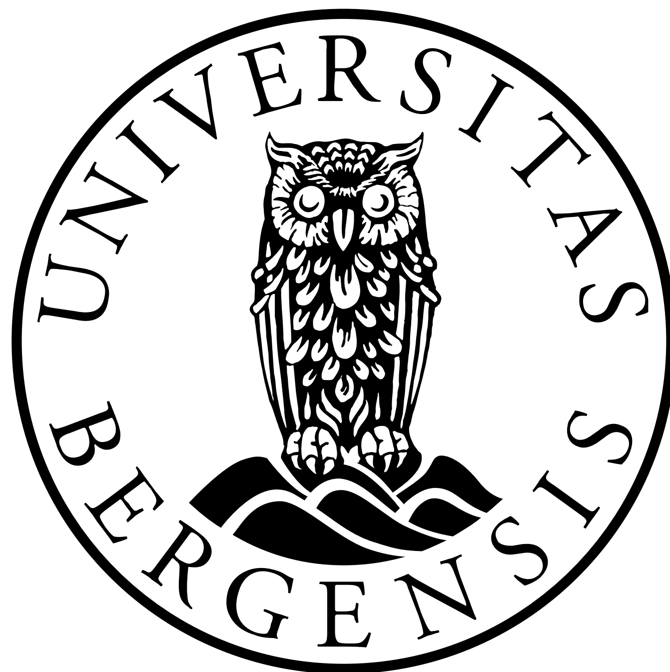
“A cleaner break” – Genetic analysis identifies two major groups of ballan wrasse (*Labrus bergylta*) within Scandinavia and evidence of phenotype derived divergence

Gaute Wilhelmsen Seljestad



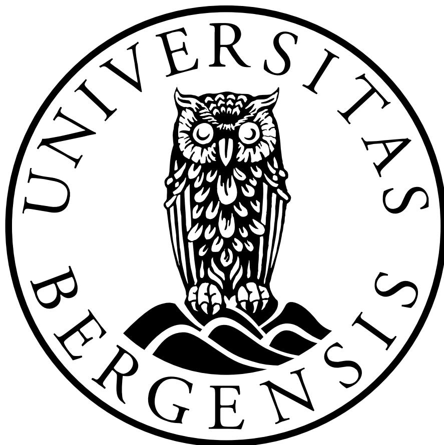
Master’s Thesis
Department of Biological Sciences

UNIVERSITY OF BERGEN

June 2019

“A cleaner break” – Genetic analysis identifies two major groups of ballan wrasse (*Labrus bergylta*) within Scandinavia and evidence of phenotype derived divergence

Gaute Wilhelmsen Seljestad



Master’s Thesis
Department of Biological Sciences

UNIVERSITY OF BERGEN

June 2019

© Gaute Wilhelmsen Seljestad

2019

“A cleaner break” – Genetic analysis identifies two major groups of ballan wrasse (Labrus bergylta) within Scandinavia and evidence of phenotype derived divergence

Gaute Wilhelmsen Seljestad

http://bora.uib.no/

Front page photo by Jeanette Ekren

Abstract

For a long time, sea lice have represented the major financial and fish welfare challenge of the salmonid farming industry. In the late 1980s, the use of parasite-eating cleaner fish, first and foremost wrasses (family *Labridae*), was put into practice. The use of wrasse declined in the 2000s as reliance on chemotherapeutants grew. However, sea lice quickly evolved a high level of resistance to chemotherapeutants, and therefore, the salmonid farming industry needed to come up with alternative control measures. This included, among other things, a resurge in the use of cleaner fish. These cleaner fish are primarily harvested from wild populations, and often translocated over large distances, such as from Sweden to mid-Norway. This provides at least two potential environmental challenges. First, possible overharvest in the source populations, and second, genetic interactions with recipient populations via inadvertent translocation (i.e., losses or deliberate release of non-local wrasse from farms into the surrounding coast).

To ensure sustainable use of cleaner fish in aquaculture, a better understanding of the population genetic structure of the key species, including ballan wrasse, is needed. In this thesis, an extensive population genetic analysis based on newly developed single nucleotide polymorphisms (SNPs), and previously existing microsatellites (MSATs), was conducted on >1000 individuals collected from 18 locations along the Scandinavian coast as well as one location in Galicia, Spain. Some of these samples were also morphotyped as plain or spotted. The following main results and conclusions were obtained: 1. Within Scandinavia, ballan wrasse are characterized by two distinct genetic groups, a north-western (NW) and south-eastern (SE) component, with very little within-cluster variation. The clusters are divided by a clear genetic break located in the south-western part of Norway; an area associated with a long continuous stretch of sandy substrate. It is concluded that the two genetic groups are likely to be shaped by differences in colonization histories and are held partially isolated by a combination of restricted adult movement, restricted benthic egg dispersal (i.e., non-pelagic), and habitat discontinuities in south-western Norway. 2. Less genetic diversity was observed within samples from Scandinavia than in the Galician sample. Additionally, Galician ballan wrasse were more similar to ballan wrasse from NW than SE Scandinavia. Furthermore, NW Scandinavian ballan wrasse were genetically more similar to the plain morphotype wrasse than to the spotted morphotype wrasse from Galicia. In contrast, this was not the case for SE Scandinavian ballan wrasse where no difference in the magnitude of genetic differentiation was observed with any of the Galician morphotypes. These results suggest that founder effects and genetic bottlenecks have characterized the colonization process of Scandinavia from southern

latitudes, and that NW Scandinavia was colonized with ballan wrasse of both spotted and plain morphotypes from southern latitudes prior to southern Norway and western Sweden. This route of colonization is also in agreement with glacial retreat models after the last glacial maxima (~21 kya), that propose an eastward melting pattern, with western Norway being liberated from the ice prior to eastern parts of the Scandinavian peninsula. 3. Very distinct genetic differentiation was identified between the spotted and plain wrasse in Galicia. This was not observed in Scandinavia where the phenotypic variation was also less distinct. It is therefore concluded that in Galicia, spotted and plain ballan wrasse are genetically distinct and are thus may represent sympatric sub-species as has been suggested earlier. However, the mechanisms and processes that could maintain the clear genetic differences between morphotypes in Galicia, for example assortative mating, are at best cryptic or non-existent in Scandinavia where these two potential sub-species appear undifferentiated.

From these findings it is proposed that care should be taken, and subsequently changes should be made to the current management of ballan wrasse in Scandinavian waters to minimize possibilities of future impact of escaped fish on local populations.

Acknowledgements

And so here we are, at the end of an era and the beginning of another.

I've been saying that I would become a marine biologist since the summer of 2003, and now after 16 years, well over half my life, I can actually say that I am. Pretty amazing, if I do say so myself. But I wouldn't be here if it wasn't for a bunch of other people:

Firstly, I need to thank my supervisors Kevin A. Glover, Maria Quintela and Arild Folkvord, for taking care of, believing in and encouraging me, and especially for letting me take on this project when the original failed. It's been a great pleasure working with you!

I'd also like to thank everyone at the molecular lab at IMR for letting me become part of the "crew" and for all the breaks with coffee and cakes. Special thanks to Anne Grete E. Sørvik, Eeva Jansson and Bjørghild B. Seliussen for all the help in teaching me the ways of the lab and answering my every question! In addition, a special thanks to Geir Dahle for quality-checking the raw data (again and again), even in the middle of the Christmas break, and to Francois Besnier for stepping up as IT consultant in my times of need!

A very special thank you to my mother, Inger Wilhelmsen Seljestad, and my father, Lars Ove Seljestad, since I wouldn't be here at all if it weren't for them. Also, great work on supplying genes and an upbringing that have given me an academic interest, a pretty integral part in this undertaking.

Of course, I haven't gone through these studies alone. I therefore extend a thank you to my entire master's group, you guys are the best! Special thanks to my "sisters at the south side of the Wall", Karoline Viberg and Julie Trollebø Kvalheim, without you this road would have been a lot harder!

Rounding off, I want to thank my brother Vetle V.W. Seljestad and my main man and flat mate Tore Thiesen for all the interesting discussions, and for listening to me ramble about biology (and every other science) at any time of day (though more often than not at night, during conversations that lasts hours longer than intended). Hope you never stop asking questions!

Finally, we live on a planet (ironically enough called Earth) that is covered by more than 70% ocean, representing 99% of the planets habitable space. There is in other words, more down there than we can think of. I therefore wish to finish this section with a Norwegian quote that quite well describes why marine organisms deserve our care and fascination:

“Nesten alt liv der nede har forbløffende egenskaper, som om de hører hjemme på en annen planet eller er skapt i en fjern fortid der andre regler gjaldt og enhver fantasi kunne virkeliggjøres.”
- Morten A. Strøksnes (Havboka, eller Kunsten å fange en kjempehai fra en gummibåt på et stort hav gjennom fire årstider).
Think about that the next time you chow down on some sea food…

TAVAHA!

Bergen, 01. June 2019 Gaute Wilhelmsen Seljestad

Table of contents

1. INTRODUCTION 15

1.1. SEA LICE IN AQUACULTURE 15
1.2. CLEANER FISH 15
1.2.1. Source of labrid cleaners and potential issues of current management 16
1.2.1.1. Ballan wrasse (Labrus bergylta) 17
1.3. POPULATION GENETICS 18
1.3.1. Genetics and population structure of labrids 20
1.3.2. Genetic groups and possible sub-speciation 20
1.4. THESIS AIM AND OBJECTIVES 21

2. MATERIALS AND METHODS 22

2.1. OVERALL STUDY DESIGN 22
2.2. SAMPLING 23
2.3. MORPHOLOGICAL ANALYSIS 24
2.4. DNA EXTRACTION, AMPLIFICATION AND GENOTYPING 27
2.5. QUALITY CONTROL OF GENETIC DATA 28
2.6. STATISTICAL ANALYSIS OF THE GENETIC DATA 30
2.6.1. Summary statistics, Hardy-Weinberg Equilibrium (HWE), Linkage Disequilibrium (LD), number of alleles, and allelic richness. 30
2.6.2. Pairwise F_{ST} (SNPs and MSATs) 31
2.6.3. Bayesian clustering - STRUCTURE analysis (SNPs and MSATs) 32
2.6.4. Discriminant Analysis of Principle Components (SNPs and MSATs) 33
2.6.5. Principle Components Analysis PCA (SNPs) 33
2.6.6. Outlier detection (SNPs) 33
2.6.7. Allele frequency plots (SNPs) 34
2.6.8. Cline analysis (SNPs) 34

3. RESULTS 36

3.1. OUTLIER TESTS (SNPS) 36
3.2. GENETIC DIFFERENTIATION AMONG SAMPLES (SNPS AND MSATS) 37
3.3. PATTERNS IN ALLELE FREQUENCIES (SNPS) 38
3.4. DISCRIMINANT ANALYSIS OF PRINCIPLE COMPONENTS (SNPS AND MSATS) 42
3.4.1. Unclustered Principle Component Analysis (SNPs and MSATs) 43
3.5. BAYESIAN CLUSTERING – STRUCTURE (SNPS AND MSATS) 43
3.6. CLINE ANALYSIS (SNPS) 45
3.7. GENETIC DIFFERENTIATION AMONG MORPHOTYPES (SNPS AND MSATS) 47
3.8. BAYESIAN CLUSTERING OF MORPHOTYPES (SNPS AND MSATS) 50

4. DISCUSSION 52

4.1. GENETIC DIVERSITY IN NORTHERN EUROPE 53
4.2. COLONIZATION HISTORY OF SCANDINAVIAN BALLAN WRASSE 54
4.3. COMPARISON WITH OTHER SPECIES IN SCANDINAVIA; WITH A SPECIAL FOCUS ON WRASSE 55
4.4. MORPHOLOGICAL VARIATION AND ASSOCIATED GENETICS 56
4.5. MANAGEMENT IMPLICATIONS 57
4.5.1. Potential inadvertent translocation via aquaculture 58
4.5.2. Impact of size specific fishery 59
4.6. FUTURE WORK 60
4.7. CONCLUDING REMARKS 61

5. REFERENCES 62

6. APPENDICES 70

6.1. APPENDIX 1: ADDITIONS TO MATERIALS AND METHODS 70

6.2. APPENDIX 2: ADDITIONS TO RESULTS 76

1. Introduction

1.1. Sea lice in aquaculture

Sea lice, and in particular the salmon louse *Lepeophtheirus salmonis* (Krøyer, 1837), have for a long time been the major challenge in the farming of Atlantic salmon (*Salmo salar* L., 1758) and rainbow trout (*Oncorhynchus mykiss* Walbaum, 1792) in the Northern hemisphere. In Norway alone, the control and treatment of sea lice is estimated to cost the salmonid farming industry multiple billion Norwegian kroner per year (Iversen et al. 2015, Brooker et al. 2018), in addition to being a major animal welfare issue for both farmed (Øverli et al. 2014) and wild fish. The overall increased number of sea lice resulting from widespread salmonid production in open sea cages (Torrissen et al. 2013) increases exposure of migrating wild Atlantic salmon post-smolts, as well as wild sea trout (*Salmo trutta* L., 1758) and Arctic char (*Salvelinus alpinus* L., 1758) that stay in coastal waters (Heuch and Mo 2001, Heuch et al. 2005, Finstad and Bjørn 2011, Jones and Beamish 2011). In turn, it is likely that this has caused greater marine mortality of wild salmonids in aquaculture regions, and possibly population-regulatory effects (Vollset et al. 2016)

Thus far, the most common way of handling lice infestations on commercial farms has been the use of chemotherapeutants. However, salmon lice have rapidly evolved a high level of resistance to all the conventional delousing agents (Pike and Wadsworth 1999, Denholm et al. 2002, Kaur et al. 2017). Together with concerns over the environmental impact of the chemicals used for delousing (Davies et al. 1998, Burridge et al. 2010), this situation has urged the salmon farming industry to come up with other methods to control lice infestations.

1.2. Cleaner fish

One of the alternative delousing methods has been the introduction of cleaner fish to farm pens. Cleaner fish are examples of animals that display cleaning behavior, a specialization in which the cleaner eats e.g., ectoparasites off other animals. Worldwide, cleaning behavior is especially frequent among fish in the family *Labridae,* commonly known as wrasses, where 50 species have been shown to display this trait (Côté 2000). Wrasses consist mainly of temperate to tropical shallow water reef dwelling species (Westneat and Alfaro 2005). In Norwegian waters, the family is represented by six coastal dwelling species, four of which, the goldsinny (*Ctenolabrus rupestris* L., 1758); corkwing (*Symphodus melops* L., 1758); rock cook (*Centrolabrus exoletus* L., 1758); and ballan wrasse (*Labrus bergylta* Ascanius, 1767), are used

as cleaner fish in aquaculture, in addition to the lumpfish (*Cyclopterus lumpus* L., 1758) (Darwall et al. 1992, Skiftesvik et al. 2014, Powell et al. 2017, Blanco Gonzalez and de Boer 2017).

The practice of using wild-captured wrasse as cleaner fish was first proposed scientifically and introduced to Norwegian salmon farms in 1988 (although experimental use of wrasse had been tested much earlier), followed by Scotland in 1989, and England and Ireland in 1990 (Bjordal 1988, Darwall et al. 1992, Rae 2002). But the use of wrasse dwindled in the late 1990s and mid-2000s after the introduction of the agent Slice (i.e., emamectin benzoate), an orally administered and highly effective chemical against lice (Stone et al. 1999, 2000). However, due to development of resistance to this (Espedal et al. 2013, Ljungfeldt et al. 2014) and other chemicals (e.g., Treasurer et al. 2000, Fallang et al. 2004, 2005, Kaur et al. 2017) used for delousing, there has been a renewed interest in the use of labrids in Norwegian aquaculture. This has led to increased fishing pressure to capture wild individuals, as well as catalyzing the initiation of breeding programs to produce cultured wrasse for delousing (Espeland et al. 2010, Ottesen et al. 2011, Skiftesvik et al. 2013, Helland et al. 2014, Norwegian Directorate of Fisheries 2018).

1.2.1. Source of labrid cleaners and potential issues of current management

After the resurge in cleaner fish interest, the deployment of fish to net pens increased rapidly. In 2017 approximately 22 million wild-captured labrids were reported to have been deployed as cleaner fish. Labrids have in other words relatively quickly become of commercial interest.

Because of the newfound nature of this interest, little work has been done on the species before, and as such very little is known about the labrid populations in Scandinavian waters. Even so, the Norwegian wrasse fishery has amounted to around 20 million individuals being caught every year since 2013 (Norwegian Directorate of Fisheries; www.fiskeridir.no), with no species-specific quotas being set (i.e., all labrids in Norwegian waters are fished under the same quota). In 2017, the labrid fishery had no total quota, but an individual vessel quota of 75 000 fish was set. This resulted almost 28 million fish being caught, the largest catch since the fishery started and close to 10 million higher than the quota recommended by the Norwegian Institute of Marine Research (IMR; Grefsrud et al. 2018). In 2018 and 2019 recommendations were followed and total quotas of 18 million fish were set (Norwegian Directorate of Fisheries; www.fiskeridir.no).

The labrid species present in Norwegian waters all tend to display strong site affiliation and small home ranges (Darwall et al. 1992; Gjøsæter 2002; Skiftesvik et al. 2015; Villegas-

Ríos et al. 2013). However, they have quite differing life-history strategies with big differences in growth, maturation and spawning (Darwall et al. 1992). Along with the negatively buoyant benthic eggs shown in all species found in Scandinavia, with the exception of goldsinny, site fidelity could potentially result in genetically distinct populations (Dipper et al. 1977, Dipper and Pullin 1979). Failing to take this and other underlying components of the fishery into consideration, such as the spatial extent of populations as well as the mixing of populations in time and space, can lead to differential exploitation and potentially overexploitation of resources (Allendorf et al. 2008, Kerr et al. 2017). In fact, concern that labrids, and especially ballan wrasse, are disappearing from multiple areas have already been reported (Grefsrud et al. 2018). However, scientific studies validating such claims are still to be conducted. Furthermore, the current practice of free transport of wrasse to fish farms throughout the distribution range and the potential later release or escape into the wild (thus inadvertent translocation mediated through aquaculture practice), may lead to gene-flow between local and non-local wrasses as suggested for goldsinny by Jansson et al. (2017) and found for corkwing by Faust et al. (2018).

The intense non-species-specific labrid fishery and the inadvertent translocation of labrids over large distances through aquaculture collectively represent a threat to the sustainability of wild populations, as both removal and release of fish can result in changes in allele frequencies of wild populations, diminution of genetic variation, eradication of local adaptations and/or breakdown of population structure (McGinnity et al. 2003, Hutchinson 2008, Laikre et al. 2010). There is thus an urgent need to study the population genetic structure of wild wrasses in order to identify the potential for such impacts.

1.2.1.1. Ballan wrasse (Labrus bergylta)

Of the wrasse species found in Norwegian waters, ballan wrasse has been proposed as one of the most promising cleaner fish due to its large size and high resilience (Blanco Gonzalez and de Boer 2017). Reaching an age of more than 25 years (Dipper et al. 1977, Dipper and Pullin 1979) and a size of up to 60 cm (Quignard and Pras 1986), it is the largest and longest lived of the labrids found in northern Europe. This large size makes it especially useful as a cleaner in pens with larger second year salmonids (Skiftesvik et al. 2013).

Ballan wrasse is a monandric protogynous hermaphrodite, meaning that all fish are born female before they undergo sex change and become males later in life. This usually happens after the age of 6 years and most often before reaching 40 cm in length (Dipper and Pullin 1979, Darwall et al. 1992, Muncaster et al. 2013). All larger individuals are in other words usually male. These males maintain harems of females with which they reproduce (Darwall et al. 1992),

and court these females before pairwise spawning of negatively buoyant eggs on the benthic substrate.

Ballan wrasse show highly variable body coloration (Dipper and Pullin 1979). In general, individuals may be morphologically characterized as plain (individuals that are characterized by a uniform body color, mainly greenish, reddish, or brownish); or spotted (individuals that are characterized by a dark orange or reddish body with white dots). Individuals may also display intermediate and marmorated patterns.

Ballan wrasse are distributed from the Canary Islands and Azores in the south to a northern terminus in Trøndelag in central Norway as well as surrounding the British Isles (Quignard and Pras 1986), and thus overlaps with the most important areas for salmonid aquaculture in Europe (Eurostat n.d., 2018, Norwegian Directorate of Fisheries 2018).

1.3. Population genetics

While there are several definitions, "Population Genetics" may be regarded as the study of genetic variation within and among populations in time and space and the underlying evolutionary mechanisms. Primary questions typically revolve around: "is this species made up of one or multiple genetically isolated populations", "what is the level of connectivity among these populations", "are there functional genetic differences between these populations", "what mechanisms are causing the observed differences" etc. The types of genetic markers implemented in population genetic studies has evolved over the years, e.g., from allozymes (protein variants), RAPDs (Random Amplified Polymorphic DNA), AFLPs (Amplified Fragment Length Polymorphisms), mini- and microsatellites (short non-coding tandem repeats), to Single Nucleotide Polymorphisms (SNPs) (Charlesworth and Charlesworth 2017). Additionally, the density of genetic information has evolved in tandem, from a hand-full of markers in early studies to tens of markers and in some cases even whole-genome coverage (termed genomics) in advanced studies (Luikart et al. 2003, Charlesworth and Charlesworth 2017).

Primarily, there are four evolutionary forces shaping the genetics of organisms and populations. These include mutations, which create variation; genetic drift, which is the random inheritance of alleles from one generation to the next; selection, which broadly speaking is the environments impact on the survival of genes; and gene-flow, the movement of genes between different locations. Collectively, these forces work together to shape the population genetic structure of species. Typically, small isolated populations exposed to divergent evolutionary

forces will vary greatly to each other, while large, highly overlapping populations will result in small genetic differentiation. By determining allele frequencies of multiple genetic markers in a number of individuals in multiple locations, one can thus determine the population genetic structure and level of connectivity.

In population genetic studies, putatively neutral loci were often preferred over loci under selection, as differentiation among populations at neutral loci represents a random process through the interaction between genetic drift, mutations and gene-flow. Differentiation between putative populations at neutral loci should therefore indicate population divergence (Rowe et al. 2017). Differentiation at loci known to be under selection on the other hand, results from differing selection pressures in differing environments, in addition to the before mentioned factors (Luikart et al. 2003). Differences in allele frequencies of non-neutral markers among groups of fish will thus not necessarily only indicate location disconnectedness, but also the response of selection. Therefore, while it is virtually impossible to differentiate the degree to which each factor has influenced allele frequencies, the use of non-neutral markers for population genetics studies has become more commonplace in recent times as the combination of selection and neutral processes tend to result in greater differentiation, revealing population genetic structure where otherwise weak or non-detectable (Besnier et al. 2014).

In marine environments with few obvious barriers to migration and gene-flow, the degree of connectivity among populations has historically been thought to be very high, resulting in homogenous populations spanning large areas (Hauser and Carvalho 2008). This view has been challenged by new highly specific information gained through the use of genetic methods, as multiple studies (e.g., Knutsen et al. 2003, 2007, Gysels et al. 2004, Almada et al. 2012, Blanco Gonzalez et al. 2016, Jansson et al. 2017) have shown that the degree of population genetic differentiation for many species is way higher than historically believed, even over relatively short distances. Still, genetic differentiation is expected to increase with increasing geographic distance (termed isolation-by-distance; IBD) as well as in highly divided biomes such as the Norwegian fjord landscape, resulting from less gene flow and drift between locations in addition to potential differences in selection pressure (Wright 1943, Slatkin 1993, Rousset 1997). It should therefore be expected that even in the marine realm, population genetic differentiation is likely to be the norm rather than the exception.

