-containing 17-like [*Crassostrea gigas*]	scaffold41994_94179	0.051292527	4.954787343	1.46E-06	0.04303124	Intron	CGI_10005879	0
scaffold204_117376	Exon	Trans	CGI_10026094	disks large-associated 4-like isoform X1 [*Crassostrea gigas*]	scaffold469_660273	-0.077956901	-4.932902624	1.61E-06	0.045603201	Intron	CGI_10024691	Patatin-like phospholipase domain-containing 7
scaffold201_38576	Exon	Cis	CGI_10004417	dedicator of cytokinesis 7-like isoform X4	scaffold201_38432	-0.055006521	-4.930944248	1.63E-06	0.045839184	Intron	CGI_10004417	dedicator of cytokinesis 7-like isoform X4

Chapter 4

General Discussion

Conclusion and Perspectives

Chapter 4: General Discussion

The main objectives of this thesis here to identify genetic and epigenetic signatures of oyster resistance to POMS and to quantify the relative weight of both mechanisms in the phenotypic expression of the resistance. These objectives were addressed by sampling wild oyster populations exposed to two different environment (farming and non-farming) and phenotyping them by experimental infection that mimic the natural route of infection. The genetic variation (SNPs) and one component of the epigenetic variation (DNA methylation at CG context; CpGs hereafter) were jointly obtained by optimizing a whole exome capture approach. The SeqCap Epi Enrichment System have been used for my thesis for its ability to produce information within a functional interest (e.g. exon) and the study of genetic and epigenetic information on a high number of samples at a reasonable cost. To set up this experiment in oysters several steps were optimized from the bench to the bioinformatics pipelines. Genome and Epigenome Wide Association Studies (GWAS and EWAS) were used to identify signature of oyster resistance to POMS. Correlation between the genetic and epigenetic variation, MethQTL (Methylation Quantitative Trait Loci) and variance partition methods were used to quantify the relative involvement of genetic and epigenetic variations in phenotypic expression.

Overall, the work carried out during this thesis has enabled to show:

1- That natural oyster populations differentially exposed to the emerging disease named Pacific Oyster Mortality Syndrome (POMS) display signatures of selections both in their genome (SNPs) and in their epigenome (CpGs).
2- These signatures are localized in different genes but most of them belong to the immune related biological processes.
3- Genetic and epigenetic variations are partly correlated and the former was associated with a large fraction of the second.
4- We also showed that most of the epigenetic variations significantly associated to oyster resistance were independent and explained a higher part of the phenotypic variation.

These results confirmed that host populations facing infectious disease emergence could rely on genetic and epigenetic variation to adapt rapidly to emerging disease.

In this chapter, I will discuss some of key observations generated from the results obtained in this thesis:

i) The variation in mortality rates within non-farming population.
ii) The UBA2 gene, a putative key genetic actor of oyster resistance.
iii) The JAK/STAT and TLR/NF-κB pathways and their link with the genetic and epigenetic variation.
iv) The application of these results in marker-assisted selection.

Variation in mortalities rates within non-farming population

Defining oyster phenotype represent an important step in answering the thesis objectives. For qualifying each oyster phenotypes (resistant vs susceptible or maximum time (hours) of oyster being alive), we used a randomized complete block design composed of eight tanks (replicates). Then, we used a cohabitation method to induce the POMS event, where we put the donor oysters (injected with viral suspension) in contact with recipient oysters (sampled wild juvenile oysters). The disease was then transmitted naturally from donor oysters to the recipient oysters (mimicking the natural route of infection).

With the purpose of phenotyping, results obtained from our experimental design was efficient for populations of Bay of Brest. In total, we have phenotyped 356 oysters from six populations, of which 150 oysters died (42%; susceptible phenotype) while 206 remained alive (58%; resistant phenotype). The mortality rates, kinetics of mortalities and virus load in the seawater (tank water) were similar to previous experiments (de Lorgeril et al., 2018; D Schikorski et al., 2011), which validated our approach (de Lorgeril et al., 2018; B Petton et al., 2019; Bruno Petton et al., 2013). Additionally, there were no significant differences between tank ("tank effect") in terms of mortalities or amount of virus. Further results showed that almost all the oyster coming from farming areas survived the experimental infection, which is expected as these oysters are supposed to be confronted to a POMS event.

On the other hand, oysters coming from non-farming areas, showed rates of survival ranging from 30% to 45%. This high mortality rates are obviously explained by the absence of selection pressure. However, the variability in survival within non-farming areas raises questions about the source of this variation. I propose below some non-exclusive hypotheses to explain this phenomenon beyond which we would found some non-genetic explanations.

These variabilities could be attributed to environmental differences (e.g. the difference in temperature, plankton composition, habitat etc.). Oysters are sessile organisms living in intertidal zones that are characterized by environmental variability. In such habitat, oysters are in direct contact to surrounding environmental pressures and are in constant interaction with other organisms. Additionally, oysters live with a microbiota that could be mutualistic, opportunistic or pathogenic and most of these associations are under fine control involving the oyster immune system.

It has been previously demonstrated that resistance to POMS can be influenced by biotic factors acting directly or indirectly on the oyster status of resistance/susceptibility (Bruno Petton et al., 2021b). In terms of direct factors, recent results showed that host immune system could be enhanced through biotic interaction called immune priming (Lafont et al., 2017, 2020) and immune shaping (Fallet et al., 2022). The former can be broadly defined as an increased protection to a pathogen following a previous exposure to a pathogen or an immune elicitor. The second consists in the modulation of the immune capabilities of an individual by an interaction with microorganisms during its early life.

In the case of immune priming in oysters, injections of Poly(I:C), a molecule mimicking viral infection, into susceptible oysters prior to POMS infection led to the induction of a resistant phenotype. This is followed by a protection that can reach 100% even for susceptible families (Lafont et al., 2017). This phenotypic inversion was characterized at the transcriptomic level by a strong antiviral response that impaired OsHV-1 replication and POMS disease development (Lafont et al., 2020).

In the case of immune shaping, the exposure to non-pathogenic but rich microorganisms' flora during the early life (larval stages) has enhanced the immune capabilities of oysters. The phenotypes obtained are less contrasted since the resistance level of a susceptible family increased by 9 to 13%. This increase was also characterized by significant differences at the transcriptomic level in several antiviral response genes (Fallet et al., 2022).

In terms of indirect biotic factors involved in POMS resistance, a recent study investigating the effect of seaweeds on the susceptibility to the POMS were performed (Dugeny et al., 2022). The study showed that seaweeds influenced the microbiota composition of oysters and probably induce a modification in the susceptibility to the POMS (Dugeny et al., 2022). However, the effects of seaweeds on the host transcriptome or epigenome remain unknown. This last study shows the importance of the influence of the natural environment on the disease outcome. In addition, it highlights the importance of considering the natural environment of the host to completely understand the disease.

Still through some indirect biotic effect, different oyster populations can be colonized by different microbiome (pathogenic and non-pathogenic). There is now evidence that certain microbes could be associated to oyster resistance to POMS. The roles of bacterial microbiota

have been recently studied using 16S metabarcoding technique. For example, the study of (Clerissi et al., 2020) found an association between the *Mycoplasmataceae*, *Rhodospirillaceae*, *Vibrionaceae* and *Photobacterium* genera and the susceptible oyster families. Interestingly, oyster families that survived in the field to the infectious period of POMS showed higher proportion of specific taxa including *Cyanobacteriaceae*, *Colwelliaceae*, and *Rhodobacteraceae*. In this last study, the authors showed that susceptible oysters had low abundance of *Cyanobacteria* (Subsection III, family I) comparing to resistant oysters. Thus, suggesting a potential endosymbiotic relationship between the identified *Cyanobacteria* and oyster with a link to the resistance. The mechanism of action would be a role of barrier against the secondary bacterial infection, which kill oysters during POMS (de Lorgeril et al., 2018; Lucasson et al., 2020).

In conclusion, the variation in mortalities within non-farming area could be a result from environmental factors that influence the host through epigenetic variation. Additionally, these variations in non-farming areas could be associated with differences in the microbiota, which would help oysters to better respond to the POMS event. Finally, we cannot exclude that these variations in the morality are also result of some small genetic structure (microstructure).

UBA2 a key genetic actor of oyster resistance?

Genome wide association study (GWAS) analyses identified two SNPs that were significantly associated to POMS resistance. One of these SNPs was located in a gene with an unknown function, while the other was located in an intron of a gene encoding a SUMO-activating enzyme subunit 2 (SAE2; CGI_10018487). This gene is also known as UBA2 (Ubiquitin Like Modifier Activating Enzyme 2). The role of this gene in the sumoylation pathway and its location on the chromosome 6 makes it a particularly interesting candidate to explain oyster resistance.

UBA2 is a key actor of the sumoylation pathway (Everett et al., 2013). The Sumoylation is a post-translational modification brought by the action of Small ubiquitin-like modifier (SUMO) proteins. Sumoylation regulate target protein function through modifications of their interaction, stability or activity properties. The sumoylation occurs in three steps i) the activation; ii) the conjugation; iii) the ligation (Shuai & Liu, 2005), which are mediated through the activity of three enzymes namely SUMO E1 activating enzyme [E1], SUMO E2 conjugating enzyme [E2] and SUMO E3 ligase [E3]) (Figure 4.1). The first step, starts by the removal of a carboxyl (C)-terminal residue on the SUMO to expose the di-glycine motif needed for the conjugation. Then, the enzyme E1, a heterodimer composed by a SUMO-activating enzyme subunit 1 (SAE1) and a SUMO-activating enzyme subunit 2 (SAE2/**UBA2**), activates SUMO. Then the activated SUMO is transferred from E1 to the cysteine residue of the E2 enzyme (UBC9), which help to target the specific substrate. Finally, the enzyme E3, ligate the SUMO from E2 to the target protein. This ligation involved a covalent conjugation to the lysine (K) residues exposed on the target proteins (Adorisio et al., 2017; Lork et al., 2021; Shuai & Liu, 2005).

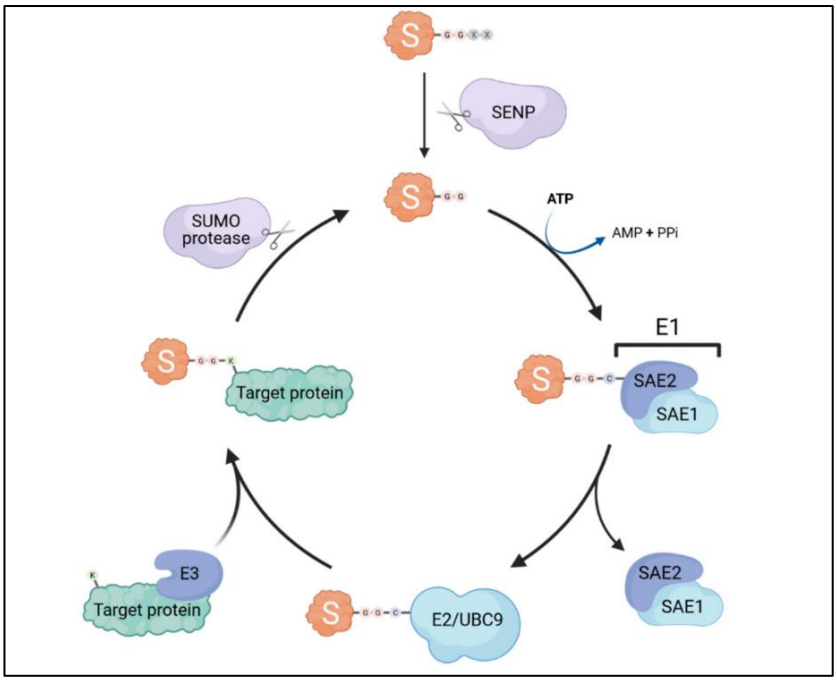


Figure 4.1 : The sumoylation pathway showing the three main steps in the sumoylation. Figure adapted from (Lork et al., 2021)

Interestingly, sumoylation was shown to be tightly associated to host-virus interaction, either through its role in the regulation of the antiviral response (Adorisio et al., 2017), as a key regulator of innate antiviral immunity (Lork et al., 2021) or through its manipulation by the virus itself (Mattoscio et al., 2013).

In terms of regulation of the antiviral pathway, studies have shown the role of the sumoylation of STAT1 (Signal transducer and activator of transcription 1) a key actor of a key antiviral pathway named JAK/STAT (Figure 4.2). This pathway start by cytokine stimulation that activate JAKs that in return phosphorylate STAT. The dimer of JAK and STAT is then translocated to the nucleus where it induces the transcription of interferon-stimulated genes (ISGs). Thus, the translocation led to expression of different ISGs that target different stages of virus life cycle (Kotenko et al., 2003; M. M. H. Li et al., 2015), resulting in a strong antiviral response. One of the key ISGs, is the Viperin (Virus Inhibitory Protein), which have been shown to be overexpressed during the OshV-1 infection in resistant oyster (de Lorgeril et al., 2018). The DNA binding activity of STAT1 is however modulated through the SUMO conjugation to STAT1, which in return negatively affect the STAT-mediated gene expression (Begitt et al., 2011; Grönholm et al., 2012) which lead to a differential regulation of the Janus kinase (JAK) and so to the down regulation of the entire pathway.

On the other hand, it is today well-known that several viruses interfere with the sumoylation pathway to either escape host immunity or to hijack host cell machinery (Fan et al., 2022). The sumoylation process could be used by the virus to inhibits the host immune response or to enhance viral replication and macromolecular synthesis and assembly (Cheng et al., 2017; Fan et al., 2022; Müller & Dejean, 1999; Tripathi et al., 2021; Wilson, 2017). Interestingly, a member of the herpes virus order, the human cytomegalovirus (HMCV) has been shown to use this pathway. The immediate-early 1 (1E1) protein, is the first virus protein expressed during infection and this protein was shown to be sumoylated. This expression is needed for the regulation of the viral genes expression and the disruption of the host immune response. Interestingly, the HMCV replication is much lower when the IE1 is sumoylation-deficient, which negatively affect the IE2 expression (Nevels et al., 2004). Other viruses have been shown to use different mechanisms based on the sumoylation. In the case of the Adenovirus, the viral protein (Gam1) can target the SAE1/UBA2 complex to induce its ubiquitination that results in the degradation of SAE1 by the proteasome. Once the SAE1 is degraded, the UBA2 remain unstable and will also be degraded later by the proteasome (Boggio et al., 2007; Fan et al., 2022). Therefore, resulting in the disruption of the host sumoylation process.

The UBA2 gene was the only gene with a genome-wide significant SNP, a SNP localized in the intron part of this gene. This localization within the intron part suggests that its biological effect is probably associated to the gene expression and not to the structure of the encoded protein. The full length of UBA2 was characterized by the exome capture but no significant nor suggestive SNP were detected in its coding sequence. The other hypothesis is that this SNP picture (by linkage disequilibrium) the presence of other genetic variations in an uncharacterized genomic portion of UBA2, its regulatory region.

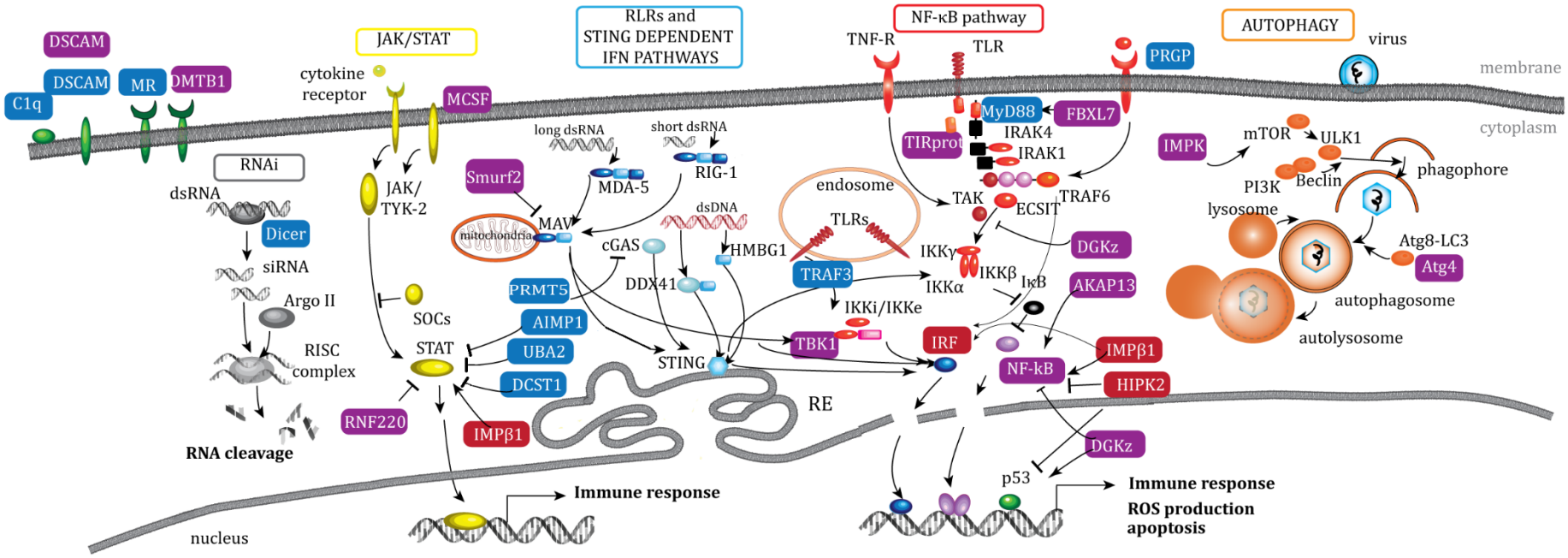


Figure 4.2 : Innate immune pathways genes display genetic and/ epigenetic variation.

