



Comparative population genomics provides insights on the evolutionary history of marine fishes

Le Moan, Alan Jacques

Publication date:
2019

Document Version
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

Citation (APA):
Le Moan, A. J. (2019). *Comparative population genomics provides insights on the evolutionary history of marine fishes*. Technical University of Denmark.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Comparative population genomics provides insights on the evolutionary history of marine fishes

PhD thesis by
Alan Le Moan
May 2019

National Institute of Aquatic Resources
Section for Marine Living Resources
Technical University of Denmark

Table of Contents

Preface	i
Thesis summary	iii
Dansk resumé	iv
Introduction	1
Population genetics under neutral evolutionary forces	2
<i>Allelic frequency and effective population size</i>	3
<i>Migration rate and population structure</i>	4
<i>The genomic revolution: increased statistical power to detect cryptic population structure</i>	8
Natural selection	9
<i>The local genomic imprinting of natural selection</i>	9
<i>Local adaptation</i>	13
<i>Speciation processes and the multiple origins of islands of divergence</i>	16
<i>Demographic inferences as a tool to disentangle the processes of divergence</i>	19
The studied system and framework	20
References	24
Chapter I	28
Manuscript I	29
Manuscript II	73
Chapter II	98
Manuscript III	99
Manuscript IV	134

Preface

This thesis was submitted for the partial fulfillment of the requirement for the doctor of Philosophy Degree (PhD) at the Technical University of Denmark (DTU). The research of the present thesis was conducted at the National Institute for Aquatic Resources (DTU aqua) under the supervision of Senior Scientist Jakob Hemmer-Hansen (main supervisor) and Senior Scientist Dorte Bekkevold (co-supervisor) from the population genetics team (Silkeborg, Denmark) of the Marine Living Resources section. Professor Oscar Gaggiotti from the Scottish Oceans Institute provided additional advice for this thesis during an external research stay of 3 months at the University of St. Andrews (Scotland). This work was made possible thanks to the funding of the PhD school of DTU aqua, the European Regional Development Fund (Interreg V-A, project “MarGen”). The external research stay was partially funded by the Otto Mønstedts fund, for which I am sincerely grateful.

The aim of this PhD was to understand the evolutionary history of several marine fishes from the Northeast of the Atlantic, with a more specific focus on populations from the North Sea and the Baltic Sea. This framework benefited from a rich literature published in the past decades, of which the scientists and previous students of DTU Aqua (Silkeborg) conducted a significant part. I had the opportunity to start this project as the use of genomic resources was booming, providing the possibility to test previously elaborated hypotheses *in situ*.

I would like to thank Jakob Hemmer-Hansen for offering me the chance to conduct this challenging work under his supervision, for his trust in my ability to carry this project from beginning to end, and for his tolerance towards my tendency to consider my beliefs as facts. Many thanks also to Dorte Bekkevold for her co-supervision during this project, and Oscar Gaggiotti for hosting me in his lab; this work would have looked very different without their valuable advice and suggestions. This work would have been much more difficult without the help of Dorte Meldrup, from whom I discovered that pipetting time could also mean a fun time. I am truly grateful for her supportive and positive energy during the difficult periods of my PhD. I am grateful for all the support I received from the people in Silkeborg: Einar Eg Nielsen for sharing his enthusiasm for science; Tine Kastrup for her help with many little things which meant a whole lot in the end; Henrik Baktoft for his valuable advice on statistics (sorry again for the ice-cream incident!); Kim Aarestrup for the insightful and hilarious

conversations over lunch; Michael Holm for his help with fish sampling; and to all the others who have contributed, in small ways, to making this PhD a reality.

This PhD was also part of a collaborative project, which gave me the opportunity to travel around all of Scandinavia. This international experience made this PhD one of the richest experiences of my life. I will remember it for a long time and I am thankful for all the insightful discussions about population genetics I had with Michael Møller Hansen, Carl André and all the other researchers, but also the memorable moment with other students of the project (Martin, Ellika, Ann-Elin, Katinka). I am looking forward to continue my Scandinavian adventure. I also spent a few months abroad in Scotland, and I wish to thank Oscar Gaggiotti for all the great advice, and also Marie Louis for helping me settle and for offering this amazing day of field work surrounded by dozens of puffins and dolphins. I would also like to thank François Bonhomme, Pierre-Alexandre Gagnaire and Nicolas Bierne for hosting me in their lab in Montpellier for a few weeks, as well as Maud for her little help at the end.

I wish also to thank the growing community of students and post-docs here in Silkeborg, including those that visited even for a short period, which made life in this little town much funnier. Kim, the greatest roommate ever; Alice, my best Italian friend; Romina, our big sister here (thanks a lot for having my back at the very end of my PhD!); Brian, the douchebag of the office; Belen, the positive attitude; Kristi, the second best roommate; Magnus, Sebastian, Hugo, Casper, Michael and Thomas.

I would not be able to submit this work if I had not met all these wonderful friends from Britany, and from Montpellier. My everyday work abroad is much easier knowing that every time I come back, nothing has changed. Thanks to all of you. Et je terminerai par le plus important des remerciements, pour ma famille, maman, Anaëlle. Si je suis arrivé à finir cette thèse, c'est avant tout parce que vous m'avez soutenu, merci pour tout.

Alan Le Moan

A handwritten signature in black ink, appearing to read 'Alan Le Moan', with a stylized flourish extending from the bottom.

Thesis summary

The study of evolutionary mechanisms that shape the divergence continuum whereby populations diverge, adapt to new environments, accumulate reproductive barriers to gene flow and evolve into new species, is a central topic in evolutionary biology. With the arrival of new sequencing techniques, we have the opportunity to advance our knowledge within this field, as well as to ask and answer fundamental research questions previously impossible to explore. While previous decades of research mainly relied on specific case studies with relatively few model species, comparative frameworks are now commonly used to understand the divergence continuum and to derive general patterns behind divergence processes.

This PhD thesis provides an insight into the evolutionary history involved during the divergence process of populations in marine fishes. These species are often characterised by their ability to disperse, with larvae drifting with ocean currents for several weeks, and their large population sizes, which together result in limited population structure. As such, these species provide interesting models to study the role of evolutionary forces promoting population divergence, for example linked to demographic histories or local environmental adaptations.

The North Sea – Baltic Sea transition zone is an ideal framework to study these evolutionary forces, as the Baltic Sea was connected to the Atlantic Ocean only 8 000 years ago, and presents steep environmental gradients (e.g. salinity) along the transition zone. This environmental gradient is associated with major breaks in the population genetic structure of several marine species. Yet, little is known about the demographic history and the genomic architecture that underlie population divergence and adaptation to this environment in marine fishes. The main goal of this PhD was to fill this knowledge gap by using a comparative population genomics approach on several marine flatfishes that have colonized the Baltic Sea. We found that this colonization was associated with a diversity of divergence patterns, with heterogeneous signals of divergence found across the genome of each species, but also in the genomic architecture of the divergence across species. Therefore, the populations from the North Sea and the Baltic Sea appear to have diverged through species-specific factors. Nevertheless, the presence of ancient lineages and ancient standing genetic variation seem to have fuelled the local adaptation of these species during the colonization of the Baltic Sea.

Dansk resumé

Studiet af de evolutionære mekanismer, der danner grundlaget for de processer hvorved populationer divergerer, tilpasser sig nye miljøforhold, akkumulerer reproduktive barrierer mod genflow og udvikler nye arter er et centralt tema indenfor evolutionsbiologien. De nye sekventeringsteknologier giver os muligheder for at forbedre vores forståelse inden for dette forskningsfelt, samt at stille og besvare fundamentale spørgsmål, som det tidligere ikke var muligt at undersøge. Mens forskningen i de seneste årtier har benyttet specifikke casestudier med relativt få modelorganismer, bliver komparative tilgange nu ofte benyttet for at forstå og udlede generelle mønstre for ovenstående kontinuum af divergensprocesser.

Denne PhD-afhandling giver ny indsigt i den evolutionære historie i forbindelse med divergensprocesser hos populationer af marine fisk. Disse arter er ofte karakteriserede ved deres store spredningsevne, hvor larver driver med havstrømme i flere uger, og deres store populationsstørrelser. Tilsammen resulterer disse faktorer i begrænset populationsstruktur. Derfor er disse arter interessante modeller til at studere rollen af de evolutionære kræfter, der øger de genetiske forskelle mellem populationer, for eksempel koblet til demografisk historie eller tilpasning til lokale miljøforhold.

Transitionszonen mellem Nordsøen og Østersøen er et ideelt system til studiet af disse evolutionære kræfter, da Østersøen blev forbundet med Atlanterhavet for kun 8 000 år siden, og da der findes kraftige miljøgradienter (eksempelvis salinitet) i transitionszonen. Denne miljøgradient er sammenfaldende med større ændringer i populationsstrukturen hos flere marine arter. Vi ved dog stadig meget lidt om den demografiske historie og den genomiske arkitektur, der er associeret med populationers divergens og tilpasning til dette miljø i marine fisk. Hovedformålet med denne PhD var at udfylde dette vidensgab ved at benytte en komparativ populations-genomisk tilgang på flere marine fladfisk, der har koloniseret Østersøen. Vi fandt, at koloniseringen var associeret med en høj diversitet af divergensmønstre, med heterogene signaler over genomerne indenfor arterne men også i den genomiske arkitektur mellem arterne. Det tyder derfor på, at populationerne i Nordsøen og Østersøen divergerede via artsspecifikke faktorer. Dog ser det også ud til, at ældre evolutionære linjer og eksisterende genetiske variation har været et vigtigt brændstof for disse arters evne til at kolonisere Østersøen og tilpasse sig lokale miljøforhold.

Introduction

Evolutionary biology describes the patterns and processes through which heritable changes within a group of organisms are transmitted from one generation to another, and it is studied by investigating the genetic variation (polymorphisms) between organisms across different time and spatial scales. This individual variability can be described at the morphological level (phenotype), but also by screening the genetic variation (genotype) along the molecule of heredity: the Deoxyribose Nucleic Acid (DNA). Genetic polymorphisms (alleles, i.e. different versions of the genetic code at a specific location in the genome) arise from random modifications in the DNA sequences (mutations) that may have consequences for the viability and fertility (i.e. fitness) of the organism carrying them. Hence, individuals with higher fitness have higher chances of transmitting their alleles to the next generation, while mutations inducing a lower fitness tend to disappear over time. This process is called natural selection, and corresponds to the main mechanism behind the “struggle for life” introduced by Charles Darwin (1859). Natural selection is responsible for the rapid sorting of standing genetic variation by keeping only the alleles with higher fitness, and therefore within-species variation should represent temporary phases on the path to optimal fitness, and should rarely be observed in wild populations. However, the first studies describing genetic polymorphisms discovered extremely high variability across individuals of the same species (Hubby and Lewontin, 1966; Lewontin and Hubby, 1966). This observed paradox led Motoo Kimura to publish an extension of the evolution theory in 1968, considering most genetic variation as “neutral” and thus not under the direct action of natural selection (Kimura, 1968). This neutrality is due to the redundancy of the genetic code and the presence of non-functional DNA along the chromosomes, resulting in mutations that have no direct effect on fitness of an individual. Therefore, the random evolutionary force known as genetic drift controls the evolutionary pathway of this type of mutation within a population. In general, both natural selection and genetic drift should reduce the level of polymorphism (except in the particular case of balancing selection) by keeping only the “best” or the “lucky” allele, and removing the “worst” (i.e., deleterious mutations) or “unlucky” allele. Therefore, the study of the equilibrium between the mutation process, which generates variation, and selection-drift, which sorts through this variation, represent the state of art in any field of evolutionary biology.