1.3.1. Genetics and population structure of labrids

As the interest for northern labrids continues to increase, elucidating population genetic structure and connectivity becomes highly important to ensure sustainable harvest in the future.

Although already started in the late 1990s (Sundt and Jørstad 1998), genetic studies of labrids did not really gain traction until the resurge in labrid interest in the late 2000s. As such, multiple studies of genetic composition and population structure have been published in later years, for both corkwing (Robalo et al. 2011, Knutsen et al. 2013, Blanco Gonzalez et al. 2016, Faust et al. 2018), goldsinny (Jansson et al. 2017), and ballan wrasse (D'Arcy et al. 2013, Almada et al. 2016, 2017, Quintela et al. 2016), all of which have indicated a decrease in genetic variation with increasing latitude, as well as significant population genetic differentiation.

This trend in decreasing genetic variation can likely be ascribed to labrids most likely being tropical in origin (Hanel et al. 2002, Westneat and Alfaro 2005) and the relatively short time period for northward expansion since the recession of the last glacial maxima (~21 kya; Lambeck et al. 2010). As some individuals migrated northwards after the receding ice front, they represented only a subset of the total genepool, this genetic subset would then become the genepool from which future generations could arise. Known as the founder effect (representing a genetic bottleneck), this new genepool by chance would display less variability than the source populations.

1.3.2. Genetic groups and possible sub-speciation

In addition to the general decrease in genetic variation with increasing latitude suggested for ballan wrasse (D'Arcy et al. 2013, Quintela et al. 2016, Almada et al. 2017), a decrease in phenotypic variation to the north also seems to be present. In southern parts of the distribution range (Galicia, Spain) two distinct morphotypes are known: spotted and plain. In Galicia, these two morphotypes carry different common names, are marketed separately, and display subtle but different life-histories, with spotted morphotype fish growing larger and faster than plain morphotype fish that invest more in reproduction (Villegas-Ríos 2013, Villegas-Ríos et al. 2013a). Microsatellite DNA analysis of ballan wrasse in Galicia have indicated that there are two sympatric genetic groups present that seem to somewhat coincide with the two morphological groups (Quintela et al. 2016), although this relationship could not be confirmed through the use of mitochondrial (18S, COI and Control Region) and nuclear (S7) DNA markers (Almada et al. 2016). A genetic component of this relationship is thus highly likely, even if not directly causative. One possibility is the presence of two genetic groups in which the frequencies of morphotypes differ, i.e., where one group is mostly spotted while the other

is mainly plain (Quintela et al. 2016), and that these genetic groups are maintained by assortative mating as proposed in Villegas-Ríos (2013) as well as in Villegas-Ríos et al. (2013) and Almada et al. (2016).

The potential presence of the two sympatric genetic groups of ballan wrasse is still to be investigated in northern parts of the distribution range, where the morphotypes plain and spotted have also been reported (Villegas-Ríos et al. 2013a) but are possibly less distinct (Maria Quintela, personal communication, 18 October 2018). As such, further studies of Scandinavian ballan wrasse are needed to determine the presence of the genetic groups, potentially complicating the population structure of ballan wrasse further, with implications for future species management.

1.4. Thesis aim and objectives

The overall aim of this thesis is to expand our knowledge of the population genetic structure of ballan wrasse in Scandinavia.

Specific objectives include:

- To elucidate population genetic structure of ballan wrasse in Scandinavia, as this species is currently treated as a single panmictic population and fish are freely transported from the south to the north for use as cleaner fish in salmonid aquaculture.
- To determine and compare population genetic characteristics of ballan wrasse in Scandinavia and Galicia (Spain), and investigate the colonization history of Scandinavian ballan wrasse.
- To investigate genetic differences between the spotted and plain morphotypes in both Scandinavia and Galicia, following previous indications of two sympatric genetic groups related to morphotype in Galicia. The two morphotypes have been reported as being present in Scandinavian waters, but the genetics of these putative groups is yet to be examined.

2. Materials and methods

2.1. Overall study design

A total of 1143 ballan wrasse were sampled from 36 locations. Some samples were collected in very close proximity to each other (i.e., just a few kilometers) and were therefore pooled. Additionally, ten locations had fewer than 20 individuals and were removed from the final dataset as they did not provide enough statistical power. This resulted in a final dataset of 16 locations in Norway, two locations in Sweden and one location in Galicia, Spain (Scandinavian sampling locations shown in Fig. 1, Galician location in Fig. A1 in appendix 1).

All of the above samples (1025 individuals total; 943 Scandinavian) were genotyped for a set of SNP markers (SNP dataset; Table 1). In addition, a sub-set of 8 of the samples (513 individuals total; 429 Scandinavian) was also genotyped for microsatellites (MSAT dataset; Table 2). Both marker types were used, due to each having its own advantages (see Box 1).

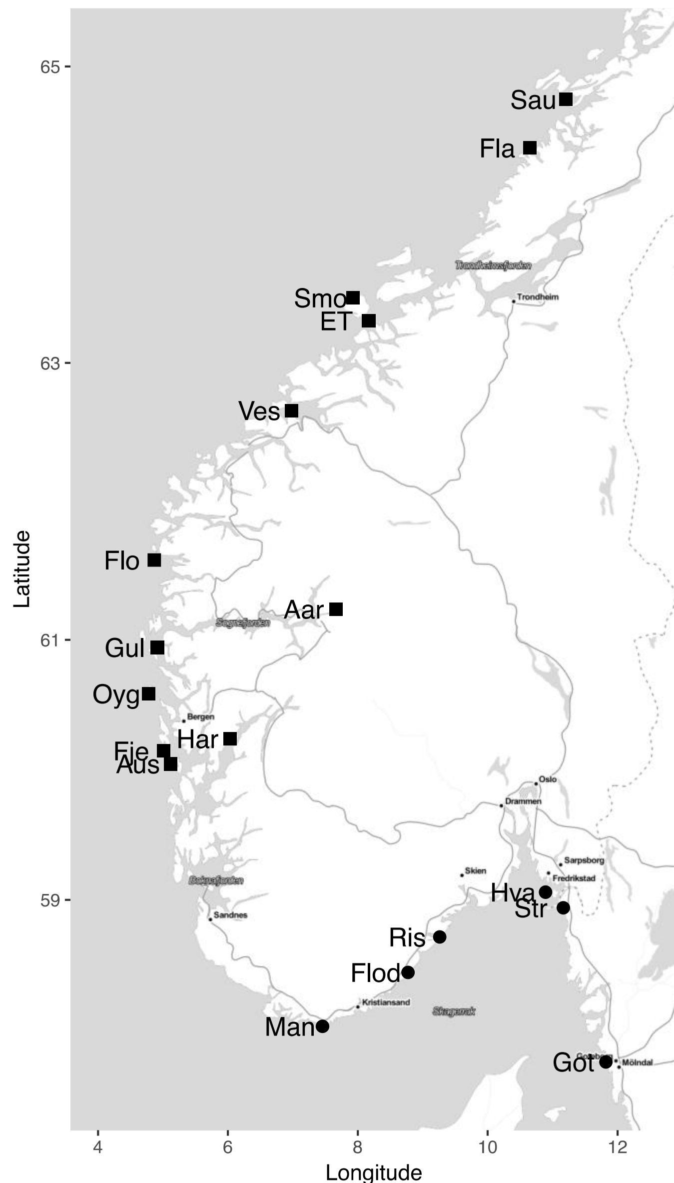


Figure 1 *Sampling locations of ballan wrasse along the Scandinavian coast. North-western samples marked with squares, south-eastern by circles. Sampling location abbreviations as given in Table 1. Map made with ggmap (Kahle and Wickham 2013).*

Photographs of individual fish were available from five locations. These individuals were visually sorted according to morphotype and have been kept separate for certain analyses as appropriate (Table 3).

Galician samples were used as an outgroup in the comparisons, as well as serving as a basis for morphotype genetics.

Box 1 Single Nucleotide Polymorphisms (SNPs) and Microsatellite (MSATs) DNA
Single Nucleotide Polymorphisms, or SNPs (pronounced "snip"), are loci (chromosome positions) at which genes differ in a single base (nucleotide) in more than 1% of the population, while remaining gene sequences are the same. SNP loci are most often bi-allelic, meaning there are two possible versions of each SNP, and as such, individuals can be either homo- or heterozygotic for a given marker. It has been shown that SNPs are relatively common throughout genomes and can be found both within coding and non-coding regions of DNA. It is technically easier to genotype more SNPs than MSATSs. This class of marker is therefore favored in modern population genetics and genomics studies where larger numbers of loci tend to reveal greater clarity in population connectivity. Microsatellites (MSATs) are short segments (1-10 nucleotides) of usually non-coding multiallelic tandem repeats, where different alleles are defined by the length of the repeating segment (i.e., different alleles have a different number of repetitions). The multiallelic nature of MSATs make them useful for determining differences in genetic diversity through measures such as total number of alleles and allelic richness (number of alleles adjusted for minimum sample size), in addition to their usefulness in determining population genetic structure through allele frequency analysis.

2.2. Sampling

Scandinavian fish were caught in coastal waters using fyke nets or pots. All samples were collected by trained individuals in accordance with EU Directive 2010/63/EU and the national legislations in the given country. Fish were killed upon catch, after which samples were either taken immediately or the fish was frozen down and stored until sampling could be conducted in laboratory facilities. Galician samples were obtained from a local fish market in La Coruña (43.359 N, 8.402 W). As the fish were obtained from commercial artisanal inshore fishing, no specific permission was required for sampling. All fin clips were stored in 96% ethanol for later analysis.

2.3. Morphological analysis

Photographs of individual fish were available from five sampling locations. All individuals from these samples were assigned to one of the morphotype groups; plain, spotted or intermediate, through visual inspection (Table 3). The intermediate group was introduced, due to less distinct and highly variable coloration patterns in Scandinavian samples, as well as variable image quality.

To investigate potential genetic differences among morphotypes, individual pairwise F_{ST} and STRUCTURE were run both on SNP and MSAT datasets for each of the morphotype sorted locations. For SNP results, a DAPC was also conducted (these methods are described in detail elsewhere in the materials and methods section).

Table 1 *Summary statistics for the Single Nucleotide Polymorphism dataset (SNPs): Number of individuals per sampling location (N), number of deviations from Hardy-Weinberg (HWE) and Linkage equilibrium (LE), expected and observed heterozygosity (H_{exp}, H_{obs}), and inbreeding coefficient (F_{IS}). Numbers in parenthesis after sequential Bonferroni corrections (B). Galician sample listed both as total sample, and as sorted by morphotype (Plain/Spotted).*

		Geographic location								
Sampling location	**Abbreviation**	**Latitude**	**Longitude**	**Year(s)**	**N**	**No deviations from HWE (B)**	**No deviations from LE (B)**	H_{exp}	H_{obs}	F_{IS}
Sauhestøya	Sau	64.784 N	11.205 E	2017	33	6 (1)	183 (1)	0.414	0.416	0.019
Flatanger	Fla	64.465 N	10.649 E	2017, 2018	57	9 (1)	168 (0)	0.410	0.401	0.040
Smøla	Smo	63.451 N	7.929 E	2017, 2018	36	5 (0)	162 (0)	0.413	0.407	0.025
Edøy-Tustna	ET	63.293 N	8.168 E	2017	25	5 (0)	136 (0)	0.403	0.399	0.033
Vestnes	Ves	62.666 N	6.982 E	2017	81	8 (1)	152 (0)	0.409	0.411	0.007
Florø	Flo	61.590 N	4.867 E	2018	29	6 (0)	143 (0)	0.412	0.414	0.010
Årdal	Aar	61.228 N	7.667 E	2018	22	4 (0)	150 (0)	0.404	0.426	-0.025
Gulen	Gul	60.943 N	4.914 E	2018	29	6 (0)	145 (0)	0.406	0.413	0.006
Øygarden	Oyg	60.595 N	4.783 E	2018	23	7 (0)	130 (0)	0.393	0.414	-0.027
Hardanger	Har	60.254 N	6.032 E	2012-2013	50	6 (0)	150 (0)	0.403	0.411	-0.012
Fjell	Fje	60.160 N	5.011 E	2018	26	6 (0)	124 (0)	0.397	0.399	0.023
Austevoll	Aus	60.062 N	5.115 E	2013	89	11 (0)	153 (0)	0.409	0.414	-0.007
Mandal	Man	57.982 N	7.457 E	2018	50	6 (0)	163 (1)	0.398	0.427	-0.056
Flødevigen	Flod	58.420 N	8.774 E	2013, 2018	197	7 (4)	144 (0)	0.398	0.393	0.022
Risør	Ris	58.704 N	9.261 E	2018	50	5 (1)	157 (1)	0.391	0.402	-0.018
Hvaler	Hva	59.061 N	10.891 E	2012	50	6 (0)	146 (0)	0.390	0.375	0.040
Strömstad, Sweden	Str	58.937 N	11.164 E	2018	47	7 (0)	163 (0)	0.395	0.389	0.025
Gothenburg, Sweden	Got	57.691 N	11.821 E	2018	49	6 (0)	162 (1)	0.395	0.390	0.026
Galicia, Spain	Gal	43.389 N	8.432 W	2014	82	9 (2)	281 (4)	0.412	0.385	0.067
Galicia Plain	GalP	43.389 N	8.432 W	2014	43	8 (2)	287 (0)	0.410	0.396	0.049
Galicia Spotted	GalS	43.389 N	8.432 W	2014	39	4 (2)	136 (0)	0.391	0.373	0.054

Table 2 *Summary statistics for the MSAT data set: Number of individuals (N) and number of alleles (N_A) per sampling location, allelic richness (A_R), number of deviations from Hardy-Weinberg (HWE) and Linkage equilibrium (LE), expected and observed heterozygosity (H_{exp}, H_{obs}), and inbreeding coefficient (F_{IS}). Numbers in parenthesis after sequential Bonferroni corrections (B). Galician sample listed both as total sample, and as sorted by morphotype (Plain/Spotted).*

Location	**N**	**N_A**	**A_R**	**No deviations from HWE (B)**	**No deviations from LE (B)**	**H_{exp}**	**H_{obs}**	**F_{IS}**
Vestnes	85	99	4.3	3 (2)	11 (0)	0.566	0.540	0.059
Hardanger	38	72	3.8	2 (0)	5 (0)	0.567	0.525	-0.016
Austevoll	88	94	4.0	2 (1)	10 (0)	0.544	0.536	0.023
Flødevigen	102	92	3.9	3 (3)	9 (0)	0.513	0.512	-0.001
Hvaler	22	69	3.8	1 (1)	6 (0)	0.538	0.493	-0.004
Strömstad	44	81	3.9	1 (1)	9 (0)	0.515	0.497	0.047
Gothenburg	50	84	4.0	3 (1)	10 (0)	0.521	0.508	0.035
Galicia	89	139	5.9	3 (1)	17 (1)	0.629	0.620	0.023
Galicia Plain	48	127	5.9	1 (1)	17 (0)	0.638	0.640	0.002
Galicia Spotted	41	121	5.7	2 (0)	10 (1)	0.600	0.600	0.022

Table 3 *Morphotype sorted locations and associated number of Plain (N_P), Spotted (N_S), and Intermediate (N_I) individuals.*

Location	**N_P**	**N_S**	**N_I**
Smøla	13	8	-
Flødevigen	31	32	63
Strømstad	21	15	11
Gothenburg	12	20	17
Galicia	43	39	-

2.4. DNA extraction, amplification and genotyping

Total genomic DNA was extracted from fin clips using the Qiagen DNeasy Blood & Tissue Kit following the manufacturer's instructions. After extraction, DNA concentration was measured using a NanoDrop 8000 spectrophotometer (ThermoFisher Scientific), and DNA was diluted to an approximate concentration of 15 ng/µl. A Selma pipetting robot was used to transfer samples from 96- to 384-well plates.

Sequences containing SNPs had been previously obtained by staff at IMR via the mining of sequence data from the ballan wrasse genome available from NCBI (BioSample: SAMEA3939555, BioProject: PRJEB13687, https://www.ncbi.nlm.nih.gov/assembly/GCF_900080235.1/). These segments made up the basis for a panel of 106 SNPs that were used in this study. The total SNP panel was divided into 4 multiplexes each consisting of 24-29 markers. Each forward-reverse primer set, un-extended extend primer (UEP), and according markers were named BaWr-1 to -106 (see Table A1-4 in appendix 1).

SNPs were amplified in 5 ml of PCR mastermix consisting of 0.5 µM Multiplex primer mix (one of four multiplexes), 25 mM dNTP, 25 mM $MgCl_2$, PCR Buffer, 5U/µl PCR Enzyme, ~15 ng/µl DNA and HPLC Grade water. Cycling conditions were 2 min denaturing at 95 °C followed by 45 cycles of 30 s at 95 °C, 30 s at 56 °C and 1 min at 72 °C, and finished with 5 minutes at 72 °C. After initial PCR, remaining primers and nucleotides were removed using Shrimp alkaline phosphatase (SAP) and followed by iPLEX (Agena Bioscience) Extension Reaction, with extension primer (UEP) mix according to multiplex (see Table A1-4 in appendix 1). SAP cycling conditions were 40 min at 37 °C, followed by 5 min at 85 °C, while iPLEX cycling conditions were 30 sec at 94 °C, followed by 40 cycles consisting of 5 sec at 94 °C and 5 cycles of 5 sec at 52 °C, 5 sec at 80 °C. The iPLEX extension reaction was finished off by 3 min at 72 °C. All thermocycler programs were set infinitely to 4 °C after the cycling program was run to prevent product degradation.

Extension products were desalted using Clear Resin before transfer onto a SpectroCHIP Array by automated nanodispenser (Agena Bioscience). SpectroCHIP arrays were analyzed using a MALDI-TOF Mass Spectrometer (Agena Bioscience MassARRAY Dx Analyzer; Box 2), that assigns genotypes to sample fragments according to mass, and the results were control checked in TyperAnalyzer (Agena Bioscience).

Box 2 MALDI-TOF Mass Spectrometer
Within a MALDI-TOF Mass Spectrometer an electric field is established in a vacuum. When running the analysis, each sample on a SpectroCHIP array is irradiated one by one by a laser, inducing ionization and desorption. The now positively charged molecules accelerate through the flight tube until hitting a detector. Different molecules are then separated by time-of-flight, which is proportional to the mass of the molecule. The detector records the relative time of flight for each sample and the results are automatically displayed using Typer software as identified SNP alleles (homozygous or heterozygous for diploid organisms).

For the microsatellite analysis, nineteen loci (WrA103, WrA107, WrA111, WrA112, WrA113, WrA203, WrA223, WrA224, WrA236, WrA237, WrA254, WrA255, WrA256, WrA259, WrA261, WrB102, WrB212, WrB213, WrB215) were used as described in Quintela et al. (2014). Microsatellites (six multiplexes) were amplified in 10 µL of PCR master mix, consisting of 50 ng DNA template, 1× buffer, 2 mM $MgCl_2$, 1.25 mM dNTPs, 0.06–0.12 µM of each primer and 1U Go-Taq polymerase. Cycling conditions were 4 min denaturation at 94 °C followed by 24 cycles of 50 sec at 94 °C, 90 sec at an annealing temperature of 56 °C, 1 min of extension at 72 °C and a final extension of 72 °C for 10 min. Forward primers were labeled with fluorescent dyes and PCR products were electrophoresed on an ABI Prism 377 Genetic Analyzer (Applied Biosystems). The 500LIZ size standard (Applied Biosystems) was used to accurately determine the size of the fragments and allelic variation. Fragments were analyzed with the software GeneMapper v5 (Applied Biosystems).

2.5. Quality control of genetic data

The microsatellite markers used here have previously been used and validated (Quintela et al. 2014, 2016). This is in contrast to the SNP markers that have been used here for the first time, and thus needed validation before using the data. Therefore, SNP multiplexes were tested and control-checked using multiple approaches described below to assess if all markers had amplified to a satisfactory degree and produced relevant, easily scorable genotypes. During the initial rounds of testing done before this study, a number of markers failed to give reliable information and were thus removed from the final multiplex designs (Table 3; Marker test).

After running the multiplexes for all samples, the genotype results were analyzed using cluster analysis in TyperAnalyzer. In the cluster analysis, it was shown that some markers failed

to properly amplify or were non-informative, i.e., showing no distinct clusters. This resulted in nine markers being removed before genotypes were exported (Table 4: TyperAnalyzer).

After importing the dataset to R (v.3.5.0; R Core Team 2008), quality control was initiated by the removal of missing data using the function missingno in the package *poppr* (Kamvar et al. 2014) with thresholds set to 30% for markers and 50% for individuals. This resulted in the removal of markers BaWr-14 and -61 (Table 4: Missing data) as well as 22 individuals.

Following the removal of the loci with missing data above the threshold, Hardy-Weinberg (HW) tests of every sampling location and SNP marker were conducted. During this analysis, markers BaWr-49 and 53 were shown to be out of HW equilibrium for most of the locations and were thus removed (Table 4: Hardy-Weinberg). In addition, marker BaWr-53 also showed non-Mendelian inheritance in a small paternity test that was conducted using a sample set of known parents and offspring from Marine Harvest Labrus AS (see Table A5 appendix 1), although this might have been affected by the small sample size.

In multiple of the marker tests, loci BaWr-57 and -79 were shown to deviate to a significant degree and were thus reanalyzed in TyperAnalyzer. The new analysis of these markers revealed that there had been technical implications during the genotyping, resulting in the decision to remove these markers from further analysis.

The extensive data quality control described above resulted in a total of 24 markers being removed, leaving a final set of 82 SNPs to be used for population genetic analysis.

Table 4 *Overview of quality control methods and corresponding removed markers.*

	Pre-genotyping marker test	TyperAnalyzer	Missing data (>30%)	Hardy-Weinberg	Paternity test
Markers removed; BaWr	2, 16, 24, 27 38, 45, 51, 56 71 106	7, 18, 26, 41, 64, 66, 73, 92 57, 79	14, 61	49, 53	53

As described above, the MassARRAY system makes genotype calls in real time, and automatically transfers the results to a database. Using TyperAnalyzer, the resulting genotypes were analyzed by cluster analysis and checked individually if necessary. Following analysis, changes were done to maximize the number of correct assignments and minimize the risk of false positives. False positives are genotype results of very low intensity (intensity lower than 0.4) or that are outside of clusters, that have still been assigned by TyperAnalyzer. Such low

intensity and non-clustering results should be removed, as they have a high chance of being wrongly assigned.

2.6. Statistical analysis of the genetic data

In order to address the primary objectives of the study, i.e., investigate population genetic structure and possible genetic differences among morphotypes, both the SNP and MSAT datasets were subject to a suite of statistical analyses described below. Most of these were computed in a suite of R packages, while others were computed in their respective packages available online (see respective sections).

2.6.1. Summary statistics, Hardy-Weinberg Equilibrium (HWE), Linkage Disequilibrium (LD), number of alleles, and allelic richness.

For all samples in both SNP and MSAT datasets, the following summary statistics were computed using the packages *adegenet* and *hierfstat* (Goudet 2005, Jombart 2008) as implemented in R: sample size, observed and expected heterozygosity as well as inbreeding coefficient (F_{IS}).

Hardy-Weinberg equilibrium (HWE; Box 3) was calculated for each sample using the *pegas* package (Paradis 2010). This tests the hypothesis that the observed genotype frequencies follow Hardy-Weinberg equilibrium for each marker in the dataset either using a chi-square test or, if a number of permutations is given, based on Monte Carlo permutations of alleles, and results in a table of chi-square values (and corresponding P values) for each SNP. A significant p-value from the test, signifies that the population is out of Hardy-Weinberg equilibrium for the given marker (i.e., in Hardy-Weinberg disequilibrium). Depending on the genotype frequencies, this can mean a number of things, e.g., selection pressure for the given allele, or that you are in fact sampling from multiple populations (termed the Wahlund effect; a deficiency of heterozygotes signifies mechanical mixing of two populations).