Highlighting genes of innate immune signalling pathways displaying genetic variation (blue rectangle), epigenetic variation (red rectangle) or a mix of genetic and epigenetic variation (SNP plus CpG or MethQTL, violet rectangle).

Another interesting point about UBA2 gene is its location in the chromosome 6 of the oyster genome. Previous studies have identified association of this chromosome with oyster resistance to POMS (Gutierrez et al., 2018; Sauvage et al., 2010). However, none of these studies has identified an association between UBA2 and the resistance to POMS. We further estimated the exact distance of this gene from other SNPs and QTL identified. Interestingly, the UBA2 gene is located close to the QTL region and SNPs (Figure 4.3) identified by (Gutierrez et al., 2018; Sauvage et al., 2010).

All these results suggests that this region, and maybe UBA2 specifically could harbour an essential genetic role in the explanation of oyster resistance to POMS. Future studies focusing on the function of UBA2 and sumoylation process will be needed to confirm this hypothesis. In addition, the recent development of gene invalidation methods in oyster would be very interesting to address this question.

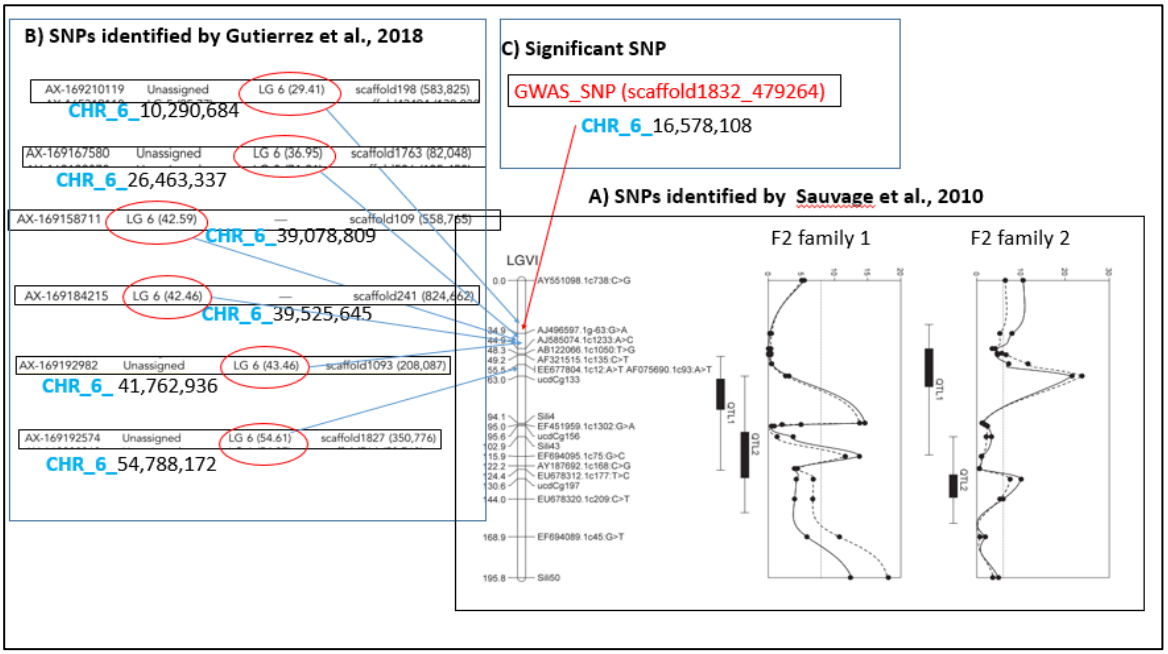


Figure 4.3: Chromosome 6 (linkage group 6; LG6) showing QTL and SNPs associated to resistance to POMS.

A) Figure adapted from; Sauvage et al. (2010), showing the LG6 and the graphic of significant QTLs identified in two different families F2 families. B) The SNPs identified in the LG6 from Gutierrez et al. (2018). C) The Only SNP identified in the UBA2 gene

Key genes with genetic and epigenetic variation are involved in JAK/STAT and TLR/NF-KB pathways.

The GWAS and EWAS approaches identified 320 putative genes and among these genes, only six genes were in common. Interestingly, 31 genes (10%) were immune-related genes and are known to be involved in different immune pathways. While genes with genetic variation were involved mostly in JAK/STAT and TLR/NF-KB pathways, the genes with epigenetic variation were mostly involved in the TLR/NF-KB pathway only.

Genetic markers involved in JAK/STAT pathway:

JAK/STAT pathway is activated once JAK is stimulated by cytokine signal through the JAK receptor. The activation of JAKs is followed by the phosphorylation of STATs that dimerize and translocate to the nucleus, where it activate gene transcription that mediate various responses including antiviral response. The STATs activation can be regulated by various protein modification processes including the sumoylation process and ubiquitination (Shuai & Liu, 2003).

In our GWAS study, we focused on the top SNPs associated to POMS. We identified 186 SNPs suggestively associated to POMS resistance. Three SNPs were located in three different genes involved in the regulation of the JAK/STAT pathway (Figure 4.2). These genes are AIMP1, DCST1 and UBA2 (see above for the description of tUBA2 function):

- One SNP was located in the exon part of a DC-STAMP domain-containing 1-like (DCST1; CGI_10007724) gene. This SNP induced a non-synonymous mutation that may affect the function of the encoded protein. DCST1 was shown to negatively regulate the interferon signalling pathway by ubiquitination-mediated degradation of STAT2 (Nair et al., 2016). The identified DC-STAMP gene did not display the RING domain that was reported to be needed for the degradation of STAT2 (Nair et al., 2016) however this question would be further studied.
- One SNP was found in the exon part of a gene encoding an aminoacyl tRNA synthase complex-interacting multifunctional 1 (AIMP1; CGI_10018862). Interestingly, AIMP1 have also been associated to the negative regulation of STATs protein (Zheng Zhou et al., 2020). AIMP1 was shown to be the precursor of endothelial monocyte activating polypeptide II (EMAP II). The role of AIMP1 in the occurrence and development of

cancer have been reported (Lee et al., 2019). As for DCST1, the SNP induced a non-synonymous mutation potentially inducing functional changes.

Overall, the three genes, participating in the regulation of the JAK/STAT pathway displayed a SNP. This pathway was previously shown to be associated with the resistance of oyster to POMS. Several genes (e.g. 2',5'- oligoadenylate synthetase (2',5'-OAS), suppressor of cytokine signalling (SOCS2) and STAT) of the JAK/STAT pathway were shown to display a higher basal expression and or a higher responsiveness in response to the diseases in resistant oyster by comparison to susceptible (de Lorgeril et al., 2018, 2020). The results previously obtained added to the one we provide here let us to hypothesis that the negative regulation derived from UBA2, DSCT1 and AIMP1 could negatively affect STAT1/2 activation, and in return the expression of essential antiviral effectors such as ADAR (double-stranded RNA-specific adenosine deaminase) or Viperin (virus inhibitory protein, endoplasmic reticulum-associated, IFN-inducible). This would result in an overall decreased effectiveness of the antiviral response in susceptible oysters. To confirm this hypothesis the role of UBA2, DSCT1 and AIMP1 needs to be further examined using genome editing techniques.

Genetic markers involved in TLR/NF-κB

The TLR/NF-κB (Toll-Like Receptor/ Nuclear Factor Kappa B) is a signalling pathway involved in innate immune system (Figure 4.2). The pathway is activated by the interaction between pathogen associated molecular patterns (PAMPs) with pattern recognition receptors (PRRs) such as the TLRs. TLRs act as primary sensors detecting different component of pathogens and initiate an innate immune response (L. Wang et al., 2018). TLRs send a signal transmit the MyD88 protein through the homophilic interactions between the TIR-TIR domains. The Death domain present on MyD88 allow the association of Myd88 with death domain of the serine threonine protein kinase IL-1R-associated kinase (IRAK). The activated IRAK then interact with TRAF6, which involves its oligomerization that lead to the activation of IkappaB kinase (IKK). The IKK further phosphorylates IκBα, which triggers an ubiquitin-dependent IκBα degradation in the proteasome and therefore releasing and translocating NF-κB into the nucleus (Figure 4.2). Once NF-κB is translocated into the nucleus, it induce the transcription of target genes (Horng & Medzhitov, 2001) including different inflammatory cytokine genes (Kawai & Akira, 2007) that will activate the JAK/STAT pathway.

Several genes involved in the backbone of the NF-κB pathway or in its regulation displayed genetic variation:

i- Myeloid differentiation primary response 88 (MyD88; CGI_10007490), displayed a SNP located in the intron part of this gene. MyD88 is a cytosolic adaptor involved in the activation of NF-κB signalling pathway (L. Wang et al., 2018). In *C. gigas*, there are 10 MyD88 genes, suggesting a specific role for each MyD88 coupled with their coupled Toll-like receptors (TLRs) (L. Zhang et al., 2015). Interestingly, two of these genes were found over expressed in resistant oyster families at the basal transcriptomic level (de Lorgeril et al., 2020). This suggests that a higher expression of MyD88 gene may enhance the efficiency (speed/strength) of the antiviral response by increasing the signal transduction from the TLR to the translocation of NF-κB (Sang et al., 2020). As previously said, a localisation in the intronic part of the gene suggest that the causal SNP is located upstream, probably in the regulatory portion of the gene which would explain a differential expression level between susceptible and resistant oysters. Identifying this gene in previous studies and our study further strengthen the involvement of this gene in oyster resistance to POMS.

ii- TANK-binding kinase 1 (TBK1; CGI_10016954), displayed a SNP located in the intron part. TBK1 is known to be a key kinase required for the phosphorylation of the transcription factor IRF3. This phosphorylation is essential for the translocation of IRF3 to the nucleus where it mediate the transcription of IFN and other co-regulated genes (B. Huang et al., 2019; Tang et al., 2016). This gene along with another TBK1 gene were higher expressed under basal condition in the resistant family sampled from natural population of farming area (de Lorgeril et al., 2020).

iii- TNF receptor-associated factor 3-like (TRAF3; CGI_10019401), displayed a SNP in the intron part of the gene. TRAF3 is an adaptor protein that is recruited by PRRs that subsequently bind to TBK1 to activate it (B. Huang et al., 2021). Out of the seven members of TRAF family that found in mammals (TRAF1-7), only TRAF1-3 and TRAF7 have been identified in oyster (L. Wang et al., 2018). In response to OsHV-1 virus, the cgTRAF2 and cgTRAF3 were shown to response to *Vibrio alginolyticus* or OsHV-1 challenges (B. Huang et al., 2014, 2016; X. De Huang et al., 2012). More recently, in orange-spotted grouper (*Epinephelus coioides*), a Poly (I:C)

and RGNNV (red-spotted grouper nervous necrosis virus) stimulation was shown to increase the expression level of ecTRAF3 (Wu et al., 2022). This study showed that the overexpression of ecTRAF3 significantly induced NF-κB activity. Similarly, the association of TRAF3 and NF-κB activation have been observed, for example in black carp, where the authors found that TRAF3 could activate NF-κB (Xu Wang et al., 2018). Controversially, the overexpression of TRAF3 have been associated with the suppression of the NF-KB pathway (Cai et al., 2015). Interestingly, in Pacific oyster, TRAF were overexpressed in susceptible oyster compared to resistant oyster (de Lorgeril et al., 2020), suggesting that the overexpression of TRAF3 could negatively affect the NF-κB activity.

iv- N-acetylmuramoyl-L-alanine amidase (or Peptidoglycan Recognition Protein; PGRP; CGI_10001975), the SNP was located in the intron part of this gene. PGRPs are conserved PRRs that recognise the peptidoglycans present in the cell wall of the bacteria and then activate the prophenoloxidase cascade (Royet et al., 2011). PGRPs have been found to be upregulated in the response to *Marinococcus halophilus* and *Vibrio tubiashii* (Itoh & Takahashi, 2009). Additionally, the PGRPs are expressed at the highest levels in the digestive gland of *C. gigas*, suggesting that this organ could act as first-line of defence against pathogens propagation (X. Guo & Ford, 2016; G. Zhang et al., 2012). However, during the POMS event, PGRPs have not been reported, but we could suggest that it play a role in the elimination of pathogens during the later stage of POMS.

In conclusion, the genetic variation within the TLR/NF-κB pathway were located within the intron part, which could indicate a change in the gene expression and/or the regulatory functions. MyD88, TBK1 and TRAF3 were all related to the IRF3 and the NF-κB factors activation. These two factors are vital for an efficient antiviral response. These two factors are the main transcription factors that induce type I interferons (IFN) and other inflammatory and antimicrobial molecules as well as interferon-stimulated genes (ISGs) (Schoggins & Rice, 2011).

These genes identified by GWAS along genes previously identified by other studies can be used as target for future studies investigating the causative genes for disease resistance and to better understand the resistance mechanism. These results confirm the previous conclusion

stating that POMS resistance is polygenic and imply that oyster could be resistant by different ways. In addition, it highlights the potential gene candidates for further validation and functional studies.

Epigenetic markers of resistance involved in TLR/NF-κB pathway

In this thesis, we hypothesised that epigenetic variation (DNA methylation) could be implicated in the resistance of oyster to POMS diseases. To answer this hypothesis, we used Epigenome wide association study (EWAS) that lead to the identification of 305 CpGs significantly associated to POMS resistance. In total 171 genes displayed at least one differentially methylated CpG. From these genes, 95% (164 genes) had at least one CpGs hypomethylated in resistant oysters comparing to susceptible oysters. Interestingly, some of the genes identified here, have been previously identified to be associated to POMS through gene expression analysis. Several of these genes were involved in antiviral pathways including the TLR/NF-κB signaling pathway. As previously mentioned, the TLR/NF-κB is a crucial signalling pathway involved in innate immune system. Several CpGs identified were located in genes of the backbone of the TLR/NF-κB pathway (Figure 2):

i- Toll-like receptor 4, also named here "Protein with epidermal growth factor and TIR domains (EGF-TIR)" (TIRprot; CGI_10003134), is a gene that had one CpG hypomethylated in the resistant oysters compared to the susceptible. TLRs are type I transmembrane proteins, which contain an amino-terminal leucine-rich repeat (LRR) domain and a carboxyl-terminal Toll-interleukin-1 receptor (TIR) domain (Takeda et al., 2003). In a previous study focusing on the basal transcriptomic level of oyster families resistant to POMS a Toll-like receptor 13 (TLR13), was shown to be over expressed in the resistant oysters.

ii- NF-kappa B (NF-κB; CGI_10021567), a gene with one CpG associated to POMS that was hypomethylated in the resistant oysters comparing to susceptible. This gene encode a transcription factor involved in the expression of many innate immune genes and is one of the main gene of the TLR/NF-κB pathways (Montagnani et al., 2004). As mentioned in previous section, NF-κB is certainly one of the main transcription factors activating the expression of type-I interferons (IFN), the most effective antiviral immune responses (Abrahao et al., 2019). The expression of IFN are needed for the

expression of interferon-stimulated genes (ISGs that includes antiviral effectors) through the JAK/STAT pathways.

iii- Interferon regulatory factor 1 (IRF; CGI_10021170), a gene that had one CpGs associated to POMS. The resistant oyster were hypermethylated in this gene comparing to susceptible oysters. This is another transcription factor involved in the TLR/NF-κB and RLR/STING pathways for regulating the expression of IFN and ISGs (Honda et al., 2006). This gene contribute to antiviral immune response *in C. gigas* by functioning as an activator of IFN expression (Lu et al., 2018). This gene was one of the genes with a higher and earlier expression characterizing the response of resistant oysters facing POMS (early 6 and 12 hours post infection; de Lorgeril *et al.*, 2018). IRF regulate the early control of viral replication by regulating the expression of IFN and ISGs. Interestingly, this gene was one of the few gene that was hypermethylated in resistant oysters compared to susceptible oysters.

iv- A-kinase anchor 13 (AKAP13; CGI_10007280), a gene with two CpGs associated to POMS resistance, showed hypomethylation in resistant oyster compared to susceptible ones. this gene is a members of the guanine exchange factor (GEF) family, which acts as a scaffolder protein that is associated with TLR2-mediated NF-κB activation (Shibolet et al., 2007; Xiaojun Zhang et al., 2019).

v- Homeodomain-interacting kinase 2 (HIPK2; CGI_10006263), this gene contained one CpG, which was hypomethylated in the susceptible oysters by comparison to susceptible oysters. This gene have been associated to the inhibition NF-κB activity (Y. Feng et al., 2017).

vi- Diacylglycerol kinase zeta (DGKz; CGI_10008117), a gene with three hypomethylated CpGs in resistant oyster compared to susceptible ones. This gene have been associated to NF-κB activation, the downregulation of DGKz results in a faster phosphorylation of the p65 subunit and to its nuclear translocation (Tanaka et al., 2016).