Population genetics under neutral evolutionary forces

Allelic frequency and effective population size

Different fields of evolutionary biology arose to understand the diversification of life across different time-scales. The micro-evolutionary processes occurring in a more recent timeframe are studied within the field of population genetics, to which this PhD dissertation belongs. Population genetics describes the genetic variability within and among populations of a species. In this field, populations are defined as groups of individuals with higher probability of interbreeding than crossbreeding with individuals from other populations (Mayr, 1970). The genetic variability within a population can be quantified by counting the number of allele along sequences of DNA and calculating their frequencies in the population.

For eukaryotes carrying two sets of chromosomes, the frequency of two alleles A (p) and a (q) in a bi-allelic marker, such as single nucleotide polymorphisms (SNPs), in a population is calculated by:

$$p = \frac{2 \times N_{AA} + N_{Aa}}{N_{total}} \quad \& \quad q = 1 - p$$

Where N_{total} is the total number of individuals sampled in the population, N_{AA} the number of samples with two A alleles (i.e homozygous), N_{Aa} the number of individuals carrying both alleles (i.e heterozygous).

To understand how allelic frequencies change through time, the population genetics research field relies on a long-standing theoretical framework that was coined in the beginning of the 19th century by Sewall Wright, Ronald. Fisher and John Haldane, after the re-discovery of the work of Gregor Mendel in the 19th century. At the beginning, this theoretical framework described the sampling of alleles in a (diploid) ideal population with no selection, no mutation, no migration, non-overlapping generations and random mating (i.e. Wright-Fisher population) (Ewens, 2012). In the ideal situation, such populations have an infinite number of individuals evolving in total isolation. Under these conditions, populations can quickly reach equilibrium, where their allelic and genotypic makeup does not change over time. This stasis is called Hardy-Weinberg equilibrium (HWE) (Mayr, 1970) and here, the homozygous individuals AA , aa and the heterozygous Aa are expected to be found at the constant frequencies of p^2 , q^2 and $2pq$, respectively.

In reality, populations are composed of a finite number of individuals, from which only a fraction successfully reproduces (i.e. leaves descendants in the next generation). This fraction is named Effective Population Size (N_E) and corresponds to the size of the idealized population losing variation due to genetic drift at the same rate as the actual population (Soulé, 1987). This finite N_E is directly proportional to the strength of the genetic drift contributing to the variation of allele frequencies from one generation to another, as illustrated Figure 1.

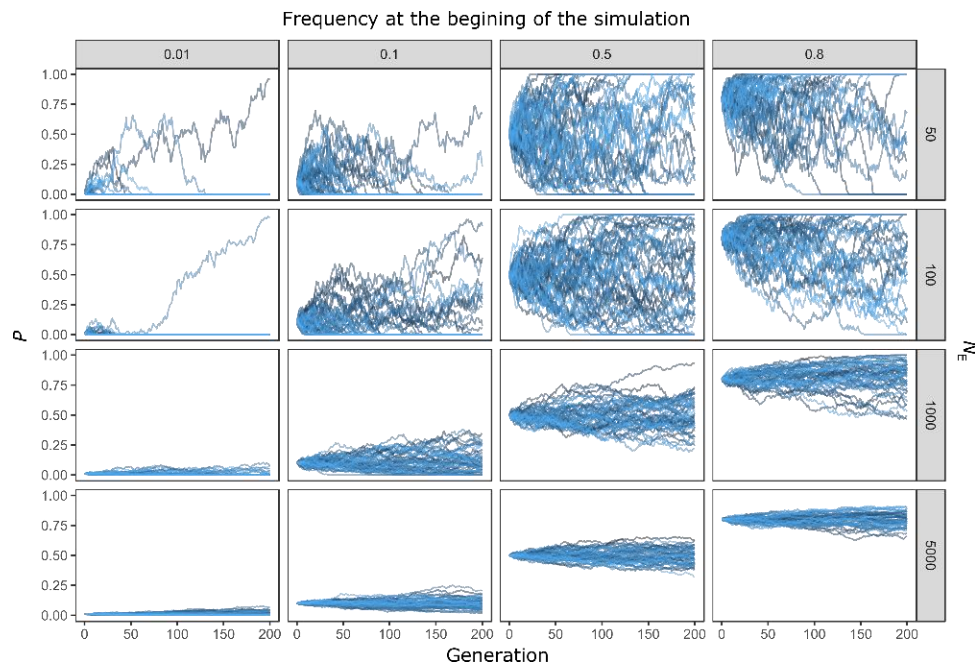


Figure 1: Illustration of the effect of genetic drift on the evolution of the allelic frequencies over 200 generations in 50 independent bi-allelic markers for different values of N_E (rows), and different frequencies (p) of allele A at the beginning of the simulation (columns)

Standing polymorphisms in populations with small N_E are quickly sorted with one allele, generally the major allele, reaching fixation within a short number of generations (Figure1, two first rows). However, polymorphisms are maintained over a longer period in populations with larger N_E (Figure1, two last rows). Indeed, a neutral mutation takes on average $N_E \times$ generations to reach fixation in a given population, and its probability to reach fixation at a given generation is equal to its frequency (Kimura, 1962). Therefore, within the time-scales studied in population genetics, populations with large N_E show less variation in allele frequencies across generations than populations with small N_E . Thus, when comparing different populations, species with larger N_E accumulate fewer differences in allelic

frequencies (i.e., are more genetically similar) than species with smaller N_E . However, this assumption does not imply that species with larger N_E have lower evolutionary rates, since new mutations appear in more individuals. The neutral rate of evolution is thus assumed to be equal to the mutation rate μ (i.e. molecular clock) for species of both small and large N_E (Kimura, 1968).

Migration rate and population structure

The individuals of a given species rarely behave as a homogenous unit, but are often divided into several populations that can be more or less connected by migration events. The resulting gene flow between the population is a neutral force and the fourth evolutionary mechanism shaping the evolutionary history of populations, after mutation, selection and genetic drift. When migration also results in successful reproduction in the receiving population, it results in gene flow between populations, which acts against the process of divergence (Slatkin, 1987). Altogether, these four evolutionary forces are necessary to understand how population are structured. The core of population genetics is therefore to find the limits between populations, to quantify their differences and to evaluate the gene flow underlying their connectivity to understand the structure of the populations. One of the most used metrics to study patterns of population structure rely on the F -statistics developed by Sewall Wright (1949). F -statistics divide the observed genetic variability into three components associated with different hierarchical levels of the structured population: i. the local inbreeding coefficient F_{IS} , ii. the fixation index F_{ST} and iii. the global inbreeding coefficient F_{IT} . The relationship among these three components can be represented by:

$$(1 - F_{IT}) = (1 - F_{ST})(1 - F_{IS})$$

The F_{IS} characterises the probability of two alleles from any random individual (I) in a subpopulation (S) to be inherited from the same ancestor (i.e. inbreeding). It provides insights into the reproduction dynamics of a local population and is calculated by quantifying the departure of the population from HWE:

$$F_{IS} = \frac{H_E - H_O}{H_E}$$

Where H_O is the observed heterozygosity and H_E is the expected heterozygosity under HWE. Low H_O in comparison to H_E leads to high F_{IS} , which means that individuals with similar genetic frequencies are more likely to reproduce with each other than with genetically dissimilar individuals. Biologically, high F_{IS} is caused by inbreeding, self-fertilization (Wright, 1949), or can result from a Wahlund effect, i.e. when two structured populations (i.e. with different allele frequencies) co-occur in the same area and are sampled together as one (Wahlund, 1928).

The differences between the genetic makeups of a subpopulation (S) and that of the total populations (T) is assessed by F_{ST} . This metric estimates the level of structure in the structured population, and represents a proxy for population differentiation (Wright, 1949). Several methods exist to estimate F_{ST} . Its most basic equation is:

$$F_{ST} = \frac{var(p)}{p(1-p)}$$

Where $var(p)$ is the variance of the focal allele frequency (p) across the sub-populations and $p(1-p)$ is the H_E in the structured population under HWE. If two populations are considered, high allele frequency differences among them result in a F_{ST} value close to one. Similarly, differentially fixed alleles lead to a maximum value of $F_{ST} = 1$. Alternatively, populations with similar genetic background will have an F_{ST} close to 0. In absence of selective pressures, population differentiation is shaped by the equilibrium between genetic drift and migration (illustrated Figure 2) where F_{ST} has reached the following stable value (Wright, 1931):

$$F_{ST} = \frac{1}{1 + 4 N_E * m}$$

Where N_E is the effective population size and m is the migration rate. Hence, poorly connected and/or small N_E populations show higher F_{ST} value (Figure 2, two first columns and first rows) than highly connected and/or large N_E populations (Figure 2, two last columns and last rows). Therefore, the structure of a population in given species is influenced by biological traits responsible for their dispersal capacity and their population density.