Box 3 Hardy-Weinberg
A Hardy-Weinberg population is a theoretical population based on the assumptions of no selection, no mutation, no migration (i.e., no geneflow), infinitely large population size, and totally random mating (i.e., all alleles have the same chance of being inherited by the next generation). For a Hardy-Weinberg population, the genotype frequencies for the offspring generation is then purely a function of the allele frequencies in the parental generation. If the allele frequencies of the alleles A_1 and A_2 are said to be p and q, it follows that the genotype frequencies in the offspring generation would then be p^2 for A_1A_1, q^2 for A_2A_2, and 2pq for A_1A_2. A natural population that (more or less) follows these proportions is said to be in Hardy-Weinberg equilibrium.

Linkage describes the degree to which to two loci are inherited together. If the two loci are independent of each other (i.e., not inherited together in any higher proportion than they would be by chance), they are said to be in linkage equilibrium. In contrast, if two loci are inherited together at a proportion higher than they would by chance they are said to be in linkage disequilibrium (LD). LD between all locus pairs per population as well as across all populations was examined using the program GENEPOP 7 (Rousset 2008) with the following Markov chain parameters: 10000 steps of dememorization, 100 batches and 5000 iterations per batch. Signification was assessed after the *post hoc* sequential Bonferroni correction (Holm 1979) implemented in the calculator developed by Gaetano (2013).

For the MSAT dataset, the number of alleles per loci and allelic richness was calculated following the method described in El Mousadik and Petit (1996) as implemented in the *hierfstat* R-package. Allelic richness is a measure of genetic diversity that is adjusted for the minimum sampling size as opposed to the total number of alleles.

2.6.2. Pairwise F_{ST} (SNPs and MSATs)

Pairwise F_{ST} (Box 4) matrices were produced for both SNPs and MSATs, for all sampling locations and samples split into morphotypes (i.e., spotted, plain and intermediate), following the method described by Nei (1987) as implemented in the program Arlequin v.3.5.1.2. (significance estimated after 10000 permutations; Excoffier et al. (2005), Excoffier and Lischer (2010)). For the location sorted datasets all 82 SNP markers were used, while only 61 markers were used for the morphotype sorted datasets (due to the requirements of Arlequin). For MSATs, both location and morphotype sorted datasets, all 19 markers were used.

Box 4 F_{ST}
F_{ST}, sometimes called “fixation index”, is a measure of the genetic distance between populations, based upon the loss of heterozygosity that results from allele frequencies in finite populations drifting towards fixation (Herron and Freeman 2015). F_{ST} is calculated by the formula $F_{ST} = \frac{H_T - H_S}{H_T}$, where H_T is the heterozygosity in the total population expected under Hardy-Weinberg equilibrium, and H_S is the heterozygosity in the sample (i.e., potential subpopulation). F_{ST} is thus a ratio varying between 0 and 1, where 0 represents a population with no substructuring (i.e., no difference in heterozygosity between samples), while 1 represents a totally divided population (i.e., subpopulations fixed for different alleles), i.e., the higher the F_{ST}, the larger the difference between sample groups (Allendorf et al. 2013, Herron and Freeman 2015).

2.6.3. Bayesian clustering - STRUCTURE analysis (SNPs and MSATs)

Bayesian clustering analysis uses genetic data to organize a set of individuals into a user-defined number of genetic groups without taking sampling location into consideration. This permits the identification of cryptic genetic variation within and among population samples. Bayesian clustering analysis was performed using STRUCTURE v.2.3.4 (Pritchard et al. 2000), using an admixture model and correlated allele frequency without population information. In an admixture model, all individuals are expected to have some fraction of their genome originating from a combination of one to all genetic clusters (K). The analysis was conducted using the program ParallelStructure (Besnier and Glover 2013) which distributes the jobs between parallel processors in order to significantly speed up the computational time.

Ten runs with a burn-in period consisting of 100000 replications and a run length of 1000000 Markov chain Monte Carlo (MCMC) iterations were performed for K = 1 to 5 clusters. To determine the number of genetic groups in which the samples could be divided into, STRUCTURE output was analyzed using two approaches. Firstly, the *ad hoc* summary statistic ΔK of Evanno et al. (2005) was calculated, which is based on the rate of change of the ‘estimated likelihood’ between successive K values and allows the determination of the uppermost hierarchical level of structure in the data. Secondly, StructureSelector (Li and Liu 2018) was used to estimate four alternative statistics (MedMed, MedMean, MaxMed and MaxMean; see example in appendix 1 (Fig. A4)), which have been described as more accurate than the previously used methods to determine the real number of clusters, for both even and uneven sampling data. Finally, the ten runs for the selected Ks were averaged with CLUMPP v.1.1.1 (Jakobsson and Rosenberg 2007) using the FullSearch algorithm and the G’ pairwise matrix similarity statistic, and graphically displayed using bar plots.

In the resulting plot, each bar represents a single individual and the proportion of the genome that originated from each K is indicated by different colors. STRUCTURE was run for both SNP and MSAT datasets, as well as for single locations where morphotype information was available (with both sets of markers).

2.6.4. Discriminant Analysis of Principle Components (SNPs and MSATs)

Discriminant Analysis of Principle Components (DAPC; Jombart et al. 2010) were conducted in *adegenet* with clustering based on the sampling locations for both the SNP and MSAT datasets with and without Galician samples. For SNPs, DAPCs were also conducted with and without three outlier loci. For all DAPCs (both SNPs and MSATs), 60 Principle Components (PCs) were kept (explaining > 80% of the variation). Number of Discriminant Analyses (DAs) kept varied between datasets and is indicated by shading of DA eigenvalues in the plot. DAPC optimizes between cluster variation while minimizing within group variation, thus giving good indications for the clustering of the samples.

2.6.5. Principle Components Analysis PCA (SNPs)

As DAPC minimizes within group variation, an unclustered PCA was also conducted for the SNP dataset (without Galicia) using the R packages *adegenet* and *ade4* (Jombart 2008, Dray and Dufour 2015). Due to the high number of individuals and relatively low differentiation, it was determined to split the dataset into four. Each individual was therefore assigned to one of four groups. Individual PCA plots were then produced, with each plot representing approximately ¼ of each sample. Color and plot symbols were based upon sampling locations, to be able to determine if any individuals clustered outside their sampling location.

2.6.6. Outlier detection (SNPs)

Three different approaches were used to detect putatively outlier, i.e., non-neutral, loci (Beaumont and Nichols 1996, Foll and Gaggiotti 2006, Excoffier et al. 2009). A consensus approach was followed to minimize the risk of detecting false positives, retaining only loci that more than two of the three approaches highlighted as deviating from neutrality. The first approach to detect outliers was the hierarchical Bayesian method described in Beaumont and Balding (2004) as implemented in BayeScan (v.2.1; Foll and Gaggiotti 2008). This method is based on the principle that genetic differentiation among populations living in contrasting environments is expected to be different for loci under selection than for the rest of the genome. BayeScan estimates population-specific F_{ST} coefficients and uses a cut-off based on the mode

of the posterior distribution (Foll and Gaggiotti 2008). The program was run by setting sample size to 10000 and the thinning interval to 50 as suggested by Foll and Gaggiotti (2008) resulting in a total chain length of 550000 iterations. Loci with a posterior probability over 0.99 were retained as outliers, corresponding to a Bayes Factor >2 (i.e., "decisive selection" (Foll and Gaggiotti 2006)) and providing substantial support for the model.

The second approach to detect outliers was using the Fdist approach by Beaumont and Nichols (1996) implemented in LOSITAN (Antao et al. 2008). In this approach, loci with an unusually high F_{ST} are putatively under directional selection, while loci with low F_{ST} value are considered to be potentially under stabilizing selection. A neutral distribution of F_{ST} with 100000 iterations was simulated, with forced mean F_{ST} at a significance level of 0.05 under an infinite alleles model. This method also implements a multiple test correction based on false discovery rates (FDR) that is fundamental to avoid high overestimation of the percentage of outliers (e.g., 1% of false positive with a threshold of 99%).

The third approach to detect outliers was based on the allele frequency method implemented in PCAdapt (Luu et al. 2017). In this approach, population structure is first assessed via PCA, and outliers are thereafter detected with respect to how they relate to population structure. Cattell's graphical rule was used to select the number of principal components (K) to identify SNPs deviating from neutrality. Outliers were selected by performing the q-value procedure at false discovery rate (FDR) 0.0.

2.6.7. Allele frequency plots (SNPs)

Allele frequencies for all SNP loci were found through the use of the function basic.stats in *hierfstat*. A single allele ("the first") was then chosen for plotting. Loci that best showed the structuring of the stock was used to indicate structuring between north-west and south-east Scandinavia, as well as loci showing latitudinal trends. Frequencies of loci that were shown to be under selection in Scandinavia and Galicia, respectively, were plotted for all locations. Allele frequencies of loci under selection in Galicia were also plotted for Scandinavian morphotype split locations. All plots were produced in *ggplot2* (Wickham 2009) for better presentation.

2.6.8. Cline analysis (SNPs)

Hybrid zones between genetically-divergent populations or species are often characterized by a sigmoid curve showing the expected allele frequencies along a geographic axis. Cline analysis is used to estimate the shape, center and width of the sigmoid curves generated by molecular, phenotypic or environmental markers and to test for concordance and coincidence between

markers (Gay et al. 2008). A geographic cline analysis was conducted with the R package *HZAR* (Derryberry et al. 2014) over a 1500 km transect starting in Sauhestøya and finishing in Gothenburg (i.e., the Scandinavian samples). The fifteen models implemented in *HZAR* were fitted to the normalized loading on the first principal component analysis (PCA) axis based both on the panel of 82 SNPs as well as to the allele frequency of every individual locus to determine the position, width and shape of clines over the total geographic distance. The reference cline was built using STRUCTURE Q-score for the total dataset and, in both cases, the best cline model was decided upon AIC scores. Clines were considered significantly displaced if the two log-likelihood unit support limits of the cline center did not overlap with the STRUCTURE Q-score ($Qb = 1 - Qs$).

3. Results

The final datasets consisted of 82 SNP markers genotyped on 1025 ballan wrasse from 19 locations, and 19 microsatellite markers genotyped on a sub-set of 513 ballan wrasse from 8 locations. Wrasse from five of these locations were also characterized according to morphotype (plain, spotted, intermediate).

Overall observed heterozygosity was fairly similar across sampling locations for SNPs, with values ranging from 0.375 to 0.426, with a value of 0.385 observed for the Galician sample. Within Scandinavia, the mean was found to be 0.405. In general, a slightly higher degree of heterozygosity was observed in north-western (NW) compared to south-eastern (SE) Scandinavia (means of 0.410 and 0.396, respectively (two-sample $t(6.0) = 1.9$, $p = 0.104$); Table 1).

Similar variation was observed for microsatellites, with values ranging from 0.493 to 0.540 in Scandinavia (mean = 0.516). Between NW and SE, a significant difference (two-sample $t(4.8) = 4.9$, $p = 0.005$) was observed with means of 0.534 and 0.503, respectively. A slightly higher observed heterozygosity of 0.620 was observed in the Galician sample.

Within the MSAT dataset, the total number of alleles (N_A) and allelic richness (A_R) per sample in Scandinavia ranged from 69 to 99 (mean = 84) and 3.8 to 4.0 (mean = 4.0) respectively. No differentiation was observed between the two Scandinavian clusters (two sample t-tests: N_A: $t(3.3) = 0.7$, $p = 0.522$; A_R: $t(2.2) = 0.9$, $p = 0.46$), while values were considerably higher in Galicia ($N_A = 139$, $A_R = 5.9$), thus demonstrating greater genetic diversity in southern latitudes (N_A: $p < 0.0001$; A_R: $p < 0.0001$).

3.1. Outlier tests (SNPs)

The three approaches used to detect outliers in the SNP data set produced partially overlapping results (Table 5). Within Scandinavia, all three programs were in consensus about markers BaWr-46 and -82. In addition, BayeScan and LOSITAN both flagged marker BaWr-60. Within Galicia, no loci were flagged as under selection by BayeScan, while the loci BaWr-22, -32, -58, and -97 were in consensus between LOSITAN and PCAdapt. Pairwise F_{ST}, STRUCTURE and DAPC were all conducted with and without inclusion of markers under selection. As no significant change in clustering was observed in the results with or without outliers, all markers were retained for the other analyses.

Table 5 *Overview of SNP markers under directional selection within Scandinavia and within Galicia from outlier tests using three different programs/approaches.*

Program	Scandinavia; BaWr-	Galicia; BaWr-
BayeScan	46, 60, 82	-
LOSITAN	46, 60, 82	22, 32, 46, 52, 58, 70, 97, 105
PCAdapt	29, 40, 46, 47, 82	22, 32, 58, 80, 82, 93, 97, 99
Consensus (2/3)	46, 60, 82	22, 32, 58, 97

3.2. Genetic differentiation among samples (SNPs and MSATs)

Within Scandinavia, genetic differentiation among samples was low to moderate with pairwise F_{ST} levels ranging from 0.001 ($p = 0.287$) to 0.041 for SNPs (Table 6) and 0.025 to 0.040 for MSATs (Table 7). However, two very distinct genetic clusters representing north-western (NW) and south-eastern (SE) Scandinavia were observed in both datasets (Table 6 and 7). Within the two identified clusters in Scandinavia, pairwise genetic differentiation was generally very low with few values reaching statistical significance. Looking at SNPs, maximum pairwise F_{ST} values were 0.009 within the NW cluster and 0.004 within the SE cluster. This demonstrates significant and very clear genetic differences between, but not within the two major groups identified in Scandinavia.

A very high degree of genetic differentiation was observed between samples from Scandinavia and Galicia, with all pairwise F_{ST} values ranging from 0.106 to 0.148 (Table 6). Slightly higher genetic differentiation was observed between the SE Scandinavian cluster and Galicia (pairwise F_{ST} = 0.139 – 0.148), than between the NW cluster and Galicia (pairwise F_{ST} = 0.106 – 0.144). Additionally, differentiation was especially high towards the Galician plain morphotype (NW: 0.104 – 0.144; SE: 0.143 – 0.154). To test if the differentiation between the two Scandinavian clusters and the two Galician morphotypes differed from each other, a significance test based on resampling 1000 combinations of 30 random SNPs was conducted for all combinations of NW, SE, GalP, and GalS. These significance tests demonstrated that the Galician Plain morphotype (GalP) were genetically more similar to the NW Scandinavian samples, than to the SE Scandinavian samples ($p = 0.011$), while all other combinations between the two groups in Scandinavia and Galicia morphotyped samples were non-significant ($p > 0.05$).

Although with a reduced set of samples, the general patterns reported for SNPs as described above were also reflected in results from the MSAT dataset (Table 7).

3.3. Patterns in allele frequencies (SNPs)

Looking at the allele frequencies, some SNPs displayed striking trends along the north-south gradient, and/or between the three genetic clusters identified from pairwise F_{ST} analysis (NW and SE Scandinavia, and Galicia). Six markers (~7%) showed low allele frequencies in Scandinavia (with relatively little variation) in contrast with very high frequencies in the Galician samples (Fig. 2a). Marker BaWr40 was especially interesting, displaying a very low frequency of the A allele in all Scandinavian samples while having a frequency of ~0.7 in Galicia.

Other loci displayed clear allele-frequency changes along the north-south gradient (Fig. 2b). Loci putatively identified as being under directional selection (consensus across LOSITAN, BayeScan, and/or PCAdapt) in Scandinavia were shown to vary in allele frequencies between the two Scandinavian clusters as expected (Fig. 2c). Of these, marker BaWr46 was especially noteworthy, with a frequency around 0.5 in the NW cluster increasing to a level very close to fixation in the SE cluster (~0.97). Markers under selection in Galicia showed little variation within Scandinavia both regarding latitude (Fig. 2d) and morphotype assignment (Fig. 2e).

Table 6 *Pairwise genetic differentiation (F_{ST}; below diagonal, with associated p-values after sequential Bonferroni correction above) between sampling locations genotyped at 82 SNP loci, including Galicia, Spain (Pooled, Plain morph (GalP) and Spotted morph (GalS)). Statistically significant results after sequential Bonferroni correction are indicated in bold. Color gradient represents relative values and indicate low (green) to high (red) differentiation.*

		North-western (NW) Scandinavia												South-eastern (SE) Scandinavia						Spain (SP)		
		Sau	**Fla**	**Smo**	**ET**	**Ves**	**Flo**	**Aar**	**Gul**	**Oyg**	**Har**	**Fje**	**Aus**	**Man**	**Flod**	**Ris**	**Hva**	**Str**	**Got**	**Gal**	**GalP**	**GalS**
NW	**Sau**	*	0.712	0.567	0.269	0.675	0.760	0.113	0.083	0.224	**0.049**	**0.007**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
NW	**Fla**	0.000	*	0.998	0.882	1.000	0.774	0.776	0.552	0.159	0.106	0.161	0.679	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
NW	**Smo**	0.000	0.000	*	0.503	0.493	0.352	0.061	0.208	0.285	0.239	**0.025**	0.190	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
NW	**ET**	0.002	0.000	0.000	*	0.657	0.699	0.189	0.810	0.169	0.145	0.086	**0.008**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
NW	**Ves**	0.000	0.000	0.000	0.000	*	0.660	**0.047**	**0.020**	0.131	0.060	**0.002**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
NW	**Flo**	0.000	0.000	0.001	0.000	0.000	*	0.073	0.822	0.487	0.195	0.052	**0.007**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
NW	**Aar**	0.004	0.000	0.005	0.003	**0.004**	0.005	*	0.283	0.956	0.126	0.084	1.000	**0.024**	0.089	**0.048**	**0.021**	0.225	0.287	**0.000**	**0.000**	**0.000**
NW	**Gul**	0.004	0.000	0.002	0.000	**0.005**	0.000	0.001	*	0.355	0.335	0.111	0.072	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
NW	**Oyg**	0.002	0.003	0.001	0.004	0.003	0.000	0.000	0.000	*	0.958	0.271	0.927	**0.000**	**0.000**	**0.000**	**0.000**	**0.001**	**0.000**	**0.000**	**0.000**	**0.000**
NW	**Har**	**0.004**	0.002	0.001	0.003	0.002	0.002	0.003	0.000	0.000	*	0.096	0.055	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
NW	**Fje**	**0.009**	0.002	**0.007**	0.005	**0.008**	0.006	0.005	0.004	0.002	0.003	*	0.724	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
NW	**Aus**	**0.007**	0.000	0.001	**0.006**	**0.008**	**0.006**	0.000	0.003	0.000	0.002	0.000	*	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
SE	**Man**	**0.031**	**0.020**	**0.026**	**0.031**	**0.032**	**0.034**	**0.005**	**0.031**	**0.017**	**0.035**	**0.023**	**0.011**	*	**0.007**	**0.026**	**0.013**	0.346	0.163	**0.000**	**0.000**	**0.000**
SE	**Flod**	**0.035**	**0.019**	**0.025**	**0.029**	**0.034**	**0.031**	0.003	**0.029**	**0.016**	**0.031**	**0.015**	**0.012**	**0.003**	*	0.109	0.643	0.775	0.269	**0.000**	**0.000**	**0.000**
SE	**Ris**	**0.039**	**0.025**	**0.031**	**0.032**	**0.035**	**0.033**	**0.005**	**0.028**	**0.019**	**0.034**	**0.015**	**0.010**	**0.004**	0.001	*	0.740	0.653	0.051	**0.000**	**0.000**	**0.000**
SE	**Hva**	**0.038**	**0.022**	**0.029**	**0.030**	**0.035**	**0.035**	**0.007**	**0.030**	**0.019**	**0.036**	**0.016**	**0.011**	**0.004**	0.000	0.000	*	0.540	0.447	**0.000**	**0.000**	**0.000**
SE	**Str**	**0.032**	**0.019**	**0.023**	**0.025**	**0.033**	**0.029**	0.002	**0.028**	**0.012**	**0.032**	**0.016**	**0.012**	0.000	0.000	0.000	0.000	*	0.329	**0.000**	**0.000**	**0.000**
SE	**Got**	**0.041**	**0.023**	**0.031**	**0.031**	**0.038**	**0.039**	0.001	**0.033**	**0.019**	**0.035**	**0.020**	**0.012**	0.001	0.001	0.003	0.000	0.001	*	**0.000**	**0.000**	**0.000**
SP	**Gal**	**0.116**	**0.106**	**0.115**	**0.107**	**0.109**	**0.120**	0.121	**0.122**	**0.144**	**0.129**	**0.120**	**0.122**	**0.140**	**0.139**	**0.140**	**0.142**	**0.142**	**0.148**	*	0.475	0.343
SP	**GalP**	**0.110**	**0.105**	**0.110**	**0.104**	**0.104**	**0.112**	0.125	**0.119**	**0.144**	**0.125**	**0.120**	**0.121**	**0.146**	**0.143**	**0.149**	**0.148**	**0.146**	**0.154**	0.000	*	**0.000**
SP	**GalS**	**0.132**	**0.121**	**0.133**	**0.121**	**0.124**	**0.139**	0.133	**0.136**	**0.156**	**0.143**	**0.133**	**0.133**	**0.149**	**0.144**	**0.146**	**0.149**	**0.147**	**0.153**	0.001	**0.025**	*

Table 7 *Pairwise genetic differentiation (F_{ST}; below diagonal, with associated p-values after sequential Bonferroni correction above) between sampling locations genotyped at 19 microsatellites, including Galicia, Spain (Pooled, Plain morph, Spotted morph). Statistically significant results after sequential Bonferroni correction are indicated in bold. Color gradient represents relative values and indicate low (green) to high (red) differentiation.*

		North-west. (NW) Scandinavia			South-eastern (SE) Scandinavia				Spain (SP)		
		Vestnes	**Hardanger**	**Austevoll**	**Flødevigen**	**Hvaler**	**Strömstad**	**Gothenburg**	**Galicia**	**Galicia Plain**	**Galicia Spotted**
NW	**Vestnes**	*	**0.0017**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**
	Hardanger	**0.0089**	*	**0.0017**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**
	Austevoll	**0.0094**	**0.0097**	*	**0.0000**	**0.0484**	**0.0002**	**0.0010**	**0.0000**	**0.0000**	**0.0000**
SE	**Flødevigen**	**0.0362**	**0.0402**	**0.0105**	*	0.4185	**0.0148**	**0.0254**	**0.0000**	**0.0000**	**0.0000**
	Hvaler	**0.0358**	**0.0363**	**0.0068**	0.0005	*	0.8566	0.3406	**0.0000**	**0.0000**	**0.0000**
	Strömstad	**0.0341**	**0.0372**	**0.0111**	**0.0053**	0.0000	*	0.4379	**0.0000**	**0.0000**	**0.0000**
	Gothenburg	**0.0249**	**0.0353**	**0.0084**	**0.0043**	0.0017	0.0004	*	**0.0000**	**0.0000**	**0.0000**
SP	**Galicia**	**0.0621**	**0.0685**	**0.0649**	**0.0806**	**0.0732**	**0.0830**	**0.0703**	*	0.8285	0.4793
	Galicia Plain	**0.0580**	**0.0670**	**0.0647**	**0.0836**	**0.0706**	**0.0812**	**0.0675**	0.0000	*	**0.0000**
	Galicia Spotted	**0.0799**	**0.0839**	**0.0801**	**0.0938**	**0.0906**	**0.1017**	**0.0891**	0.0001	**0.0191**	*