Importin subunit beta 1 (IMPβ1; CGI_10019491), a gene with one CpG that was hypomethylated in the resistant oysters. This gene has been implemented in the activation of the transcription factor NF-κB, IRF3 and STAT. The knockdown of the IMPβ1 activity hindered IRF3 and reduced NF-κB p65 translocation (Gagné et al., 2017). IMPβ1T is known to be targeted by the hepatitis C virus NS3/4A protein which

restricts the IRF3 and NF-κB activation and suppress the interferon-β induction (Gagné et al., 2017). In Pacific oyster, this gene was upregulated 12 hours post POMS infection in susceptible oysters, this could suggest that IMPβ1 could be involved in the induction of antiviral response to OshV-1 (de Lorgeril et al., 2018).

In conclusion, this is the first study to use EWAS to associate DNA methylation variation to POMS, which open the door for future studies in other marine species. We identified a group of genes that are involved in antiviral response. Most of the genes identified were implicated in the TLR/NF-κB pathway.

Marker-assisted selection and genomic selection

One of the main potential applications from this thesis is to build further knowledge about the genes implicated in the resistance to POMS, especially through the identification of markers of resistance at the genetic and epigenetic levels. In total, 186 SNPs and 305 CpGs were suggestively or significantly associated to POMS resistance. These markers provide a resource for future Marker-Assisted Selection (MAS) and potential source for genomic selection (GS) by integrating these markers with other traits markers of interest. While MAS relies on selection of small number of markers associated to the trait of interest, GS relies on selection of genome-wide markers associated to trait of interest.

Selective breeding provides an interesting avenue to enhance the quality of aquaculture species by improving traits of economic importance (e.g. disease resistance or growth rate). In addition, it represent a useful tool to manage and control disease in farming areas that generally localized in the wild open sea environment (Stear et al., 2001). Human selected animal and plants displaying traits of high economic importance (Gjedrem, 1983). The selection was either, unconscious by domestication, or intentional using various techniques from breeding programs, polyploidisation and genome editing tools (Dégremont, Garcia, et al., 2015). Selective breeding programs of disease resistant livestock (mainly vertebrates) started only last few decades and rapidly gained significant improvement. In marine aquaculture, selective breeding programmes are less advanced because aquaculture is more recent and convey the image of “wild” animals that many farmers want to keep.

MAS provides several advantages over the classical selective breeding programs. Although, the selective breeding offers a great deal of improving the trait of interest, two main post selection limitation can arise (Sauvage et al., 2010). First, the selective breeding focuses on specific family (families) with potential limited genetic background (Dégremont et al., 2005). Second, it is absolutely necessary to better understand the cause of mortality to identify the underlying factors behind the mortality. The phenotyped oysters need to be closely monitored, which in return would allow to describe the physiological (from host and pathogen) and immunological status of the individuals (Sauvage et al., 2010). However, these limitations in selective breeding programs can be overcome by the application of genetic enhancement programs using MAS and GS.

MAS application have been started with the availability of QTL mapping and Genome wide-association study (GWAS) analyses. These two analysis helps in identifying the genetic variation associated with phenotypic variation. Once these analyses found the genetic regions of interest and that these regions are validated these lead to the identification of causal SNPs. These causal SNPs then pave the road for the use of genomic resources for trait improvement. In addition it shed light on the biological process and molecular mechanism involved in the expression of the phenotype (Abdelrahman et al., 2017).

Many successful examples of the use of MAS in aquaculture have been reported. One of the first example is about the Japanese flounder (*Paralichthys olivaceus*) for the selection of lymphocystis disease resistant individual (Fuji et al., 2007). In this study, authors found a single major QTL using microsatellite data. This QTL called "Poli9-8TUF" had a dominant effect for resistance to lymphocystis disease. Based on this QTL, a new population was developed using the linkage information. This new population was reported to be fully resistant to lymphocystis disease, while the control group showed a diseases incidence of 4-6%. Another successful example was obtained in the Atlantic salmon (*Salmo salar*) in response to the resistance to Infectious pancreatic necrosis virus (IPNV) disease. The study found a major QTL affecting the resistance to IPNV. After applying MAS, the IPN-resistant salmon were produced with a 75% decrease in the number of IPN outbreaks (Moen et al., 2009, 2015).

However, the use of MAS is still in its infancy in the case of Pacific oyster. The reason is most probably due to the low number of GWAS and QTL mapping studies or to the level of confidence or precision achieved by the markers already identified. In oyster, this field of research as started in 2010 where QTL mapping approach has identified several QTLs associated to POMS (Sauvage et al., 2010). These results showed variations between the families and the QTLs had only a moderate accuracy rates due to the use of a low number of markers (Sauvage et al., 2010). In the case of GWAS, several significant and suggestive SNPs were identified and located in or near several genes but the function of these genes was not perfectly understood (Gutierrez et al., 2018, 2020).

The nature of many traits are controlled by a wide range of loci (polygenic) each of them wearing a small effect. Additionally, in some traits, a QTL could be identified based on a selected family (or families), and this QTL can be absent from others. It is therefore necessary

to verify the presence of such QTL in a wide range of families and more particularly in natural populations.

Therefore, the use of the MAS is limited to the QTL with the moderate to large effect, which is not the case for the traits with polygenic nature. GWAS is powerful in detecting DNA variation associated with polygenic traits that have very small effects. GWAS is further integrated with the GS that search at the whole-genome large sets of SNPs to estimate the effects of genetic variation to the trait of interest (Meuwissen et al., 2001). Thus, GS is highly advantageous for traits that are polygenic (Gutierrez et al., 2020). Unlike the GWAS, GS neglects the significance and focus on the estimation of the effect bring by the marker by the use of prediction methods. These methods include Genomic Best Linear Unbiased Prediction (GBLUP) and Bayesian estimates (BE). GBLUP assumes an equal weight of all the markers and BE assumes that only a group of markers have a non-zero effect (Daetwyler et al., 2010; Hayes et al., 2009).

The Application of GS has been mostly successfully used in livestock animals including the dairy cow and cattle beef (Hayes et al., 2009). In aquaculture, the use of GS has been done in rainbow trout for selection of resistance to bacterial cold water disease (Vallejo et al., 2016). Additionally, GS have been used to estimate the breeding values in several aquaculture species including Atlantic Salmon (Ødegård et al., 2014) and Pacific oyster (Gutierrez et al., 2020).

Additional to genetic markers, DNA methylation markers can be integrated in genomic selection as diagnostic prognostic markers of traits. Although, the changes on DNA methylation that are acquired during the life are revisable, they can be relatively stable and passed to next generations (Bishop & Ferguson, 2015; Granada et al., 2018). Recently, it is more commonly accepted along with the newly acquired empirical data that DNA methylation induced by an environmental stimuli can mediate phenotypic changes (Granada et al., 2018). With the increasing numbers of approaches used for detecting DNA methylation variation, the use of such variation is advancing in different applications. DNA methylation markers are widely used in clinical epigenetic field of research, which are promising in human disease diagnoses applications (Berdasco & Esteller, 2019; How Kit et al., 2012; P. A. Jones et al., 2016). Although, DNA methylation marker application in aquaculture are still in its infancy. The majority of studies are based on the association of DNA methylation with changes in the

environment. In addition, DNA methylation markers have been used in some cases to estimate age, for example in fish and lobsters (Anastasiadi & Piferrer, 2020; Fairfield et al., 2021). In oyster, in general, most of the studies were based on the correlation between gene expression and DNA methylation changes in response to environmental changes. This Includes the effects of a parental exposure to diuron (Akcha et al., 2021; Rondon et al., 2017), ocean acidification (Chandra Rajan et al., 2021; Lim et al., 2021; Venkataraman et al., 2020); salinity (Johnson et al., 2021), heat stress (Roberto et al., 2021; Xinxing Wang et al., 2021) and rich microflora (Fallet et al., 2022). In the context of diseases, there are few studies focusing on the DNA methylation association to disease. For example, *Perkinsus marinus* infection and gene expression have been associated to DNA methylation in the Eastern Oyster *Crassostrea virginica* (Johnson et al., 2020). Although, the application of DNA methylation marker have not yet seen in oyster, the framework developed in human clinical filed could be applied in Pacific oyster. DNA methylation markers offer a great potential for diagnoses and prognosis of diseases exposure and disease resistance/susceptibility.

In conclusion, the use of MAS is still in its beginning steps in aquaculture comparing to other domestic animal and crop plants. In the case of Pacific oyster resistance to POMS, so far no MAS has been reported. The SNPs and CpGs markers identified here will collectively offer with the other dataset published (Gutierrez et al., 2018, 2020; Sauvage et al., 2010) a great potential for future MAS application in oysters farming. From the top SNPs and CpGs, we are able to differentiate susceptible oysters from the resistant one (Figure 4.4A-B). However, it will be necessary to validate these SNP and CpGs before their use, a question that will be addressed in the perspective section of this manuscript.

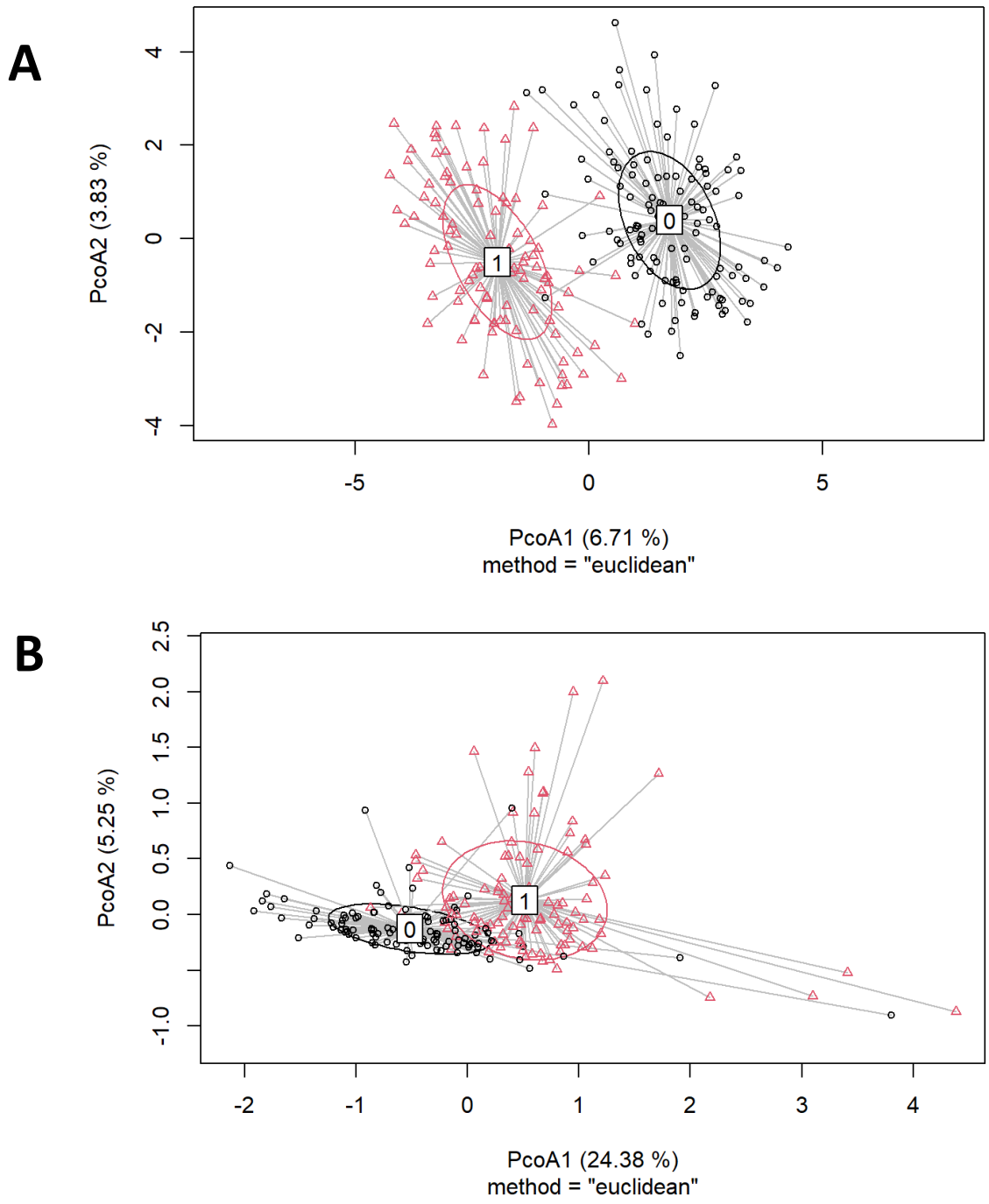


Figure 4.4: Principal Component Analysis (PCA).

Plot showing the variance between the resistant (0) and susceptible (1) oysters using a Euclidean distance-based test for homogeneity of multivariate dispersions implemented by the betadisper() function in the vegan package. A genetic variability is measured as the average distance of all oysters from the centroid of each group. B) The same as A but for epigenetic variation.

Conclusion

In this thesis, we hypothesized, that genotype and epigenotype of oyster can play a role in the resistance to POMS. We propose a framework to study simultaneously the potential role of genetic and epigenetic in shaping a phenotype by using the *C. gigas*/POMS model at the natural population level. We developed an exome capture approach to obtain genetic (SNP) and epigenetic (DNA methylation) information. We sampled natural oyster population with contrasted exposure to POMS. The "Seq Cap Epi Enrichment System" protocol allowed us to capture the genetic and epigenetic information of more than 65 % of exons. We saw that some regions were undersequenced, others oversequenced but the overall results were sufficient to provide new information. In addition, the deep characterization of genetic and epigenetic variations into functional parts of the genome for hundreds of individuals was obtained at a reasonable cost. Our study showed the importance of both genetic and epigenetic mechanisms in explaining the resistant phenotype. The results showed that natural oyster populations differentially exposed to POMS display signatures of selections both in their genome (SNPs) and epigenome (DNA methylation at CG context). These signatures were localized in different genes but most of them belong to immune related biological processes. Although, the genetic and epigenetic variations are partly correlated and while the former was associated with a large fraction of the second; we also showed that most of the epigenetic variation significantly associated to oyster resistance was independent and explained a higher part of the phenotypic variation. From one side, these results confirmed that host population facing infectious diseases emergence could rely on genetic and epigenetic variation to rapidly adapt to emerging diseases. On the other side, these results showed that using such integrative approaches (Genomic and epigenomic), has enabled us to have access to a larger portion of the mechanisms involved in the resistance. Thus, providing an inclusive understanding of oyster immunity from genotype and epigenotype to phenotype.

Perspective

Studies have shown that resistance mechanisms are complex and a wide spectrum of genes have been identified as good candidates to explain resistance to POMS (de Lorgeril et al., 2018). Therefore, it is difficult to distinguish which gene have a critical role to elucidate the resistance. Future studies would further characterize the role of the genes identified in this thesis and their downstream products (e.g. proteins, metabolites) *via* other omics approaches (e.g. proteomics and metabolomics). In addition, the use of invalidation methods are more and more needed to complete this characterization. Genome editing techniques such as the CRISPR/cas9 (clustered regularly interspaced short palindromic repeats (CRISPR) and the nuclease Cas9) or gene expression invalidation by RNAi (RNA interference) can be powerful to further study these genes. Additionally, with the dCas9 system (the updated version for DNA methylation editing) it is possible to target the epigenome alterations (Pulecio et al., 2017).

If this work provides new fundamental knowledge on oyster resistance to POMS it also provides promising outreach for oyster production through MAS and GS. However, several steps of validation are needed before the use MAS and GS. Machine Learning (ML) is very promising, since it can be used to better understand and model disease resistance and susceptibility at a multiomics level. ML method could be applied to further validate and predict the potential role of these SNPs and CpGs. The ML is composed of three different steps. First, the data is obtained and filtered to remove the noise. Second, the data is split into three datasets (training, testing and validating). In the training dataset, the model algorithm is optimised by training. In the testing dataset, the performance of the trained model algorithm is evaluated. In the validation dataset the model if further validated (Rauschert et al., 2020).

For example, the ML can be used to learn from the genetic and epigenetic data obtained in this thesis. Then, models can be build and would predict the phenotype (either binary or semi-quantitative), the model could then be used for future diagnose of phenotypes based on the signature of SNPs or CpGs or a combination of both of them.