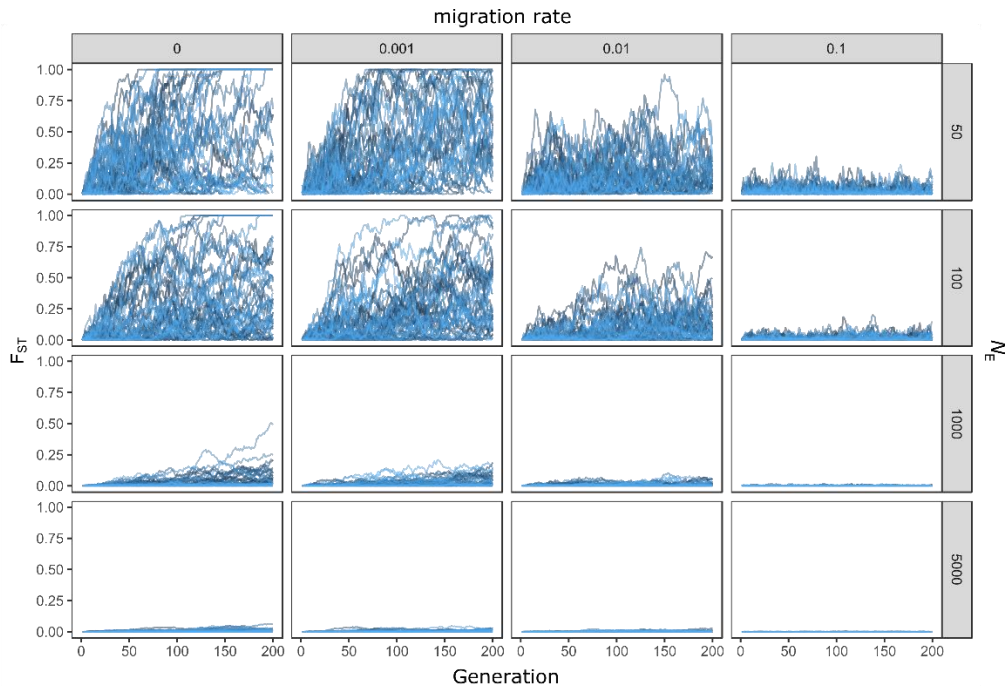


Figure 2: Variation of F_{ST} values between two populations connected by migration over 200 generations in 50 independent bi-allelic markers for different values of N_E (rows) and migration rates (columns)

Genetic metrics such as F -statistics are quite useful to study population structure in somewhat illusive species, where direct, standardized observations are difficult. In its simplest form, neutral population structure can result in a linear relationship between genetic differentiation (F_{ST}) and geographical distance (Wright, 1943). This effect is called isolation-by-distance (IBD), and its strength is proportional to the ability of an individual to disperse. Disruption of this linear relationship often highlights other relevant mechanisms responsible for shaping the population structure of a species (Slatkin, 1993). For instance, groups of populations with high F_{ST} as compared under the IBD expectation can correspond to several scenarios of disruption of gene flow, involving recent and dramatic demographic events (bottlenecks/founder effects – Nei *et al.*, 1975), physical barriers leading to isolation (e.g. mountain ranges, oceanographic fronts – Slatkin, 1987), ancient demographic history (e.g. past isolation – Hewitt, 2000) or selective processes (e.g. local adaptation, speciation – Barton, 1979). On the contrary, an absence of genetic structure and IBD can be due to peculiar breeding systems such as the one found for the European eels (*Anguilla anguilla*). The European eel is distributed across the entire eastern coasts of the northern Atlantic and migrates to the Sargasso Sea to spawn, leading to panmixia (i.e. random mating) at the species level ($F_{ST} = 0$) (Palm *et al.*, 2009). In some cases F -statistics can misrepresent true

population sub-structuring patterns, as low genetic divergence levels (and thus indicating panmixia) can be obtained between populations that evolved under strong neutral evolutionary forces, i.e. with large N_E and/or high connectivity (two bottom-left graphs, Figure 2), while corresponding nonetheless to different demographic scenarios (Figure 3).

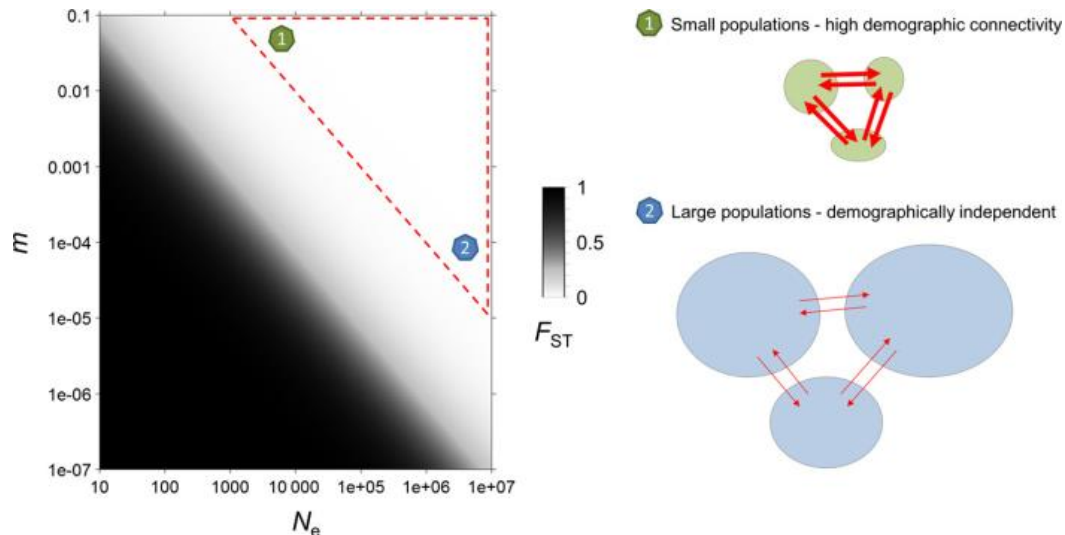


Figure 3: Illustration of the grey zone of population structure where the F -statistics are poorly informative, either due to small populations with high demographic connectivity (1) or large populations that are demographically independent (2), adapted from Gagnaire *et al.*, 2015.

In Figure 3, the grey zone represents one of the main challenges faced by population geneticists in the previous decades, as population sub-structuring becomes difficult to detect based on the F -statistic alone. This population structure “grey zone” has been reported several times in different marine species (Waples and Gaggiotti, 2006; Ward *et al.*, 1994). For example, the blue shark (*Prionace glauca*) represents one of the most extraordinary cases with a worldwide genetic panmixia reported despite occurring in all the oceans of the planet (Bailleul *et al.*, 2018). Two main reasons are likely to be responsible for the low structure in the marine environment. Firstly, marine species often have pelagic egg and larval life stages, promoting high population connectivity, and secondly, marine species generally have large N_E . This combination leads to limited population divergence (Figure 3, left). Consequently, true IBD patterns have rarely been observed in marine fishes, and most marine species could appear effectively panmictic.

The genomic revolution: increased statistical power to detect cryptic population structure

Recently, the field of genetics has been transformed by the progress of sequencing technologies (Metzker, 2010). While only a dozen of markers were used to study population genetics in the previous decades, entire genomes can be sequenced nowadays. This sequencing revolution has moved the field from the genetic to the genomic area, increasing the power to characterise genetic diversity of natural populations. In this PhD, we used such genomic tools to investigate the population structure of several marine fish species. More specifically, we applied a reduced-representation sequencing technique, by using the double-digestion Restriction Associated-site DNA (ddRAD) method (Poland and Rife, 2012), which aims to sequence only a subset of the genome across many individuals. This method uses the ability of restriction enzymes to cut genomic DNA into small fragments. The chosen enzymes target specific nucleotide patterns along the chromosome called restriction sites that are generally conserved across individuals of a given species. Thus, by selecting only sequences within a certain size range, only a fraction of the genomes can be sequenced, all coming from the same piece of chromosome across the studied individuals (Figure 4). Therefore, this technique is a cost efficient approach to study the diversity of many individuals collected from several populations and normally provides several thousands of genetic markers for the study of population structure.

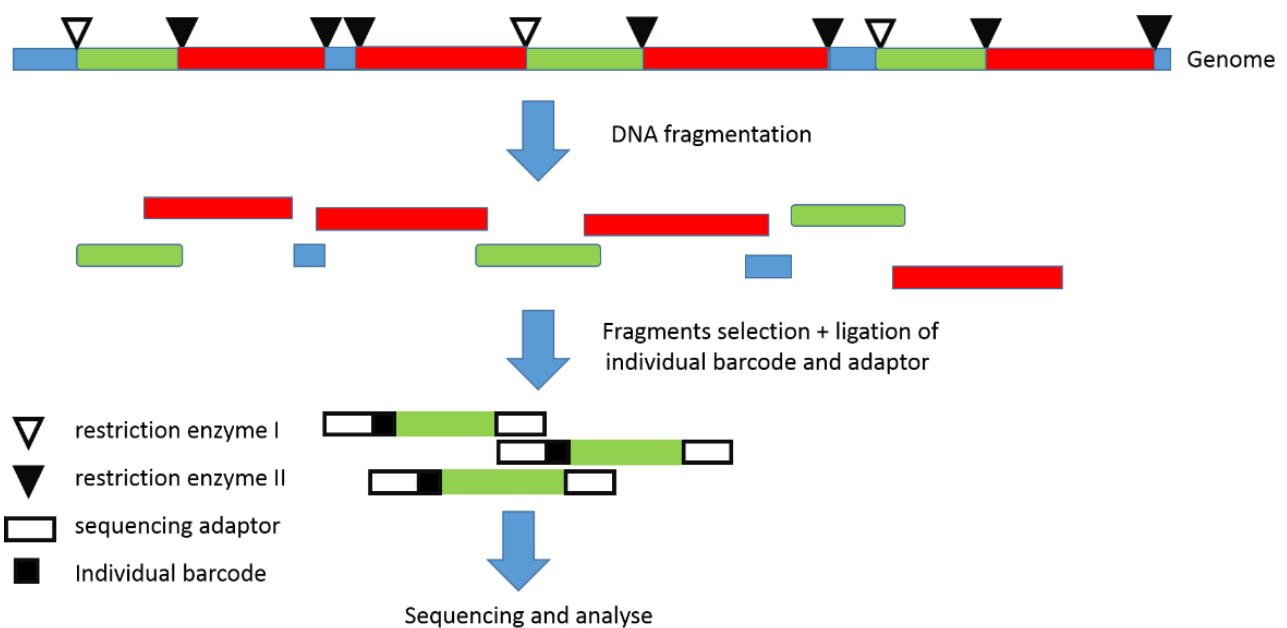


Figure 4: Summary illustration of the ddRAD protocol

With this method, we obtained thousands of independent SNPs and provided examples of the increase in statistical power to identify cryptic neutral population structure in marine fishes which have biological traits that have historically been associated with low levels of population structure, i.e. large N_E and high connectivity. In **manuscript I**, we used a comparative approach with four flatfish species sampled across the North Sea and the Baltic Sea. All the species showed a strong pattern of population structure separating the North Sea individuals from the Baltic Sea individuals, highlighting the reduced connectivity between the two seas despite their geographical proximity. The second example is described in **manuscript II**, focusing on the common sole (*Solea solea*), for which a large scale IBD had previously been described across the entire continental shelf of Europe, but where panmixia was assumed to be common at a finer geographical scale. In our study, the increase in numbers of markers applied allowed us to identify a fine scale IBD driven mainly by neutral processes along the coasts of Denmark. In **manuscript III**, which focused on the European plaice (*Pleuronectes platessa*), previous studies failed to detect population structure within most of the northeastern Atlantic. However, this was again likely due to the lack of statistical power of the previous markers used, since we found strong IBD patterns in the region. Although relatively few sampling sites were studied ($N = 8$), this example is to our knowledge one of the clearest IBD patterns reported in a marine fish, where a correlation of $r = 0.96$ was found between F_{ST} and geographical distance. Nevertheless, these three studies did not only provide evidence of the importance of neutral processes, but also showed the role of natural selection in driving some of the observed patterns of population structure (see below).