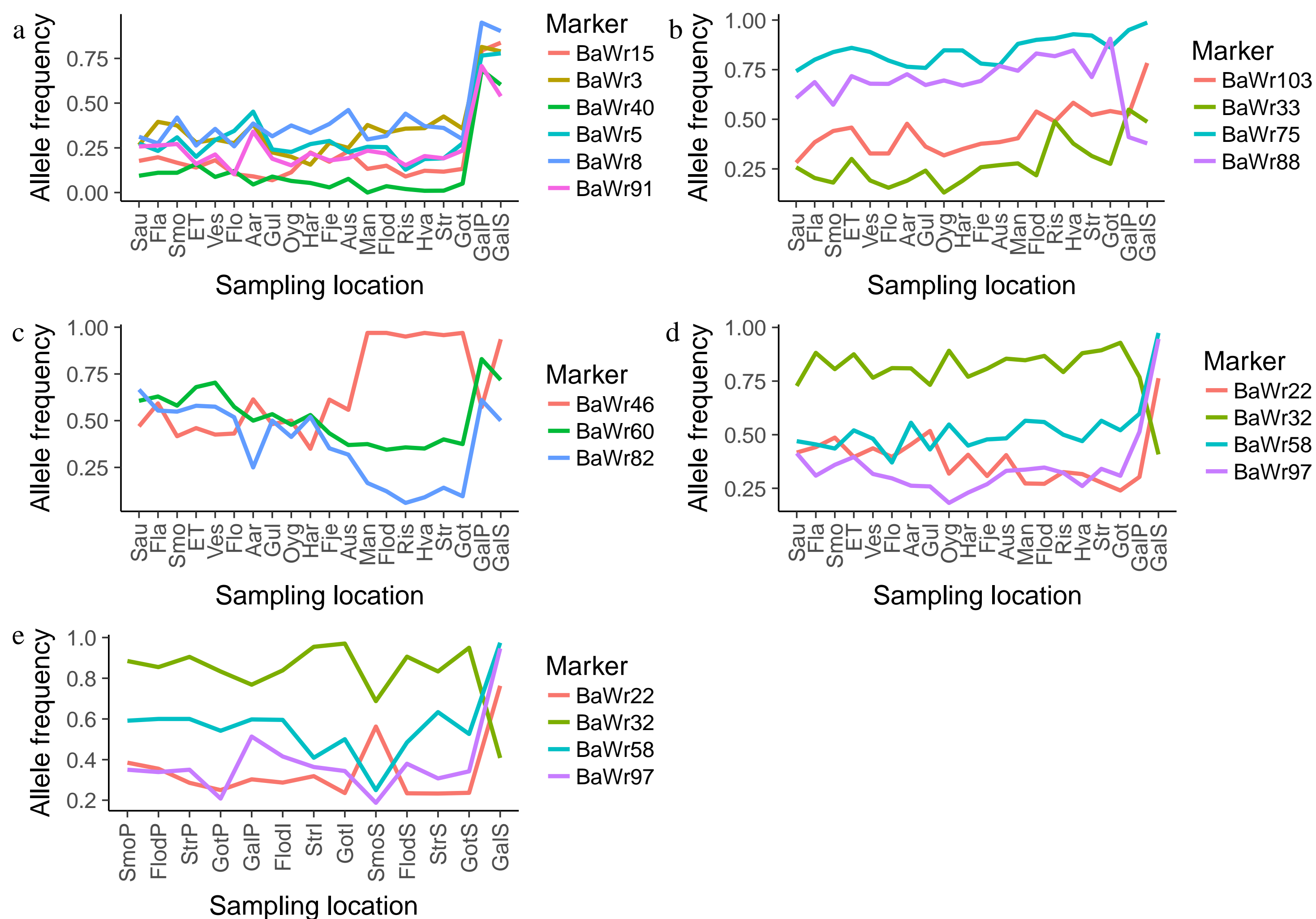


Figure 2 *Examples of allele frequencies of some SNP markers for ballan wrasse. Graphs illustrate SNPs that display clear allele frequency trends; a) between Scandinavia and Galicia, Spain, b) with latitude, c) for markers under selection in Scandinavia, d) for markers under Selection in Galicia, e) for markers under selection in Galicia in all morphotype split locations (ordered by latitude and phenotype; P: plain, I: intermediate, S: spotted).*

3.4. Discriminant Analysis of Principle Components (SNPs and MSATs)

Discriminant analysis of principle components (DAPC; Fig. 3) showed similar trends in clustering as pairwise F_{ST} (Table 7 and 8) for both SNPs (Fig. 3a, b) and MSATs (Fig. 3c, d), i.e., samples clustered into three main groups, NW Scandinavia, SE Scandinavia and Galicia. Looking at the Scandinavian samples alone (Fig. 3b, d) the resolution of the clustering is increased, thus showing that the degree of overlap between samples within clusters is generally higher than between clusters, i.e., samples within the same cluster are genetically more similar. This same pattern of clustering was found both with and without three loci under directional selection (see Fig. A2 and A3 in appendix 2).

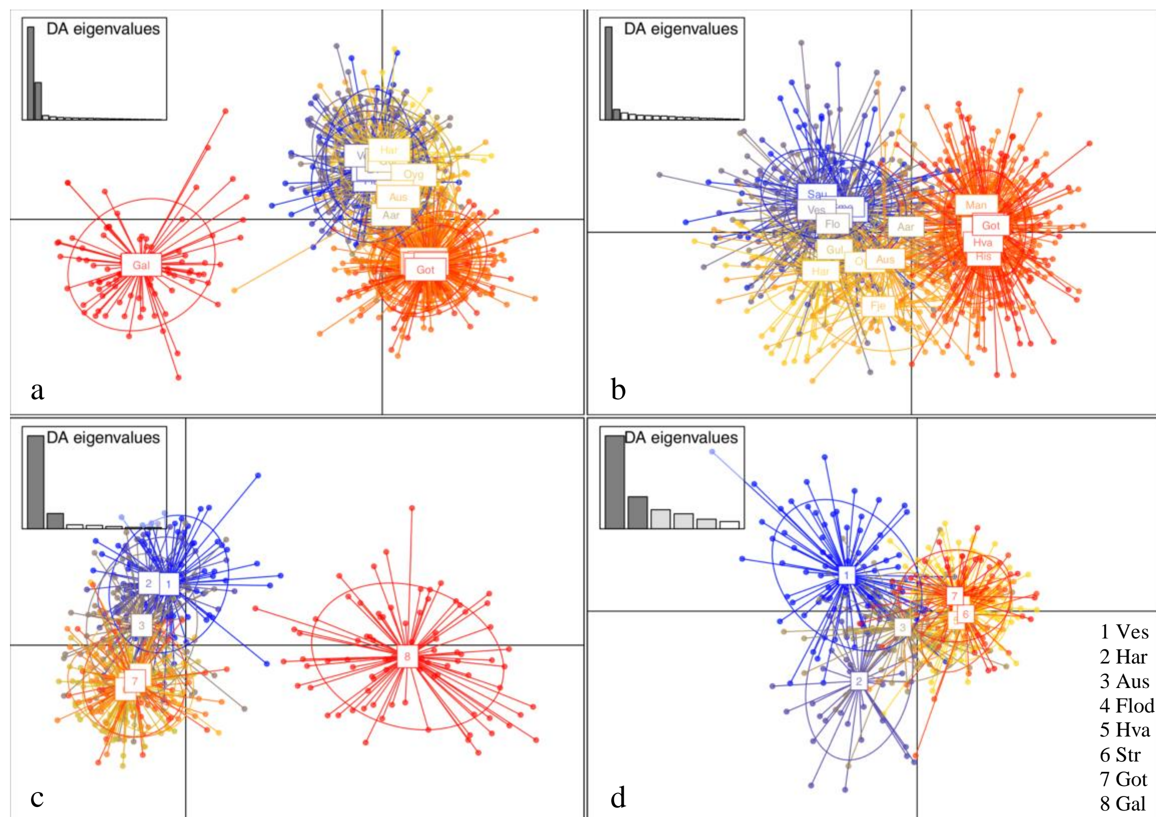


Figure 3 *Discriminant analysis of principle components (DAPC) for ballan wrasse samples. 82 markers where used for SNPs (a, b), while 19 markers where used for MSATs (c, d). Plots showing all samples (a, c) and Scandinavian samples only (b, d). DAPCs were also conducted without three markers under selection with no change in overall clustering (see Fig. A2 and A3 in appendix 2). Projected inertia % for the axes: a) PC1 59.6%, PC2 24.0%; b) PC1 61.7%, PC2 7.0%, c) PC1 76.8%, PC2 12.7%, d) PC1 53.0%, PC2 18.2%, PC3 10.9%. Color indicates sample site. Shading of DA eigenvalues indicate DAs kept.*

3.4.1. Unclustered Principle Component Analysis (SNPs and MSATs)

The overall trend of the unclustered PCA (Fig. 4) follows the lines of the other methods of analysis with individuals from NW and SE Scandinavian locations clustering largely independent of each other. All samples showed relatively little variation along both principle components, with PC1 and PC2 accounting for about 6% and 3% of the variation, respectively. In general, individuals from NW and SE Scandinavia do cluster on opposite sides, but due to the low degree of variation quite a lot of overlap is observed. It was noted that one individual from Austevoll (Fig. 4d), was shown to differ from all other individuals.

3.5. Bayesian clustering – STRUCTURE (SNPs and MSATs)

The analysis computed by STRUCTURE on the SNP data set showed K = 3 as the most likely number of clusters, both according to Evanno test (ΔK=1360) and in agreement with the four estimators (MedMedK, MedMeanK, MaxMedK, MaxMeanK) implemented in STRUCTURESelector. This outcome corroborated the results from pairwise F_{ST} and DAPC and showed that the samples can be organized into the following genetic groups; NW and SE Scandinavia, and Galicia (Fig. 5; K = 3). The removal of the three SNPs that were identified to be under directional selection reduces clarity of the three genetic groups slightly, but the general pattern still remains (Fig. 5b). Although with a reduced number of samples, the MSAT dataset also showed the same general trends (Fig. 5c).

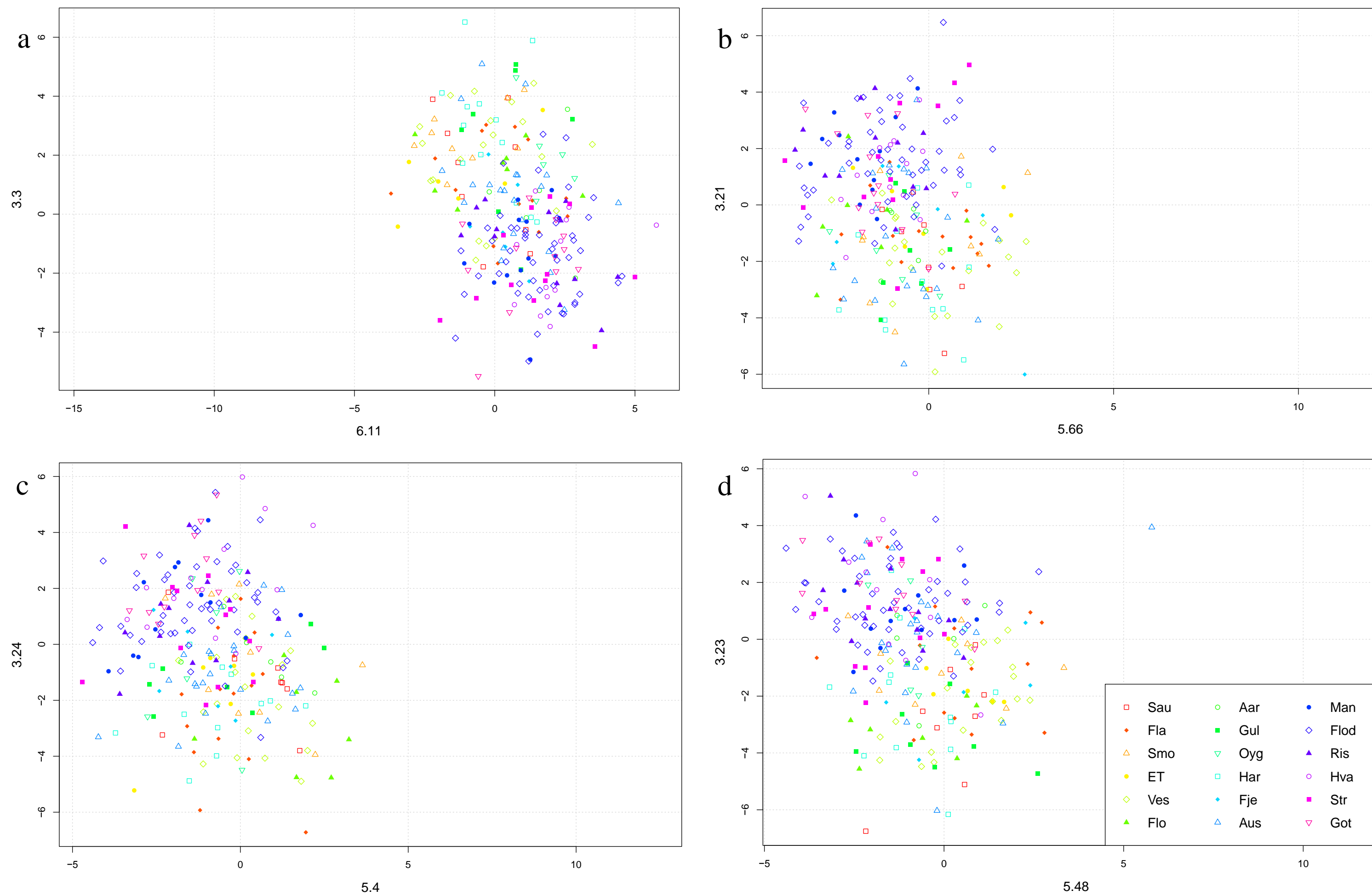


Figure 4 *Principle Component Analysis (PCA) of Scandinavian populations, with the x-axis representing the first and the y-axis representing the second principle component. Values on axes given in percent. Each plot contains approximately ¼ of individuals from each population. All plots show general clustering of individuals from north-western and south-eastern locations, respectively, clustering together. Some individuals group together with the opposite general cluster. Legend representative for all four plots (a-d).*

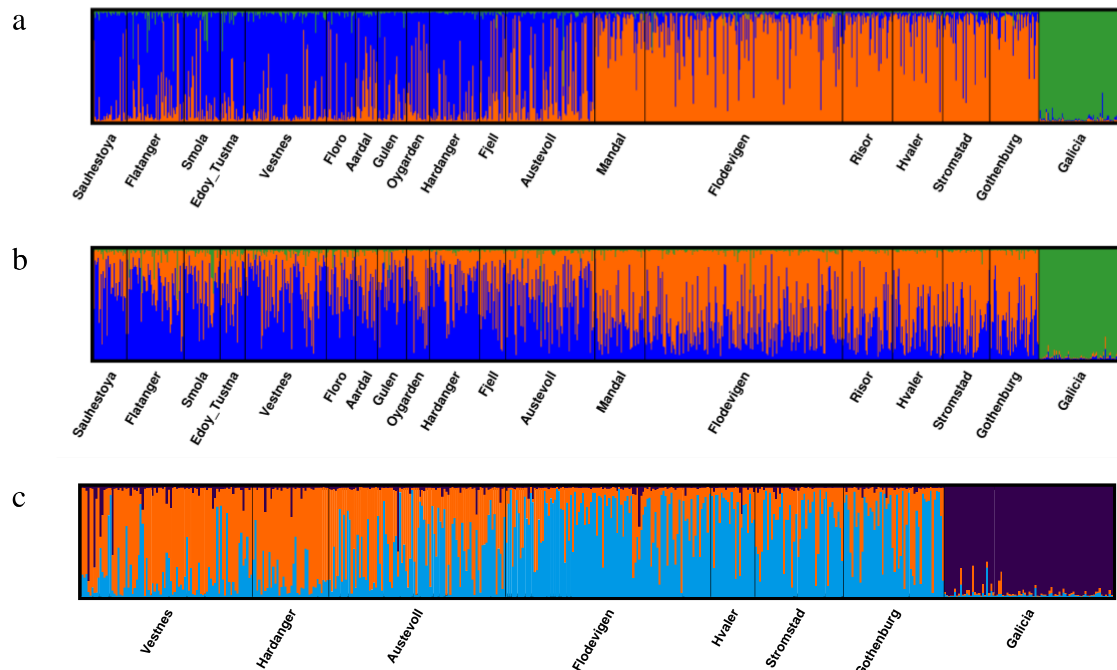


Figure 5 *STRUCTURE bar plots for K = 3 for a) 82 SNPs, b) after removal of 3 loci under directional selection (79 SNPs), and c) 19 MSATs. Each vertical line represents one individual, and colors indicate genetic clusters (i.e., groups). Majority of individuals in north-western Norway being assigned mostly to cluster 1, while south-eastern individuals are assigned to cluster 2, with Galician individuals making up cluster 3.*

3.6. Cline analysis (SNPs)

The reference cline based on the STRUCTURE Q-score fitted a fixL model with the center situated at 874 km from Sauhestøya and showing a width of 139 km (Fig. 6a). This area corresponds to the Jæren coastline, the most extensive area of continuous sandy substrate found along the Norwegian coast (Blanco Gonzalez et al. 2016). Out of the 82 SNP loci tested, 46 did not fit any of the 15 cline modes as their allele frequencies were stable throughout the 1500 km of transect separating Sauhestøya and Gothenburg. In contrast, the remaining 36 loci showed clines with centers ranging between 295 and 1386 km from the starting point (see Table A6 in appendix 2 for full details). Seven of the cline centers were found to overlap with the reference cline (Fig. 6b) with one of these markers, BaWr46, depicted as under directional selection within Scandinavia.

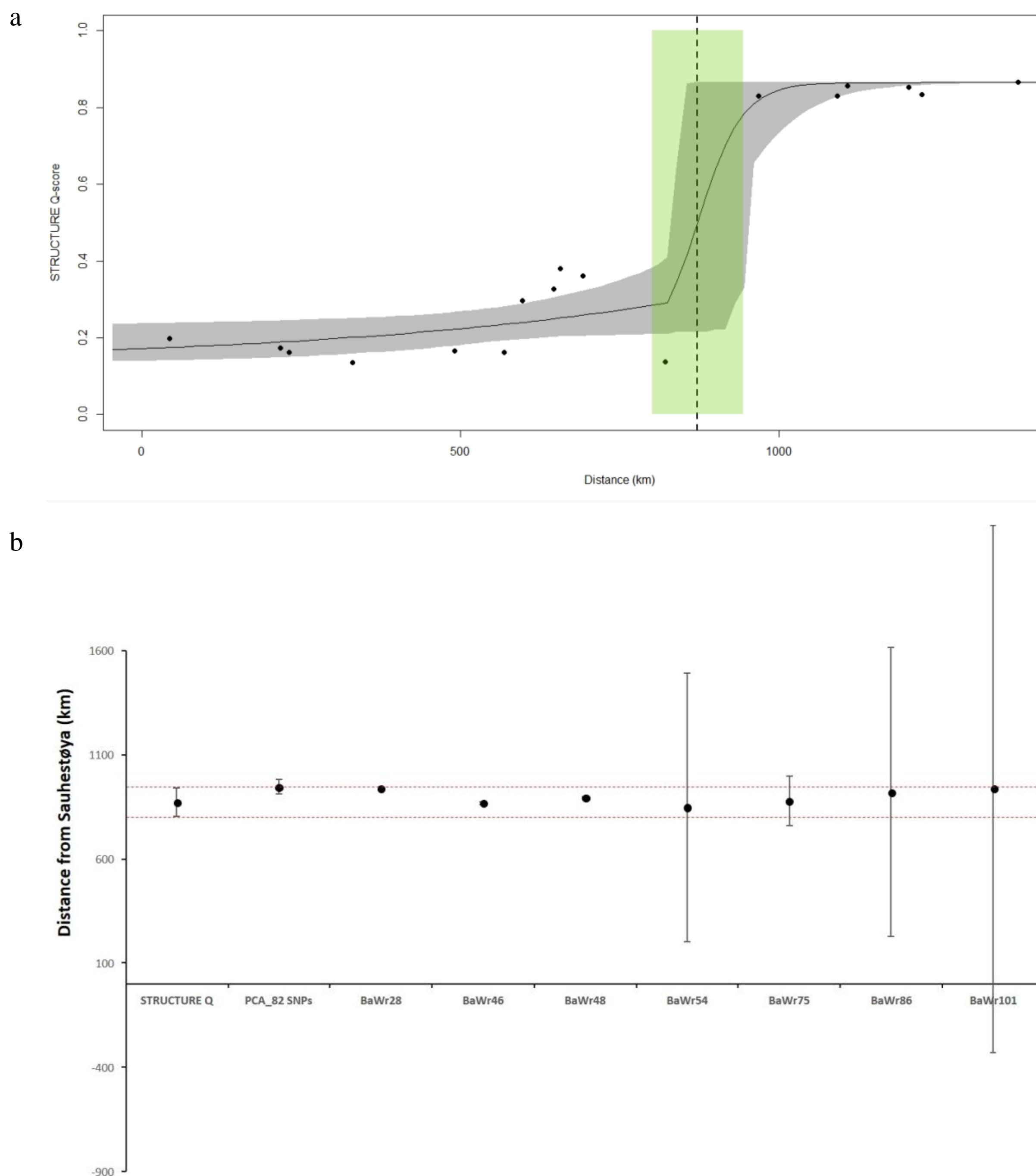


Figure 6 *Geographical cline analysis for ballan wrasse across a 1500 km transect running along the Scandinavian coast from Sauhestøya in Trøndelag, Norway to Gothenburg, Sweden for: a) Cline fitting the STRUCTURE Q-score, with the center of the cline depicted by the vertical dashed line and the width highlighted in green, b) Compilation of the position of the clines (center and width) for markers that fit within the STRUCTURE reference as cline depicted by dotted red lines.*

3.7. Genetic differentiation among morphotypes (SNPs and MSATs)

Large and highly significant genetic differentiation was observed between the spotted and plain wrasse sampled in Galicia, as revealed by SNPs: $F_{ST} = 0.031$, $p < 0.001$ (Table 8), and MSATs: $F_{ST} = 0.020$, $p < 0.001$ (Table 9). However, and in stark contrast, no genetic differentiation was revealed among spotted, intermediate or plain morphotypes in any of the Scandinavian samples (Table 8 and 9).

Overall, morphotype split data showed the same general patterns of genetic differentiation as location sorted datasets (Table 6 and 7). Meaning that genetic differentiation was observed between samples from the previously defined clusters NW and SE Scandinavia, while little to no differentiation was found between samples from the same genetic clusters. This was observed independent of morphotype assignment. Additionally, samples from both Scandinavian clusters showed high F_{ST} values towards both Galician morphotypes, though, as for location sorted datasets, samples from NW Scandinavia showed lower differentiation towards Galicia than the SE cluster, especially towards the plain morphotype.

Similar patterns were also reflected in the DAPC of morphotype sorted locations (Fig. 7), with Scandinavian samples all clustering together independent of morphotype. Additionally, Galician morphotypes clustered independent of each other (Fig. 7a). In the DAPC of only Scandinavian samples (Fig. 7b), samples from Smøla (the only NW Scandinavian sample) clustered away from the SE Scandinavian samples. It should be noted that the five first Principle Components all explained more than 10% of the variation within Scandinavia.