Additionally, the CpGs markers need to be further validated at a tissue specific level, since they can be tissue specific. Unlike the genetic variation, which are the same in all the cells, the DNA methylation shows a cell or tissue specific patterns. In the Pacific oyster, studies have been implemented either using whole tissues or specific tissue (such as the gills, mantle or

gonad tissues). While there are differences in the oocytes and different embryonic developmental stages (Riviere et al., 2013), similarity of DNA methylation levels in male gametes and gills have been found (Olson & Roberts, 2014; Riviere et al., 2013). Additionally, difference in the global DNA methylation level in the whole oyster tissues have been observed in response to diuron exposure (Akcha et al., 2021).

Finally, one limitation in our study is the absence of transcriptomic data, which would further strengthen our results interpretation. It is known that DNA methylation variation could influence the phenotypic outcome (gene expression). Previous studies showed contrasting (positive or negative) association between DNA methylation and gene expression. In the study of (Riviere et al., 2013), the author showed that DNA methylation in proximal promoter regions and first exons were associated to a decreased gene expression. In contrast, in another study, a positive correlation between DNA methylation and gene expression was found (Olson & Roberts, 2014). In study of Wang *et al.* (2020) on the Pacific oyster subjected to heat-stress treatment, authors found a posistive correlation beetween the gene bodies DNA methylation level and gene expression which suggest a divergence in the phenotype that is facilitated by the DNA methylation.

While the transcriptomic data (RNA-seq) is important for further validating the results and association between the DNA methylation and gene expression, designing such study should be carefully considered. For example in the case of the POMS, resistant oyster displayed early transcription of antiviral genes during the first 12 hours post infection, which was absent in the susceptible oyster. (de Lorgeril et al., 2018). After 12 hours, the expression of the same genes were more similar and even stronger in the susceptible oyster, unfortunately this was too late for the oyster to induce an efficient response. Therefore, the design of RNA-seq should be a time specific during the POMS infection. Such relationship between DNA methylation and gene expression should be further investigated. For example, exome capture (or WGBS) along RNA-seq (messenger RNA sequencing for gene expression) and proteomics data could be used for time specific of pre and post infection samples of oyster.

References

Abbadi M, Zamperin G, Gastaldelli M, Pascoli F, Rosani U, Milani A, *et al.* (2018). Identification of a newly described OsHV-1 µvar from the North Adriatic sea (Italy). *J Gen Virol* **99**: 693–703.

Abdelrahman H, ElHady M, Alcivar-Warren A, Allen S, Al-Tobasei R, Bao L, *et al.* (2017). Aquaculture genomics, genetics and breeding in the United States: Current status, challenges, and priorities for future research. *BMC Genomics* **18**: 1–23.

Abrahao J, Elias Marques R, Maximina Moreno-Altamirano MB, Javier Sánchez-García F, Kolstoe SE (2019). Virus Control of Cell Metabolism for Replication and Evasion of Host Immune Responses. *Front Cell Infect Microbiol | www.frontiersin.org* **1**: 95.

Adorisio S, Fierabracci A, Muscari I, Liberati AM, Ayroldi E, Migliorati G, *et al.* (2017). SUMO proteins: Guardians of immune system. *J Autoimmun* **84**: 21–28.

Akcha F, Barranger A, Bachère E (2021). Genotoxic and epigenetic effects of diuron in the Pacific oyster: in vitro evidence of interaction between DNA damage and DNA methylation. *Environ Sci Pollut Res* **28**: 8266–8280.

Allam B, Raftos D (2015). Immune responses to infectious diseases in bivalves. *J Invertebr Pathol* **131**: 121–136.

Allum F, Shao X, Guénard F, Simon MM, Busche S, Caron M, *et al.* (2015). Characterization of functional methylomes by next-generation capture sequencing identifies novel disease-associated variants. *Nat Commun* **6**.

Anastasiadi D, Piferrer F (2020). A clockwork fish: Age prediction using DNA methylation-based biomarkers in the European seabass. *Mol Ecol Resour* **20**: 387–397.

Ayres JS, Schneider DS (2009). The role of anorexia in resistance and tolerance to infections in Drosophila. *PLoS Biol* **7**.

Azéma P, Lamy JB, Boudry P, Renault T, Travers MA, Dégremont L (2017). Genetic parameters of resistance to Vibrio aestuarianus, and OsHV-1 infections in the Pacific oyster, *Crassostrea gigas*, at three different life stages. *Genet Sel Evol* **49**: 1–16.

Bachère E, Gueguen Y, Gonzalez M, De Lorgeril J, Garnier J, Romestand B (2004). Insights into the anti-microbial defense of marine invertebrates: The penaeid shrimps and the oyster *Crassostrea gigas*. *Immunol Rev* **198**: 149–168.

Bachère E, Rosa RD, Schmitt P, Poirier AC, Merou N, Charrière GM, *et al.* (2015). The new insights into the oyster antimicrobial defense: Cellular, molecular and genetic view. *Fish Shellfish Immunol* **415**: 50–64.

Badariotti F, Lelong C, Dubos MP, Favrel P (2007). Characterization of chitinase-like proteins (Cg-Clp1 and Cg-Clp2) involved in immune defence of the mollusc *Crassostrea gigas*. *FEBS J* **274**: 3646–3654.

Baggett LP, Powers SP, Brumbaugh R, Coen LD, DeAngelis B, Greene J, *et al.* (2014). Oyster habitat restoration monitoring and assessment handbook. *Nat Conserv*: 96pp.

Bagusche F (2013). Environmental effects on the physology of calcification in the Pacific oyster *Crassostrea gigas* Thunberg, 1793.

Bai CM, Morga B, Rosani U, Shi J, Li C, Xin LS, *et al.* (2019). Long-range PCR and high-throughput sequencing of Ostreid herpesvirus 1 indicate high genetic diversity and complex evolution process. *Virology* **526**: 81–90.

Bannister AJ, Kouzarides T (2011). Regulation of chromatin by histone modifications. *Nat Publ Gr* **21**: 381–395.

Banta JA, Richards CL (2018). Quantitative epigenetics and evolution. *Heredity (Edinb)* **121**:

210–224.
Barturen G, Rueda A, Oliver JL, Hackenberg M (2013). MethylExtract: High-Quality methylation maps and SNV calling from whole genome bisulfite sequencing data. *F1000Research* **2**: 217.
Bayne BL, Ahrens M, Allen SK, D'Auriac MA, Backeljau T, Beninger P, *et al.* (2017). The proposed dropping of the genus crassostrea for all pacific cupped oysters and its replacement by a new genus magallana: A dissenting view. *J Shellfish Res* **36**: 545–547.
Bédier E, D'Amico F, Annezo J-P, Auby I, Barret J, Bouget J-F, *et al.* (2009). *Observatoire national Conchylicole. Rapport 2009*.
Begitt A, Droescher M, Knobeloch KP, Vinkemeier U (2011). SUMO conjugation of STAT1 protects cells from hyperresponsiveness to IFNγ. *Blood* **118**: 1002–1007.
Ben-Haim Y, Zicherman-Keren M, Rosenberg E (2003). Temperature-regulated bleaching and lysis of the coral Pocillopora damicomis by the novel pathogen Vibrio coralliilyticus. *Appl Environ Microbiol* **69**: 4236–4242.
Bentley RA, Hahn MW, Shennan SJ (2004). Random drift and culture change. : 1443–1450.
Berdasco M, Esteller M (2019). Clinical epigenetics: seizing opportunities for translation. *Nat Rev Genet* **20**: 109–127.
Berger SL (2007). The complex language of chromatin regulation during transcription. **447**: 407–412.
Bird A (2007). Perceptions of epigenetics. *Nature* **447**: 396–398.
Bishop KS, Ferguson LR (2015). The interaction between epigenetics, nutrition and the development of cancer. *Nutrients* **7**: 922–947.
Boggio R, Passafaro A, Chiocca S (2007). Targeting SUMO E1 to ubiquitin ligases: A viral strategy to counteract sumoylation. *J Biol Chem* **282**: 15376–15382.
Bonduriansky R (2012). Rethinking heredity, again. *Trends Ecol Evol* **27**: 330–336.
Bonduriansky R, Day T (2009). Nongenetic Inheritance and Its Evolutionary Implications. *Annu Rev Ecol Evol Syst* **40**: 103–125.
Bossdorf O, Richards CL, Pigliucci M (2008). Epigenetics for ecologists. *Ecol Lett* **11**: 106–115.
Bougrier S, Raguenes G, Bachère E, Tigé G, Grizel H (1986). Essai de réimplantation de Crasscstrea angulata en France. Résistance au chambrage et comportement des hybrides C. angulata - *C. gigas*. *ICES, Copenhagen*.
Bouilly K, Chaves R, Fernandes M, Guedes-Pinto H (2010). Histone H3 gene in the Pacific oyster, *Crassostrea gigas* Thunberg, 1793: Molecular and cytogenetic characterisations. *Comp Cytogenet* **4**: 111–121.
Bowman JC (1972). GENOTYPE x ENVIRONMENT INTERACTIONS. *Ann Génétique Sélection Anim* **4**: 117–123.
Brogden KA (2005). Antimicrobial peptides: Pore formers or metabolic inhibitors in bacteria? *Nat Rev Microbiol* **3**: 238–250.
Brooke CB (2017). Population Diversity and Collective Interactions during Influenza Virus Infection. *J Virol* **91**.
Bruto M, James A, Petton B, Labreuche Y, Chenivesse S, Alunno-Bruscia M, *et al.* (2017). Vibrio crassostreae, a benign oyster colonizer turned into a pathogen after plasmid acquisition. *ISME J* **11**: 1043–1052.
Bruto M, Labreuche Y, James A, Piel D, Chenivesse S, Petton B, *et al.* (2018). Ancestral gene acquisition as the key to virulence potential in environmental Vibrio populations. *ISME J*: 2954–2966.
Buestel D, Ropert M, Prou J, Goulletquer P (2009). History, status, and future of oyster culture

in France. *J Shellfish Res* **28**: 813–820.
Bulet P, Stöcklin R, Menin L (2004). Anti-microbial peptides: From invertebrates to vertebrates. *Immunol Rev* **198**: 169–184.
Burge CA, Griffin FJ, Friedman CS (2006). Mortality and herpesvirus infections of the Pacific oyster *Crassostrea gigas* in Tomales Bay, California, USA. *Dis Aquat Organ* **72**: 31–43.
Burge CA, Judah LR, Conquest LL, Griffin FJ, Cheney DP, Suhrbier A, *et al.* (2007). Summer seed: the influence ofoyster stock, planting time, pathogens and environmental stressors. *J Shellfish Res* **26**: 163–172.
Burioli EAV, Prearo M, Houssin M (2017). Complete genome sequence of Ostreid herpesvirus type 1 µVar isolated during mortality events in the Pacific oyster *Crassostrea gigas* in France and Ireland. *Virology* **509**: 239–251.
Burioli EAV, Varello K, Lavazza A, Bozzetta E, Prearo M, Houssin M (2018). A novel divergent group of Ostreid herpesvirus 1 µVar variants associated with a mortality event in Pacific oyster spat in Normandy (France) in 2016. *J Fish Dis* **41**: 1759–1769.
Burkhardt RW (2013). Lamarck, evolution, and the inheritance of acquired characters. *Genetics* **194**: 793–805.
Cai J, Xia H, Huang Y, Tang J, Jian J, Wu Z, *et al.* (2015). Identification and characterization of tumor necrosis factor receptor (TNFR)-associated factor 3 from humphead snapper, Lutjanus sanguineus. *Fish Shellfish Immunol* **46**: 243–251.
Camara MD, Yen S, Kaspar HF, Kesarcodi-Watson A, King N, Jeffs AG, *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.
Cambi A, Koopman M, Figdor CG (2005). How C-type lectins detect pathogens. *Cell Microbiol* **7**: 481–488.
Canesi L, Gallo G, Gavioli M, Pruzzo C (2002). Bacteria-hemocyte interactions and phagocytosis in marine bivalves. *Microsc Res Tech* **57**: 469–476.
Cazaly E, Saad J, Wang W, Heckman C, Ollikainen M, Tang J (2019). Making sense of the epigenome using data integration approaches. *Front Pharmacol* **10**: 1–15.
Chandra Rajan K, Meng Y, Yu Z, Roberts SB, Vengatesen T (2021). Oyster biomineralization under ocean acidification: From genes to shell. *Glob Chang Biol* **27**: 3779–3797.
Cheng X, Xiong R, Li Y, Li F, Zhou X, Wang A (2017). Sumoylation of turnip mosaic virus RNA polymerase promotes viral infection by counteracting the host NPR1-mediated immune response. *Plant Cell* **29**: 508–525.
CIESM (2000). Ostreidae oysters - *Crassostrea gigas* (Thunberg 1793). https://www.ciesm.org/atlas/Crassostreagigas.html.
Civitello DJ, Fatima H, Johnson LR, Nisbet RM, Rohr JR (2018). Bioenergetic theory predicts infection dynamics of human schistosomes in intermediate host snails across ecological gradients. *Ecol Lett* **21**: 692–701.
Clerissi C, de Lorgeril J, Petton B, Lucasson A, Escoubas JM, Gueguen Y, *et al.* (2020). Microbiota Composition and Evenness Predict Survival Rate of Oysters Confronted to Pacific Oyster Mortality Syndrome. *Front Microbiol* **11**: 1–11.
Comps M (1969). Observations relatives à l'affection branchiale des huîtres portuguaises (*Crassostrea angulata* Lmk). *Rev des Trav Inst des Pêches Marit* **33**: 151–160.
Comps M (1983). Recherches histologiques et cytologiques sur les infections intracellulaires des mollusques bivalves marins. Thèse Doct. Etat Sci. Nat., Montpellier. 128 pp.
COMPS M (1970). La maladie des branchies chez les huitres du genre Crassostrea caracteristiques et evolution des alterations processus de cicatrisation. *Rev des Trav*

l'Institut des Pêches Marit **34**: 23–44.
Comps M, BONAMI J-R, VAGO C, CAMPILLO A (1976). Une virose de l'huître portugaise (Crassostea angulata LMK). *Comptes Rendus Académie des Sci Paris D 282 1991-1993* **282**: 139–142.
Conrath U, Beckers GJM, Langenbach CJG, Jaskiewicz MR (2015). Priming for Enhanced Defense. *Annu Rev Phytopathol* **53**: 97–119.
Corbeil S, Faury N, Segarra A, Renault T (2015). Development of an in situ hybridization assay for the detection of ostreid herpesvirus type 1 mRNAs in the Pacific oyster, *Crassostrea gigas*. *J Virol Methods* **211**: 43–50.
Cosseau C, Wolkenhauer O, Padalino G, Geyer KK, Hoffmann KF, Grunau C (2017). (Epi)genetic Inheritance in Schistosoma mansoni: A Systems Approach. *Trends Parasitol* **33**: 285–294.
Crick FH, Barnett L, Brenner S, Watts-Tobin RJ (1961). General Nature of the Genetic Code for Proteins. *Nat Rev Immunol* **192**: 1227–1232.
Daetwyler HD, Pong-Wong R, Villanueva B, Woolliams JA (2010). The impact of genetic architecture on genome-wide evaluation methods. *Genetics* **185**: 1021–1031.
Danchin É (2013). Avatars of information: Towards an inclusive evolutionary synthesis. *Trends Ecol Evol* **28**: 351–358.
Danchin É, Charmantier A, Champagne FA, Mesoudi A, Pujol B, Blanchet S (2011). Beyond DNA: Integrating inclusive inheritance into an extended theory of evolution. *Nat Rev Genet* **12**: 475–486.
Danchin É, Pocheville A, Rey O, Pujol B, Blanchet S (2019). Epigenetically facilitated mutational assimilation: epigenetics as a hub within the inclusive evolutionary synthesis. *Biol Rev* **94**: 259–282.
Danchin É, Wagner RH (2010). Inclusive heritability: Combining genetic and non-genetic information to study animal behavior and culture. *Oikos* **119**: 210–218.
Davison AJ, Trus BL, Cheng N, Steven A, Watson MS, Cunningham C, *et al.* (2005). A novel class of herpesvirus with bivalve hosts. *J Gen Virol* **86**: 41–53.
Dégremont L (2011). Evidence of herpesvirus (OsHV-1) resistance in juvenile *Crassostrea gigas* selected for high resistance to the summer mortality phenomenon. *Aquaculture* **317**: 94–98.
Dégremont L (2013a). Size and genotype affect resistance to mortality caused by OsHV-1 in *Crassostrea gigas*. *Aquaculture* **416**–**417**: 129–134.
Dégremont L (2013b). Size and genotype affect resistance to mortality caused by OsHV-1 in *Crassostrea gigas*. *Aquaculture* **416**–**417**: 129–134.
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, Garcia C, Allen SK (2015). Genetic improvement for disease resistance in oysters: A review. *J Invertebr Pathol* **131**: 226–241.
Dégremont L, Lamy JB, Pépin JF, Travers MA, Renault T (2015). New insight for the genetic evaluation of resistance to ostreid herpesvirus infection, a worldwide disease, in *Crassostrea gigas*. *PLoS One* **10**: 1–12.
Dégremont L, Nourry M, 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.
Delmotte J, Chaparro C, Galinier R, de Lorgeril J, Petton B, Stenger PL, *et al.* (2020). Contribution of Viral Genomic Diversity to Oyster Susceptibility in the Pacific Oyster