Natural selection

The local genomic imprinting of natural selection

Unlike genetic drift and migration which act on the entire genome, natural selection targets specific mutations, and its accurate detection requires a high genomic coverage. Mutations under selection may increase rapidly in the affected populations and can influence other neutral alleles occurring in their chromosomal vicinity through hitchhiking effects (Smith and Haigh, 1974). This phenomenon generates linkage disequilibrium (LD) among variants and corresponds to statistical associations of polymorphisms across individuals (Hill and Robertson, 1968). Therefore, hitchhiking results in a decrease in genetic diversity around

the selected mutation (Smith and Haigh, 1974). Regions of the genomes showing low diversity and high LD represent clear signatures left by the effects of natural selection on the genome, and are called “selective sweeps” (Figure 5). Selective sweeps should be quickly eroded by recombination, resulting in small and localized signatures of selection on the genomes. Thus, another important benefit from the progress of the sequencing technology is the possibility to look for such genomic signatures of selective sweeps.

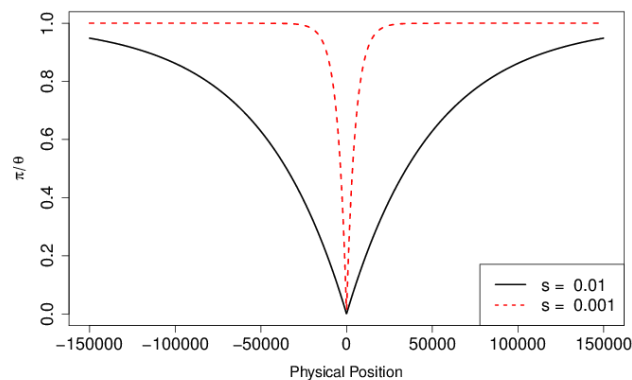


Figure 5: Visualization of the genetic diversity (π) decrease in the chromosome region surrounding a positively selected mutation (position 0) with different fitness during a selective sweep. Figure adapted from <https://cooplabs.github.io/popgen-notes/>

The fitness of a given advantageous mutation, A , appearing in a population as a bi-allelic SNP with alleles A and a with basic additive effect (i.e. the fitness effect of the two alleles in heterozygous correspond to the sum of the fitness effect of each allele) is calculated by a selection coefficient:

$$s = 1 - \omega_{aa}$$

where ω_{aa} is the relative fitness of the homozygous aa , which corresponds to the average number of descendants produced by the individuals carrying two non-advantageous alleles aa divided by the average number of descendants produced by the individuals carrying the two advantageous mutations AA .

Positively selected mutations appear randomly in the population over time. The speed of fixation will then depend on the selection coefficient and N_E of the population. In comparison to a weakly advantageous mutation, strongly advantageous mutations reach fixation at faster rates (rows, Figure 6), and lead to stronger selective sweeps responsible for a larger

reduction of diversity along the chromosome (Figure 5). However, the chance of a new mutation under selection to be lost by random processes is higher in populations with smaller rather than larger N_E . This effect is illustrated by the first column in Figure 6, where most of the advantageous mutations are lost in the beginning of the simulation for small population ($N_E = 100$), while the largest population ($N_E = 5\,000$) rarely lose any of the advantageous mutations (column 3, Figure 6). Therefore, species with large N_E , such as marine fishes, should theoretically display a greater adaptive potential with more mutations appearing at each generation, and the advantageous mutations more likely to reach fixation than in species with smaller N_E .

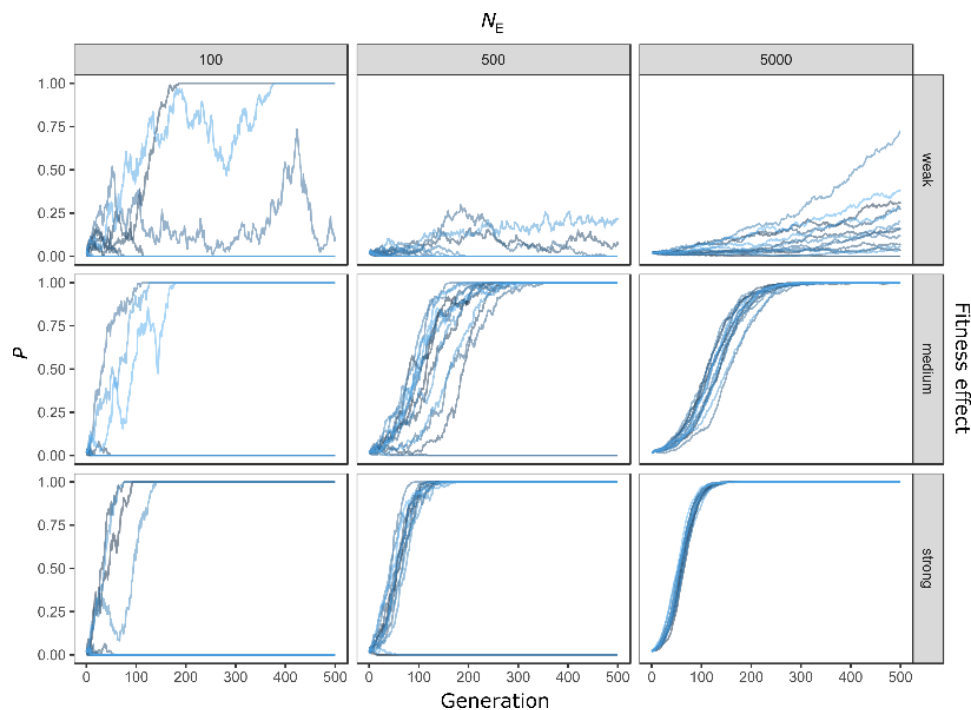


Figure 6: The effect of positive selection on the frequency of favorable mutations for variable fitness effect (rows) and for different values of effective population size (columns – N_E).

Nevertheless, despite this strong theoretical expectation, genomic studies have so far not identified clear patterns linking the number of adaptive mutations and effective population size across species. Furthermore, Rousselle *et al.* (2019) have recently compared 50 species with a large range of effective population sizes from various taxonomic groups and have described a slightly negative relationship between N_E and adaptive substitutions. This negative relationship is illustrated in Figure 7 with a dashed line linking the within-species genetic diversity π_S (proxy for N_E , see part one of the introduction) and the number of

adaptive substitutions. This relationship was however positive at lower taxonomic resolution (Figure 7, solid lines) but the slope linking N_E and adaptive mutations was again less steep in taxonomic groups with large N_E (molluscs, insects) compared to species with small N_E (primates, social insects). To explain this surprising departure of the data from the theoretical expectation described in the previous paragraph, the authors emphasized that large N_E populations are less loaded with deleterious mutations (mutations with negative effect on fitness) and therefore are likely to stay closer to their fitness optimum. Therefore, the mutation rate does not appear to be a limiting factor for the species with large N_E to maintain their fitness optimum. In contrast, small N_E populations tend to accumulate more adaptive mutations over similar time scales, which could be explained by a faster fixation rate of adaptive mutation in order to counteract an important mutation load (i.e. accumulation of deleterious mutation due to drift). Moreover, mutations arise in each generation in more individuals in larger N_E populations, which could involve a larger range of fitness effects across mutations. Thus, large N_E populations could potentially reach their fitness optimum with fewer adaptive mutations but with larger fitness effects, while small N_E populations will require several mutations with small fitness effects to reach the same optimum.

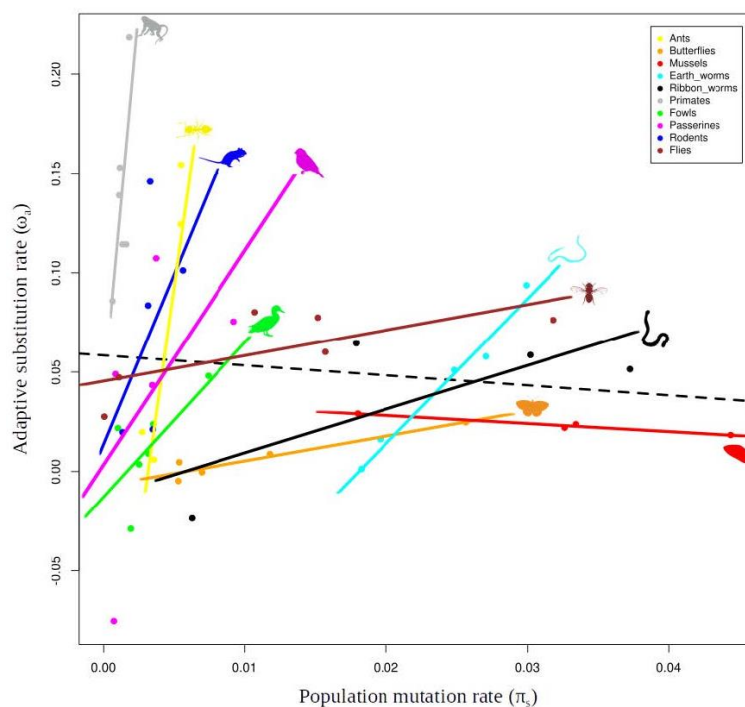


Figure 7: Relationship between N_E approximated by the population mutation rate (π_s , the number of polymorphic synonymous substitutions in the population) and the proportion of adaptive substitutions across 50 species sampled across 10 different taxa. Figure adapted from Rousselle *et al.* (2019).

Local adaptation

Local adaptation is the process where natural selection is acting in opposite directions on the alleles of a locus depending on the environmental context in which they occur (Williams, 1966). Thus, organisms carrying locally adapted alleles will have a fitness advantage in a specific habitat condition, which may be different from the optimum of the organism living in a different habitat. Local adaptation can therefore maintain polymorphisms at the species level when populations occur in various ecological contrasts that contribute to the observed population structure. Here, the signature of the selective sweep resulting from the increase in frequency of the locally adapted mutation should coincide with a peak of F_{ST} between the differentially adapted populations. The resulting populations that are adapted to their local environment are generally defined as ecotypes.

If migration is weak relative to selection ($s \gg m$), a mutation involved in local adaptation should theoretically spread and show a positive association with causative environmental factors resulting in a cline of allelic frequencies along the environmental gradient (Figure 7; Fisher, 1937). The strength of the allelic cline will depend on the equilibrium between drift, selection and gene flow, and at equivalent rates of gene flow, the strength of the allelic cline will depend of the strength of selection (as shown on Figure 8 for 3 independent loci with different fitness effects linked to the same environmental factor across the same populations).