Table 8 *Pairwise genetic differentiation (F_{ST}; below diagonal, with associated p-values after sequential Bonferroni correction above) between morphotype split samples (from north-western (NW) and south-eastern (SE) Scandinavia, as well as one sample from Galicia, Spain) based on 61 SNP loci (after Arlequin purging of loci not meeting the 0.05 allowed level of missing data). Results statistically significant after sequential Bonferroni corrections indicated in bold. Color gradient represents relative values and indicate low (green) to high (red) differentiation. P: plain, I: intermediate, S: spotted.*

		NW Scandinavia		**South-eastern (SE) Scandinavia**									**Spain (SP)**	
		SmoP	**SmoS**	**FlodP**	**FlodS**	**FlodI**	**StrP**	**StrS**	**StrI**	**GotP**	**GotS**	**GotI**	**GalP**	**GalS**
NW	**SmoP**	*	0.687	0.002	0.003	0.006	0.025	0.003	0.064	0.001	**0.001**	**0.000**	**0.000**	**0.000**
	SmoS	0.000	*	0.254	0.097	0.183	0.036	0.431	0.350	0.056	0.030	0.072	**0.000**	**0.000**
SE	**FlodP**	0.017	0.002	*	0.665	0.479	0.257	0.823	0.591	0.039	0.243	0.149	**0.000**	**0.000**
	FlodS	0.018	0.010	0.000	*	0.048	0.127	0.442	0.884	0.019	0.374	0.216	**0.000**	**0.000**
	FlodI	0.014	0.007	0.000	0.005	*	0.513	0.742	0.954	0.110	0.404	0.338	**0.000**	**0.000**
	StrP	0.015	0.016	0.002	0.005	0.000	*	0.453	0.706	0.347	0.265	0.055	**0.000**	**0.000**
	StrS	0.022	0.001	0.000	0.000	0.000	0.001	*	0.710	0.523	0.450	0.281	**0.000**	**0.000**
	StrI	0.014	0.003	0.000	0.000	0.000	0.000	0.000	*	0.493	0.594	0.577	**0.000**	**0.000**
	GotP	0.030	0.016	0.009	0.012	0.007	0.002	0.000	0.000	*	0.036	0.112	**0.000**	**0.000**
	GotS	**0.027**	0.019	0.002	0.001	0.001	0.004	0.001	0.000	0.013	*	0.377	**0.000**	**0.000**
	GotI	**0.026**	0.014	0.004	0.003	0.002	0.009	0.003	0.000	0.008	0.002	*	**0.000**	**0.000**
SP	**GalP**	**0.103**	**0.109**	**0.136**	**0.146**	**0.132**	**0.147**	**0.138**	**0.153**	**0.165**	**0.152**	**0.152**	*	**0.000**
	GalS	**0.126**	**0.123**	**0.141**	**0.153**	**0.137**	**0.154**	**0.141**	**0.158**	**0.171**	**0.157**	**0.153**	**0.031**	*

Table 9 *Pairwise genetic differentiation (*F_{ST}*; below diagonal, with associated p-values after sequential Bonferroni correction above) between morphotype split samples (from south-eastern (SE) Scandinavia, as well as one sample from Galicia, Spain) based on 19 microsatellite loci. Results statistically significant after sequential Bonferroni corrections indicated in bold. Color gradient represents relative values and indicate low (green) to high (red) differentiation. P: plain, I: intermediate, S: spotted.*

		South-eastern (SE) Scandinavia									**Spain (SP)**	
		FlodP	**FlodS**	**FlodI**	**StrP**	**StrS**	**StrI**	**GotP**	**GotS**	**GotI**	**GalP**	**GalS**
SE	**FlodP**	*	0.078	0.318	0.209	0.197	0.791	0.010	0.295	0.023	**0.000**	**0.000**
	FlodS	0.014	*	0.034	0.020	0.014	0.212	0.019	0.064	0.038	**0.000**	**0.000**
	FlodI	0.003	0.013	*	0.604	0.361	0.509	0.152	0.445	0.119	**0.000**	**0.000**
	StrP	0.007	0.017	0.000	*	0.812	0.575	0.289	0.279	0.694	**0.000**	**0.000**
	StrS	0.009	0.023	0.002	0.000	*	0.766	0.778	0.380	0.332	**0.000**	**0.000**
	StrI	0.000	0.010	0.000	0.000	0.000	*	0.334	0.314	0.578	**0.000**	**0.000**
	GotP	0.030	0.025	0.008	0.005	0.000	0.006	*	0.261	0.317	**0.000**	**0.000**
	GotS	0.005	0.014	0.001	0.004	0.004	0.006	0.006	*	0.040	**0.000**	**0.000**
	GotI	0.019	0.015	0.007	0.000	0.003	0.000	0.003	0.013	*	**0.000**	**0.000**
SP	**GalP**	**0.098**	**0.083**	**0.086**	**0.084**	**0.080**	**0.081**	**0.066**	**0.091**	**0.060**	*	**0.000**
	GalS	**0.104**	**0.088**	**0.099**	**0.103**	**0.105**	**0.094**	**0.094**	**0.105**	**0.081**	**0.020**	*

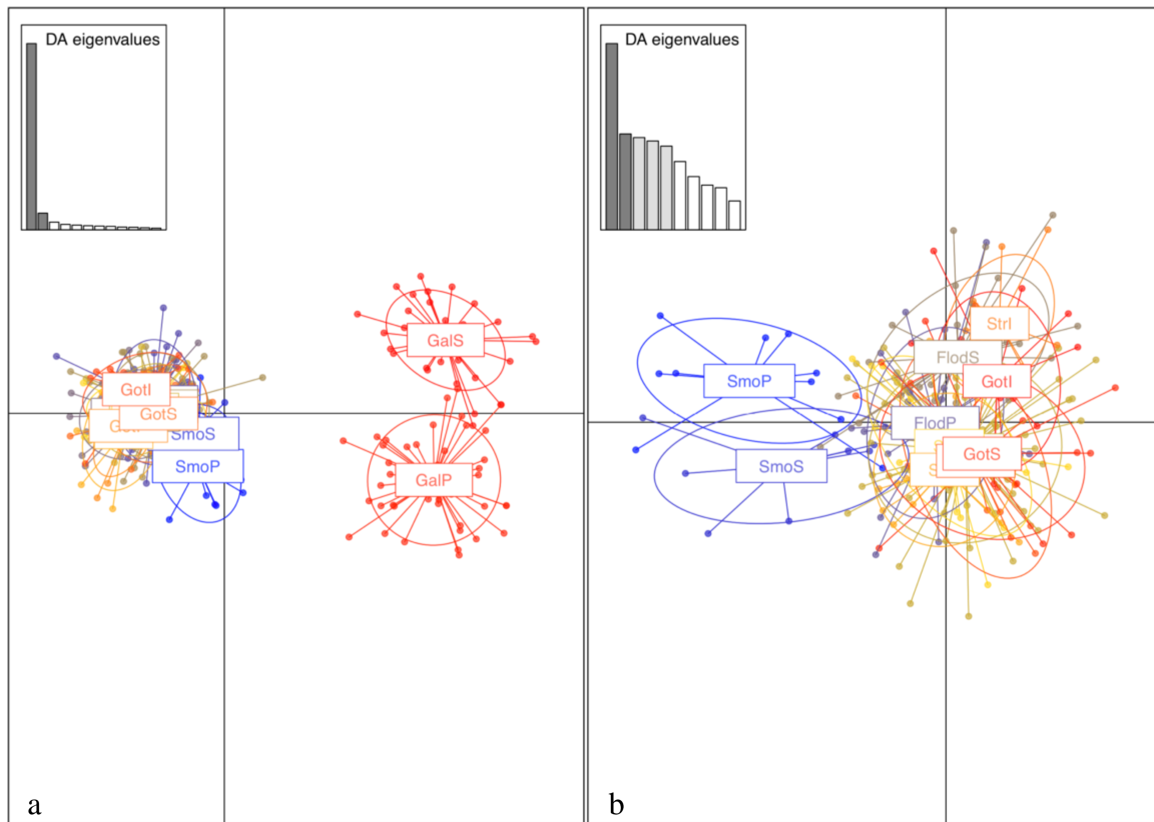


Figure 7 *DAPC of morphotype sorted locations, a) with and b) without Galicia. Color indicates sampling location. Projected inertia % for the axes: a) PC1 77.1%, PC2 6.9%; b) PC1 23.7%, PC2 12.2%, PC3 12.0%, PC4 11.3%, PC5 10.7%. Color indicates sample site. Shading of DA eigenvalues indicate DAs kept. P: plain, I: intermediate, S: spotted.*

3.8. Bayesian clustering of morphotypes (SNPs and MSATs)

Running individual STRUCTURE (K = 2) for the different morphotype sorted samples showed distinct clustering in Galicia using SNPs with plain individuals being assigned to cluster 1 to a very high degree while spotted individuals are assigned to cluster 2 in a similarly consistent manner (Fig. 8a). This pattern was however not reflected in the MSATs results (Fig. 8b).

For the Scandinavian morphotype sorted samples, all individuals were assigned to both clusters at much the same degree, independent of morphotype assignment. These results were mirrored for both types of markers (see example in Fig. 9).

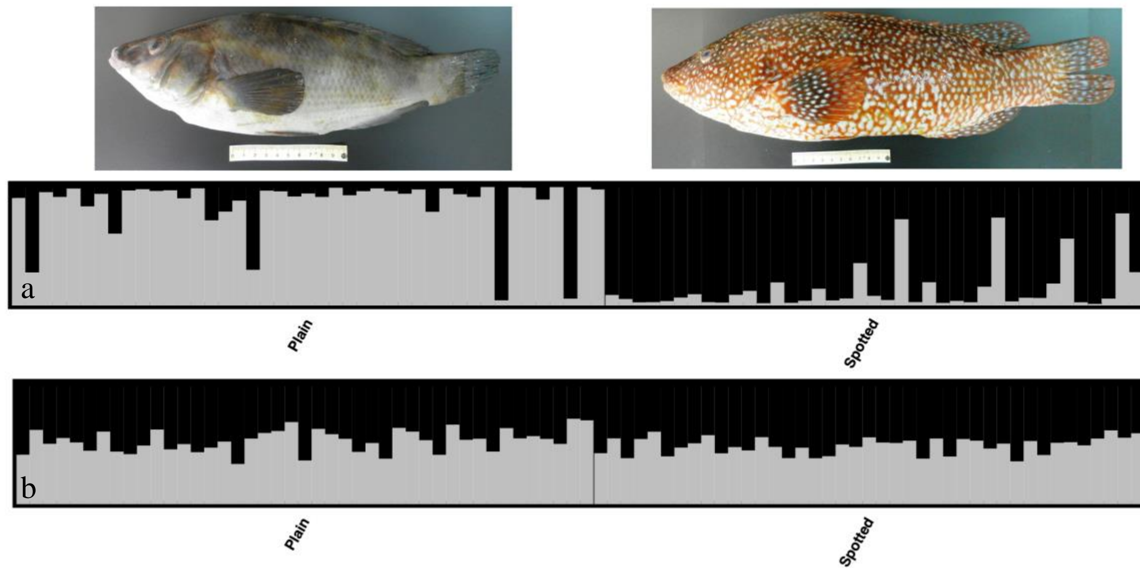


Figure 8 *Examples of plain and spotted fish sampled in Galicia, Spain. Structure bar plot at K = 2 for both morphotypes, for a) SNPs and b) MSATs. Each vertical line represents an individual, while color represents genetic clusters (K).*

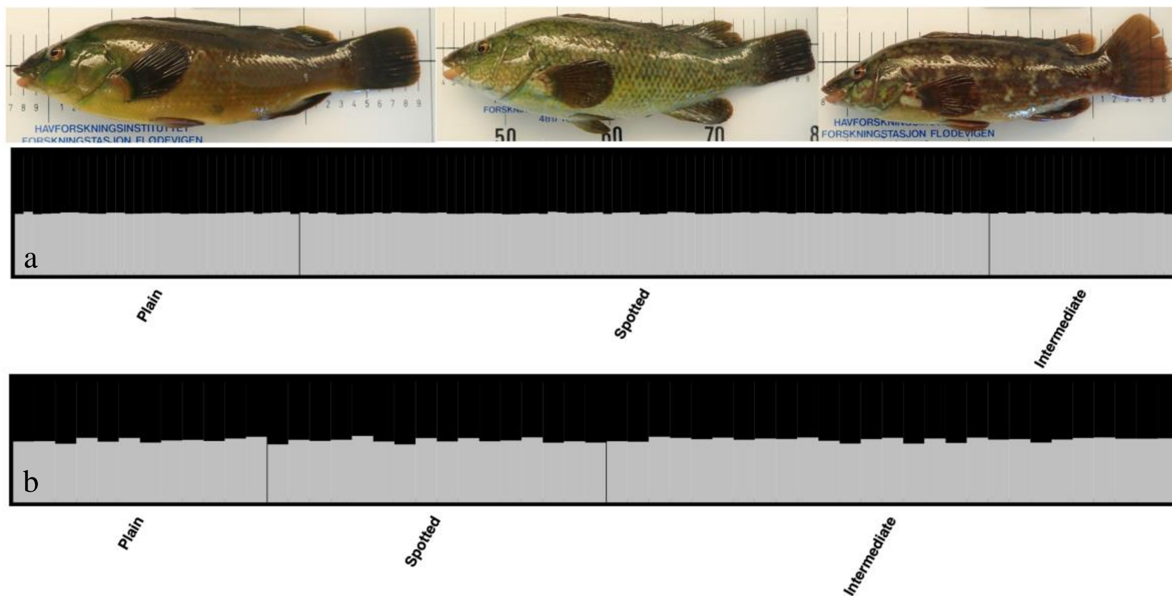


Figure 9 *Examples of plain, spotted and intermediate morph fish sampled in Flødevigen, Norway. STRUCTURE bar plot at K = 2 for individuals from this site that could be unequivocally classified into any of these categories, for a) SNPs and b) MSATs. Each vertical line represents an individual, while color represents genetic clusters (K). The same general trend, i.e., lack of morphotype-differentiation, is representative for all available Scandinavian samples (not presented).*

4. Discussion

This study represents the most extensive analysis of ballan wrasse genetics thus far. Following the analysis of newly developed SNPs and previously existing MSATs on >1000 individuals collected from 19 locations, the following main results and conclusions were obtained:

1. Two distinct genetic groups of ballan wrasse were identified in Scandinavia, with a clear break centered in the south-western part of Norway. Despite clear genetic differences between these two groups, within them, very little evidence of population genetic structuring was identified (best illustrated in Fig. 5, Table 6). ***It is therefore suggested that these two genetic groups are shaped by differences in colonization histories and are held partially isolated by a combination of restricted adult movement, restricted egg dispersal (due to benthic eggs), and habitat discontinuities in south-western Norway.***

2. All Scandinavian samples displayed lower genetic variation than samples from Galicia. Ballan wrasse from Galicia were also (marginally) more similar genetically to those from north-western (NW) Scandinavia than to the wrasse from south-eastern (SE) Scandinavia. Furthermore, ballan wrasse from NW Scandinavia were genetically more similar to the plain wrasse from Galicia, than to the spotted wrasse from Galicia. In contrast however, no difference in the magnitude of genetic differentiation was observed between the ballan wrasse from SE Scandinavia and any of the Galician morphotypes (best illustrated in Table 6). ***These results suggest that founder effects and genetic bottlenecks have characterized the colonization process of Scandinavia from southern latitudes, and, that NW Scandinavia was colonized with ballan wrasse of both spotted and plain morphotypes from southern latitudes prior to southern Norway and western Sweden. This is also in accordance with proposed deglaciation patterns of the Scandinavian peninsula (Stroeven et al. 2016), which show that western Norway became ice free earlier than SE Scandinavia.***

3. Highly distinct genetic differentiation was identified between the spotted and plain morphotypes in Galicia, but not among morphotypes in Scandinavia where the phenotypic variation was also less distinct. ***It is therefore concluded that in Galicia, spotted and plain ballan wrasse are genetically distinct and are thus likely to represent sympatric sub-***

species as tentatively suggested earlier (Quintela et al. 2016). However, the mechanisms and processes maintaining the genetic differences between the clear morphotypes in Galicia, for example assortative mating, are at best cryptic or non-existent in Scandinavia where they are genetically similar and phenotypic variation is less distinct.

4.1. Genetic diversity in northern Europe

For much of the last glacial maximum (LGM: ~21 kya; Lambeck et al. (2010)), the majority of the North Sea was covered by an ice sheet (CLIMAP Project Members 1984), with the polar front reaching as far south as the Bay of Biscay (Hayes et al. 2005) and possibly even the western Iberian coast (CLIMAP Project Members 1984). Thus, with most of northern European habitats likely being unavailable during the LGM, species currently inhabiting northern European waters must have survived in glacial refugia. For ballan wrasse, areas along the Iberian and North-African coasts could have served such as purpose as these areas upheld surface temperatures within the species thermal limits during the LGM (Almada et al. 2012, 2017). In addition, multiple possible glacial refugia have been proposed for other marine species, such as around the British Isles (Almada et al. 2012) and even in the North Sea (Gysels et al. 2004, Robalo et al. 2011). The observed differences in genetic variation and degree of genetic differentiation between ballan wrasse in different parts of Europe might therefore be caused by recolonization from different glacial refugia as proposed by Almada et al. (2017). Especially as both D'Arcy et al. (2013) and Almada et al. (2017) found ages of spatial expansion of Scandinavian ballan wrasse predating LGM in their analyses of mutation rate of mtDNA (assumed 5%/My).

A decrease in mtDNA haplotype and nucleotide diversity was found by D'Arcy et al. (2013) when comparing ballan wrasse from southern Norway to Great Britain (GB), and by Almada et al. (2017) when comparing ballan wrasse from southern Norway to GB and continental Europe (France and western Iberia) as well as the Azores. In the current study, Scandinavian samples consistently displayed a lower number of alleles and showed ~50% decrease in allelic richness for the polymorphic microsatellite markers in comparison with the sample from Galicia. Based on results of the present and earlier studies, the higher degree of genetic variation observed in southern and central Europe indicates that populations in these parts of the distribution range have undergone differentiation over a longer period than northern populations and/or that northern populations have gone through genetic bottlenecks resulting from founder effects and the process of colonization. Additionally, similar genetic structuring

between southern Scandinavia and the Atlantic (British Isles and western Iberia), as well as decrease in genetic variation, has also been found in corkwing wrasse (Robalo et al. 2011, Knutsen et al. 2013). For both these species, Scandinavia represents the northern limit of their distribution. Therefore, a decrease in genetic diversity is expected to the north, as populations at the outskirts of geographical distributions generally hold lower genetic variation, especially if the environment in these areas also represent extremes of the species tolerance (Johannesson and André 2006).

4.2. Colonization history of Scandinavian ballan wrasse

From the observed genetic patterns (current study as well as D'Arcy et al. (2013) and Almada et al. (2017)), labrids mainly being tropical to temperate in distribution (highest diversity of labrid species found in tropical areas (Hanel et al. 2002)), and newly proposed thermal optima of ballan wrasse activity (Yuen et al. 2019), it is derived that Scandinavian ballan wrasse originate from southern latitudes, possibly from around the Iberian peninsula. As such, with ballan wrasse being a shallow water coastal species, colonization of Scandinavia would most likely have followed the coastline northward, opening up for two possible main routes.

A potential route of colonization of Scandinavia could have followed the coastline of continental Europe to southern Scandinavia and further along the Norwegian coast. Colonization along this route should likely have resulted in a latitudinal gradient of increasing genetic differentiation from southern samples (as well as decreasing genetic variation). However, this was not observed in the current study. Contrarily, the genetic cluster covering south-eastern Scandinavia showed a higher degree of genetic differentiation towards Galicia compared to the north-western cluster. It is therefore suggested that an alternative route of colonization was probably followed. Similar to what was proposed for the Scandinavian colonization of corkwing wrasse by Knutsen et al. (2013), ballan wrasse colonization could have moved along the coasts of the British isles and from there followed the major ocean currents to the western coast of Norway. Expansion along the Scandinavian coast could thus have spread from a center on the Norwegian west coast, including migration of a few individuals or drifting progeny over what is today the Jæren coastline, that subsequently established populations along the coastline of present-day SE Scandinavia. Depending on timing of colonization of the Scandinavian peninsula this second colonization route might also fit better with climatological records (e.g. Stroeven et al. 2016) and structure of the European

coastline after the LGM (as most of what is the North Sea today was dry land during and immediately following the LGM, due to lower sea levels (Anonymous 2018)).

As the deglaciation of the Scandinavian peninsula after the LGM took place over more than 10 000 years, the extent of ice sheet varied greatly and different areas were liberated from the ice at different times (Stroeven et al. 2016). Looking at the proposed deglaciation pattern of Fennoscandia by Stroeven et al. (2016), the general trend of melting seems to have followed an eastward gradient. According to this melting pattern, the western and southern coasts of Norway became ice free 19-20 kya while the eastern Norwegian and western Swedish coasts became ice free sometime between 17 and 14 kya. Following the assumption that ballan wrasse recolonization followed deglaciation (either closely or with a delay), it is therefore logical that western areas may have been colonized first, before eastern areas were available. The possibility of population expansion and subsequent founder effect from western to eastern Scandinavia thus presents itself as the most prominent route of colonization and plausible explanation of genetic differentiation within Scandinavia.

The observed genetic break between ballan wrasse samples in Scandinavia as described here, as well as previously reported breaks between southern Scandinavia and both the British Isles and continental Europe (D'Arcy et al. 2013, Almada et al. 2017), illustrates an interesting colonization history that is in further need of elucidation. Future studies of ballan wrasse genetics would benefit from including samples from the entire distribution range, especially the British Isles and Denmark to Atlantic France, in addition to employing multiple types of markers.

4.3. Comparison with other species in Scandinavia; with a special focus on wrasse

As previously mentioned, a genetic break was found within Scandinavia splitting the ballan wrasse into a NW and SE Scandinavian cluster. The cline analysis centered this break around the coast of southern Rogaland, south-west Norway, an area of coastline defined by the longest continuous stretch of sandy bottom substrate along the Norwegian coast (the Jæren beaches; ~26 km; Blanco Gonzalez et al. (2016)) and a relatively wide stretch of open water (Boknafjorden). As ballan wrasse is a rocky bottom species with demersal spawning (Quignard and Pras 1986, Darwall et al. 1992), as well as displaying a very short home range (Villegas-Ríos et al. 2013b), this area with its associated habitat discontinuity would likely work as an expansion barrier resulting in and maintaining the two observed genetic clusters (i.e., reducing

migration and gene-flow across this divide). These types of genetic breaks in fish associated with discontinuities in rocky substrates by sandy areas are not especially common in the scientific record, but examples are known (see e.g., Bernardi 2000, Riginos and Nachman 2001). In addition, similar clustering centered around the same area has previously been reported in studies of the related labrid corkwing (Blanco Gonzalez et al. 2016, Faust et al. 2018), as well as in other hard bottom species inhabiting the same waters e.g., sugar kelp (*Saccharina latissima* L., 1753; Evankow et al. (2019), Halvor Knutsen personal communication, 15 May, 2019).

Interestingly, the genetic break revealed here for the ballan wrasse across south-western Norway was not observed in a study of goldsinny wrasse by Jansson et al. (2017) covering the same area. Those authors revealed a genetic isolation-by-distance relationship along the Norwegian and Swedish coastlines without any distinct genetic breaks. However, and in contrast to both ballan and corkwing wrasse that spawn in algal nests on rocky substrates (Hilldén 1984 as cited in Muncaster et al. 2010, Quignard and Pras 1986, Darwall et al. 1992), goldsinny wrasse is a pelagic spawner (Quignard and Pras 1986). Thus, Jansson et al. (2017) concluded that the habitat discontinuity across the Jæren coastline is not enough to induce a similar degree of population structuring in goldsinny wrasse, as pelagic eggs could easily drift across the divide. It is therefore likely that non-pelagic spawning in ballan and corkwing wrasse is a key reason for the clear genetic differentiation observed in these two species across this region, while not in goldsinny wrasse.