Mortality Syndrome. *Front Microbiol* **11**.
Diederich S, Nehls G, van Beusekom JE, Reise K (2005). Introduced Pacific oysters (*Crassostrea gigas*) in the northern Wadden Sea: Invasion accelerated by warm summers? *Helgol Mar Res* **59**: 97–106.
Dixon G, Liao Y, Bay LK, Matz M V (2018). Role of gene body methylation in acclimatization and adaptation in a basal metazoan. *Proc Natl Acad Sci U S A* **115**: 13342–13346.
Dolinoy DC, Weidman JR, Waterland RA, Jirtle RL (2006). Maternal genistein alters coat color and protects Avy mouse offspring from obesity by modifying the fetal epigenome. *Environ Health Perspect* **114**: 567–572.
Dugeny E, de Lorgeril J, Petton B, Toulza E, Gueguen Y, Pernet F (2022). Seaweeds influence oyster microbiota and disease susceptibility. *J Anim Ecol* **00**: 1–14.
Dunkelberger JR, Song WC (2010). Complement and its role in innate and adaptive immune responses. *Cell Res* **20**: 34–50.
Duperthuy M, Schmitt P, Garzón E, Caro A, Rosa RD, Le Roux F, *et al.* (2011). Use of OmpU porins for attachment and invasion of *Crassostrea gigas* immune cells by the oyster pathogen Vibrio splendidus. *Proc Natl Acad Sci U S A* **108**: 2993–2998.
ECOSCOPA (2021). https://wwz.ifremer.fr/observatoire_conchylicole/.
Edris A, den Dekker HT, Melén E, Lahousse L (2019). Epigenome-wide association studies in asthma: A systematic review. *Clin Exp Allergy* **49**: 953–968.
EFSA PAHW (2015). Oyster mortality. *FASEB J* **13**: 4u22–4n/au22.
Eirin-Lopez JM, Putnam HM (2019). Marine environmental epigenetics. *Ann Rev Mar Sci* **11**: 335–368.
Enríquez-Díaz M, Pouvreau S, Chávez-Villalba J, Le Pennec M (2009). Gametogenesis, reproductive investment, and spawning behavior of the Pacific giant oyster *Crassostrea gigas*: Evidence of an environment-dependent strategy. *Aquac Int* **17**: 491–506.
Escoubas JM, Briant L, Montagnani C, Hez S, Devaux C, Roch P (1999). Oyster IKK-like protein shares structural and functional properties with its mammalian homologues. *FEBS Lett* **453**: 293–298.
Evans O, Hick P, Dhand N, Whittington R (2015). Transmission of Ostreid herpesvirus-1 in *Crassostrea gigas* by cohabitation: effects of food and number of infected donor oysters. *Aquac Environ Interact* **7**: 281–295.
Everett RD, Boutell C, Hale BG (2013). Interplay between viruses and host sumoylation pathways. *Nat Rev Microbiol* **11**: 400–411.
Fairfield EA, Richardson DS, Daniels CL, Butler CL, Bell E, Taylor MI (2021). Ageing European lobsters (Homarus gammarus) using DNA methylation of evolutionarily conserved ribosomal DNA. *Evol Appl* **14**: 2305–2318.
Fallet M, Luquet E, David P, Cosseau C (2020). Epigenetic inheritance and intergenerational effects in mollusks. *Gene* **729**: 144166.
Fallet M, Montagnani C, Petton B, Ird UBOC, Umr L, Chaparro C, *et al.* (2022). Early life microbial exposures shape the *Crassostrea gigas* immune system for lifelong and intergenerational disease protection. *BMC microbiome*.
Fan Y, Li X, Zhang L, Zong Z, Wang F, Huang J, *et al.* (2022). SUMOylation in Viral Replication and Antiviral Defense. *Adv Sci* **2104126**: 1–14.
FAO (2021). *Fisheries and Aquaculture Information and Statistics Branch - 07/09/2021 - http://www.fao.org/figis/servlet/TabSelector*.
Fellous A, Favrel P, Guo X, Riviere G (2014). The Jumonji gene family in *Crassostrea gigas* suggests evolutionary conservation of Jmj-C histone demethylases orthologues in the

oyster gametogenesis and development. *Gene* **538**: 164–175.
Fellous A, Favrel P, Riviere G (2015). Temperature influences histone methylation and mRNA expression of the Jmj-C histone-demethylase orthologues during the early development of the oyster *Crassostrea gigas*. *Mar Genomics* **19**: 23–30.
Fellous A, Lefranc L, Jouaux A, Goux D, Favrel P, Rivière G (2019). Histone Methylation Participates in Gene Expression Control during the Early Development of the Pacific Oyster *Crassostrea gigas*. *Genes (Basel)* **10**.
Feng SY (1988). Cellular defense mechanisms of oysters and mussels. *Am Fish Soc Spec Pub* **18**: 153–168.
Feng Y, Zhou L, Sun X, Li Q (2017). Homeodomain-interacting protein kinase 2 (HIPK2): A promising target for anti-cancer therapies. *Oncotarget* **8**: 20452–20461.
Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC (1998). Potent andspecific genetic interferenceby double-strandedRNA in Caenorhabditiselegans. *Nature* **1**: 7.
Fleury E, Barbier P, Petton B, Normand J, Thomas Y, Pouvreau S, *et al.* (2020). Latitudinal drivers of oyster mortality: deciphering host, pathogen and environmental risk factors. *Sci Rep* **10**.
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.
Fleury E, Moal J, Boulo V, Daniel JY, Mazurais D, Hénaut A, *et al.* (2010). Microarray-based identification of gonad transcripts differentially expressed between lines of pacific oyster selected to be resistant or susceptible to summer mortality. *Mar Biotechnol* **12**: 326–339.
Friedman CS, Estes RM, Stokes NA, Burge CA, Hargove JS, Barber BJ, *et al.* (2005). Herpes virus in juvenile Pacific oysters *Crassostrea gigas* from Tomales Bay, California, coincides with summer mortality episodes. *Dis Aquat Organ* **63**: 33–41.
Frommer M, Mcdonald LE, Millar DS, Collist CM, Wattt F, Griggt GW, *et al.* (1992). A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Genetics* **89**: 1827–1831.
Fuji K, Hasegawa O, Honda K, Kumasaka K, Sakamoto T, Okamoto N (2007). Marker-assisted breeding of a lymphocystis disease-resistant Japanese flounder (*Paralichthys olivaceus*). *Aquaculture* **272**: 291–295.
Gagnaire PA, Lamy JB, Cornette F, Heurtebise S, Dégremont L, Flahauw E, *et al.* (2018). Analysis of genome-wide differentiation between native and introduced populations of the cupped oysters *Crassostrea gigas* and *Crassostrea angulata*. *Genome Biol Evol* **10**: 2518–2534.
Gagné B, Tremblay N, Park AY, Baril M, Lamarre D (2017). Importin β1 targeting by hepatitis C virus NS3/4A protein restricts IRF3 and NF-κB signaling of IFNB1 antiviral response. *Traffic* **18**: 362–377.
Le Gall J-L, Raillard O (1988). Influence de la température sur la physiologie de l'huître *Crassostrea gigas*. *Oceanis* **14**: 603–608.
Gauthier P (1990). Does Weismann's Experiment Constitute a Refutation of the Lamarckian Hypothesis? *Bios* **61**: 6–8.
Gavery MR, Roberts SB (2010). DNA methylation patterns provide insight into epigenetic regulation in the Pacific oyster (*Crassostrea gigas*). *BMC Genomics* **11**.
Gavery MR, Roberts SB (2013). Predominant intragenic methylation is associated with gene expression characteristics in a bivalve mollusc. *PeerJ* **2013**: 1–15.

Gavery MR, Roberts SB (2017). Epigenetic considerations in aquaculture. *PeerJ* **2017**: e4147.
Gayon J (2016). De Mendel à l'épigénétique : histoire de la génétique. *Comptes Rendus - Biol* **339**: 225–230.
Geoghegan JL, Spencer HG (2013). Exploring epiallele stability in a population-epigenetic model. *Theor Popul Biol* **83**: 136–144.
Gerdol M, Venier P, Pallavicini A (2015). The genome of the Pacific oyster *Crassostrea gigas* brings new insights on the massive expansion of the C1q gene family in Bivalvia. *Dev Comp Immunol* **49**: 59–71.
Gjedrem T (1983). Genetic variation in quantitative traits and selective breeding in fish and shellfish. *Aquaculture* **33**: 51–72.
Gonzalez M, Romestand B, Fievet J, Huvet A, Lebart MC, Gueguen Y, *et al.* (2005). Evidence in oyster of a plasma extracellular superoxide dismutase which binds LPS. *Biochem Biophys Res Commun* **338**: 1089–1097.
Granada L, Lemos MFL, Cabral HN, Bossier P, Novais SC (2018). Epigenetics in aquaculture – the last frontier. *Rev Aquac* **10**: 994–1013.
Grant-Downtown RT, Dickinson HG (2005). Epigenetics and its Implications for Plant Biology . 1 . The Epigenetic Network in Plants. *Ann Bot* **96**: 1143–1164.
Green TJ, Helbig K, Speck P, Raftos DA (2016). Primed for success: Oyster parents treated with poly(I:C) produce offspring with enhanced protection against Ostreid herpesvirus type I infection. *Mol Immunol* **78**: 113–120.
Green TJ, Montagnani C (2013). Poly I: C induces a protective antiviral immune response in the Pacific oyster (*Crassostrea gigas*) against subsequent challenge with Ostreid herpesvirus (OsHV-1 µvar). *Fish Shellfish Immunol* **35**: 382–388.
Green TJ, Raftos D, Speck P, Montagnani C (2015). Antiviral immunity in marine molluscs. *J Gen Virol* **96**: 2471–2482.
Green TJ, Robinson N, Chataway T, Benkendorff K, Connor WO, Speck P (2014). Evidence that the major hemolymph protein of the Pacific oyster , *Crassostrea gigas* , has antiviral activity against herpesviruses. *Antiviral Res* **110**: 168–174.
Grizel H, Héral M (1991). Introduction into France of the Japanese oyster (*Crassostrea gigas*). **47**: 399–403.
Grönholm J, Vanhatupa S, Ungureanu D, Väliaho J, Laitinen T, Valjakka J, *et al.* (2012). Structure-function analysis indicates that sumoylation modulates DNA-binding activity of STAT1. *BMC Biochem* **13**: 1–12.
Guerrero-bosagna C, Settles M, Lucker B, Skinner MK (2010). Epigenetic Transgenerational Actions of Vinclozolin on Promoter Regions of the Sperm Epigenome. **5**.
Guo K, Eid SA, Elzinga SE, Pacut C, Feldman EL, Hur J (2020). Genome-wide profiling of DNA methylation and gene expression identifies candidate genes for human diabetic neuropathy. *Clin Epigenetics*: 1–16.
Guo X, Ford SE (2016). Infectious diseases of marine mollusks and host responses as revealed by genomic tools. *Philos Trans R Soc B Biol Sci* **371**.
Guo X, He Y, Zhang L, Lelong C, Jouaux A (2015). Immune and stress responses in oysters with insights on adaptation. *Fish Shellfish Immunol* **46**: 107–119.
Guo X, Hedgecock D, Hershberger WK, Cooper K, Allen SK (1998). Genetic determinants of protandric sex in the Pacific oyster, *Crassostrea gigas* Thunberg. *Evolution (N Y)* **52**: 394–402.
Gutierrez AP, Bean TP, Hooper C, 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, Genet* **8**: 1273–1280.
Gutierrez AP, Symonds J, King N, Steiner K, Bean TP, Houston RD (2020). Potential of genomic selection for improvement of resistance to ostreid herpesvirus in Pacific oyster (*Crassostrea gigas*). *Anim Genet* **51**: 249–257.
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, Genet* **7**: 2209–2218.
Ha M, Kim VN (2014). Regulation of microRNA biogenesis. *Nat Publ Gr* **15**: 509–524.
Haig D (2007). Weismann Rules! OK? Epigenetics and the Lamarckian temptation. *Biol Philos* **22**: 415–428.
Haldane JB (1946). The interaction of nature and nurture. *Ann Eugen* **13**: 197–205.
Hall SR, Simonis JL, Nisbet RM, Tessier AJ, Cáceres CE (2009). Resource ecology of virulence in a planktonic host-parasite system: An explanation using dynamic energy budgets. *Am Nat* **174**: 149–162.
Hanington PC, Zhang SM (2011). The primary role of fibrinogen-related proteins in invertebrates is defense, not coagulation. *J Innate Immun* **3**: 17–27.
Harvell CD, Mitchell C, Ward J, Altizer S, Dobson A, Ostfeld R, *et al.* (2002). Climate warming and disease risks for terrestrial and marine biota. *Science (80-)* **2158**: 2158–2162.
Hauser MT, Aufsatz W, Jonak C, Luschnig C (2011). Transgenerational epigenetic inheritance in plants. *Biochim Biophys Acta - Gene Regul Mech* **1809**: 459–468.
Hayes BJ, Bowman PJ, Chamberlain AC, Verbyla K, Goddard ME (2009). Accuracy of genomic breeding values in multi-breed dairy cattle populations. *Genet Sel Evol* **41**: 1–9.
He Y, Jouaux A, Ford SE, Lelong C, Sourdaine P, Mathieu M, *et al.* (2015). Transcriptome analysis reveals strong and complex antiviral response in a mollusc. *Fish Shellfish Immunol* **46**: 131–144.
Helanterä H, Uller T (2010). The Price Equation and Extended Inheritance. *Philos Theory Biol* **2**.
Héral M, Prou J, et Deslous-Paoli JM (1986). Influence des facteurs climatiques sur la production conchylicole du bassin de Marennes-Oléron. Haliotis. **15**: 193–207.
Hill B (2002). National and international impacts of white spot disease of shrimp. *Bull Eur Assoc Fish Pathol* **22**: 58–65.
Hine PM (1999). The inter-relationships of bivalve haemocytes. *Fish Shellfish Immunol* **9**: 367–385.
His E (1972). Premiers elements de comparaison entre l'huitre portugaise et l'huitre japonaise. *Sci Pech Bull Inst Pech marit* **219**: 1-9.
Hodges E, Xuan Z, Balija V, Kramer M, Molla MN, Smith SW, *et al.* (2007). Genome-wide in situ exon capture for selective resequencing. *Nat Genet* **39**: 1522–1527.
Holmskov U, Thiel S, Jensenius JC (2003). Collectins and ficolins: Humoral lectins of the innate immune defense. *Annu Rev Immunol* **21**: 547–578.
Honda K, Takaoka A, Taniguchi T (2006). Type I Inteferon Gene Induction by the Interferon Regulatory Factor Family of Transcription Factors. *Immunity* **25**: 349–360.
Horng T, Medzhitov R (2001). Drosophila MyD88 is an adapter in the Toll signaling pathway. *Proc Natl Acad Sci U S A* **98**: 12654–12658.
van den Hove DLA, Riemens RJM, Koulousakis P, Pishva E (2020). Epigenome-wide association studies in Alzheimer's disease; achievements and challenges. *Brain Pathol* **30**: 978–983.
How Kit A, Nielsen HM, Tost J (2012). DNA methylation based biomarkers: Practical considerations and applications. *Biochimie* **94**: 2314–2337.

Huang X De, Liu WG, Guan YY, Shi Y, Wang Q, Zhao M, *et al.* (2012). Molecular cloning and characterization of class I NF-κB transcription factor from pearl oyster (Pinctada fucata). *Fish Shellfish Immunol* **33**: 659–666.

Huang B, Sang X, Dong J, Li L, Wang X, Yang B, *et al.* (2021). Oyster TBK1/IKKε responds to bacterial and viral challenges and participates in the innate immune signaling. *Aquaculture* **534**: 736276.

Huang B, Zhang L, Du Y, Li L, Qu T, Meng J, *et al.* (2014). Alternative splicing and immune response of *Crassostrea gigas* tumor necrosis factor receptor-associated factor 3. *Mol Biol Rep* **41**: 6481–6491.

Huang B, Zhang L, Du Y, Li L, Tang X, Zhang G (2016). Molecular characterization and functional analysis of tumor necrosis factor receptor-associated factor 2 in the Pacific oyster. *Fish Shellfish Immunol* **48**: 12–19.

Huang B, Zhang L, Du Y, Xu F, Li L, Zhang G (2017). Characterization of the Mollusc RIG-I/MAVS Pathway Reveals an Archaic Antiviral Signalling Framework in Invertebrates. *Sci Rep* **7**: 1–13.