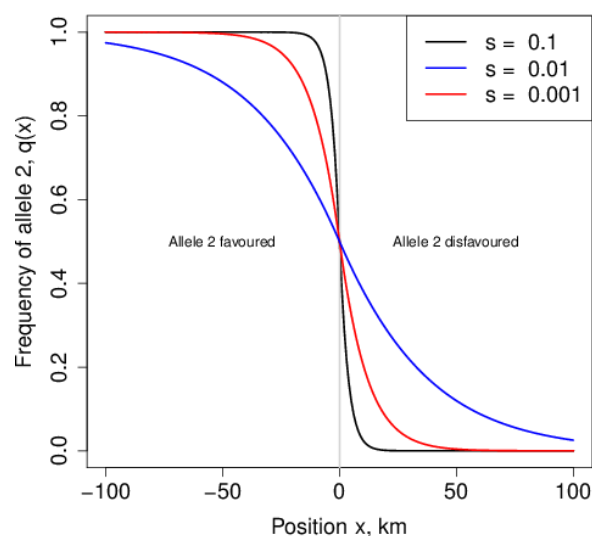


Figure 8: Effects of the selection coefficient on the allelic frequency cline between two locally adapted populations. Figure adapted from <https://cooplab.github.io/popgen-notes/>

However, if migration is too strong (i.e. $s < m$), the gene flow from a differentially adapted background can prevent adaptive mutations from spreading in a given population (Haldane, 1930). The process where locally adapted mutations are swamped by gene flow is called the migration load (Lenormand, 2002). The effects of the migration load can be reduced if specific heritable features evolve, reducing the recombination between adapted and maladapted backgrounds (Otto and Michalakis, 1998). Such features can evolve through various changes of the structure of the genomes (i.e. structural variants) which are frequently identified nowadays with the progress of sequencing technologies (Wellenreuther *et al.*, 2019). For example, structural variants can arise through a chromosomal inversion (i.e. flipping of a fragment of chromosome) which results in the absence of recombination with the un-flipped part of the chromosome (Kirkpatrick, 2010). Chromosomal inversion are responsible for major phenotypic changes relevant for local adaptation of ecotypes in various organisms (e.g. crab vs wave ecotype of *Littorina saxatilis*, Faria *et al.*, 2019; migratory and stationary ecotype of *Gadus morhua*, Kirubakaran *et al.*, 2016; mimicry in *heliconius* butterflies, Jay *et al.*, 2018).

Although other forms of adaptation than those based on loci with strong fitness effect are possible, such adaptation relies on the redundancy of the genetic code and selection on highly polygenic traits, maintaining locally adapted mutations at low frequencies in the population (Pritchard *et al.*, 2010; Yeaman, 2015). Therefore, it may be challenging to detect polygenic adaptation that require very high genomic coverage of the transcriptome, which is difficult to obtain with reduced representation techniques such as ddRAD (see Gouy *et al.*, 2017). Altogether, with the type of data used in this thesis, the detection of genetic variation involved in local adaptation between populations of marine fishes with both high N_E and high m should theoretically be expected to be biased towards the mutations with strong selective advantage or genomic features limiting recombination between adapted and maladapted backgrounds. The four manuscripts of the present PhD thesis provide some evidence for the role of selection and local adaptation acting on few loci with strong effects (**manuscripts I and II**) and on structural variants (**manuscript I, III and IV**) across species of fish.

The most basic method to identify evidence of local adaptation is to translocate individuals from a given population into the environmental condition of another population, and then to evaluate the fitness effects by comparing the survival rates and number of offspring, or other traits linked to fitness, of the translocated vs. resident individuals (Kawecki and Ebert, 2004). However, this method does not provide information about the genetic basis of local adaptation and it may be logistically challenging to perform for some organisms such as the marine fishes studied in this PhD. Thus, alternative methods can be used to study local adaptation without estimating fitness effects of mutations directly, for instance by performing association studies between genetic variation and environmental factors.

Studies of natural replicates with population divergence linked to similar ecological contrast provide relevant frameworks to perform association studies by identifying alleles that are repeatedly associated with a common environmental gradient across populations or species (Rundle and Nosil, 2005). The threespine stickleback (*Gasterosteus aculeatus*) is an iconic model to study such replicates of ecological divergence with freshwater ecotypes that have repetitively evolved from the marine ancestral populations in several isolated geographical areas (McKinnon and Rundle, 2002). This repeated evolution of ecotypes is underlined by strong phenotypic and genetic parallelism for which genomic sequencing has allowed the identification of the great complexity of this adaptation (Jones *et al.*, 2012) with several regions of the genome carrying clusters of high divergence (high F_{ST}) in so-called “islands of divergence”. These islands of divergence are sometimes composed of large inversions showing complex interaction with the environment (Jones *et al.*, 2012; Roesti *et al.*, 2014) and also by loci leading to strong phenotypic modifications relevant for local adaptation, such as the gene ectodysplasin-A (*Eda*) which is responsible for the loss of lateral armour plates in freshwater stickleback (Colosimo *et al.*, 2005). Moreover, these islands of divergence appear to be not only linked to local adaptation but also to loci involved in reproductive isolation (RI) of the ecotypes by reducing fitness of hybrid individuals (i.e. intrinsic incompatibilities; McKinnon and Rundle, 2002). This great complexity of traits under selection, and of the genomic architecture of these traits, is characteristic of the process of speciation (Wu, 2001) and is identified in an increasing number of non-model organisms (Ravinet *et al.*, 2017).

Speciation processes and the multiple origins of islands of divergence

A primary zone of divergence refers to the evolution of ecotypes occurring in the face of continuous gene flow, as described in the stickleback. It starts by the acquisition of a locus involved in local adaptation, which can then initiate a process of ecological speciation if some association (LD) evolves between the local adaptation and the RI loci of the ecotypes. Such associations can evolve if a new mutation involved in RI appears in the chromosomal vicinity of the local adaptation locus (Butlin, 1987). Reproductive isolation can evolve through two processes of incompatibility, which can be either exogenous or endogenous. Exogenous incompatibilities correspond to factors limiting the chances of reproduction between ecotypes which can evolve through behavioural modifications, such as assortative mating or habitat choice (Mayr, 1970). Furthermore, endogenous incompatibilities correspond to intrinsic incompatibilities responsible for the fitness decrease in hybrids, which then could act to reinforce the divergence process (speciation by reinforcement, Blair, 1955). Thus, an island of divergence will appear around the locus involved in local adaptation or RI while the rest of the genome will still be affected by gene flow and will remain undifferentiated (Wu, 2001). The size of the island around the selected loci will depend on their fitness effects (Figure 5), the chromosomal proximity among the loci under selection (Figure 9A) and the recombination rate around the loci (Figure 9A, Faria *et al.*, 2019; Roesti *et al.*, 2013).

The conditions under which primary zones of speciation can evolve are relatively restrictive (Butlin, 1987) and their importance for population/ecotype divergence is still debated (Bierne *et al.*, 2011; Cruickshank and Hahn, 2014; Le Moan *et al.*, 2016; Roux *et al.*, 2016). However, the process of speciation may be facilitated if populations are experiencing some degree of isolation. Indeed, the isolated populations follow at least partially independent evolutionary pathways, promoting the evolution of reproductive barriers to gene flow, such as Bateson-Dobzansky-Muller incompatibilities (BDMi, Figure 10, Dobzhansky, 1970). Examples of BDMis have been described with mito-nuclear co-adaptation (Burton and Barreto, 2012) but can also occur through nuclear co-adaptation and highly polygenic pathways (Simon *et al.*, 2018).

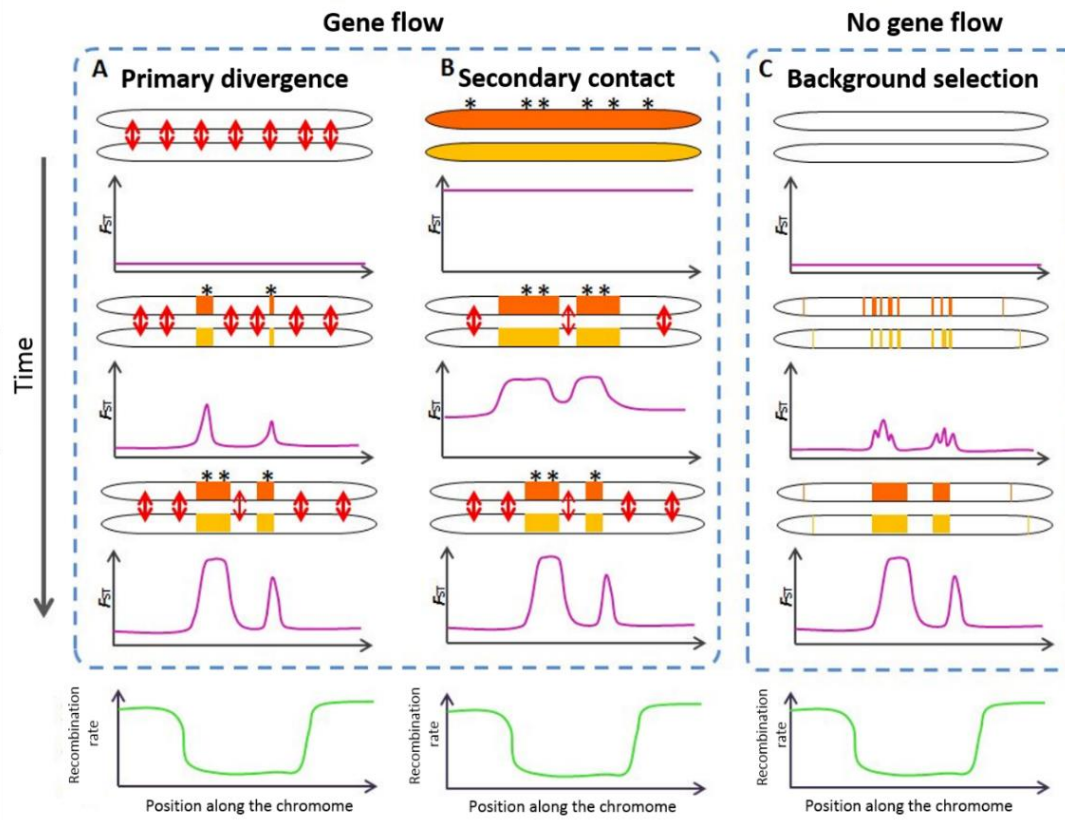


Figure 9: Schematic representation of the evolution of island of divergence for A. a primary zone of divergence, B. a secondary zone of divergence and C. allopatric divergence. The asterisks represents mutations under some sort of selection along the genome while the reds arrows represent the part of the genome exhibiting gene flow. Figure modified with permission from Maud Duranton (oral presentation of Marine Evolution, Strömstad, 2018).

