LICENTIATE THESIS

Genetic Identification of Corkwing Wrasse Cleaner Fish Escaping from Norwegian Aquaculture

Ellika Faust

DEPARTMENT OF MARINE SCIENCES

UNIVERSITY OF GOTHENBURG

Cover photo: Corkwing wrasse nesting male
Credit: Paul Naylor at marinephoto.co.uk

Genetic identification of corkwing wrasse cleaner fish escaping from Norwegian aquaculture

Ellika Faust

Licentiate Thesis

UNIVERSITY OF GOTHENBURG

Department of Marine Sciences
Faculty of Science
2020

Supervisor:
Prof. Carl André

Co-Supervisor:
Dr. Pierre De Wit

Examiner:
Prof. Helle Ploug

Abstract

The genetic impact of farmed fish escaping aquaculture is a highly debated issue. However, non-target species, such as cleaner fish that are used in fish farms to remove parasitic sea lice, are rarely considered. Here, we report that wild corkwing wrasse (*Symphodus melops*), which are transported long distances to be used as cleaner fish in salmon farms, escape and hybridize with local wrasse populations. Recently, increasing numbers of corkwing wrasse have been reported north of its described distribution range, in Flatanger in Trøndelag in Norway, an area heavily relying on the import of cleaner fish from Skagerrak. Using a high number of nuclear genetic markers identified with 2bRAD sequencing, we show that, although the Flatanger population is largely a result of a northward range expansion, there is also evidence of considerable gene flow from southern populations in Skagerrak. Of the 40 corkwing wrasses first sampled in Flatanger, we discovered two individuals with clear southern genotypes, one first-generation hybrid, and 12 potential second-generation hybrids. Thus, we found clear evidence of gene flow from source populations of translocated cleaner fish at the edge of an ongoing northwards range expansion.

To better understand the extent of gene flow we then greatly expanded our sampling. Based on patterns of genetic divergence and homogeneity, we identified a smaller battery of 84 SNPs which is able to detect escapees with a Skagerrak origin as well as first and second-generation hybrids with high accuracy and power. We then used these SNPs to investigate the magnitude and geographic extent of escaping and hybridizing cleaner fish along the Norwegian coast. We found that escapees and hybrids may constitute up to 20 % of the local populations at the northern edge of the species distribution. In other parts of the Norwegian coast where salmon farming is also common, we found surprisingly few escapees and hybrids. Possible causes for few escapees and hybrids found in these areas are difficult to evaluate with the current lack of reporting of translocations by aquaculture operators.

Overall, these findings provide critical information both for aquaculture management and conservation of wild populations of non-target species, and have implications for the increasing use of cleaner fish as parasite control in fish farms, that is both poorly documented and regulated. Moving genetic material between isolated populations could drastically alter the genetic composition and erode population structure, potentially resulting in loss of local adaptation and hampering natural range expansion. Although the ecological and evolutionary significance of escapees warrant further investigation, these results should be taken into consideration in the use of translocated cleaner fish.

Keywords: Conservation, Population structure, Genetics, Aquaculture, Hybridization, Corkwing wrasse, Cleaner fish, Sea lice, *Symphodus melops*, Escapee, Range expansion

Populärvetenskaplig sammanfattning.

Läppfiskar är så kallade putsarfiskar, vilket betyder att de har ett naturligt beteende var de plockar och äter parasiter som sitter på huden på andra större fiskar. På 1980-talet upptäckte man att läppfiskar även kan äta laxlus, en vanlig parasit som orsakar stora problem inom laxodling. I slutet av 2000-talet började laxlusen bli resistent mot kemiska bekämpningsmedel, vilket ökade efterfrågan på putsarfisk. Sedan 2008 har användningen av putsarfiskar i norsk laxodling ökat exponentiellt, och nu används ca 50 miljoner putsarfiskar inom norsk laxodling varje år. Många av de läppfiskar som används som putsarfisk fångas i områden långt från de odlingar de används i. Framförallt fiskas mycket läppfisk i Skagerrak, för att sedan transporteras levande i tankbilar till odlingar på norska västkusten där lokala populationer saknas eller inte kan möta efterfrågan.

Under de senaste åren har norska fiskare sett att skärsnultror, en av de mest använda arterna av putsarfisk, har etablerat sig i nya områden, norr om deras normala utbredning. Då det finns många odlingar i området, uppstod frågan om detta kunde vara ett resultat av att importerade fiskar hade rymt. Med hjälp av genetiska metoder kunde vi undersöka 40 individer från området och jämförde dem med sydligare populationer i Norge, och från Sverige. Utifrån resultaten kan vi se att den nya populationen i Flatanger norr om Trondheim verkar vara ett resultat av att arten har börjat expandera norrut, men också att importerade individer rymt från laxodlingar och börjat förökat sig med de lokala populationerna.

Att putsarfiskar som ursprungligen kommer från Skagerrak blandar sig med populationer längs den norska kusten kan få både genetiska och ekologiska konsekvenser. Lokala populationers tillstånd riskeras att försämras om gener som är sämre anpassade till den lokala miljön sprids i populationen. Då kan den lokala anpassningen, som tagit tusentals år att utveckla, under kort tid gå förlorad. Rymlingar kan också påverka andra arter i form av ökad konkurrens om föda och boplatser. Men också genom att introducera nya sjukdomar och parasiter till området som lokala arter och populationer inte har utvecklat något skydd mot.