Huang B, Zhang L, Xu F, Tang X, Li L, Wang W, *et al.* (2019). Oyster Versatile IKKα/βs Are Involved in Toll-Like Receptor and RIG-I-Like Receptor Signaling for Innate Immune Response. *Front Immunol* **10**: 1826.

Huneman P, Whalsh DM (2017). *Challenging the Modern Synthesis. Adaptation, Development, and Inheritance. Oxford University Press, New York.*

Huvet A, Herpin A, Dégremont L, Labreuche Y, Samain JF, Cunningham C (2004). The identification of genes from the oyster *Crassostrea gigas* that are differentially expressed in progeny exhibiting opposed susceptibility to summer mortality. *Gene* **343**: 211–220.

Iiyama C, Yoneda F, Tsutsumi M, Tsutsui S, Nakamura O (2021). Mannose-binding C-type lectins as defense molecules on the body surface of the sea urchin Pseudocentrotus depressus. *Dev Comp Immunol* **116**: 103915.

Itoh N, Takahashi KG (2009). A novel peptidoglycan recognition protein containing a goose-type lysozyme domain from the Pacific oyster, *Crassostrea gigas*. *Mol Immunol* **46**: 1768–1774.

Ittiprasert W, Knight M (2012). Reversing the resistance phenotype of the biomphalaria glabrata snail host schistosoma mansoni infection by temperature modulation. *PLoS Pathog* **8**.

Jablonka E, Raz G (2009). Transgenerational Epigenetic Inheritance: Prevalence, Mechanisms, and Implications for the Study of Heredity and Evolution. *Q Rev Biol* **84**: 131–176.

Jablonka E, Noble D (2019). Systemic integration of different inheritance systems. *Curr Opin Syst Biol* **13**: 52–58.

Jeannin P, Jaillon S, Delneste Y (2008). Pattern recognition receptors in the immune response against dying cells. *Curr Opin Immunol* **20**: 530–537.

Jiang S, Jia Z, Zhang T, Wang L, Qiu L, Sun J, *et al.* (2016). Functional characterisation of phagocytes in the Pacific oyster *Crassostrea gigas*. *PeerJ*.

Johannes F, Colot V, Jansen RC (2008). Epigenome dynamics: A quantitative genetics perspective. *Nat Rev Genet* **9**: 883–890.

Johannsen W (2014). The genotype conception of heredity1. *Int J Epidemiol* **43**: 989–1000.

Johnson KM, Sirovy KA, Casas SM, La Peyre JF, Kelly MW (2020). Characterizing the Epigenetic and Transcriptomic Responses to *Perkinsus marinus* Infection in the Eastern Oyster Crassostrea virginica. *Front Mar Sci* **7**: 1–13.

Johnson KM, Sirovy KA, Kelly MW (2021). Differential DNA methylation across environments

has no effect on gene expression in the eastern oyster. *J Anim Ecol*: 1–13.
Jones PA (2012). Functions of DNA methylation: Islands, start sites, gene bodies and beyond. *Nat Rev Genet* **13**: 484–492.
Jones PA, Issa JJ, Baylin S (2016). Targeting the cancer epigenome for therapy. *Nat Publ Gr* **17**: 630–641.
Jouaux A, Lafont M, Blin J, Houssin M, Mathieu M, Lelong C (2013). Physiological change under OsHV-1 contamination in Pacific oyster *Crassostrea gigas* through massive mortality events on fields. : 1–14.
Kawai T, Akira S (2007). Signaling to NF-κB by Toll-like receptors. *Trends Mol Med* **13**: 460–469.
Kimes NE, Grim CJ, Johnson WR, Hasan NA, Tall BD, Kothary MH, *et al.* (2012). Temperature regulation of virulence factors in the pathogen Vibrio coralliilyticus. *ISME J* **6**: 835–846.
King W, Siboni N, Williams N, Kahlke T, Nguyen K, Jenkins C, *et al.* (2019). Variability in the Composition of Pacific Oyster Microbiomes Across Oyster Families Exhibiting Different Levels of Susceptibility to OsHV-1 µvar Disease. *Front Microbiol* **10**.
Kingsolver MB, Huang Z, Hardy RW (2013). Insect antiviral innate immunity: Pathways, effectors, and connections. *J Mol Biol* **425**: 4921–4936.
Klironomos FD, Berg J, Collins S (2013). How epigenetic mutations can affect genetic evolution: Model and mechanism. *BioEssays* **35**: 571–578.
Klosin A, Casas E, Hidalgo-Carcedo C, Vavouri T, Lehner B (2017). Transgenerational transmission of environmental information in C. elegans. *Science (80-)* **356**: 320–323.
Knecht AL, Truong L, Marvel SW, Reif DM, Garcia A, Lu C, *et al.* (2017). Transgenerational inheritance of neurobehavioral and physiological de fi cits from developmental exposure to benzo [a] pyrene in zebra fi sh. *Toxicol Appl Pharmacol* **329**: 148–157.
Kotenko S V., Gallagher G, Baurin V V., Lewis-Antes A, Shen M, Shah NK, *et al.* (2003). IFN-λs mediate antiviral protection through a distinct class II cytokine receptor complex. *Nat Immunol* **4**: 69–77.
Krajcsi P, Wold WSM (1998). Inhibition of tumor necrosis factor and interferon triggered responses by DNA viruses. *Semin Cell Dev Biol* **9**: 351–358.
Kronholm I, Collins S (2016). Epigenetic mutations can both help and hinder adaptive evolution. *Mol Ecol* **25**: 1856–1868.
Kumar H, Kawai T, Akira S (2009). Pathogen recognition in the innate immune response. *Biochem J* **420**: 1–16.
Kurdyukov S, Bullock M (2016). DNA Methylation Analysis: Choosing the Right Method. *Biology (Basel)* **5**: 3.
Lafont M (2017). Mécanismes et spécificité du priming immunitaire antiviral chez un Lophotrochozoaire, l'huître creuse *Crassostrea gigas* Soutenance.
Lafont M, Petton B, Vergnes A, Pauletto M, Segarra A, Gourbal B, *et al.* (2017). Long-lasting antiviral innate immune priming in the Lophotrochozoan Pacific oyster, *Crassostrea gigas*. *Sci Rep* **7**: 1–14.
Lafont M, Vergnes A, Vidal-Dupiol J, De Lorgeril J, Gueguen Y, Haffner P, *et al.* (2020). A sustained immune response supports long-term antiviral immune priming in the pacific oyster, *Crassostrea gigas*. *MBio* **11**.
Laine VN, Gossmann TI, Van Oers K, Visser ME, Groenen MAM (2019). Exploring the unmapped DNA and RNA reads in a songbird genome. *BMC Genomics* **20**: 1–12.
Laland K, Uller T, Feldman M, Sterelny K, Müller GB, Moczek A, *et al.* (2014). Does evolutionary theory need a rethink? - Point Yes, urgently. *Nature* **514**: 161–164.

Lamarck JBPA (1809). *Phylosophie Zoologique. Dentus, Paris*.

Lapègue 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.

Lapègue S, Heurtebise S, Cornette F, Guichoux E, Gagnaire PA (2020). Genetic characterization of cupped oyster resources in Europe using informative single nucleotide polymorphism (SNP) panels. *Genes (Basel)* **11**.

Lasa A, di Cesare A, Tassistro G, Borello A, Gualdi S, Furones D, *et al.* (2019). Dynamics of the Pacific oyster pathobiota during mortality episodes in Europe assessed by 16S rRNA gene profiling and a new target enrichment next-generation sequencing strategy. *Environ Microbiol* **21**: 4548–4562.

Lea AJ, Vilgalys TP, Durst PAP, Tung J (2017). Maximizing ecological and evolutionary insight in bisulfite sequencing data sets. *Nat Ecol Evol* **1**: 1074–1083.

Lee DD, Hochstetler A, Murphy C, Lowe CW, Schwarz MA (2019). A distinct transcriptional profile in response to endothelial monocyte activating polypeptide II is partially mediated by JAK-STAT3 in murine macrophages. *Am J Physiol - Cell Physiol* **317**: C449–C456.

Lemire A, Goudenège D, Versigny T, Petton B, Calteau A, Labreuche Y, *et al.* (2015). Populations, not clones, are the unit of vibrio pathogenesis in naturally infected oysters. *ISME J* **9**: 1523–1531.

Li Q, Eichten SR, Hermanson PJ, Zaunbrecher VM, Song J, Wendt J, *et al.* (2014). Genetic perturbation of the maize methylome. *Plant Cell* **26**: 4602–4616.

Li MMH, MacDonald MR, Rice CM (2015). To translate, or not to translate: Viral and host mRNA regulation by interferon-stimulated genes. *Trends Cell Biol* **25**: 320–329.

Li Y, Song X, Wang W, Wang L, Yi Q, Jiang S, *et al.* (2017). The hematopoiesis in gill and its role in the immune response of Pacific oyster *Crassostrea gigas* against secondary challenge with Vibrio splendidus. *Dev Comp Immunol* **71**: 59–69.

Li Q, Suzuki M, Wendt J, Patterson N, Eichten SR, Hermanson PJ, *et al.* (2015). Post-conversion targeted capture of modified cytosines in mammalian and plant genomes. *Nucleic Acids Res* **43**: 1–16.

Li H, Zhang H, Jiang S, Wang W, Xin L, Wang H, *et al.* (2015). A single-CRD C-type lectin from oyster *Crassostrea gigas* mediates immune recognition and pathogen elimination with a potential role in the activation of complement system. *Fish Shellfish Immunol* **44**: 566–575.

Lim YK, 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. *Mar Environ Res* **163**: 105214.

Lionel Degremont (2003). Genetic basis of summer mortality and relationship with growth in juvenile Pacific cupped oysters *Crassostrea gigas*. *Phd Thesis*: 1–335.

Lionel D, Guyader T, Tourbiez D, Pépin JF (2013). Is horizontal transmission of the Ostreid herpesvirus OsHV-1 in *Crassostrea gigas* affected by unselected or selected survival status in adults to juveniles? *Aquaculture* **408**–**409**: 51–57.

Liu L, Yang J, Qiu L, Wang L, Zhang H, Wang M, *et al.* (2011). A novel scavenger receptor-cysteine-rich (SRCR) domain containing scavenger receptor identified from mollusk mediated PAMP recognition and binding. *Dev Comp Immunol* **35**: 227–239.

Liu C, Zhang T, Wang L, Wang M, Wang W, Jia Z, *et al.* (2016). The modulation of extracellular superoxide dismutase in the specifically enhanced cellular immune response against secondary challenge of Vibrio splendidus in Pacific oyster (*Crassostrea gigas*). *Dev Comp*

Immunol **63**: 163–170.
Lochmiller RL, Deerenberg C (2000). Trade-offs in evolutionary immunology: Just what is the cost of immunity? *Oikos* **88**: 87–98.
Loker ES, Adema CM, Zhang SM, Kepler TB (2004). Invertebrate immune systems - Not homogeneous, not simple, not well understood. *Immunol Rev* **198**: 10–24.
Lokmer A, Goedknegt MA, Thieltges DW, Fiorentino D, Kuenzel S, Baines JF, *et al.* (2016). Spatial and temporal dynamics of pacific oyster hemolymph microbiota across multiple scales. *Front Microbiol* **7**: 1–18.
Lokmer A, Wegner KM (2015). Hemolymph microbiome of Pacific oysters in response to temperature, temperature stress and infection. *ISME J* **9**: 670–682.
de Lorgeril J, Lucasson A, Petton B, Toulza E, Montagnani C, Clerissi C, *et al.* (2018). Immune-suppression by OsHV-1 viral infection causes fatal bacteraemia in Pacific oysters. *Nat Commun* **9**: 4215.
de Lorgeril J, Petton B, Lucasson A, Perez V, Stenger PL, Dégremont L, *et al.* (2020). Differential basal expression of immune genes confers *Crassostrea gigas* resistance to Pacific oyster mortality syndrome. *BMC Genomics* **21**: 1–14.
Lork M, Lieber G, Hale BG (2021). Proteomic approaches to dissect host sumoylation during innate antiviral immune responses. *Viruses* **13**.
Lu M, Yang C, Li M, Yi Q, Lu G, Wu Y, *et al.* (2018). A conserved interferon regulation factor 1 (IRF-1) from Pacific oyster *Crassostrea gigas* functioned as an activator of IFN pathway. *Fish Shellfish Immunol* **76**: 68–77.
Lucasson A, Luo X, Mortaza S, de Lorgeril J, Toulza E, Petton B, *et al.* (2020). A core of functionally complementary bacteria colonizes oysters in Pacific Oyster Mortality Syndrome. *bioRxiv*.
Luviano N, Lopez M, Gawehns F, Chaparro C, Arimondo PB, Ivanovic S, *et al.* (2021). The methylome of Biomphalaria glabrata and other mollusks: enduring modification of epigenetic landscape and phenotypic traits by a new DNA methylation inhibitor. *Epigenetics and Chromatin* **14**: 1–25.
Lynch M (2010). Evolution of the mutation rate. *Trends Genet* **26**: 345–352.
Maher B (2008). The Case of the Missing Heritability. *Nature* **456**: 18–21.
Mameli M (2004). Nongenetic selection and nongenetic inheritance. *Br J Philos Sci* **55**: 35–71.
Mao ZQ, He R, Sun M, Qi Y, Huang YJ, Ruan Q (2007). The relationship between polymorphisms of HCMV UL144 ORF and clinical manifestations in 73 strains with congenital and/or perinatal HCMV infection. *Arch Virol* **152**: 115–124.
Martenot C, Gervais O, Chollet B, Houssin M, Renault T (2017). Haemocytes collected from experimentally infected Pacific oysters, *Crassostrea gigas*: Detection of ostreid herpesvirus 1 DNA, RNA, and proteins in relation with inhibition of apoptosis. *PLoS One* **12**: 1–19.
Martenot C, Oden E, Travaillé E, Malas JP, Houssin M (2011). Detection of different variants of Ostreid Herpesvirus 1 in the Pacific oyster, *Crassostrea gigas* between 2008 and 2010. *Virus Res* **160**: 25–31.
Mattoscio D, Segré C V., Chiocca S (2013). Viral manipulation of cellular protein conjugation pathways: The SUMO lesson. *World J Virol* **2**: 79.
Mayr E (1982). *The growth of biological thought diversity, evolution, and inheritance. Cambridge, Massachusetts : Belknap Press*.
Mayr E (1996). The modern evolutionary theory. *J Mammal* **77**: 1–7.
Meissner A, Gnirke A, Bell GW, Ramsahoye B, Lander ES, Jaenisch R (2005). Reduced

representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res* **33**: 5868–5877.

Meister G, Tuschl T (2004). Mechanisms of gene silencing by double-stranded RNA. *Nature* **431**: 343–349.

de Mendoza A, Lister R, Bogdanovic O (2020). Evolution of DNA Methylome Diversity in Eukaryotes. *J Mol Biol* **432**: 1687–1705.

Meuwissen THE, Hayes BJ, Goddard ME (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157**: 1819–1829.

Miller LK (1999). An exegesis of IAPs: Salvation and surprises from BIR motifs. *Trends Cell Biol* **9**: 323–328.

Milutinović B, Kurtz J (2016). Immune memory in invertebrates. *Semin Immunol* **28**: 328–342.

Miossec L, Le Deuff R-M, Goulletquer P (2009). Alien Species Alert: *Crassostrea gigas* (Pacific oyster). *ICES Coop Res Rep* **299**: 42.

Mizuta DD, Wikfors GH (2019). Seeking the perfect oyster shell: a brief review of current knowledge. *Rev Aquac* **11**: 586–602.

Moen T, Baranski M, Sonesson AK, Kjøglum S (2009). Confirmation and fine-mapping of a major QTL for resistance to infectious pancreatic necrosis in Atlantic salmon (Salmo salar): Population-level associations between markers and trait. *BMC Genomics* **10**: 1–14.

Moen T, Torgersen J, Santi N, Davidson WS, Baranski M, Ødegård J, *et al.* (2015). Epithelial cadherin determines resistance to infectious pancreatic necrosis virus in Atlantic salmon. *Genetics* **200**: 1313–1326.

Mogensen TH (2009). Pathogen recognition and inflammatory signaling in innate immune defenses. *Clin Microbiol Rev* **22**: 240–273.

Montagnani C, Kappler C, Reichhart JM, Escoubas JM (2004). Cg-Rel, the first Rel/NF-κB homolog characterized in a mollusk, the Pacific oyster *Crassostrea gigas*. *FEBS Lett* **561**: 75–82.

Montagnani C, Labreuche Y, Escoubas JM (2008). Cg-IκB, a new member of the IκB protein family characterized in the pacific oyster *Crassostrea gigas*. *Dev Comp Immunol* **32**: 182–190.

Moreau P, Moreau K, Segarra A, Tourbiez D, Travers MA, Rubinsztein DC, *et al.* (2015). Autophagy plays an important role in protecting Pacific oysters from OsHV-1 and Vibrio aestuarianus infections. *Autophagy* **11**: 516–526.