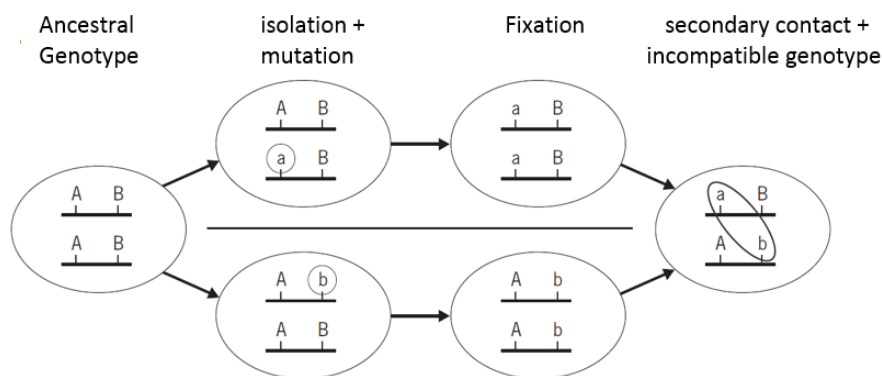


Figure 10: Schematic view of the evolution of a Bateson-Dobzhansky-Muller incompatibility at two loci A and B between two populations. The solid line represents the isolation of the populations. The allele 'a' is fixed at the locus A in the first population and allele 'b' is fixed at locus B in the second population during their isolation. During the secondary contact the occurrence of alleles 'a' and 'b' within the same individual leads to reduced fitness compared to the pure parental genotypes. Figure adapted from (Veen, 2008)

Therefore, populations experiencing a secondary contact after a temporary phase of isolation are likely to show some degree of RI (Abbott *et al.*, 2013). Under this scenario of secondary contact, gene flow will only be reduced in the region of the genome carrying BDMi while the rest of the genome will be quickly homogenised (Wu, 2001). Hence, this heterogeneous gene flow is expected to lead to the formation of islands of divergence (Figure 9B). In this situation, the size of the island will depend on the fitness effect of the BDMi in the hybrid individuals, the number of loci involved in the RI and, again, on the local recombination rate around the incompatible loci (Tine *et al.*, 2014). Moreover, loci responsible for RI are expected to display allelic frequency clines in the secondary zone of hybridization with similar shapes as the clines associated with adaptation (Figure 8). However, these RI clines are mobile in space and tend to be trapped by other gene flow barriers, such as physical obstacles (e.g. oceanic fronts) or environmental gradients (e.g. salinity) through coupling effects (Barton, 1979b). In the case of coupling in environmental gradients, loci responsible for incompatibilities are expected to be coupled with locally adapted alleles, as illustrated in Figure 11 (Bierne *et al.*, 2011). Therefore, it may be difficult to distinguish secondary clines and primary clines, which are both expected to lead to islands of divergence (Figure 9A, B) located along environmental gradients (Figure 11).

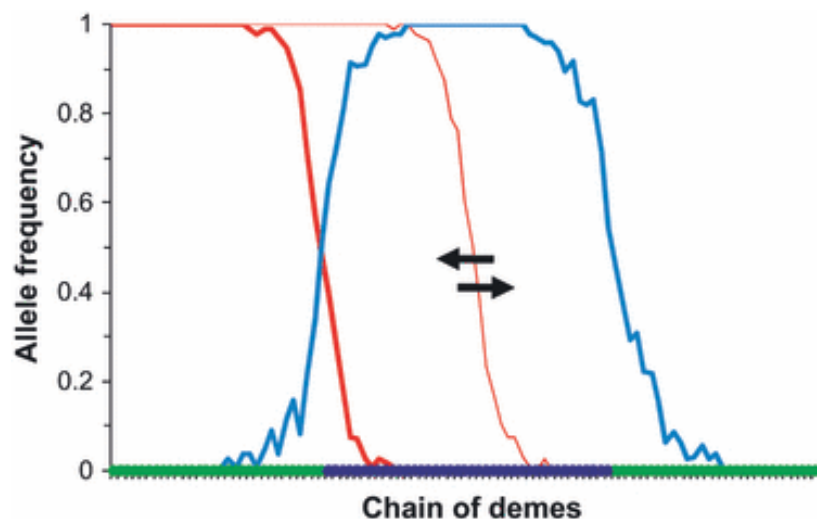


Figure 11: Illustration of the coupling effect between a cline of intrinsic incompatibility (red) and a cline of local adaptation (blue) in populations (dots along x axis) connected by gene flow in a stepping-stone model. At the beginning of the simulation, the populations are already adapted to two environmental conditions (blue and green, respectively), and the intrinsic incompatibility is found in the middle of the “blue” environment (light red cline). After some generations, the coupling occurs and the intrinsic cline (bold red) co-localize with the cline of local adaptation (blue). Figure adapted from Bierne *et al.*, (2011)

The progress of sequencing technology has recently revealed the great variability of the genome structures leading to variation of recombination rates along the chromosomes. This variability has important consequences for the formation of the islands of divergence. However, by considering the interaction between the variation of recombination rate and background selection, Cruikshank and Hahn (2014) re-analysed islands of divergence presumably involved in primary or secondary zones of divergence in various species and showed that most of these islands could be reproduced in scenarios of complete isolation without involving positive selection (illustrated in Figure 9C). Altogether, various evolutionary processes are expected to result in similar genomic signatures, and it can therefore be challenging to disentangle the underlying mechanisms based solely on the interpretation of statistics of divergence (Cruikshank and Hahn, 2014) or through association studies (Bierne *et al.*, 2011).

Demographic inferences as a tool to disentangle the processes of divergence

The development of sequencing technologies has been accompanied by a progress in computational power, which has also facilitated the use of methods for model-based demographic inference, thus limiting the consequences of subjective data interpretation (Emerson *et al.*, 2001). The aims behind inferring demography from genetic data is to reproduce statistics of population diversity and divergence under various demographic scenarios, which represent simplified visions of evolution (Sousa and Hey, 2013). Then, the model(s) producing the best reproduction of the observed dataset can be considered as the most likely scenario at the origin of population divergence. As for any model-based analysis, the interpretation of the data is limited to the specific assessed models (Fraïsse *et al.*, 2018). Thus, conclusions will be biased if the real evolutionary scenarios are not considered. Nonetheless, by comparing a large number of scenarios, such analyses can provide important means to improve our understanding of the evolutionary history of the studied populations (Roux *et al.*, 2016). In my PhD, I have used the approximation of maximum likelihood by equation of diffusion (Gutenkunst *et al.*, 2010) from which contrasted models of evolution can be derived to infer the demographic history of population divergence. I have applied these models to understand the demographic history linked to the colonization of the Baltic Sea.

The studied system and framework

The North Sea - Baltic Sea area provides an interesting opportunity to study replicates of ecotype divergence with the presence of an important environmental gradient associated with major genetic changes in the marine fauna. Evidence for local adaptation underlined by shared genetic basis has been generated using candidate gene based approaches (i.e. through the analyses of genes believed *a priori* to be important for local adaptation). For instance, the polymorphisms of heat-shock protein genes, proteins responding to environmental stress, show repeated association with the salinity gradient of the Baltic Sea across several species of fish (Hemmer-Hansen *et al.*, 2007 for the flounder; Nielsen *et al.*, 2009 for the cod; Limborg *et al.*, 2012 for the herring). However, little is known about the genomic architecture underlying the North Sea – Baltic Sea population divergence (but see Hemmer-Hansen *et al.*, 2013, for an example from the cod). Furthermore, other evolutionary processes than local adaptation are also expected to be involved in the maintenance of population structure (reviewed in Johannesson and André, 2006 and more detailed in **manuscript I and II**).

In this PhD project, I attempted to evaluate the extent to which the contrasted scenarios of evolution described in this introduction (neutral processes, complex demographic history, local adaptation, and speciation) can explain patterns of population structure. I have focused on the North Sea-Baltic Sea transition zone, but have also included cases from within these two geographical areas. These analyses were performed under a comparative population genomics approach where the population structure from several marine fishes were analysed independently through similar pipelines, providing independent replicates of population divergence. The results of these analyses were then compared in order to identify common evolutionary patterns across species.

The main results from the comparative approach are presented in the first chapter: **“Comparative population genomics of flatfishes along the North Sea – Baltic Sea transition zone”**. This chapter is composed of two manuscripts:

Manuscript I:

Beyond parallel evolution: when several species colonize the same environmental gradient

In this manuscript, I focused on the population divergence of four species of flatfish that are successfully established within the Baltic Sea: turbot (*Scophthalmus maximus*), common dab (*Limanda limanda*), European plaice (*Pleuronectes platessa*) and European flounder (*Platichthys flesus*). Data from these species were aligned to a reference genome assembled to the chromosome level, providing a framework to analyse the genomic architecture underlying population divergence. Moreover, three of these species were aligned to the same reference genome, which provided the opportunity to evaluate the degree of parallelism involved during these replicated cases of divergence in closely related species. Altogether, we found a great heterogeneity of divergence pattern, both across the genomes of the different species, but also across species, which were nonetheless clearly associated with the environmental gradient. The genetic differences were found genome-wide in two species, while they were clustered in specific regions of the genome for the other species. Most of these heterogeneous patterns of differentiation were inferred to be the result of complex demographic history of secondary contact and adaptation to the environmental gradient. In one case, the divergence was inferred to be initiated in the face of continuous gene flow, which was also the only case with strong evidence for the presence of larger structural variants driving most of the patterns of divergence.

Manuscript II

Fine scale population structure linked to neutral divergence in the common sole (*Solea solea*), a marine fish with high dispersal capacity

In this second paper, we analysed the fine scale population structure of the common sole (*Solea solea*), a species that cannot be considered as successful within the Baltic Sea as the other species studied in **manuscript I** since it rarely occurs after the main environmental break. This species was analysed separately as no reference genome is currently available for this species. We showed that even without a reference genome, we could draw important inferences on the evolutionary processes involved during the population divergence of sole. We found a weak population structure in the area that nonetheless followed a pattern of isolation-by-distance (IBD). The demographic inferences suggest that this IBD is mostly maintained by neutral process, for example related to the smaller effective population sizes

of the populations from the transition zone and asymmetrical gene flow. Few genetic markers showed evidence for selection. These inferences were further corroborated by a comparative IBD analyses by including the others species from **manuscript I**. Only the plaice populations, which shared the same inferred demographic history with the sole, showed an IBD pattern similarly low, especially when the structural variants were removed from the plaice IBD analyses. Altogether, this second manuscript provides support for the great diversity and the species-specific patterns of divergence involved during the colonization of the Baltic Sea.