4.4. Morphological variation and associated genetics

There are two very distinct morphotypes of ballan wrasse known to be present in the southern parts of its distribution range; the spotted and plain. In Galicia, these morphotypes are known and marketed under different common names (Villegas-Ríos et al. 2013a). Also, in that region the two morphotypes have been shown to display overlapping yet differing life-history strategies, with plain fish investing more in reproduction while spotted fish grow larger and faster (Villegas-Ríos et al. 2013a). These observations indicate that the two morphotypes represent different genetic groups, and the possibility of speciation has been put forward (Villegas-Ríos 2013, Quintela et al. 2016).

Quintela et al. (2016), were the first authors to identify a genetic difference between ballan wrasse morphotypes. These authors used microsatellites and found weak but significant genetic differentiation between sympatric plain and spotted morphotypes in Galicia. The

present study expanded upon their work by analyzing morphotyped samples from five separate locations with two independent sets of genetic markers (SNPs and MSATs). Looking at samples from Galicia, differentiation of microsatellites between morphotypes was similar here to what was presented by Quintela et al. (2016), i.e., weak but statistically significant differentiation between morphotypes. Significantly however, the SNPs developed and implemented here revealed very clear genetic differentiation between the two morphotypes in Galicia. This was clearly illustrated by the results from STRUCTURE, pairwise F_{ST} as well as DAPCs. Therefore, the present study both confirms suggestions from earlier studies, and indeed provides stronger evidence that in the southern part of its distribution range, two morphotype-associated genetic groups exist in sympatry, and may potentially be regarded as sub-species.

In contrast to the samples from Galicia, no evidence of genetic differentiation was found between morphotypes in any of the four Scandinavia samples, using either SNPs or MSATs. Furthermore, while the two morphotypes have been reported as being present in Norwegian waters (Villegas-Ríos et al. 2013a), sorting fish into clear morphotypes from the pictures was highly challenging (personal observations), and a third category, intermediate, was required. It is therefore proposed that the two genetic groups may be upheld through assortative mating in southern latitudes, but not so in the northern latitudes. Spotting patterns in fish may be both environmentally and genetically-determined (Jørgensen et al. 2018). Therefore, it is speculated that the prevailing climatic conditions (temperature/light/population density) in the northern regions hampers the development of clear phenotypic differences between the groups, thus preventing phenotype-assortative mating. In turn, this leads to a break-down of genetic differentiation among morphotypes in this geographic region. It is also possible that due to the founder effect, whereby a higher number of plain than spotted morphotype ballan wrasses colonized the waters along the Scandinavian peninsula as weakly suggested by the presented data, may have contributed to the lack of morphological differentiation in Scandinavia. Furthermore, alone or in addition to the above mentioned-plasticity mechanisms, it is possible that lower density of northern populations has resulted in the breakdown of assortative mating between these two morphotypes in Scandinavian waters.

4.5. Management implications

There has been a recent move towards a more sustainable wrasse fishery via multiple approaches. These include: species-specific minimum sizes in 2015 (14 cm for ballan wrasse); setting a total catch quota of 18 million fish; dividing the distribution area into three

management zones with specific catch quotas (2018 and 2019: south – 4 million; west – 10 million; central Norway – 4 million individuals) (Norwegian Directorate of Fisheries; www.fiskridir.no). Even so, the results from the present study have clear implications to the current management of the species.

4.5.1. Potential inadvertent translocation via aquaculture

Currently, aquaculture demand for cleaner fish exceeds the local supply in areas of high salmonid production, and very extensive movements of cleaner fish from southern to middle and northern parts of the Scandinavian distribution range are taking place (Svåsand et al. 2016, 2017). While most of the wrasse used as cleaner fish in commercial farms are killed after a production cycle, some may be accidently or deliberately released into the wild. This represents potentially extensive inadvertent translocation via aquaculture practice. That such translocations occur has already been demonstrated in corkwing wrasse by Faust et al. (2018), and suggested in goldsinny wrasse by Jansson et al. (2017), and a similar problem may exist for ballan wrasse. The results of the genetic analysis here demonstrate a clear genetic difference between ballan wrasse from these two regions, and therefore, inadvertent translocation via aquaculture may result in mixing of genetically-differentiated populations. In turn, this may challenge local adaptation in the recipient populations. As such, care should be taken and the movement of fish from southern to north-western/central Norway should be minimized to decrease the possibility of inadvertent translocation. This information of genetic background also needs to be taken into account in the current and future establishment of cleaner fish farming, as the genetic differences, and thus potential genetic impact between aquaculture escapees and wild populations, is likely to be greater when rearing fish of non-local origin (Glover et al. 2017).

A decrease in genetic diversity was also observed in Scandinavian samples, as such, even though genetic diversity in Scandinavia was not low, care should be taken as populations harboring lower genetic diversity could be more vulnerable to future anthropogenic impact. Additionally, the high current fishing pressure in Scandinavian waters, and the fishery being based on highly localized fishing methods, might result in the removal of ballan wrasse from local ecosystems. Reports of concern that this is already underway are known from local societies as well as from scientists, divers and fishermen. Additionally, fishermen also report having to move to new locations as catches are declining (Grefsrud et al. 2018). The high connectivity observed within the Scandinavian genetic clusters could work to counteract this,

but care should still be taken both to uphold ecological processes, genetic diversity and a sustainable fishery for the future.

4.5.2. Impact of size specific fishery

Before the introduction of the species-specific minimum catch sizes in 2015, the minimum size was set to 11 cm for all labrids. However, currently, the minimum sizes are 14 cm ballan wrasse, 12 cm for corkwing and 11 cm for the remaining labrids (Halvorsen et al. 2017). Ballan wrasse reaches minimum catch size before age 3 (Dipper et al. 1977) while first spawning happens at age 6-9 (for females (Dipper and Pullin 1979). Hence, fish are targeted by the fishery even before first spawning. Taking into consideration that ballan wrasse is a monandric protogynous hermaphrodite with a natural 1:9 female skewed sex ratio (Dipper 1987 as cited in Grant et al. 2016), where sex change happens after year 5 (Dipper and Pullin 1979), it is likely that the fishery is primarily targeting females. The impact of such a skewed fishery pressure on wild populations is still to be determined, but if the sex ratio is changed towards fewer females per male, a higher degree of intrasexual competition might be expected (as seen in Two-spotted gobies (*Gobiusculus flavescens* Fabricius, 1779); Wacker et al. 2013).

The current practice of targeting everything from relatively small immature females to large, and thus male, wrasse could induce inadvertent selection pressures resulting in changes in maturation and/or age at sex change. Previous studies have indicated that increased fishing mortality on all fish as well as specific for immature fish both cause selection for earlier maturation, additionally there are indications that size-independent harvest decreases overall biomass yield in iteroparous species (Heino et al. 2015). Thus, the type of targeting currently going on in ballan wrasse, could therefore have similar effects as seen in e.g., cod (Jørgensen 1990, Barot et al. 2004) and California sheephead (*Semicossyphus pulcher* Ayres, 1854) (Hamilton et al. 2007), where a reduction in size at maturation and maximum size has been observed. Additionally, similar to ballan wrasse, California sheephead is a protogynous hermaphrodite and displays sexual size dimorphism with males being larger. Thus, similar to observations in sheephead (Hamilton et al. 2007), if large ballan wrasse are targeted (such as in the Galician fishery (Almada et al. 2016)), it could result in the removal of males from the populations, and subsequent change of sex ratio towards fewer males.

A change in the current management is therefore urgent, and proposed measures to secure future generations include maximum catch size as well as dividing the management zone into smaller areas to be fished in non-overlapping intervals (Hosteland 2017). Such measures

should work to uphold numbers of mature fish and ensure that a larger proportion of the populations are able to spawn.

4.6. Future work

Future studies on the ballan wrasse need to be targeted to expand upon the major findings from this study. Two major types of study are proposed below.

1 – The population genetic structure and colonization history/route of the ballan wrasse needs elucidating throughout its entire range. In order to achieve this, further samples of from the British Isles, as well as coastal areas of continental Europe (Denmark to Atlantic France, as well as Portugal) are needed to "fill the gaps" present in the current study. Furthermore, although the SNP markers presented here have clearly elucidated population genetic structure, the proliferation of next-generation sequencing techniques has permitted the transition from population genetics to population genomics. Good examples of population genomics approaches revealing genetic structure where previously undetected when using low numbers of markers exist (e.g. Besnier et al. 2014, Sodeland et al. 2016). Thus, a combination of greater numbers of samples and better geographical coverage, together with whole-genome information (either by sequencing or genotyping by sequencing) represents an important next-step. Finally, whole genome methods do not only provide additional clarity for population structure, they permit detection of important evolutionary processes such as adaptation (e.g., Besnier et al. 2014, Sodeland et al. 2016). This may in turn help elucidate the functional genetic differences, and potential adaptations, among these genetically isolated populations.

2 – The underlying genetic basis of the morphological variation observed in this species, i.e., spotted and plain morphotypes, needs further investigation. Is the ballan wrasse to be regarded as two separate sub-species in the south, but one single species in the north, or is this just environmental plasticity? In addition to analyzing phenotypically-characterized population samples in combination with follow-up work described above, it is proposed that crossing experiments, to look at genetic and environmental influences on this phenotype, are required. This is the sort of work that has been conducted in salmon to reveal that spotting patterns in that species are largely determined by environmental triggers but have an underlying genetic basis (Jørgensen et al. 2018). Examples of crossing experiments could include crosses as below,

reared at differing temperatures and feeding regimes (and could be combined with pedigree and genomic information):

- Within genetic group crossings
- Crossing NW and SE Scandinavian ballan wrasse
- Crossing both NW and SE Scandinavian ballan wrasse with plain morphotype ballan wrasse from southern Europe
- Crossing both NW and SE Scandinavian ballan wrasse with spotted morphotype ballan wrasse from southern Europe
- If present, similar crossings with genetic groups from other parts of the distribution range

4.7. Concluding remarks

This work represents the first study of ballan wrasse population genetics throughout Scandinavia. As such, it has resulted in new knowledge about the population genetic structure and connectivity of the Scandinavian ballan wrasse, with two major genetic clusters, each inhabiting north-western and south-eastern Scandinavian coastlines, respectively. Additionally, this study represents the first look at the genetics of phenotypic diversity in northern latitudes, as well as providing data further questioning the taxonomic status of ballan wrasse in Galicia (with two putative sub-species; spotted and plain).

Despite this study revealing significant advances in our knowledge of this species, multiple questions remain. For example, there is a need to further elaborate the colonization history of ballan wrasse throughout Europe. Additionally, questions regarding and the functional genetic control and expression of phenotypic traits relating to coloration patterns, as well as the taxonomic status of the species (at least in southern latitudes) still remain unanswered. Finally, current management of the labrid fishery should be questioned. This includes researching the impact of targeting long-lived sequentially hermaphroditic species with slow and complex life-histories through most of their life (starting before maturation and sex change), as well as questioning the justification for targeting all labrids in Scandinavia under one quota, given their largely differing life-histories. As such, the work described in this thesis could serve as a conduit for future research on ballan wrasse and other labrids in the years to come.

5. References

Allendorf, F. W., P. R. England, G. Luikart, P. A. Ritchie, and N. Ryman. 2008. Genetic effects of harvest on wild animal populations. Trends in Ecology and Evolution 23:327–337.

Allendorf, F. W., G. Luikart, and S. N. Aitken. 2013. Conservation and the genetics of populations. 2nd editio. John Wiley & Sons, Ltd.

Almada, F., L. Casas, S. M. Francisco, D. Villegas-Ríos, F. Saborido-Rey, X. Irigoien, and J. I. Robalo. 2016. On the absence of genetic differentiation between morphotypes of the ballan wrasse *Labrus bergylta* (Labridae). Marine Biology 163:86.

Almada, F., S. M. Francisco, C. S. Lima, R. Fitzgerald, L. Mirimin, D. Villegas-Ríos, F. Saborido-Rey, P. Afonso, T. Morato, S. Bexiga, and J. I. Robalo. 2017. Historical gene flow constraints in a northeastern Atlantic fish: Phylogeography of the ballan wrasse *Labrus bergylta* across its distribution range. Royal Society Open Science 4:160773.

Almada, V. C., F. Almada, S. M. Francisco, R. Castilho, and J. I. Robalo. 2012. Unexpected High Genetic Diversity at the Extreme Northern Geographic Limit of *Taurulus bubalis* (Euphrasen, 1786). PLoS ONE 7.

Anonymous. 2018, June. Doggerland Exploration Continues. Ocean News & Technology:32–33.

Antao, T., A. Lopes, R. J. Lopes, A. Beja-Pereira, and G. Luikart. 2008. LOSITAN: A workbench to detect molecular adaptation based on a F st -outlier method. BMC Bioinformatics 9:323.

Barot, S., M. Heino, and L. O'brien. 2004. Long-Term Trend in the Maturation Reaction Norm of Two Cod Stocks. Ecological Applications 14:1257–1271.

Beaumont, M. A., and D. J. Balding. 2004. Identifying adaptive genetic divergence among populations from genome scans. Molecular Ecology 13:969–980.

Beaumont, M. A., and R. A. Nichols. 1996. Evaluating Loci for Use in the Genetic Analysis of Population Structure. Proceedings: Biological Sciences 263:1619–1626.

Bernardi, G. 2000. Barriers to Gene Flow in *Embiotoca jacksoni*, a Marine Fish Lacking a Pelagic. Evolution 54:226–237.

Besnier, F., and K. A. Glover. 2013. ParallelStructure: A R Package to Distribute Parallel Runs of the Population Genetics Program STRUCTURE on Multi-Core Computers. PLoS ONE 8:70651.

Besnier, F., M. Kent, R. Skern-Mauritzen, S. Lien, K. Malde, R. B. Edvardsen, S. Taylor, L. E. Ljungfeldt, F. Nilsen, and K. A. Glover. 2014. Human-induced evolution caught in action: SNP-array reveals rapid amphi-atlantic spread of pesticide resistance in the salmon ecotoparasite *Lepeophtheirus salmonis*. BMC Genomics 15.

Bjordal, Å. 1988. Cleaning symbiosis between wrasse (Labridae) and lice infested salmon (*Salmo salar*) in mariculture. ICES 188.

Blanco Gonzalez, E., and F. de Boer. 2017. The development of the Norwegian wrasse fishery and the use of wrasses as cleaner fish in the salmon aquaculture industry. Fisheries Science 83:661–670.

Blanco Gonzalez, E., H. Knutsen, and P. E. Jorde. 2016. Habitat discontinuities separate genetically divergent populations of a rocky shore marine fish. PLoS ONE 11:e0163052.

Brooker, A. J., R. Skern-Mauritzen, and J. E. Bron. 2018. Production, mortality, and infectivity of planktonic larval sea lice, *Lepeophtheirus salmonis* (Krøyer, 1837): current knowledge and implications for epidemiological modelling. ICES Journal of Marine Science 75:1214–1234.

Burridge, L., J. S. Weis, F. Cabello, J. Pizarro, and K. Bostick. 2010. Chemical use in salmon aquaculture: A review of current practices and possible environmental effects.

Aquaculture 306:7–23.
Charlesworth, B., and D. Charlesworth. 2017. Population genetics from 1966 to 2016. Heredity 118:2–9.
CLIMAP Project Members. 1984. The Last Interglacial Ocean. Quarternary Research 21:123–224.
Côté, I. M. 2000. Evolution and ecology of cleaning symbioses in the sea. Pages 311–355 *in* R. N. Gibson and M. Barnes, editors. Oceanography and Marine Biology: an annual review. Taylor & Francis Ltd., London.
D'Arcy, J., L. Mirimin, and R. FitzGerald. 2013. Phylogeographic structure of a protogynous hermaphrodite species, the ballan wrasse *Labrus bergylta*, in Ireland, Scotland, and Norway, using mitochondrial DNA sequence data. ICES Journal of Marine Science 70:685–693.
Darwall, W. R. T., M. J. Costello, R. Donnelly, and S. Lysaght. 1992. Implications of life-history strategies for a new wrasse fishery. Journal of Fish Biology 41:111–123.
Davies, I. M., P. A. Gillibrand, J. G. McHenery, and G. H. Rae. 1998. Environmental risk of ivermectin to sediment dwelling organisms. Aquaculture 163:29–46.
Denholm, I., G. J. Devine, T. E. Horsberg, S. Sevatdal, A. Fallang, D. V Nolan, and R. Powell. 2002. Analysis and management of resistance to chemotherapeutants in salmon lice, *Lepeophtheirus salmonis* (Copepoda: Caligidae). Pest Management Science 58:528–536.
Derryberry, E. P., G. E. Derryberry, J. M. Maley, and R. T. Brumfield. 2014. hzar : hybrid zone analysis using an R software package. Molecular Ecology Resources 14:652–663.
Dipper, F. 1987. British Sea Fishes. 1st edition. Underwater World Publications Ltd., London.
Dipper, F. A., C. R. Bridges, and A. Menz. 1977. Age, growth and feeding in the ballan wrasse Labrus bergylta Ascanius 1767. Journal of Fish Biology 11:105–120.
Dipper, F. A., and R. S. V. Pullin. 1979. Gonochorism and sex-inversion in British Labridae (Pisces). Journal of Zoology 187:97–112.
Dray, S., and A.-B. Dufour. 2015. The ade4 Package: Implementing the Duality Diagram for Ecologists . Journal of Statistical Software 22.
Espedal, P. G., K. A. Glover, T. E. Horsberg, and F. Nilsen. 2013. Emamectin benzoate resistance and fitness in laboratory reared salmon lice (Lepeophtheirus salmonis). Aquaculture 416–417:111–118.
Espeland, S. H., K. Nedreaas, S. Mortensen, A. B. Skiftesvik, A.-L. Agnalt, C. Durif, L. S. Harkestad, E. Karlsbakk, H. Knutsen, T. Thangstad, K. Jørstad, Å. Bjordal, and J. Gjøsaeter. 2010. Kunnskapsstatus leppefisk: Utfordringer i et økende fiskeri. Fisken og havet:[in Norwegian].
Eurostat. (n.d.). Aquaculture statistics - Statistics Explained. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Fishery_statistics#Aquaculture_statistics.
Eurostat. 2018. Agriculture, forestry and fishery statistics. Luxembourg.
Evankow, A., H. Christie, K. Hancke, A. K. Brysting, C. Junge, S. Fredriksen, and J. Thaulow. 2019. Genetic heterogeneity of two bioeconomically important kelp species along the Norwegian coast. Conservation Genetics 0:0.
Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software structure: a simulation study. Molecular Ecology 14:2611–2620.
Excoffier, L., T. Hofer, and M. Foll. 2009. Detecting loci under selection in a hierarchically structured population. Heredity 103:285–298.
Excoffier, L., G. Laval, and S. Schneider. 2005. Arlequin (version 3.0): An integrated software package for population genetics data analysis. Page 117693430500100 Evolutionary Bioinformatics. SAGE PublicationsSage UK: London, England.

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

Fallang, A., I. Denholm, T. E. Horsberg, and M. S. Williamson. 2005. Novel point mutation in the sodium channel gene of pyrethroid-resistant sea lice *Lepeophtheirus salmonis* (Crustacea : Copepoda). Diseases of Aquatic Organisms 65:129–136.

Fallang, A., J. M. Ramsay, S. Sevatdal, J. F. Burka, P. Jewess, K. L. Hammell, and T. E. Horsberg. 2004. Evidence for occurrence of an organophosphate-resistant type of acetylcholinesterase in strains of sea lice (Lepeophtheirus salmonis Krøyer). Pest Management Science 60:1163–1170.

Faust, E., K. T. Halvorsen, P. Andersen, H. Knutsen, and C. André. 2018. Cleaner fish escape salmon farms and hybridize with local wrasse populations. Royal Society Open Science 5:171752.

Finstad, B., and P. A. Bjørn. 2011. Present status and implications of salmon lice on wild salmonids in Norwegian Coastal zones. Pages 545–597 Salmon lice : An integrated approach to understanding parasite abundance and distribution. 1st Edt. John Wiley & Sons, Ltd, Hoboken.

Foll, M., and O. Gaggiotti. 2006. Identifying the environmental factors that determine the genetic structure of populations. Genetics 174:875–891.

Foll, M., and O. Gaggiotti. 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. Genetics 180:977–93.

Gaetano, J. 2013. Holm-Bonferroni sequential correction: an EXCEL calculator–Ver. 1.2.

Gay, L., P.-A. Crochet, D. A. Bell, and T. Lenormand. 2008. Comparing Clines on Molecular and Phenotypic Traits in Hybrid Zones: A Window on Tension Zone Models. Evolution 62:2789–2806.

Gjøsæter, J. 2002. Distribution and density of goldsinny wrasse (*Ctenolabrus rupestris*) (Labridae) in the Risør and Arendal areas along the Norwegian Skagerrak coast. SARSIA 87:75–82.

Glover, K. A., M. F. Solberg, P. McGinnity, K. Hindar, E. Verspoor, M. W. Coulson, M. M. Hansen, H. Araki, Ø. Skaala, and T. Svåsand. 2017. Half a century of genetic interaction between farmed and wild Atlantic salmon: Status of knowledge and unanswered questions. Fish and Fisheries 18:890–927.

Goudet, J. 2005. hierfstat, a package for R to compute and test hierarchical F-statistics. Molecular Ecology Notes 5:184–186.

Grant, B., A. Davie, J. B. Taggart, S.-L. C. Selly, N. Picchi, C. Bradley, P. Prodohl, E. Leclercq, and H. Migaud. 2016. Seasonal changes in broodstock spawning performance and egg quality in ballan wrasse (*Labrus bergylta*). Aquaculture 464:505–514.

Grefsrud, E. S., K. Glover, B. E. Grøsvik, V. Husa, Ø. Karlsen, T. Kristiansen, B. O. Kvamme, S. Mortensen, O. B. Samuelsen, L. H. Stien, and T. Svåsand. 2018. Risikorapport norsk fiskeoppdrett 2018 (Risk Report Norwegian Fish Farming 2018). Fisken og havet særnr. 1-2:166-174. [in Norwegian].

Gysels, E. S., B. Hellemans, C. Pampoulie, and F. A. M. Volckaert. 2004. Phylogeography of the common goby, Pomatoschistus microps, with particular emphasis on the colonization of the Mediterranean and the North Sea. Molecular Ecology 13:403–417.

Halvorsen, K. T., T. Larsen, T. Knutsen Sørdalen, L. A. Vøllestad, H. Knutsen, E. M. Olsen, and S. Matic-Skoko. 2017. Impact of harvesting cleaner fish for salmonid aquaculture assessed from replicated coastal marine protected areas. Marine Biology Research 0.