Morga B, Renault T, Faury N, Lerond S, Garcia C, Chollet B, *et al.* (2017). Contribution of in vivo experimental challenges to understanding flat oyster *Ostrea edulis* resistance to Bonamia ostreae. *Front Cell Infect Microbiol* **7**: 1–13.

Müller S, Dejean A (1999). Viral Immediate-Early Proteins Abrogate the Modification by SUMO-1 of PML and Sp100 Proteins, Correlating with Nuclear Body Disruption. *J Virol* **73**: 5137–5143.

Nair S, Bist P, Dikshit N, Krishnan MN (2016). Global functional profiling of human ubiquitome identifies E3 ubiquitin ligase DCST1 as a novel negative regulator of Type-I interferon signaling. *Sci Rep* **6**: 1–13.

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. *Rev Aquac* **13**: 2289–2304.

Netea MG, Domínguez-Andrés J, Barreiro LB, Chavakis T, Divangahi M, Fuchs E, *et al.* (2020). Defining trained immunity and its role in health and disease. *Nat Rev Immunol* **20**: 375–388.

Nevels M, Brune W, Shenk T (2004). SUMOylation of the Human Cytomegalovirus 72-Kilodalton IE1 Protein Facilitates Expression of the 86-Kilodalton IE2 Protein and Promotes Viral Replication. *J Virol* **78**: 7803–7812.

Nicolas J, Comps M, Cochennec N (1992). Herpes-like virus infecting Pacific-oyster larvae, *Crassostrea gigas*. *Bull Eur Assoc Fish Pathol* **12**: 11–13.

Nonaka M, Miyazawa S (2002). Evolution of the initiating enzymes of the complement system. *Genome Biol* **3**: 1–5.

Ødegård J, Moen T, Santi N, Korsvoll SA, Kjøglum S, Meuwisse THE (2014). Genomic prediction in an admixed population of Atlantic salmon (Salmo salar). *Front Genet* **5**: 1–8.

Olson CE, Roberts SB (2014). Genome-wide profiling of DNA methylation and gene expression in *Crassostrea gigas* male gametes. *Front Physiol* **5 JUN**: 1–7.

Oyanedel D, Labreuche Y, Bruto M, Amraoui H, Robino E, Haffner P, *et al.* (2020). Vibrio splendidus O-antigen structure: a trade-off between virulence to oysters and resistance to grazers. *Environ Microbiol* **22**: 4264–4278.

Pajares MJ, Palanca-Ballester C, Urtasun R, Alemany-Cosme E, Lahoz A, Sandoval J (2021). Methods for analysis of specific DNA methylation status. *Methods* **187**: 3–12.

Paul-Pont I, Dhand NK, Whittington RJ (2013). Influence of husbandry practices on OsHV-1 associated mortality of Pacific oysters *Crassostrea gigas*. *Aquaculture* **412–413**: 202–214.

Peeler EJ, Allan Reese R, Cheslett DL, Geoghegan F, Power A, Thrush MA (2012). Investigation of mortality in Pacific oysters associated with Ostreid herpesvirus-1??Var in the Republic of Ireland in 2009. *Prev Vet Med* **105**: 136–143.

Peñaloza C, Gutierrez AP, Eöry L, Wang S, Guo X, Archibald AL, *et al.* (2021). A chromosome-level genome assembly for the Pacific oyster *Crassostrea gigas*. *Gigascience* **10**: 1–9.

Pennisi E (2008). Modernizing the modern synthesis. *Science (80-)* **321**: 196–197.

Pepin JF, Riou A, Renault T (2008). Rapid and sensitive detection of ostreid herpesvirus 1 in oyster samples by real-time PCR. **149**: 269–276.

Perales C, Moreno E, Domingo E (2015). Clonality and intracellular polyploidy in virus evolution and pathogenesis. *Proc Natl Acad Sci U S A* **112**: 8887–8892.

Pernet F, Barret J, Le Gall P, Corporeau C, Dégremont L, Lagarde F, *et al.* (2012). Mass mortalities of Pacific oysters *Crassostrea gigas* reflect infectious diseases and vary with farming practices in the Mediterranean Thau lagoon, France. *Aquac Environ Interact* **2**: 215–237.

Pernet F, Barret J, Marty C, Moal J, Gall P Le, Boudry P (2010). Environmental anomalies, energetic reserves and fatty acid modifications in oysters coincide with an exceptional mortality event. *Mar Ecol Prog Ser* **401**: 129–146.

Pernet F, Fuhrmann M, Petton B, Mazurié J, Bouget JF, Fleury E, *et al.* (2018). Determination of risk factors for herpesvirus outbreak in oysters using a broad-scale spatial epidemiology framework. *Sci Rep* **8**: 1–11.

Pernet F, Lagarde F, Jeanné N, Daigle G, Barret J, Le Gall P, *et al.* (2014). Spatial and temporal dynamics of mass mortalities in oysters is influenced by energetic reserves and food quality. *PLoS One* **9**.

Pernet F, Lupo C, Bacher C, Whittington RJ (2016). Infectious diseases in oyster aquaculture require a new integrated approach. *Philos Trans R Soc B Biol Sci* **371**.

Pernet F, Tamayo D, Fuhrmann M, Petton B (2019). Deciphering the effect of food availability, growth and host condition on disease susceptibility in a marine invertebrate. *J Exp Biol* **222**: jeb210534.

Pernet F, Tamayo D, Petton B (2015). Influence of low temperatures on the survival of the

Pacific oyster (*Crassostrea gigas*) infected with ostreid herpes virus type 1. *Aquaculture* **445**: 57–62.

Petton B, Boudry P, Alunno-Bruscia M, Pernet F (2015). Factors influencing disease-induced mortality of Pacific oysters *Crassostrea gigas*. *Aquac Environ Interact* **6**: 205–222.

Petton B, Bruto M, James A, Labreuche Y, Alunno-Bruscia M, Le Roux F (2015). *Crassostrea gigas* mortality in France: The usual suspect, a herpes virus, may not be the killer in this polymicrobial opportunistic disease. *Front Microbiol* **6**: 1–10.

Petton B, Destoumieux-Garzón D, Pernet F, Toulza E, de Lorgeril J, Degremont L, *et al.* (2021a). The Pacific Oyster Mortality Syndrome, a Polymicrobial and Multifactorial Disease: State of Knowledge and Future Directions. *Front Immunol* **12**: 1–10.

Petton B, Destoumieux-Garzón D, Pernet F, Toulza E, de Lorgeril J, Degremont L, *et al.* (2021b). The Pacific Oyster Mortality Syndrome, a Polymicrobial and Multifactorial Disease: State of Knowledge and Future Directions. *Front Immunol* **12**.

Petton B, de Lorgeril J, Mitta G, Daigle G, Pernet F, Alunno-Bruscia M (2019). Fine-scale temporal dynamics of herpes virus and vibrios in seawater during a polymicrobial infection in the Pacific oyster *Crassostrea gigas*. *Dis Aquat Organ* **135 2**: 97–106.

Petton B, Pernet F, Robert R, Boudry P (2013). Temperature influence on pathogen transmission and subsequent mortalities in juvenile pacific oysters *Crassostrea gigas*. *Aquac Environ Interact* **3**: 257–273.

Pfeiffer JK, Kirkegaard K (2005). Increased fidelity reduces poliovirus fitness and virulence under selective pressure in mice. *PLoS Pathog* **1**: 0102–0110.

Pigliucci M (2007). Do we need an extended evolutionary synthesis? *Evolution (N Y)* **61**: 2743–2749.

Pigliucci M, Müller GB (2010). Evolution: The Extended Synthesis edited. *MIT Press Cambridge*: 137–174.

Pinaud S, Portela J, Duval D, Nowacki FC, Olive MA, Allienne JF, *et al.* (2016). A Shift from Cellular to Humoral Responses Contributes to Innate Immune Memory in the Vector Snail Biomphalaria glabrata. *PLoS Pathog* **12**: 1–18.

Pocheville A, Danchin É (2017). Genetic assimilation and the paradox of blind variation. In: *Challenging the Modern Synthesis: Adaptation, Development, and Inheritance*, Oxford University press, Oxford: Oxford, pp 111–136.

Poirier EZ, Vignuzzi M (2017). Virus population dynamics during infection. *Curr Opin Virol* **23**: 82–87.

Portela J, Duval D, Rognon A, Galinier R, Boissier J, Coustau C, *et al.* (2013). Evidence for specific genotype-dependent immune priming in the lophotrochozoan biomphalaria glabrata snail. *J Innate Immun* **5**: 261–276.

Pulecio J, Verma N, Mejía-Ramírez E, Huangfu D, Raya A (2017). CRISPR/Cas9-Based Engineering of the Epigenome. *Cell Stem Cell* **21**: 431–447.

Qi H, Li L, Zhang G (2021). Construction of a chromosome-level genome and variation map for the Pacific oyster *Crassostrea gigas*. *Mol Ecol Resour* **21**: 1670–1685.

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**: 1–16.

Qiao X, Wang L, Song L (2021). The primitive interferon-like system and its antiviral function in molluscs. *Dev Comp Immunol* **118**: 103997.

Qiao X, Zong Y, Liu Z, Li Y, Wang J, Wang L, *et al.* (2021). A novel CgIFNLP receptor involved in regulating ISG expression in oyster *Crassostrea gigas*. *Dev Comp Immunol* **124**.

Qu T, Zhang L, Wang W, Huang B, Li Y, Zhu Q, *et al.* (2015). Characterization of an inhibitor of

apoptosis protein in *Crassostrea gigas* clarifies its role in apoptosis and immune defense. *Dev Comp Immunol* **51**: 74–78.

Quayle D (1988). *Pacific Oyster Culture in British Columbia. Fisheries Research Board of Canada*. Bulletin.

Randall RE, Goodbourn S (2008). Interferons and viruses: An interplay between induction, signalling, antiviral responses and virus countermeasures. *J Gen Virol* **89**: 1–47.

Rauschert S, Raubenheimer K, Melton PE, Huang RC (2020). Machine learning and clinical epigenetics: A review of challenges for diagnosis and classification. *Clin Epigenetics* **12**.

Reikine S, Nguyen JB, Modis Y (2014). Pattern recognition and signaling mechanisms of RIG-I and MDA5. *Front Immunol* **5**: 1–7.

Renault T, Le Deuff RM, Chollet B, Cochennec N, Gerard A (2000). Concomitant herpes-like virus infections in hatchery-reared larvae and nursery-cultured spat *Crassostrea gigas* and *Ostrea edulis*. *Dis Aquat Organ* **42**: 173–183.

Renault T, Le Deuff RM, Cochennec N, Maffart P (1994). Herpesvirus associated with mortalities among Pacific oyster. *Crassostrea gigas*, in France comparative study. *Rev Méd Vét* **145**: 735–742.

Renault T, Moreau P, Faury N, Pepin JF, Segarra A, Webb S (2012). Analysis of Clinical Ostreid Herpesvirus 1 (Malacoherpesviridae) Specimens by Sequencing Amplified Fragments from Three Virus Genome Areas. *J Virol* **86**: 5942–5947.

Renault T, Novoa B (2004). Viruses infecting bivalve molluscs. *Aquat Living Resour* **17**: 394–409.

Renault T, Tchaleu G, Faury N, Moreau P, Segarra A, Barbosa-Solomieu V, *et al.* (2014). Genotyping of a microsatellite locus to differentiate clinical Ostreid herpesvirus 1 specimens. *Vet Res* **45**: 14–19.

Richards EJ (2006). Inherited epigenetic variation - Revisiting soft inheritance. *Nat Rev Genet* **7**: 395–401.

Richards CL, Alonso C, Becker C, Bossdorf O, Bucher E, Colomé-Tatché M, *et al.* (2017). Ecological plant epigenetics: Evidence from model and non-model species, and the way forward. *Ecol Lett* **20**: 1576–1590.

Riviere G, He Y, Tecchio S, Crowell E, Gras M, Sourdaine P, *et al.* (2017). Dynamics of DNA methylomes underlie oyster development. *PLoS Genet* **13**.

Riviere G, Wu GC, Fellous A, Goux D, Sourdaine P, Favrel P (2013). DNA Methylation Is Crucial for the Early Development in the Oyster *C. gigas*. *Mar Biotechnol* **15**: 739–753.

Robalino J, Browdy CL, Prior S, Metz A, Parnell P, Gross P, *et al.* (2004). Induction of Antiviral Immunity by Double-Stranded RNA in a Marine Invertebrate. *J Virol* **78**: 10442–10448.

Roberto AE, Ana M. I, Steven RB, Maria Teresa SG, Cristina EF (2021). Differentially methylated gene regions between resistant and susceptible heat-phenotypes of the Pacific oyster *Crassostrea gigas*. *Aquaculture* **543**: 736923.

Rodgers CJ, Carnegie RB, Chávez-Sánchez MC, Martínez-Chávez CC, Furones Nozal MD, Hine PM (2015). Legislative and regulatory aspects of molluscan health management. *J Invertebr Pathol* **131**: 242–255.

Rohfritsch A, Bierne N, Boudry P, Heurtebise S, Cornette F, Lapègue S (2013). Population genomics shed light on the demographic and adaptive histories of European invasion in the Pacific oyster, *Crassostrea gigas*. *Evol Appl* **6**: 1064–1078.

Rondon R, Grunau C, Fallet M, Charlemagne N, Sussarellu R, Chaparro C, *et al.* (2017). Effects of a parental exposure to diuron on Pacific oyster spat methylome. *Environ Epigenetics* **3**.

Rosani U, Pallavicini A, Venier P (2016). The miRNA biogenesis in marine bivalves. *PeerJ* **2016**.
Royet J, Gupta D, Dziarski R (2011). Peptidoglycan recognition proteins: Modulators of the microbiome and inflammation. *Nat Rev Immunol* **11**: 837–851.
Rubio T, Oyanedel D, Labreuche Y, Toulza E, Luo X, Bruto M, *et al.* (2019). Species-specific mechanisms of cytotoxicity toward immune cells determine the successful outcome of Vibrio infections. *Proc Natl Acad Sci U S A* **116**: 14238–14247.
Salvi D, Mariottini P (2017). Molecular taxonomy in 2D: A novel ITS2 rRNA sequencestructure approach guides the description of the oysters' subfamily Saccostreinae and the genus Magallana (Bivalvia: Ostreidae). *Zool J Linn Soc* **179**: 263–276.
Salvi D, Mariottini P (2021). Revision shock in Pacific oysters taxonomy: The genus Magallana (formerly Crassostrea in part) is well-founded and necessary. *Zool J Linn Soc* **192**: 43–58.
Sang X, Dong J, Chen F, Wei L, Liu Y, Zhang M, *et al.* (2020). Molecular cloning and immune function study of an oyster IκB gene in the NF-κB signaling pathway. *Aquaculture* **525**: 735322.
Sarda S, Zeng J, Hunt BG, Yi S V. (2012). The evolution of invertebrate gene body methylation. *Mol Biol Evol* **29**: 1907–1916.
Saulnier D, de Decker S, Haffner P, Cobret L, Robert M, Garcia C (2010). A large-scale epidemiological study to identify bacteria pathogenic to Pacific Oyster *Crassostrea gigas* and correlation between virulence and metalloprotease-like activity. *Microb Ecol* **59**: 787–798.
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*. *Gene* **406**: 13–22.
Sauvage C, Boudry P, De Koning DJ, Haley CS, Heurtebise S, Lapègue S (2010). QTL for resistance to summer mortality and OsHV-1 load in the Pacific oyster (*Crassostrea gigas*). *Anim Genet* **41**: 390–399.
Sauvage C, Boudry P, Lapègue S (2009). Identification and characterization of 18 novel polymorphic microsatellite makers derived from expressed sequence tags in the Pacific oyster *Crassostrea gigas*. *Mol Ecol Resour* **9**: 853–855.
Schikorski D, Faury N, Pepin F, Saulnier D, Tourbiez D, Renault T (2011). Experimental ostreid herpesvirus 1 infection of the Pacific oyster *Crassostrea gigas*: Kinetics of virus DNA detection by q-PCR in seawater and in oyster samples. *Virus Res* **155**: 28–34.
Schikorski D, Renault T, Saulnier D, Faury N, Moreau P, Pépin JF (2011). Experimental infection of Pacific oyster *Crassostrea gigas* spat by ostreid herpesvirus 1: Demonstration of oyster spat susceptibility. *Vet Res* **42**: 27.
Schirmer M, Ijaz UZ, D'Amore R, Hall N, Sloan WT, Quince C (2015). Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. *Nucleic Acids Res* **43**.
Schmitt P, Duperthuy M, Montagnani C, Bachère E, Destoumieux-Garzón D (2012). *Immune responses in the Pacific oyster Crassostrea gigas: An overview with focus on summer mortalities*.
Schmitt P, Rosa RD, Duperthuy M, de Lorgeril J, Bachère E, Destoumieux-Garzón D (2012). The antimicrobial defense of the Pacific oyster, *Crassostrea gigas*. How diversity may compensate for scarcity in the regulation of resident/pathogenic microflora. *Front Microbiol* **3**.
Schoggins JW, Rice CM (2011). Interferon-stimulated genes and their antiviral effector functions. *Curr Opin Virol* **1**: 519–525.