The second chapter of this PhD is entitled “**Influence of genomic variability in structural variants on patterns of population sub-structuring of marine fishes**”, and is also divided in two manuscripts:

Manuscript III

Evolution at two-time frames shape structural variants and population structure of European plaice (*Pleuronectes platessa*)

This third manuscript presents further detailed analyses of the large structural variants (SV) identified in the North Sea and the Baltic Sea populations of European plaice (**manuscript I**). Here, we focus on a larger geographical scale in order to evaluate if these SVs have evolved in response to local environmental gradients or are ancient standing variants recently associated with the gradients. Although the genome-wide differentiation showed a strong pattern of IBD, the pairwise F_{ST} at the SVs were found to be poorly correlated with the geographical distance. Nevertheless, the alleles of the SVs were associated with the edge of the plaice distribution, with the derived allele increasing in frequency within the Baltic Sea and within the Barents Sea, while the ancestral allele were mostly found in the southwest of Norway or in the North Sea. Interestingly, the population sampled from Iceland was clearly differentiated from the other populations sampled along the continental shelf. This divergence was inferred to be due to a complex demographic history of past isolation followed by a secondary phase of gene flow. However, one of the structural variants was also polymorphic in this isolated population suggesting that these SVs are rather old. Through phylogenetic analyses, we were able to date the origin of the structural variants to be more than 225 000 year old. This study suggests that the colonisation of Baltic Sea by

the plaice mostly relied on the presence of ancient standing variants potentially already adapted to the local environment.

Manuscript IV

Weak genetic structure despite strong genomic signal in Lesser Sandeel in the North Sea

In this last manuscript from which I was a co-author, we applied the same genotyping protocol as in the other manuscripts to study the population structure of the Lesser Sandeel, focusing on samples collected within the North Sea. This study aimed to understand the population connectivity among the different management areas of this fish which has important economic value for the Danish fisheries. Despite the weak overall genetic structure, we detected a number of markers maintained under strong linkage disequilibrium (LD). This strong LD seemed to reveal the presence of genomic regions of reduced recombination, which then maintain divergent haplotypes segregating in the populations of the sandeel. One potential explanation for this pattern is the presence polymorphic structural variants, such as those described in manuscripts I and III. Thus, such SVs represent a plausible hypothesis for explaining three discrete clusters of individuals identified through multivariate analysis, which likely correspond to two pure homozygous clusters and one heterozygous cluster. When focusing on the frequency of these groups only, we found a relatively clear signal of population structure separating the Scottish population from the rest of the sampling sites, suggesting some degree of isolation of the different spawning grounds of sandeel in the North Sea. Interestingly, the frequencies of these clusters was also related to one environmental factor (Sea Bottom temperature), which could suggest some adaptive role. This study provides an interesting example of the use of putative SVs to detect fine scale population structure in species with high dispersal capacities and an absence of genome wide genetic differentiation.

References

- Abbott, R., Albach, D., Ansell, S., Arntzen, J.W., Baird, S.J.E., Bierne, N., Boughman, J., Brelsford, A., Buerkle, C.A., Buggs, R., Butlin, R.K., Dieckmann, U., Eroukhmanoff, F., Grill, A., Cahan, S.H., Hermansen, J.S., Hewitt, G., Hudson, A.G., Jiggins, C., Jones, J., Keller, B., Marczewski, T., Mallet, J., Martinez-Rodriguez, P., Möst, M., Mullen, S., Nichols, R., Nolte, A.W., Parisod, C., Pfennig, K., Rice, A.M., Ritchie, M.G., Seifert, B., Smadja, C.M., Stelkens, R., Szymura, J.M., Väinölä, R., Wolf, J.B.W., Zinner, D., 2013. Hybridization and speciation. *Journal of Evolutionary Biology* 26, 229–246. <https://doi.org/10.1111/j.1420-9101.2012.02599.x>
- Bailleul, D., Mackenzie, A., Sacchi, O., Poisson, F., Bierne, N., Arnaud-Haond, S., 2018. Large-scale genetic panmixia in the blue shark (*Prionace glauca*): A single worldwide population, or a genetic lag-time effect of the “grey zone” of differentiation? *Evolutionary Applications* 11, 614–630. <https://doi.org/10.1111/eva.12591>
- Barton, N.H., 1979a. Gene flow past a cline. *Heredity* 43, 333. <https://doi.org/10.1038/hdy.1979.86>
- Barton, N.H., 1979b. The dynamics of hybrid zones. *Heredity* 43, 341–359. <https://doi.org/10.1038/hdy.1979.87>
- Bierne, N., Welch, J., Loire, E., Bonhomme, F., David, P., 2011. The coupling hypothesis: why genome scans may fail to map local adaptation genes. *Molecular Ecology* 20, 2044–2072. <https://doi.org/10.1111/j.1365-294X.2011.05080.x>
- Burton, R.S., Barreto, F.S., 2012. A disproportionate role for mtDNA in Dobzhansky–Muller incompatibilities? *Molecular Ecology* 21, 4942–4957. <https://doi.org/10.1111/mec.12006>
- Butlin, R.K., 1987. Species, Speciation, and Reinforcement. *The American Naturalist* 130, 461–464.
- Colosimo, P.F., Hosemann, K.E., Balabhadra, S., Villarreal, G., Dickson, M., Grimwood, J., Schmutz, J., Myers, R.M., Schluter, D., Kingsley, D.M., 2005. Widespread Parallel Evolution in Sticklebacks by Repeated Fixation of Ectodysplasin Alleles. *Science* 307, 1928–1933. <https://doi.org/10.1126/science.1107239>
- Cruickshank, T.E., Hahn, M.W., 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology* 23, 3133–3157. <https://doi.org/10.1111/mec.12796>
- Csilléry, K., Blum, M.G.B., Gaggiotti, O.E., François, O., 2010. Approximate Bayesian Computation (ABC) in practice. *Trends Ecol. Evol. (Amst.)* 25, 410–418. <https://doi.org/10.1016/j.tree.2010.04.001>
- Darwin, C., 1859. *On the origin of species by means of natural selection.*
- Dobzhansky, T., 1970. *Genetics of the evolutionary process.* Columbia University Press.
- Emerson, B.C., Paradis, E., Thébaud, C., 2001. Revealing the demographic histories of species using DNA sequences. *Trends in Ecology & Evolution* 16, 707–716. [https://doi.org/10.1016/S0169-5347\(01\)02305-9](https://doi.org/10.1016/S0169-5347(01)02305-9)
- Ewens, W.J., 2012. *Mathematical Population Genetics 1: Theoretical Introduction.* Springer Science & Business Media.
- Faria, R., Chaube, P., Morales, H.E., Larsson, T., Lemmon, A.R., Lemmon, E.M., Rafajlović, M., Panova, M., Ravinet, M., Johannesson, K., Westram, A.M., Butlin, R.K., 2019. Multiple chromosomal rearrangements in a hybrid zone between *Littorina saxatilis* ecotypes. *Molecular Ecology* 28, 1375–1393. <https://doi.org/10.1111/mec.14972>
- Fisher, R.A., 1937. The Wave of Advance of Advantageous Genes. *Annals of Eugenics* 7, 355–369. <https://doi.org/10.1111/j.1469-1809.1937.tb02153.x>
- Fraïsse, C., Roux, C., Gagnaire, P.-A., Romiguier, J., Faivre, N., Welch, J.J., Bierne, N., 2018. The divergence history of European blue mussel species reconstructed from Approximate Bayesian Computation: the effects of sequencing techniques and sampling strategies. *PeerJ* 6, e5198. <https://doi.org/10.7717/peerj.5198>
- Gagnaire, P.-A., Broquet, T., Aurelle, D., Viard, F., Souissi, A., Bonhomme, F., Arnaud-Haond, S., Bierne, N., 2015. Using neutral, selected, and hitchhiker loci to assess connectivity of marine populations in the genomic era. *Evol Appl* 8, 769–786. <https://doi.org/10.1111/eva.12288>
- Gouy, A., Daub, J.T., Excoffier, L., 2017. Detecting gene subnetworks under selection in biological pathways. *Nucleic Acids Res.* 45, e149. <https://doi.org/10.1093/nar/gkx626>
- Gutenkunst, R.N., Hernandez, R.D., Williamson, S.H., Bustamante, C.D., 2010. Diffusion Approximations for Demographic Inference: DaDi. *Nature Precedings.* <https://doi.org/10.1038/npre.2010.4594.1>
- Haldane, J.B.S., 1930. A mathematical theory of natural and artificial selection. *Proc. Camb. Phil. Soc.* 26, 220–230.