Hamilton, S. L., J. E. Caselle, J. D. Standish, D. M. Schroeder, M. S. Love, J. A. Rosales-Casian, and O. Sosa-Nishizaki. 2007. Size-Selective Harvesting Alters Life Histories of

a Temperate Sex-Changing Fish. Ecological Applications 17:2268–2280.
Hanel, R., M. W. Westneat, and C. Sturmbauer. 2002. Phylogenetic relationships, evolution of broodcare behavior, and geographic speciation in the wrasse tribe Labrini. Journal of Molecular Evolution 55:776–789.
Hauser, L., and G. R. Carvalho. 2008. Paradigm shifts in marine fisheries genetics: ugly hypotheses slain by beautiful facts. Fish and Fisheries 9:333–362.
Hayes, A., M. Kucera, N. Kallel, L. Sbaffi, and E. J. Rohling. 2005. Glacial Mediterranean sea surface temperatures based on planktonic foraminiferal assemblages. Quaternary Science Reviews 24:999–1016.
Heino, M., B. Díaz Pauli, and U. Dieckmann. 2015. Fisheries-Induced Evolution. Annual Review of Ecology, Evolution, and Systematics 46:461–480.
Helland, S., S. W. Dahle, C. Hough, and J. Borthen, editors. 2014. Production of ballan wrasse - Science and practice. The Norwegian Seafood Research Fund (FHF).
Herron, J. C., and S. Freeman. 2015. Evolutionary Analysis. 5th edition. Pearson.
Heuch, P. A., P. A. Bjørn, B. Finstad, J. C. Holst, L. Asplin, and F. Nilsen. 2005. A review of the Norwegian 'National Action Plan Against Salmon Lice on Salmonids': The effect on wild salmonids. Aquaculture 246:79–92.
Heuch, P. A., and T. A. Mo. 2001. A model of salmon louse production in Norway: effects of increasing salmon production and public management measures. Diseases of Aquatic Organisms 45:145–152.
Hilldén, N.-O. 1984. Behavioural ecology of the labrid fishes (Teleostei: Labridae) at Tjärnö on the Swedish west coast. University of Stockholm, Stockholm.
Holm, S. 1979. A Simple Sequentially Rejective Multiple Test Procedure. Scandinavian Journal of Statistics 6:65–70.
Hosteland, L. T. S. 2017. Leppefiskfangst: Foreslår å vurdere vekselbruk. https://www.kyst.no/article/leppefiskfangst-foreslaar-aa-vurdere-vekselbruk/. [in Norwegian].
Hutchinson, W. F. 2008. The dangers of ignoring stock complexity in fishery management: the case of the North Sea cod. Biology Letters 4:693–695.
Iversen, A., Ø. Hermansen, O. Andreassen, R. K. Brandvik, A. Marthinussen, and R. Nystøyl. 2015. Kostnadsdrivere i lakseoppdrett (Cost Drivers in Salmon Farming). Nofima 41:1-43. [in Norwegian].
Jakobsson, M., and N. A. Rosenberg. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23:1801–1806.
Jansson, E., M. Quintela, G. Dahle, J. Albretsen, H. Knutsen, C. André, Å. Strand, S. Mortensen, J. B. Taggart, E. Karlsbakk, B. O. Kvamme, and K. A. Glover. 2017. Genetic analysis of goldsinny wrasse reveals evolutionary insights into population connectivity and potential evidence of inadvertent translocation via aquaculture. ICES Journal of Marine Science 74:2135–2147.
Johannesson, K., and C. André. 2006. Life on the margin: Genetic isolation and diversity loss in a peripheral marine ecosystem, the Baltic Sea. Molecular Ecology 15:2013–2029.
Jombart, T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403–1405.
Jombart, T., S. Devillard, and F. Balloux. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genetics 11:94.
Jones, S., and R. Beamish. 2011. Lepeophtheirus salmonis on salmonids in the northeast Pacific Ocean. Pages 598–642 Salmon lice : An integrated approach to understanding parasite abundance and distribution. 1st Edt. John Wiley & Sons, Ltd, Hoboken.

Jørgensen, K. M., M. F. Solberg, F. Besnier, A. Thorsen, P. G. Fjelldal, Ø. Skaala, K. Malde, and K. A. Glover. 2018. Judging a salmon by its spots: environmental variation is the primary determinant of spot patterns in *Salmo salar*. BMC Ecology 18:1–13.

Jørgensen, T. 1990. Long-term changes in age at sexual maturity of Northeast Arctic cod (*Gadus morhua* L.). ICES Journal of Marine Science 46:235–248.

Kahle, D., and H. Wickham. 2013. ggmap: Spatial Visualization with ggplot2. The R Journal 5:144–161.

Kamvar, Z. N., J. F. Tabima, and N. J. Grünwald. 2014. Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ 2:e281.

Kaur, K., F. Besnier, K. A. Glover, F. Nilsen, V. T. Aspehaug, H. B. Fjørtoft, and T. E. Horsberg. 2017. The mechanism (Phe362Tyr mutation) behind resistance in Lepeophtheirus salmonis pre-dates organophosphate use in salmon farming. Nature Scientific Reports 7:1–9.

Kerr, L. A., N. T. Hintzen, S. X. Cadrin, L. W. Clausen, M. Dickey-Collas, D. R. Goethel, E. M. C. Hatfield, J. P. Kritzer, and R. D. M. Nash. 2017. Lessons learned from practical approaches to reconcile mismatches between biological population structure and stock units of marine fish. ICES Journal of Marine Science 74:1708–1722.

Knutsen, H., P. E. Jorde, C. André, and N. C. Stenseth. 2003. Fine-scaled geographical population structuring in a highly mobile marine species: the Atlantic cod. Molecular Ecology 12:385–394.

Knutsen, H., P. E. Jorde, E. B. Gonzalez, J. Robalo, J. Albretsen, and V. Almada. 2013. Climate Change and Genetic Structure of Leading Edge and Rear End Populations in a Northwards Shifting Marine Fish Species, the Corkwing Wrasse (*Symphodus melops*). PLoS ONE 8:e67492.

Knutsen, H., E. M. Olsen, L. Ciannelli, S. H. Espeland, J. A. Knutsen, J. H. Simonsen, S. Skreslet, and N. C. Stenseth. 2007. Egg distribution, bottom topography and small-scale cod population structure in a coastal marine system. Marine Ecology Progress Series 333:249–255.

Laikre, L., M. K. Schwartz, R. S. Waples, N. Ryman, T. Gem, and W. Group. 2010. Compromising genetic diversity in the wild: unmonitored large-scale release of plants and animals. Trends in Ecology & Evolution 25:520–529.

Lambeck, K., A. Purcell, J. Zhao, and N. Svensson. 2010. The Scandinavian Ice Sheet: from MIS 4 to the end of the Last Glacial Maximum. Boreas 39:410–435.

Li, Y.-L., and J.-X. Liu. 2018. StructureSelector : A web-based software to select and visualize the optimal number of clusters using multiple methods. Molecular Ecology Resources 18:176–177.

Ljungfeldt, E. L. R., F. Nilsen, M. Skern-Mauritzen, and K. A. Glover. 2014. A common-garden experiment to quantify evolutionary processes in copepods: the case of emamectin benzoate resistance in the parasitic sea louse *Lepeophtheirus salmonis*. BMC Evolutionary Biology 14:108.

Luikart, G., P. R. England, D. Tallmon, S. Jordan, and P. Taberlet. 2003. The power and promise of population genomics: From genotyping to genome typing. Nature Reviews Genetics 4:981–994.

Luu, K., E. Bazin, and M. G. B. Blum. 2017. *pcadapt* : an R package to perform genome scans for selection based on principal component analysis. Molecular Ecology Resources 17:67–77.

McGinnity, P., P. Prodöhl, A. Ferguson, R. Hynes, N. Ó Maoiléidigh, N. Baker, D. Cotter, B. O'Hea, D. Cooke, G. Rogan, J. Taggart, and T. Cross. 2003. Fitness Reduction and Potential Extinction of Wild Populations of Atlantic Salmon, *Salmo salar*, as a Result of

Interactions with Escaped Farm Salmon. Proceedings of the Royal Society of London. Series B: Biological Sciences 270:2443–2450.
El Mousadik, A., and R. J. Petit. 1996. High level of genetic differentiation for allelic richness among populations of the argan tree [*Argania spinosa* (L.) Skeels] endemic to Morocco. Pages 832–839 Theoretical and Applied Genetics. Springer-Verlag.
Muncaster, S., E. Andersson, O. S. Kjesbu, G. L. Taranger, A. B. Skiftesvik, and B. Norberg. 2010. The reproductive cycle of female Ballan wrasse *Labrus bergylta* in high latitude, temperate waters. Journal of Fish Biology 77:no-no.
Muncaster, S., B. Norberg, and E. Andersson. 2013. Natural sex change in the temperate protogynous Ballan wrasse *Labrus bergylta*. Journal of Fish Biology 82:1858–1870.
Nei, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York, NY.
Norwegian Directorate of Fisheries. (n.d.). Fangst av leppefisk. https://www.fiskeridir.no/Yrkesfiske/Statistikk-yrkesfiske/Fangst-og-kvoter/Fangst-av-leppefisk.
Norwegian Directorate of Fisheries. (n.d.). Fisket etter leppefisk. https://www.fiskeridir.no/Yrkesfiske/Tema/Leppefisk.
Norwegian Directorate of Fisheries. 2018. Statistikk for akvakultur 2017 (Statistics of aquaculture 2017).
Ottesen, B. O., J. I. M. Treasurer, R. Fitzgerald, J. D. Arcy, and J. Maguire. 2011. Ecofish: Rearing Ballan Wrasse for Sea Lice Control. Fish Farmer Magazine:12–16.
Øverli, Ø., J. Nordgreen, C. M. Mejdell, A. M. Janczak, S. Kittilsen, I. B. Johansen, and T. E. Horsberg. 2014. Ectoparasitic sea lice (*Lepeophtheirus salmonis*) affect behavior and brain serotonergic activity in Atlantic salmon (*Salmo salar* L.): Perspectives on animal welfare. Physiology & Behavior 132:44–50.
Paradis, E. 2010. pegas: an R package for population genetics with an integrated-modular approach. Bioinformatics 26:419–420.
Pike, A. W., and S. L. Wadsworth. 1999. Sealice on Salmonids: Their Biology and Control.
Powell, A., J. W. Treasurer, C. L. Pooley, A. J. Keay, R. Lloyd, A. K. Imsland, and C. Garcia de Leaniz. 2017. Use of lumpfish for sea-lice control in salmon farming: challenges and opportunities. Reviews in Aquaculture.
Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of Population Structure Using Multilocus Genotype Data. Genetics 155:945–959.
Quignard, J. P., and A. Pras. 1986. Labridae. Pages 919–942 *in* P. J. P. Whitehead, M. L. Bauchot, J. C. Hureau, J. Nielsen, and E. Torronese, editors. Fishes of the North-eastern Atlantic and the Mediterranean. Vol. II. UNESCO, Paris.
Quintela, M., E. A. Danielsen, L. Lopez, R. Barreiro, T. Svåsand, H. Knutsen, A. B. Skiftesvik, and K. A. Glover. 2016. Is the ballan wrasse (*Labrus bergylta*) two species? Genetic analysis reveals within-species divergence associated with plain and spotted morphotype frequencies. Integrative Zoology 11:162–172.
Quintela, M., E. A. Danielsen, T. Svåsand, H. Knutsen, A. B. Skiftesvik, and K. A. Glover. 2014. Isolation and characterization of twenty microsatellite loci for the ballan wrasse, *Labrus bergylta*. Conservation Genetics Resources 6:425–428.
R Core Team. 2008. R: A Language and Environment for Statistical Computing.
Rae, G. H. 2002. Sea louse control in Scotland, past and present. Pest Management Science 58:515–520.
Riginos, C., and M. W. Nachman. 2001. Population subdivision in marine environments: The contributions of biogeography, geographical distance and discontinuous habitat to genetic differentiation in a blennioid fish, Axoclinus nigricaudus. Molecular Ecology 10:1439–1453.
Robalo, J. I., R. Castilho, S. M. Francisco, F. Almada, H. Knutsen, P. E. Jorde, A. M. Pereira,

and V. C. Almada. 2011. Northern refugia and recent expansion in the North Sea: the case of the wrasse *Symphodus melops* (Linnaeus, 1758). Ecology and Evolution 2:153–164.

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

Rousset, F. 2008. genepop’007: a complete re-implementation of the genepop software for Windows and Linux. Molecular Ecology Resources 8:103–106.

Rowe, G., M. Sweet, and T. J. C. Beebee. 2017. An introduction to Molecular Ecology. 3rd edition. Oxford University Press, Oxford.

Skiftesvik, A. B., R. M. Bjelland, C. M. F. Durif, I. S. Johansen, and H. I. Browman. 2013. Delousing of Atlantic salmon (*Salmo salar*) by cultured vs. wild ballan wrasse (*Labrus bergylta*). Aquaculture 402–403:113–118.

Skiftesvik, A. B., G. Blom, A.-L. Agnalt, C. M. F. F. Durif, H. I. Browman, R. M. Bjelland, L. S. Harkestad, E. Farestveit, O. I. Paulsen, M. Fauske, T. Havelin, K. Johnsen, S. Mortensen, A. Berit Skiftesvik, and & Stein Mortensen. 2014. Wrasse (Labridae) as cleaner fish in salmonid aquaculture - The Hardangerfjord as a case study. Marine Biology Research 10:289–300.

Skiftesvik, A. B., C. M. F. Durif, R. M. Bjelland, and H. I. Browman. 2015. Distribution and habitat preferences of five species of wrasse (Family Labridae) in a Norwegian fjord. ICES Journal of Marine Science 72:890–899.

Slatkin, M. 1993. Isolation by Distance in Equilibrium and Non-Equilibrium Populations. Evolution 47:264–279.

Sodeland, M., E. Jorde, S. Lien, S. Jentoft, P. R. Berg, H. Grove, M. P. Kent, M. Arnyasi, E. M. Olsen, and H. Knutsen. 2016. "Islands of Divergence" in the Atlantic Cod Genome Represent Polymorphic Chromosomal Rearrangements. Genome Biology and Evolution 8:1012–1022.

Stone, J., I. H. Sutherland, C. Sommerville, R. H. Richards, and K. J. Varma. 2000. Field trials to evaluate the efficacy of emamectin benzoate in the control of sea lice, Lepeophtheirus salmonis (Krøyer) and Caligus elongatus Nordmann, infestations in Atlantic salmon Salmo salar L. Aquaculture 186:205–219.

Stone, J., I. H. Sutherland, C. S. Sommerville, R. H. Richards, and K. J. Varmac. 1999. The efficacy of emamectin benzoate as an oral treatment of sea lice, *Lepeophtheirus salmonis* (Krøyer), infestations in Atlantic salmon, *Salmo salar* L. Journal of Fish Diseases 22:261–270.

Stroeven, A. P., C. Hättestrand, J. Kleman, J. Heyman, D. Fabel, O. Fredin, B. W. Goodfellow, J. M. Harbor, J. D. Jansen, L. Olsen, M. W. Caffee, D. Fink, J. Lundqvist, G. C. Rosqvist, B. Strömberg, and K. N. Jansson. 2016. Deglaciation of Fennoscandia. Quaternary Science Reviews 147:91–121.

Sundt, R. C., and K. E. Jørstad. 1998. Genetic population structure of goldsinny wrasse, *Ctenolabrus rupestris* (L.), in Norway: implications for future management of parasite cleaners in the salmon farming industry. Fisheries Management and Ecology 5:291–302.

Svåsand, T., E. S. Grefsrud, Ø. Karlsen, B. O. Kvamme, K. Glover, V. Husa, and T. S. Kristiansen. 2017. Risikorapport norsk fiskeoppdrett 2017 (Risk Report Norwegian Fish Farming 2017). Fisken og havet særnr. 2-2:161-173. [in Norwegian].

Svåsand, T., Ø. Karlsen, B. O. Kvamme, L. Helge Stien, G. L. Taranger, and K. K. Boxaspen. 2016. Risikovurdering norsk fiskeoppdrett 2016 (Risk Assessment Norwegian Fish Farming 2016). Fisken og havet særnr. 2-2:173-184. [in Norwegian].

Torrissen, O., S. Jones, F. Asche, A. Guttormsen, O. T. Skilbrei, F. Nilsen, E. Hors-Berg, and D. Jackson. 2013. Salmon lice - impact on wild salmonids and salmon aquaculture. Journal of Fish Diseases 36:171–194.

Treasurer, J. W., S. Wadsworth, and A. Grant. 2000. Resistance of sea lice, Lepeophtheirus salmonis (Kroyer), to hydrogen peroxide on farmed Atlantic salmon, Salmo salar L. Aquaculture Research 31:855–860.

Villegas-Ríos, D. 2013. Life-history and behaviour of *Labrus bergylta* in Galicia. Universida de Vigo.

Villegas-Ríos, D., A. Alonso-Fernández, M. Fabeiro, R. Bañón, and F. Saborido-Rey. 2013a. Demographic Variation between Colour Patterns in a Temperate Protogynous Hermaphrodite, the Ballan Wrasse *Labrus bergylta*. PLoS ONE 8:e71591.

Villegas-Ríos, D., J. Alós, D. March, M. Palmer, G. Mucientes, and F. Saborido-Rey. 2013b. Home range and diel behavior of the ballan wrasse, *Labrus bergylta*, determined by acoustic telemetry. Journal of Sea Research 80:61–71.

Vollset, K. W., R. I. Krontveit, P. A. Jansen, B. Finstad, B. T. Barlaup, O. T. Skilbrei, M. Krkošek, P. Romunstad, A. Aunsmo, A. J. Jensen, and I. Dohoo. 2016. Impacts of parasites on marine survival of Atlantic salmon: a meta-analysis. Fish and Fisheries 17:714–730.

Wacker, S., K. Mobley, E. Forsgren, L. Cats Myhre, K. De Jong, and T. Amundsen. 2013. OPERATIONAL SEX RATIO BUT NOT DENSITY AFFECTS SEXUAL SELECTION IN A FISH. Evolution 67:1937–1949.

Westneat, M. W., and M. E. Alfaro. 2005. Phylogenetic relationships and evolutionary history of the reef fish family Labridae. Molecular Phylogenetics and Evolution 36:370–390.

Wickham, H. 2009. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag, New York, NY.

Wright, S. 1943. ISOLATION BY DISTANCE. Genetics 28:114–138.

Yuen, J. W., T. Dempster, F. Oppedal, and M. Hvas. 2019. Physiological performance of ballan wrasse (*Labrus bergylta*) at different temperatures and its implication for cleaner fish usage in salmon aquaculture. Biological Control.

6. Appendices

6.1. Appendix 1: Additions to Materials and Methods

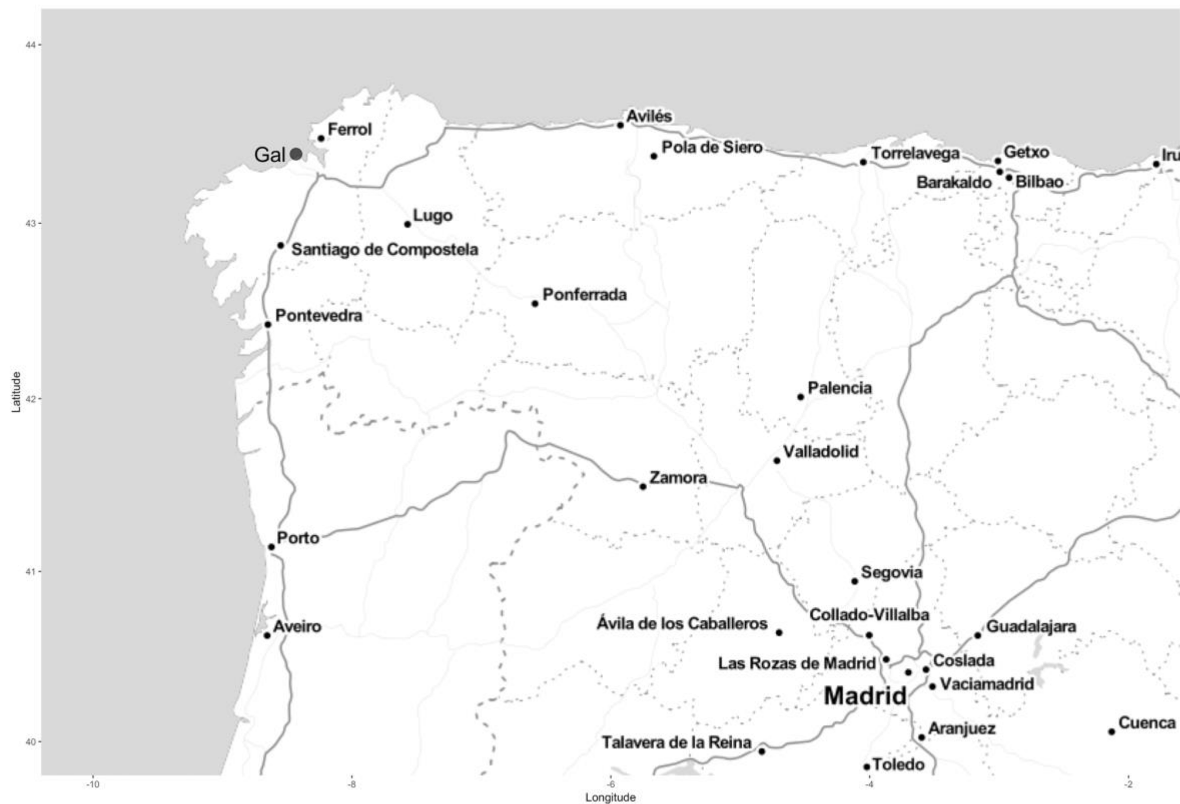


Figure A1 *Map of approximate sampling location of Galician ballan wrasse. Exact sampling location was unknown as samples were obtained from a local fish market in La Coruña (43.359 N, 8.402 W – just inland from marked sampling location).*

Table A1 Overview of multiplex (MP) 1 including volumes of PCR Primers and Unextended Extend Primers (UEP). Markers in yellow removed from multiplex before this study was initiated.

Multiplex	SNP_ID	Marker name	PCR Primers (1st & 2nd) Total volume	UEP Total volume
			1000	1500
MP1	NW_018114517.1_953305	BaWr-1	5	21
MP1	NW_018114501.1_1110519	BaWr-2		
MP1	NW_018114781.1_75751	BaWr-3	5	21
MP1	NW_018114604.1_365777	BaWr-4	5	21
MP1	NW_018114505.1_499976	BaWr-5	5	21
MP1	NW_018114434.1_45754	BaWr-6	5	21
MP1	NW_018114474.1_1618630	BaWr-7	5	21
MP1	NW_018114489.1_483994	BaWr-8	5	21
MP1	NW_018114526.1_227777	BaWr-9	5	28
MP1	NW_018115046.1_126119	BaWr-10	5	28
MP1	NW_018114560.1_1014493	BaWr-11	5	28
MP1	NW_018114681.1_636455	BaWr-12	5	28
MP1	NW_018114571.1_854545	BaWr-13	5	28
MP1	NW_018114696.1_226866	BaWr-14	5	28
MP1	NW_018114523.1_1049816	BaWr-15	5	35
MP1	NW_018114483.1_1338510	BaWr-16		
MP1	NW_018114481.1_912790	BaWr-17	5	35
MP1	NW_018114757.1_45148	BaWr-18	5	35
MP1	NW_018114486.1_1620397	BaWr-19	5	35
MP1	NW_018114444.1_1460970	BaWr-20	5	35
MP1	NW_018114436.1_1764087	BaWr-21	5	35
MP1	NW_018114461.1_1600142	BaWr-22	5	42
MP1	NW_018114430.1_251585	BaWr-23	5	42
MP1	NW_018114471.1_453074	BaWr-24		
MP1	NW_018114451.1_1926182	BaWr-25	5	42
MP1	NW_018114428.1_906974	BaWr-26	5	42
MP1	NW_018114714.1_15927	BaWr-27		
MP1	NW_018114433.1_2198459	BaWr-28	5	42
MP1	NW_018114500.1_827265	BaWr-29	5	42
Combined primer volume			**750**	**723**

Table A2 Overview of multiplex (MP) 2 including volumes of PCR Primers and Unextended Extend Primers (UEP). Markers in yellow removed from multiplex before this study was initiated.