Segarra A, Baillon L, Faury N, Tourbiez D, Renault T (2016). Detection and distribution of ostreid herpesvirus 1 in experimentally infected Pacific oyster spat. *J Invertebr Pathol* **133**: 59–65.

Segarra A, Baillon L, Tourbiez D, Benabdelmouna A, Faury N, Bourgougnon N, *et al.* (2014). Ostreid herpesvirus type 1 replication and host response in adult Pacific oysters , *Crassostrea gigas*. *Vet Res* **45**: 30–32.

Segarra A, Pépin JF, Arzul I, Morga B, Faury N, Renault T (2010). Detection and description of a particular Ostreid herpesvirus 1 genotype associated with massive mortality outbreaks of Pacific oysters, *Crassostrea gigas*, in France in 2008. *Virus Res* **153**: 92–99.

Sheldon BC, Verhulst S (1996). Ecological immunology: costly parasite defences and trade-offs in evolutionary ecology. *Trends Ecol Evo* **11**: 317–321.

Shibolet O, Giallourakis C, Rosenberg I, Mueller T, Xavier RJ, Podolsky DK (2007). AKAP13, a RhoA GTPase-specific guanine exchange factor, is a novel regulator of TLR2 signaling. *J Biol Chem* **282**: 35308–35317.

Shuai K, Liu B (2003). Regulation of JAK-STAT signalling in the immune system. *Nat Rev Immunol* **3**: 900–911.

Shuai K, Liu B (2005). Regulation of gene-activation pathways by pias proteins in the immune system. *Nat Rev Immunol* **5**: 593–605.

Simpson GG (1953). The Baldwin Effect. *Evolution (N Y)* **7**: 110–117.

Smith VH, Jones TP, Smith MS (2005). Host nutrition and infectious disease: An ecological view. *Front Ecol Environ* **3**: 268–274.

Smith ZD, Meissner A (2013). DNA methylation: Roles in mammalian development. *Nat Rev Genet* **14**: 204–220.

Song K, Li L, Zhang G (2017). The association between DNA methylation and exon expression in the Pacific oyster *Crassostrea gigas*. *PLoS One* **12**: 1–12.

Stear MJ, Bishop SC, Mallard BA, Raadsma H (2001). The sustainability, feasibility and desirability of breeding livestock for disease resistance. *Res Vet Sci* **71**: 1–7.

Takeda K, Kaisho T, Akira S (2003). Toll-like receptors. *Annu Rev Immunol* **21**: 335–376.

Tanaka T, Tsuchiya R, Hozumi Y, Nakano T, Okada M, Goto K (2016). Reciprocal regulation of p53 and NF-κB by diacylglycerol kinase ζ. *Adv Biol Regul* **60**: 15–21.

Tang X, Huang B, Zhang L, Li L, Zhang G (2016). TANK-binding kinase-1 broadly affects oyster immune response to bacteria and viruses. *Fish Shellfish Immunol* **56**: 330–335.

Teodoro JG, Branton PE (1997). Regulation of apoptosis by viral gene products. *J Virol* **71**: 1739–1746.

Terahara K, Takahashi KG, Nakamura A, Osada M, Yoda M, Hiroi T, *et al.* (2006). Differences in integrin-dependent phagocytosis among three hemocyte subpopulations of the Pacific oyster '*Crassostrea gigas*'. *Dev Comp Immunol* **30**: 667–683.

Tirnaz S, Batley J (2019). DNA Methylation: Toward Crop Disease Resistance Improvement. *Trends Plant Sci* **24**: 1137–1150.

Torda G, Donelson JM, Aranda M, Barshis DJ, Bay L, Berumen ML, *et al.* (2017). Rapid adaptive responses to climate change in corals. *Nat Clim Chang* **7**: 627–636.

Trerotola M, Relli V, Simeone P, Alberti S (2015). Epigenetic inheritance and the missing heritability. *Hum Genomics* **9**.

Tripathi V, Chatterjee KS, Das R (2021). Non-covalent Interaction With SUMO Enhances the Activity of Human Cytomegalovirus Protein IE1. *Front Cell Dev Biol* **9**: 1–11.

Vallejo RL, Leeds TD, Fragomeni BO, Gao G, Hernandez AG, Misztal I, *et al.* (2016). Evaluation of genome-enabled selection for bacterial cold water disease resistance using progeny

performance data in rainbow trout: Insights on genotyping methods and genomic prediction models. *Front Genet* **7**: 1–13.

Vanhove AS, Rubio TP, Nguyen AN, Lemire A, Roche D, Nicod J, *et al.* (2016). Copper homeostasis at the host vibrio interface: Lessons from intracellular vibrio transcriptomics. *Environ Microbiol* **18**: 875–888.

Venkataraman YR, Downey-Wall AM, Ries J, Westfield I, White SJ, Roberts SB, *et al.* (2020). General DNA Methylation Patterns and Environmentally-Induced Differential Methylation in the Eastern Oyster (<i>Crassostrea virginica<i/>). *Front Mar Sci* **7**: 1–14.

Verhoeven KJF, VonHoldt BM, Sork VL (2016). Epigenetics in ecology and evolution: What we know and what we need to know. *Mol Ecol* **25**: 1631–1638.

Vidal-Dupiol J, Dheilly NM, Rondon R, Grunau C, Cosseau C, Smith KM, *et al.* (2014). Thermal stress triggers broad Pocillopora damicornis transcriptomic remodeling, while Vibrio coralliilyticus infection induces a more targeted immuno-suppression response. *PLoS One* **9**.

Vidal-Dupiol J, Ladrière O, Destoumieux-Garzón D, Sautière PE, Meistertzheim AL, Tambutté E, *et al.* (2011). Innate immune responses of a scleractinian coral to vibriosis. *J Biol Chem* **286**: 22688–22698.

Wang X, Li Q, Lian J, Li L, Jin L, Cai H, *et al.* (2014). Genome-wide and single-base resolution DNA methylomes of the Pacific oyster *Crassostrea gigas* provide insight into the evolution of invertebrate CpG methylation. *BMC Genomics* **15**.

Wang X, Li A, Wang W, Que H, Zhang G, Li L (2021). DNA methylation mediates differentiation in thermal responses of Pacific oyster (*Crassostrea gigas*) derived from different tidal levels. *Heredity (Edinb)* **126**: 10–22.

Wang Y, Liu H, Sun Z (2017). Lamarck rises from his grave: parental environment-induced epigenetic inheritance in model organisms and humans. *Biol Rev* **92**: 2084–2111.

Wang L, Song X, Song L (2018). The oyster immunity. *Dev Comp Immunol* **80**: 99–118.

Wang X, Song X, Xie X, Li W, Lu L, Chen S, *et al.* (2018). TRAF3 enhances STING-mediated antiviral signaling during the innate immune activation of black carp. *Dev Comp Immunol* **88**: 83–93.

Wang XB, Wu Q, Ito T, Cillo F, Li WX, Chen X, *et al.* (2010). RNAi-mediated viral immunity requires amplification of virus-derived siRNAs in Arabidopsis thaliana. *Proc Natl Acad Sci U S A* **107**: 484–489.

Watson JD, Crick FHC (1953). Molecular structure of nucleic acids : A structure for deoxyribose nucleic acid. *Nature* **171**: 737–738.

Weaver ICG, Cervoni N, Champagne FA, Alessio ACD, Sharma S, Seckl JR, *et al.* (2004). Epigenetic programming by maternal behavior. **7**: 847–854.

Wegner KM, Volkenborn N, Peter H, Eiler A (2013). Disturbance induced decoupling between host genetics and composition of the associated microbiome. *BMC Microbiol* **13**.

Wendt J, Rosenbaum H, Richmond TA, Jeddeloh JA, Burgess DL (2018). Targeted bisulfite sequencing using the SeqCap epi enrichment system. In: *Methods in Molecular Biology*, Vol 1708, pp 383–405.

West-Eberhard MJ (2005). Developmental plasticity and the origin of species differences. *Exp Eye Res* **102**: 6543–6549.

Wilson VG (2017). *Advances in Experimental Medicine and Biology - SUMO Regulatio of Cellular processes*.

Winther RG (2001). August Weismann on germ-plasm variation. *J Hist Biol* **34**: 517–555.

Wray GA, Hoekstra HE, Futuyma DJ, Lenski RE, Mackay TFC, Schluter D, *et al.* (2014). Does

evolutionary theory need a rethink? - counterpoint no, all is well. *Nature* **514**: 161–164.
Wu S, Sun M, Zhang L, Kang S, Liao J, Zhu Z, *et al.* (2022). Grouper TRAF3 inhibits nodavirus infection by regulating the STING-mediated antiviral signaling pathway. *Fish Shellfish Immunol* **123**: 172–181.
Xiong A, Clarke-Katzenberg RH, Valenzuela G, Izumi KM, Millan MT (2004). Epstein-Barr virus latent membrane protein 1 activates nuclear factor-κB in human endothelial cells and inhibits apoptosis. *Transplantation* **78**: 41–49.
Xu T, Zhang X, Ruan Z, Yu H, Chen J, Jiang S, *et al.* (2019). Genome resequencing of the orange-spotted grouper (*Epinephelus coioides*) for a genome-wide association study on ammonia tolerance. *Aquaculture* **512**: 734332.
Yamaura K, Takahashi KG, Suzuki T (2008). Identification and tissue expression analysis of C-type lectin and galectin in the Pacific oyster, *Crassostrea gigas*. *Comp Biochem Physiol - B Biochem Mol Biol* **149**: 168–175.
Yoneyama M, Fujita T (2009). RNA recognition and signal transduction by RIG-I-like receptors. *Immunol Rev* **227**: 54–65.
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.
Zhang L, Li L, Guo X, Litman GW, Dishaw LJ, Zhang G (2015). Massive expansion and functional divergence of innate immune genes in a protostome. *Sci Rep* **5**: 1–11.
Zhang X, Li Q, Kong L, Yu H (2017). DNA methylation changes detected by methylation-sensitive amplified polymorphism in the Pacific oyster (<i>*Crassostrea gigas*<i/>) in response to salinity stress. *Genes and Genomics* **39**: 1173–1181.
Zhang L, Li L, Zhang G (2011). Gene discovery, comparative analysis and expression profile reveal the complexity of the *Crassostrea gigas* apoptosis system. *Dev Comp Immunol* **35**: 603–610.
Zhang R, Liu R, Wang W, Xin L, Wang L, Li C, *et al.* (2015). Identification and functional analysis of a novel IFN-like protein (CgIFNLP) in *Crassostrea gigas*. *Fish Shellfish Immunol* **44**: 547–554.
Zhang R, Liu R, Xin L, Chen H, Li C, Wang L, *et al.* (2016). A CgIFNLP receptor from *Crassostrea gigas* and its activation of the related genes in human JAK/STAT signaling pathway. *Dev Comp Immunol* **65**: 98–106.
Zhang T, Qiu L, Sun Z, Wang L, Zhou Z, Liu R, *et al.* (2014). The specifically enhanced cellular immune responses in Pacific oyster (*Crassostrea gigas*) against secondary challenge with Vibrio splendidus. *Dev Comp Immunol* **45**: 141–150.
Zhang X, Yuan J, Sun Y, Li S, Gao Y, Yu Y, *et al.* (2019). Penaeid shrimp genome provides insights into benthic adaptation and frequent molting. *Nat Commun* **10**.
Zhao X, Yu H, Kong L, Liu S, Li Q (2016). High throughput sequencing of small RNAs transcriptomes in two Crassostrea oysters identifies microRNAs involved in osmotic stress response. *Sci Rep* **6**: 1–11.
Zhou Z, Sun B, Huang S, Yu D, Zhang X (2020). Roles of aminoacyl-tRNA synthetase-interacting multi-functional proteins in physiology and cancer. *Cell Death Dis* **11**.
Zhou Z, Wang L, Song L, Liu R, Zhang H, Huang M, *et al.* (2014). The identification and characteristics of immune-related MicroRNAs in haemocytes of oyster *Crassostrea gigas*. *PLoS One* **9**: 1–9.
Zhu J-K (2011). Active DNA Demethylation Mediated by DNA Glycosylases. *Annu Rev ofGenetics* **43**: 143–166.

Résumé

L'augmentation de la population humaine s'est accompagnée d'une augmentation de la demande de nourriture. Depuis 2008, des événements de mortalité massive de juvéniles d'huîtres causés (*Crassostrea gigas*) par le syndrome de mortalité des huîtres du Pacifique (Pacific Oyster Mortality Syndrome ; POMS) ont menacé l'industrie ostréicole. La résistance de *C. gigas* au POMS a démontré qu'elle repose sur des bases génétiques. Plus récemment, il a été démontré au niveau phénotypique qu'elle repose sur une réponse transcriptomique précoce à l'infection virale. Bien que des données concernant l'implication de l'épigénétique dans la résistance au POMS soit encore rare, le rôle essentiel du transcriptome, de son niveau de base à la réponse antivirale, et l'effet de l'environnementale sur la résistance de l'huître, suggèrent collectivement que l'épigénétique peut jouer un rôle essentiel. Nous proposons ici un cadre pour étudier simultanément le rôle potentiel de la génétique et de l'épigénétique dans l'expression d'un phénotype en utilisant le modèle *C. gigas*/POMS en population naturelle. Nous avons développé une approche de capture d'exome pour obtenir des informations génétiques (Single Nucleotide Polymorphisms ; SNPs) et épigénétiques (méthylation de l'ADN au niveau du contexte CG; CpGs). Des populations naturelles d'huîtres avec une exposition contrastée au POMS ont été phénotypées par une infection expérimentale. Les résultats obtenus montrent que l'exome capture a permis de caractériser la variation génétique et épigénétique de plus de 65 % des exons. Nous avons montré que les populations d'huîtres sauvages exposées de manière différentielle au POMS présentent des signatures de sélections à la fois dans leur génome (SNPs) et leur épigénome (CpGs). Un grand nombre de ces SNPs et CpGs étaient situés dans des gènes codant pour des protéines impliqué dans la réponse immunitaire. Ces résultats confirment que les populations hôtes confrontées à l'émergence de pathogènes peuvent s'appuyer sur la variation génétique et épigénétique pour s'adapter rapidement aux maladies émergentes. Si notre étude confirme le rôle essentiel joué par la séquence d'ADN, elle montre également que d'autres mécanismes peuvent interagir avec cette séquence pour coder un phénotype résistant ; cependant ils peuvent aussi être indépendants de cette séquence et participer à l'expression de la résistance. Ces résultats confirment que des approches plus holistiques de la résistance des populations hôtes doivent être envisagées pour avoir accès à la plupart des mécanismes en jeu. Par ailleurs ils démontrent également que la sélection assistée par l'épigénétique serait un moyen d'aider l'industrie de la sélection sans effet sur la séquence d'ADN.

Mots-clefs: huitre; virus; résistance/sensibilité; hôte/pathogène; génétique/épigénétique; GWAS/EWAS

Abstract

Together with the increase of human population, there is a mathematical increase for food supply. Since 2008, mass mortality events of Pacific oysters (*Crassostrea gigas*) juveniles caused by the Pacific Oyster Mortality Syndrome (POMS) have threatened the oyster aquaculture industry. Studies on the resistance of *C. gigas* to POMS has demonstrated a genetic bases and more recently, it was shown to rely on early transcriptomic response to the viral infection. Although data about the involvement of epigenetics in POMS resistance are still scarce, the essential role of the transcriptome, from the basal level to the antiviral response, and the effect of environmental exposure on the resistance of oyster, collectively suggest that epigenetic can play an essential role. Here we propose a framework to study simultaneously the potential role of genetic and epigenetic in the expression of phenotype by using the *C.gigas*/POMS model at the natural population level. We developed an exome capture approach to obtain genetic (Single Nucleotide Polymorphisms; SNPs) and epigenetic (DNA methylation at CG context; CpGs) information. In the present thesis, the exome capture developed allowed us to capture the genetic and epigenetic variation on more than 65 % of the total exons. We showed that wild oyster populations differentially exposed to the POMS display signatures of selections both in their genome (SNPs) and in epigenome (CpGs). A high number of these SNPs and CpGs were located in genes involved in immune functions. These results confirmed that host population facing pathogen emergence could rely on genetic and epigenetic variation to rapidly adapt to emerging diseases. While our study confirms the essential role played by the DNA sequence it also shows that other mechanisms can interplay with this sequence to encode a resistant phenotype. However, they can also be independent from this DNA sequence and participate to the expression of resistance. These results confirm that holistic approaches of the resistance of host population must be envisioned to have access to most of the mechanisms at stake. In addition, it also demonstrates that epigenetic assisted selection would be a way to assist breeding industry without effects on the DNA sequence.

Key words: oyster; virus; resistance/susceptibility; host/pathogen; genetic/epigenetic; GWAS/EWAS