- Hemmer-Hansen, J., Nielsen, E.E., Frydenberg, J., Loeschcke, V., 2007. Adaptive divergence in a high gene flow environment: *Hsc70* variation in the European flounder (*Platichthys flesus* L.). *Heredity* 99, 592–600. <https://doi.org/10.1038/sj.hdy.6801055>
- Hemmer-Hansen, J., Nielsen, E.E., Therkildsen, N.O., Taylor, M.I., Ogden, R., Geffen, A.J., Bekkevold, D., Helyar, S., Pampoulie, C., Johansen, T., Carvalho, G.R., 2013. A genomic island linked to ecotype divergence in Atlantic cod. *Molecular Ecology* 22, 2653–2667. <https://doi.org/10.1111/mec.12284>
- Hewitt, G., 2000. The genetic legacy of the Quaternary ice ages. *Nature* 405, 907. <https://doi.org/10.1038/35016000>
- Hill, W.G., Robertson, A., 1968. Linkage disequilibrium in finite populations. *Theoret. Appl. Genetics* 38, 226–231. <https://doi.org/10.1007/BF01245622>
- Hubby, J.L., Lewontin, R.C., 1966. A Molecular Approach to the Study of Genic Heterozygosity in Natural Populations. I. the Number of Alleles at Different Loci in *DROSOPHILA PSEUDOOBSCURA*. *Genetics* 54, 577–594.
- Jay, P., Whibley, A., Frézal, L., Rodríguez de Cara, M.Á., Nowell, R.W., Mallet, J., Dasmahapatra, K.K., Joron, M., 2018. Supergene Evolution Triggered by the Introgression of a Chromosomal Inversion. *Current Biology* 28, 1839–1845.e3. <https://doi.org/10.1016/j.cub.2018.04.072>
- Johannesson, K., André, C., 2006. INVITED REVIEW: Life on the margin: genetic isolation and diversity loss in a peripheral marine ecosystem, the Baltic Sea. *Molecular Ecology* 15, 2013–2029. <https://doi.org/10.1111/j.1365-294X.2006.02919.x>
- Jones, F.C., Grabherr, M.G., Chan, Y.F., Russell, P., Mauceli, E., Johnson, J., Swofford, R., Pirun, M., Zody, M.C., White, S., Birney, E., Searle, S., Schmutz, J., Grimwood, J., Dickson, M.C., Myers, R.M., Miller, C.T., Summers, B.R., Knecht, A.K., Brady, S.D., Zhang, H., Pollen, A.A., Howes, T., Amemiya, C., Broad Institute Genome Sequencing Platform & Whole Genome Assembly Team, Baldwin, J., Bloom, T., Jaffe, D.B., Nicol, R., Wilkinson, J., Lander, E.S., Di Palma, F., Lindblad-Toh, K., Kingsley, D.M., 2012. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* 484, 55–61. <https://doi.org/10.1038/nature10944>
- Kawecki, T.J., Ebert, D., 2004. Conceptual issues in local adaptation. *Ecology Letters* 7, 1225–1241. <https://doi.org/10.1111/j.1461-0248.2004.00684.x>
- Kimura, M., 1968. Evolutionary rate at the molecular level. *Nature* 217, 624–626.
- Kimura, M., 1962. On the Probability of Fixation of Mutant Genes in a Population. *Genetics* 47, 713–719.
- Kirkpatrick, M., 2010. How and Why Chromosome Inversions Evolve. *PLOS Biology* 8, e1000501. <https://doi.org/10.1371/journal.pbio.1000501>
- Kirubakaran, T.G., Grove, H., Kent, M.P., Sandve, S.R., Baranski, M., Nome, T., De Rosa, M.C., Righino, B., Johansen, T., Otterå, H., Sonesson, A., Lien, S., Andersen, Ø., 2016. Two adjacent inversions maintain genomic differentiation between migratory and stationary ecotypes of Atlantic cod. *Mol. Ecol.* 25, 2130–2143. <https://doi.org/10.1111/mec.13592>
- Lenormand, T., 2002. Gene flow and the limits to natural selection. *Trends in Ecology & Evolution* 17, 183–189. [https://doi.org/10.1016/S0169-5347\(02\)02497-7](https://doi.org/10.1016/S0169-5347(02)02497-7)
- Lewontin, R.C., Hubby, J.L., 1966. A Molecular Approach to the Study of Genic Heterozygosity in Natural Populations. II. Amount of Variation and Degree of Heterozygosity in Natural Populations of *DROSOPHILA PSEUDOOBSCURA*. *Genetics* 54, 595–609.
- Limborg, M.T., Helyar, S.J., De Bruyn, M., Taylor, M.I., Nielsen, E.E., Ogden, R., Carvalho, G.R., FPT Consortium, Bekkevold, D., 2012. Environmental selection on transcriptome-derived SNPs in a high gene flow marine fish, the Atlantic herring (*Clupea harengus*). *Mol. Ecol.* 21, 3686–3703. <https://doi.org/10.1111/j.1365-294X.2012.05639.x>
- Mayr, E., 1970. *Populations, Species, and Evolution: An Abridgment of Animal Species and Evolution*. Harvard University Press.
- McKinnon, J.S., Rundle, H.D., 2002. Speciation in nature: the threespine stickleback model systems. *Trends in Ecology & Evolution* 17, 480–488. [https://doi.org/10.1016/S0169-5347\(02\)02579-X](https://doi.org/10.1016/S0169-5347(02)02579-X)
- Metzker, M.L., 2010. Sequencing technologies — the next generation. *Nature Reviews Genetics* 11, 31–46. <https://doi.org/10.1038/nrg2626>
- Nei, M., Maruyama, T., Chakraborty, R., 1975. The Bottleneck Effect and Genetic Variability in Populations. *Evolution* 29, 1–10. <https://doi.org/10.1111/j.1558-5646.1975.tb00807.x>
- Nielsen, E.E., Hemmer-Hansen, J., Poulsen, N.A., Loeschcke, V., Moen, T., Johansen, T., Mittelholzer, C., Taranger, G.-L., Ogden, R., Carvalho, G.R., 2009. Genomic signatures of local directional selection in a high gene flow marine organism; the Atlantic cod (*Gadus morhua*). *BMC Evol Biol* 9, 276. <https://doi.org/10.1186/1471-2148-9-276>

- Nikula, R., Strelkov, P., Väinölä, R., 2007. Diversity and Trans-Arctic Invasion History of Mitochondrial Lineages in the North Atlantic *Macoma Balthica* Complex (bivalvia: Tellinidae). *Evolution* 61, 928–941. <https://doi.org/10.1111/j.1558-5646.2007.00066.x>
- Otto, S.P., Michalakis, Y., 1998. The evolution of recombination in changing environments. *Trends in Ecology & Evolution* 13, 145–151. [https://doi.org/10.1016/S0169-5347\(97\)01260-3](https://doi.org/10.1016/S0169-5347(97)01260-3)
- Palm, S., Dannewitz, J., Prestegard, T., Wickström, H., 2009. Panmixia in European eel revisited: no genetic difference between maturing adults from southern and northern Europe. *Heredity* 103, 82–89. <https://doi.org/10.1038/hdy.2009.51>
- Poland, J.A., Rife, T.W., 2012. Genotyping-by-Sequencing for Plant Breeding and Genetics. *The Plant Genome* 5, 92–102. <https://doi.org/10.3835/plantgenome2012.05.0005>
- Pritchard, J.K., Pickrell, J.K., Coop, G., 2010. The Genetics of Human Adaptation: Hard Sweeps, Soft Sweeps, and Polygenic Adaptation. *Current Biology* 20, R208–R215. <https://doi.org/10.1016/j.cub.2009.11.055>
- Ravinet, M., Faria, R., Butlin, R.K., Galindo, J., Bierne, N., Rafajlović, M., Noor, M. a. F., Mehlig, B., Westram, A.M., 2017. Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. *Journal of Evolutionary Biology* 30, 1450–1477. <https://doi.org/10.1111/jeb.13047>
- Riginos, C., Cunningham, C.W., 2005. INVITED REVIEW: Local adaptation and species segregation in two mussel (*Mytilus edulis* × *Mytilus trossulus*) hybrid zones. *Molecular Ecology* 14, 381–400. <https://doi.org/10.1111/j.1365-294X.2004.02379.x>
- Roesti, M., Moser, D., Berner, D., 2013. Recombination in the threespine stickleback genome--patterns and consequences. *Mol. Ecol.* 22, 3014–3027. <https://doi.org/10.1111/mec.12322>
- Rousselle, M., Simion, P., Tilak M.K., Figuet E., Nabholz B., Galtier N., 2019. Is adaptation limited by mutation? A timescale dependent effect of genetic diversity on the adaptive substitution rate in animals. *BiorXive*. <https://doi.org/10.1101/643619>
- Roux, C., Fraïsse, C., Romiguier, J., Anciaux, Y., Galtier, N., Bierne, N., 2016. Shedding Light on the Grey Zone of Speciation along a Continuum of Genomic Divergence. *PLOS Biology* 14, e2000234. <https://doi.org/10.1371/journal.pbio.2000234>
- Rundle, H.D., Nosil, P., 2005. Ecological speciation. *Ecology Letters* 8, 336–352. <https://doi.org/10.1111/j.1461-0248.2004.00715.x>
- Simon, A., Bierne, N., Welch, J.J., 2018. Coadapted genomes and selection on hybrids: Fisher's geometric model explains a variety of empirical patterns. *Evolution Letters* 2, 472–498. <https://doi.org/10.1002/evl3.66>
- Slatkin, M., 1993. Isolation by Distance in Equilibrium and Non-Equilibrium Populations. *Evolution* 47, 264–279. <https://doi.org/10.1111/j.1558-5646.1993.tb01215.x>
- Slatkin, M., 1987. Gene flow and the geographic structure of natural populations. *Science* 236, 787–792. <https://doi.org/10.1126/science.3576198>
- Smith, J.M., Haigh, J., 1974. The hitch-hiking effect of a favourable gene. *Genetics Research* 23, 23–35. <https://doi.org/10.1017/S0016672300014634>
- Soulé, M.E., 1987. *Viable Populations for Conservation*. Cambridge University Press.
- Sousa, V., Hey, J., 2013. Understanding the origin of species with genome-scale data: modelling gene flow. *Nature Reviews Genetics* 14, 404–414. <https://doi.org/10.1038/nrg3446>
- Tine, M., Kuhl, H., Gagnaire, P.-A., Louro, B., Desmarais, E., Martins, R.S.T., Hecht, J., Knaust, F., Belkhir, K., Klages, S., Dieterich, R., Stueber, K., Piferrer, F., Guinand, B., Bierne, N., Volckaert, F.A.M., Bargelloni, L., Power, D.M., Bonhomme, F., Canario, A.V.M., Reinhardt, R., 2014. European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. *Nature Communications* 5, 5770. <https://doi.org/10.1038/ncomms6770>
- Veen, T., 2008. Mating decisions in a hybrid zone. s.n.
- Wahlund, S., 1928. Zusammensetzung Von Populationen Und Korrelationserscheinungen Vom Standpunkt Der Vererbungslehre Aus Betrachtet. *Hereditas* 11, 65–106. <https://doi.org/10.1111/j.1601-5223.1928.tb02483.x>
- Waples, R.S., Gaggiotti, O., 2006. INVITED REVIEW: What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology* 15, 1419–1439. <https://doi.org/10.1111/j.1365-294X.2006.02890.x>
- Ward, R.D., Woodwark, M., Skibinski, D.O.F., 1994. A comparison of genetic diversity levels in marine, freshwater, and anadromous fishes. *Journal of Fish Biology* 44, 213–232. <https://doi.org/10.1111/j.1095-8649.1994.tb01200.x>

- Wellenreuther, M., Mérot, C., Berdan, E., Bernatchez, L., 2019. Going beyond SNPs: the role of structural genomic variants in adaptive evolution and species diversification. *Molecular Ecology* 0.
<https://doi.org/10.1111/mec.15066>
- Williams, G.C., 2018. *Adaptation and Natural Selection: A Critique of Some Current Evolutionary Thought*. Princeton University Press.
- Wright, S., 1949. The genetical structure of populations. *Annals of eugenics* 15, 323–354.
- Wright, S., 1943. Isolation by Distance. *Genetics* 28, 114–138.
- Wright, S., 1931. Evolution in Mendelian Populations. *Genetics* 16, 97–159.
- Wu, C.-l., 2001. The genic view of the process of speciation. *Journal of Evolutionary Biology* 14, 851–865.
<https://doi.org/10.1046/j.1420-9101.2001.00335.x>
- Yeaman, S., 2015. Local Adaptation by Alleles of Small Effect. *The American Naturalist* 186, S74–S89.
<https://doi.org/10.1086/682405>