Multiplex	SNP_ID	Marker name	PCR Primers (1st & 2nd) Total volume	UEP Total volume
			1000	**1500**
MP2	NW_018114480.1_686536	BaWr-30	5	21
MP2	NW_018114648.1_639468	BaWr-31	5	21
MP2	NW_018114431.1_1403930	BaWr-32	5	21
MP2	NW_018114885.1_120757	BaWr-33	5	21
MP2	NW_018114418.1_2113518	BaWr-34	5	21
MP2	NW_018114458.1_853662	BaWr-35	5	21
MP2	NW_018114580.1_402395	BaWr-36	5	28
MP2	NW_018114519.1_401347	BaWr-37	5	28
MP2	NW_018114435.1_1416097	BaWr-38		
MP2	NW_018114475.1_1475041	BaWr-39	5	28
MP2	NW_018114728.1_554763	BaWr-40	5	28
MP2	NW_018114480.1_1226704	BaWr-41	5	28
MP2	NW_018114455.1_84460	BaWr-42	5	28
MP2	NW_018114465.1_196310	BaWr-43	5	28
MP2	NW_018114441.1_653994	BaWr-44	5	35
MP2	NW_018114447.1_1776371	BaWr-45		
MP2	NW_018114416.1_56373	BaWr-46	5	35
MP2	NW_018114511.1_656939	BaWr-47	5	35
MP2	NW_018114487.1_1240633	BaWr-48	5	35
MP2	NW_018114532.1_341495	BaWr-49	5	35
MP2	NW_018114593.1_501629	BaWr-50	5	42
MP2	NW_018114492.1_409355	BaWr-51		
MP2	NW_018114468.1_189838	BaWr-52	5	42
MP2	NW_018114449.1_1926310	BaWr-53	5	42
MP2	NW_018114496.1_554438	BaWr-54	5	42
MP2	NW_018114520.1_353049	BaWr-55	5	42
MP2	NW_018114493.1_270364	BaWr-56		
	Combined primer volume		**770**	**786**

Table A3 Overview of multiplex (MP) 3 including volumes of PCR Primers and Unextended Extend Primers (UEP). Marker in yellow removed from multiplex before this study was initiated.

Multiplex	SNP_ID	Marker name	PCR Primers (1st & 2nd) Total volume	UEP Total volume
			1000	**1500**
MP3	NW_018114442.1_1221044	BaWr-57	5	21
MP3	NW_018114498.1_748966	BaWr-58	5	21
MP3	NW_018114546.1_746859	BaWr-59	5	21
MP3	NW_018114587.1_851110	BaWr-60	5	21
MP3	NW_018114552.1_430323	BaWr-61	5	21
MP3	NW_018114454.1_1274351	BaWr-62	5	21
MP3	NW_018114482.1_1351282	BaWr-63	5	21
MP3	NW_018115568.1_16432	BaWr-64	5	28
MP3	NW_018114443.1_78542	BaWr-65	5	28
MP3	NW_018114425.1_2899410	BaWr-66	5	28
MP3	NW_018114525.1_1112059	BaWr-67	5	28
MP3	NW_018114614.1_940723	BaWr-68	5	28
MP3	NW_018114499.1_643624	BaWr-69	5	28
MP3	NW_018114452.1_575674	BaWr-70	5	35
MP3	NW_018114476.1_1062158	BaWr-71		
MP3	NW_018115024.1_46771	BaWr-72	5	35
MP3	NW_018114508.1_1255457	BaWr-73	5	35
MP3	NW_018114420.1_1919501	BaWr-74	5	35
MP3	NW_018114802.1_440534	BaWr-75	5	35
MP3	NW_018114462.1_1987249	BaWr-76	5	35
MP3	NW_018114439.1_366350	BaWr-77	5	42
MP3	NW_018114453.1_383341	BaWr-78	5	42
MP3	NW_018114491.1_295879	BaWr-79	5	42
MP3	NW_018114662.1_372392	BaWr-80	5	42
MP3	NW_018114669.1_516151	BaWr-81	5	42
MP3	NW_018114565.1_470130	BaWr-82	5	42
	Combined primer volume		**750**	**723**

Table A4 Overview of multiplex (MP) 4 including volumes of PCR Primers and Unextended Extend Primers (UEP). Marker in yellow removed from multiplex before this study was initiated.

Multiplex	SNP_ID	Marker name	PCR Primers (1st & 2nd) Total volume	UEP Total volume
			1000	1500
W4	NW_018114950.1_24492	BaWr-83	5	21
W4	NW_018114537.1_993619	BaWr-84	5	21
W4	NW_018114575.1_206829	BaWr-85	5	21
W4	NW_018114427.1_1443862	BaWr-86	5	21
W4	NW_018114527.1_705419	BaWr-87	5	21
W4	NW_018114494.1_1254704	BaWr-88	5	21
W4	NW_018114426.1_930380	BaWr-89	5	28
W4	NW_018114485.1_242312	BaWr-90	5	28
W4	NW_018114522.1_178152	BaWr-91	5	28
W4	NW_018114466.1_573987	BaWr-92	5	28
W4	NW_018115213.1_20609	BaWr-93	5	28
W4	NW_018115108.1_157252	BaWr-94	5	28
W4	NW_018114557.1_105885	BaWr-95	5	35
W4	NW_018114470.1_1952717	BaWr-96	5	35
W4	NW_018114424.1_1723121	BaWr-97	5	35
W4	NW_018114472.1_763740	BaWr-98	5	35
W4	NW_018114627.1_575991	BaWr-99	5	35
W4	NW_018114528.1_354537	BaWr-100	5	35
W4	NW_018114437.1_23329	BaWr-101	5	42
W4	NW_018114464.1_749310	BaWr-102	5	42
W4	NW_018114417.1_2346164	BaWr-103	5	42
W4	NW_018114432.1_2621672	BaWr-104	5	42
W4	NW_018114457.1_1317312	BaWr-105	5	42
W4	NW_018114484.1_155855	BaWr-106		
	Combined primer volume		**770**	**786**

Table A5 *Results for marker BaWr-53 from parentage test of individuals from Marine Harvest Labrus AS*

Individual ID	Role	NW_0181144491_1926310_a BaWr-53 first allele	NW_0181144491_1926310_b BaWr-53 second allele
2226_3545_P20_A04_016	FATHER	C	C
2226_3545_P27_H04_009	MOTHER	T	T
3537_2209_10_B04_015	Offspring	T	T
3537_2209_12_D04_013	Offspring	T	T
3537_2209_20_E05_021	Offspring	-	-
3537_2209_21_F05_019	Offspring	C	C
3537_2209_23_H05_017	Offspring	C	C
3537_2209_26_C06_022	Offspring	-	-
3537_2209_27_D06_021	Offspring	C	T
3537_2209_29_F06_019	Offspring	T	T
3537_2209_34_C07_030	Offspring	C	T
3537_2209_36_E07_028	Offspring	C	C
3537_2209_37_F07_027	Offspring	C	T
3537_2209_41_B08_031	Offspring	C	T
3537_2209_43_D08_029	Offspring	T	T
3537_2209_50_D09_037	Offspring	-	-
3537_2209_55_A10_040	Offspring	-	-
3537_2209_56_B10_039	Offspring	C	C
3537_2209_57_C10_038	Offspring	C	C
3537_2209_61_G10_034	Offspring	C	C
3537_2209_63_A11_048	Offspring	T	T
Number of offspring			**19**
Number of offspring that failed to amplify			**4**
All offspring expected heterozygous (CT)			
Number of observed heterozygous (CT) offspring			**4**
Number of observed CC homozygous offspring			**6**
Number of observed TT homozygous offspring			**5**

The maps used in this thesis are all made with ggmap (Kahle and Wickham 2013) as implemented in R.

6.2. Appendix 2: Additions to results

Table A6 *Model-fitting for the different markers and parameter estimates for the geographic cline ranging from Sauhestøya to Gothenburg. Markers with clines overlapping in width with the reference Q-score are depicted in boldface type. The Table does not include the 46 loci with allele frequency not meeting cline conditions.*[1]

Marker	Model	Centre (km)	Width (km)	δ_L	τ_L	δ_M	τ_M	pmin	pmax	loglike
STRUCTURE Q	fixL	874.4	138.7	45.2495	0.0733	NA	NA	0.1350	0.8650	-11.3274
PCA_82 SNPs	typM	947.5	71.2	NA	NA	18.5849	0.0322	0.0000	1.0000	-32.8957
BaWr4	fixN	615.8	2487.8	NA	NA	NA	NA	0.1880	0.3790	-3.2808
BaWr6	fixN	517.7	1676.4	NA	NA	NA	NA	0.3850	0.6000	-3.7905
BaWr11	fixN	784.8	2888.4	NA	NA	NA	NA	0.6720	0.8650	-2.7914
BaWr20	fixN	652.8	2262.0	NA	NA	NA	NA	0.5000	0.6880	-2.9272
BaWr22	fixN	676.0	1192.1	NA	NA	NA	NA	0.2400	0.5170	-3.4777
BaWr28	optN	941.7	12.6	NA	NA	NA	NA	0.3565	0.5258	-5.9477
BaWr29	fixN	758.3	3597.4	NA	NA	NA	NA	0.2860	0.6420	-7.7079
BaWr32	fixN	629.2	3138.8	NA	NA	NA	NA	0.7320	0.9290	-6.4987
BaWr33	optN	1105.3	7.7	NA	NA	NA	NA	0.2265	0.3600	-6.8513
BaWr34	fixN	1314.5	3597.1	NA	NA	NA	NA	0.3890	0.7250	-6.5006
BaWr36	fixN	737.4	2739.8	NA	NA	NA	NA	0.8130	0.9580	-4.1455
BaWr40	fixN	449.2	1721.0	NA	NA	NA	NA	0.0000	0.1600	-5.4941
BaWr46	optN	868.3	13.4	NA	NA	NA	NA	0.4978	0.9652	-8.3079
BaWr48	optN	897.1	14.2	NA	NA	NA	NA	0.4684	0.5722	-3.9052
BaWr54	fixN	848.6	1291.0	NA	NA	NA	NA	0.1190	0.2550	-1.3241
BaWr55	optN	955.8	0.1	NA	NA	NA	NA	0.6553	0.7833	-3.0977
BaWr58	fixN	319.1	2267.4	NA	NA	NA	NA	0.3700	0.5650	-2.9698
BaWr60	fixN	567.7	672.3	NA	NA	NA	NA	0.3450	0.7040	-4.7240
BaWr62	fixN	726.5	1841.5	NA	NA	NA	NA	0.8670	0.9830	-3.3752
BaWr63	fixN	686.9	3221.5	NA	NA	NA	NA	0.2500	0.5100	-4.0062
BaWr72	fixN	519.1	1416.2	NA	NA	NA	NA	0.3330	0.5430	-3.6004
BaWr75	optN	880.1	238.1	NA	NA	NA	NA	0.8076	0.9056	-2.1963
BaWr76	fixN	449.7	2425.0	NA	NA	NA	NA	0.8240	0.9780	-4.2932
BaWr78	fixN	532.3	1511.8	NA	NA	NA	NA	0.2130	0.5420	-3.8121
BaWr82	fixN	781.4	678.8	NA	NA	NA	NA	0.0600	0.5800	-9.9821
BaWr85	fixN	1008.3	2853.4	NA	NA	NA	NA	0.5170	0.7860	-5.7847
BaWr86	fixN	922.1	1390.7	NA	NA	NA	NA	0.1170	0.3250	-3.9833
BaWr87	fixN	981.9	1510.1	NA	NA	NA	NA	0.8230	0.9550	-2.8313
BaWr88	fixN	773.0	1596.5	NA	NA	NA	NA	0.5740	0.9060	-5.8110
BaWr90	fixN	691.6	1966.8	NA	NA	NA	NA	0.6030	0.7500	-3.5275
BaWr93	fixN	295.4	2706.7	NA	NA	NA	NA	0.3650	0.6220	-3.1814
BaWr99	fixN	802.7	1452.9	NA	NA	NA	NA	0.7870	0.9040	-3.0169
BaWr101	fixN	939.2	2532.4	NA	NA	NA	NA	0.1190	0.3020	-3.3098

[1] For the given models p_{min} and p_{max} were fixed to 0 and 1 (typ models). or to their empirical values (fix models). or p_{min} and p_{max} are fitted (opt model). Tail fitting encompassed right (R). left (L). none (N) or both fitted (B). The cline width (w) was calculated as 1/maximum slope. Two log-likelihood unit support limits are presented in parentheses for centre and width. Δ and τ are the shape parameters for the left and right tails. and p_{min} and p_{max} are the character states at either end of the transect

BaWr103	optN	988.1	53.4	NA	NA	NA	NA	0.3709	0.5355	-2.3716
BaWr104	fixN	1386.3	3591.5	NA	NA	NA	NA	0.3790	0.7000	-4.5359
BaWr105	fixN	1006.6	3593.8	NA	NA	NA	NA	0.2000	0.5670	-4.6073

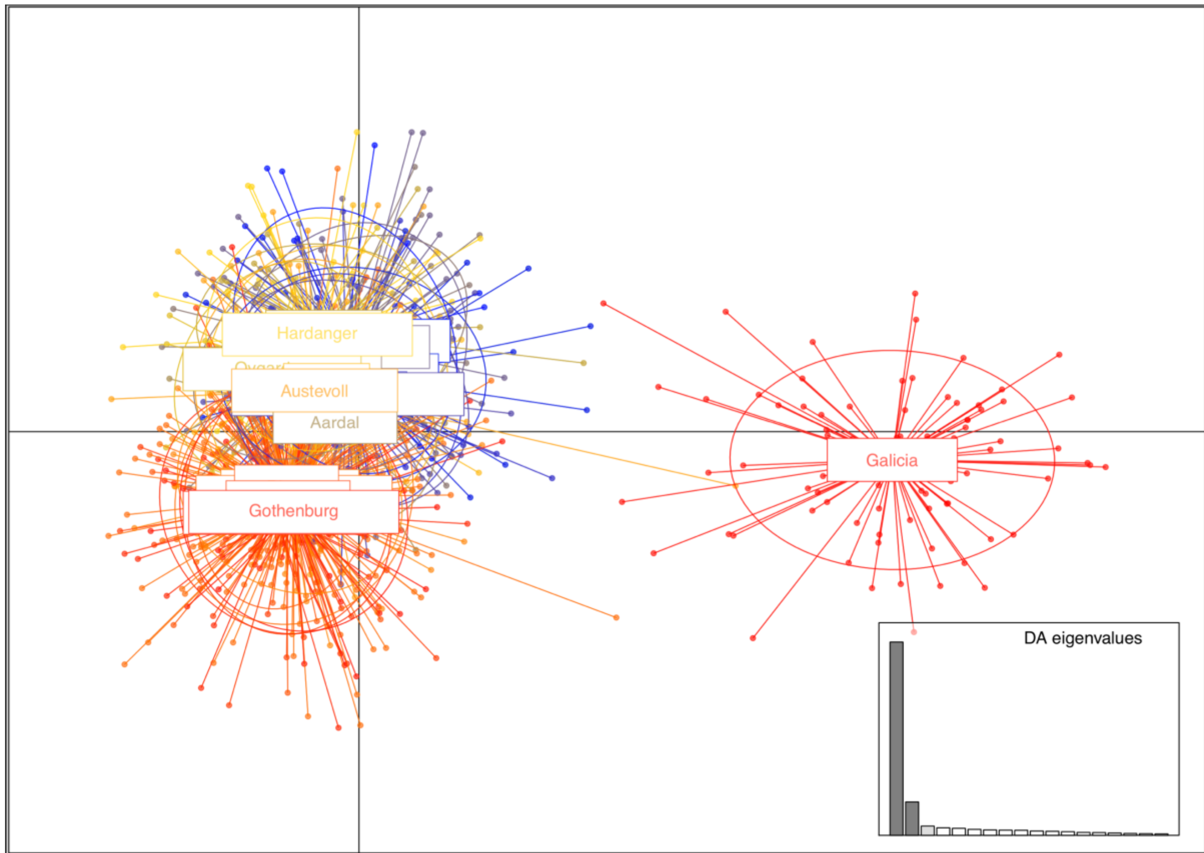


Figure A2 *Discriminant analysis of principle components (DAPC) for ballan wrasse samples, without three loci proposed to be under selection. Based on 79 SNPs. Coloration based on sampling location. Shading of DA eigenvalues signifies Discriminant analysis kept.*

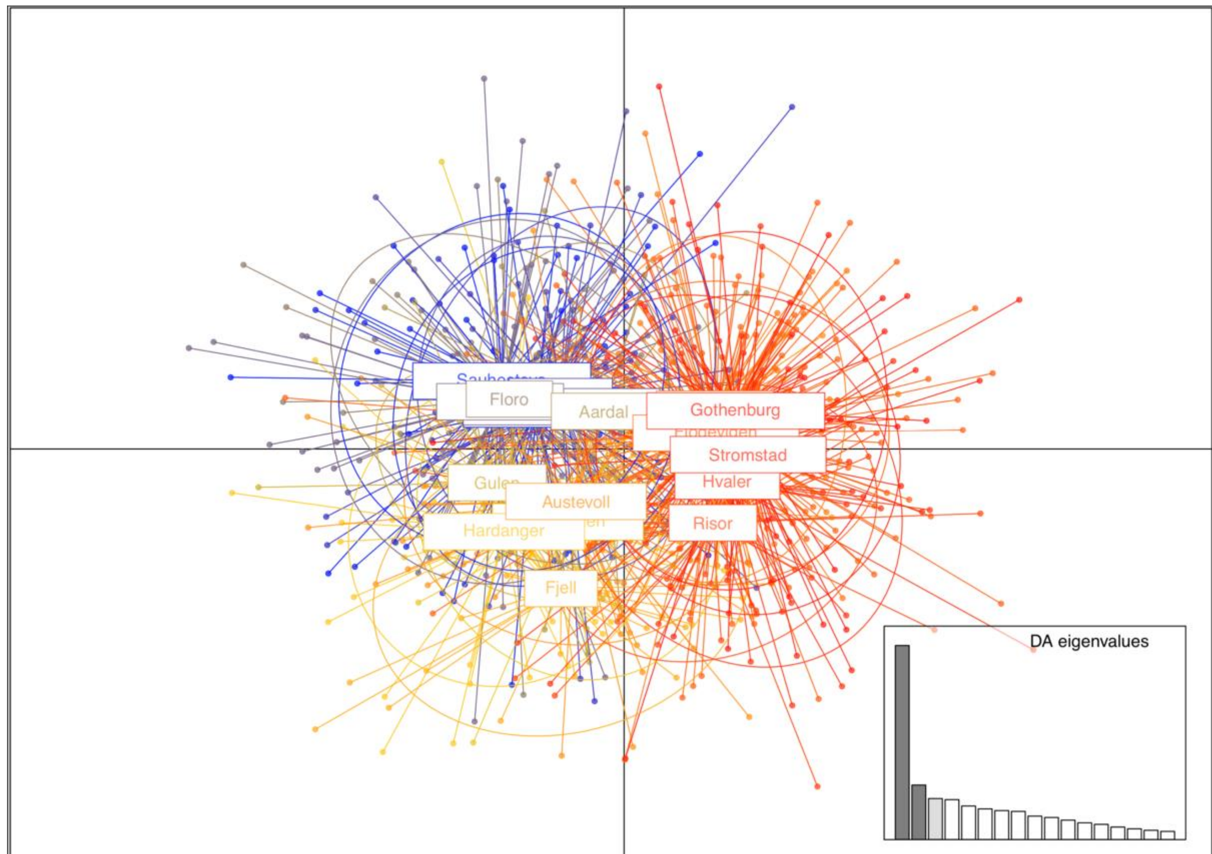


Figure A3 *Discriminant analysis of principle components (DAPC) for Scandinavian ballan wrasse samples only, without three loci proposed to be under selection. Based on 79 SNPs. Coloration based on sampling location. Shading of DA eigenvalues signifies Discriminant analysis kept.*

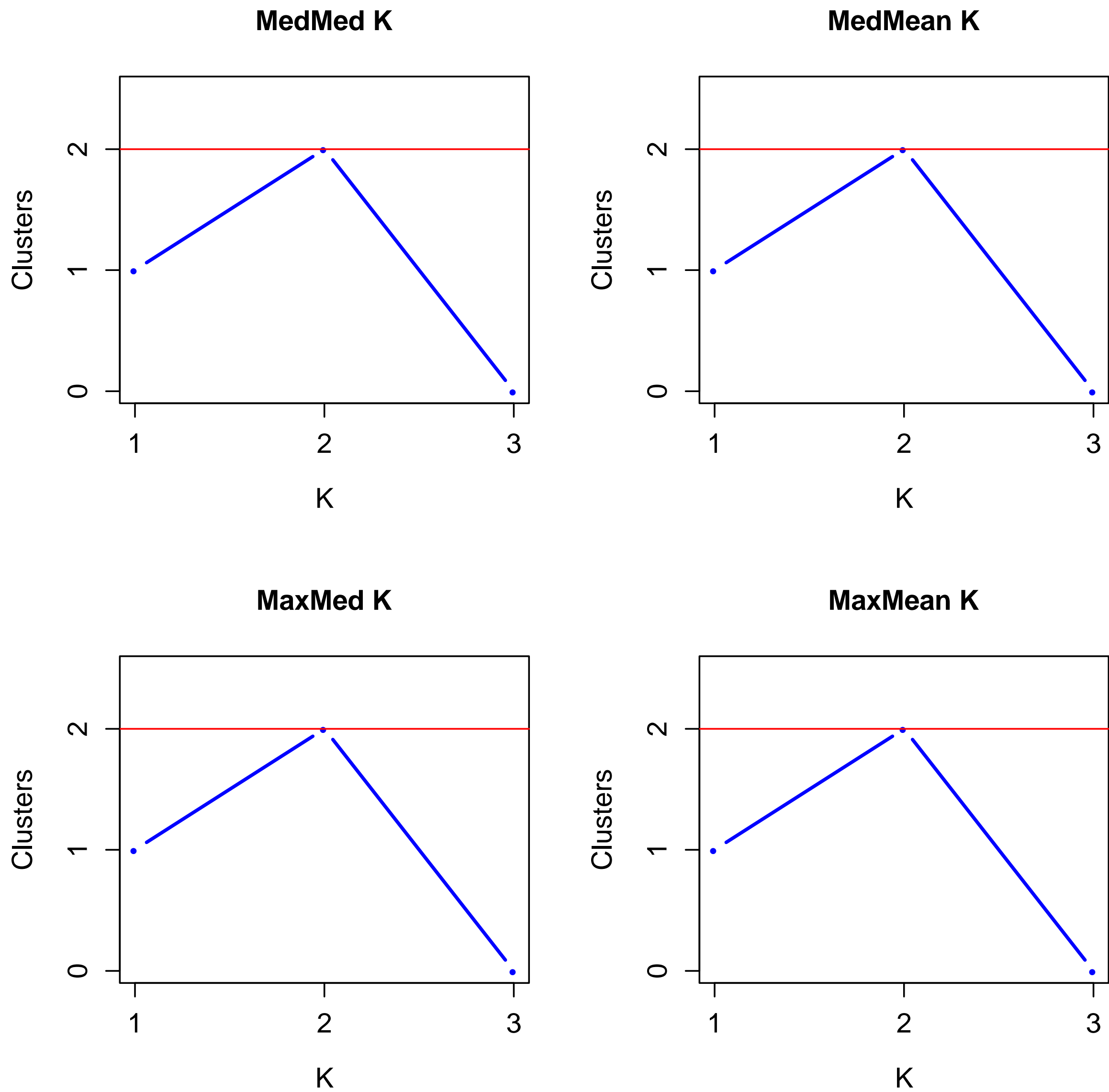


Figure A4 *Example of alternative statistics (MedMed, MedMean, MaxMed and MaxMean) for deciding number of clusters (K) in STRUCTURE results as generated by StructureSelector* (Li and Liu 2018).