För att bättre förstå hur utbrett och vanligt det är med rymlingar i vilda populationer gjorde vi en andra studie där vi ökade vår provtagning både geografiskt och i antal fiskar. Med hjälp av ett litet antal utvalda genetiska markörer analyserade vi strax under 2000 vilda skärsnultror längs den norska kusten. Resultaten visade att upp till 20% av alla individer i den nordliga populationen i Flatanger, kan vara putsarfisk som rymt eller deras avkommor. I andra delar längs den norska kusten, var laxodling också är vanligt, hittade vi förvånansvärt få individer med sydligt ursprung. Möjliga orsaker till att vi ser få rymlingar och hybrider i andra delar av utbredningsområdet är svårt att utvärdera eftersom mängden förflyttad putsarfisk inom Norge inte är känd. Även om ekologiska och evolutionära konsekvenser av rymd putsarfisk behöver vidare utredning, bör dessa resultat tas i beaktning i det framtida användandet av putsarfisk.

Att fisk som rymmer från odlingar kan ha stora effekter på vilda populationer är ett välkänt problem. För lax och öring finns det övervakningsprogram och handlingsplaner för hur man ska förebygga och hantera odlad fisk som rymmer. Detta har gjort att problemen med rymningar av dessa arter minskat kraftigt. Regelverket inkluderar dock inte putsarfiskar, och

för dessa saknas regler för att motverka rymningar. I nuläget är ett av de största hindren för en hållbar förvaltning av putsarfisk avsaknaden av dokumentation om var och hur mycket fisk som flyttas. Transportörer bör dokumentera och rapportera både källan och destinationen av fiskar som förflyttas för att det överhuvudtaget skall bli möjligt att åtgärda risken med rymlingar.

List of papers

Paper I:

Faust, E., Halvorsen, K.T., Andersen, P., Knutsen, H., André, C., 2018. Cleaner fish escape salmon farms and hybridize with local wrasse populations. Royal Society Open Science 5, 171752. https://doi.org/10.1098/rsos.171752

Paper II:

Faust, E., Jansson, E., André, C., Halvorsen, K.T., Dahle, G., Knutsen, H., Quintela, M., Glover, K.A., 2020. Large scale survey of escape and hybridisation of cleaner fish in aquaculture. *Manuscript*

Other publications not in this thesis

Faust, E., André, C., Meurling, S., Kochmann, J., Christiansen, H., Jensen, L. F., Charrier, G., Laugen, A. T., Strand, Å. (2017). Origin and route of establishment of the invasive Pacific oyster *Crassostrea gigas* in Scandinavia. Marine Ecology Progress Series, 575, 95–105. https://doi.org/10.3354/meps12219

Seljestad, G. W., Quintela, M., **Faust, E.**, Halvorsen, K. T., Besnier, F., Jansson, E., Dahle, G., Knutsen, H., André, C., Folkvord, A., Glover, K. A. (2020). "A cleaner-break": Genetic divergence between geographic groups and sympatric phenotypes revealed in ballan wrasse (*Labrus bergylta*). Ecology and Evolution. *Accepted*

Table of Contents

Introduction 7

Cleaner fish in aquaculture 7

Cleaner fish translocation 10

Study species 12

Knowledge gap 14

Thesis aims 14

Summary of Paper I 14

Summary of Paper II 17

Discussion 19

Novelty and significance 21

Acknowledgments 21

References 22

Introduction

The translocation and introduction of non-native organisms is a well known issue within management and conservation. Biological invasion in the marine environment has been highlighted as a global threat to biodiversity and biological communities, often as one of the top conservation concerns (IPCC, 2019; Molnar et al., 2008). Moving organisms outside their natural boundaries comes with many potential problems and can have a diverse range of ecological, genetic, pathogenic and socio-economic impacts (Atalah & Sanchez-Jerez, 2020). Once introduced to the wild, a successful invader can affect the whole ecosystem, by altering local food webs or community structure, through competition, predation or even by changing the abiotic environment (Crooks, 2002). For example, the Pacific oyster is able to completely alter the environment they colonise. By creating hard, and often large structures, they can change a sandy soft bottom into a completely different habitat (Troost, 2010).

Additionally, introduced organisms are seldom alone. A single individual can carry a variety of different organisms, ranging from symbionts, parasites or even pathogens. Although some of these might already exist in the environment, others will be novel and can quickly spread throughout the local ecosystem, which has not been able to create any form of resistance (Tepolt et al., 2020). Just one example is the introduction of the rinderpest virus into sub-Saharan Africa. The virus, which was transmitted through domestic cattle, decimated native ungulates (McCallum & Dobson, 1995).

Even if a species is already present, introduced individuals of the same species may not be ecologically equivalent. These newcomers may vary strongly in their ecological impacts compared to the pre-existing population, for example through differences in prey consumption (Evangelista, Cucherousset, and Lecerf 2019). If the introduced individuals are genetically divergent from the local population they may introduce unfavourable genetic material into the genepool through admixture and introgression. This can result in altered population subdivision (Glover et al., 2012), reduced genetic variation, and/or reduced fitness (Blakeslee et al., 2020; Glover et al., 2017; Laikre et al., 2010).

Genomic and genetic methods for understanding and tracking the effects of biological invasions have improved our understanding of evolutionary processes but also become an aid and a tool for management and conservation (Comtet et al., 2015; Rius et al., 2015; Viard et al., 2016; Viard & Comtet, 2015). Genetic tools can be used for understanding the route of introduction (Faust et al., 2017; Ficetola et al., 2008) as well as tracking the degree of admixture and introgression between introduced and local populations (Glover et al., 2012). Although the need for genetic information has been incorporated into many management policies, the implementation of available genetic knowledge into regulation is still limited (Lowe and Allendorf 2010; Sandström et al. 2016; Lundmark et al. 2019).

Cleaner fish in aquaculture

Farmed fish escaping aquaculture has been identified as a serious threat to wild fish populations (Atalah and Sanchez-Jerez 2020). Open-pen farming has been shown to have

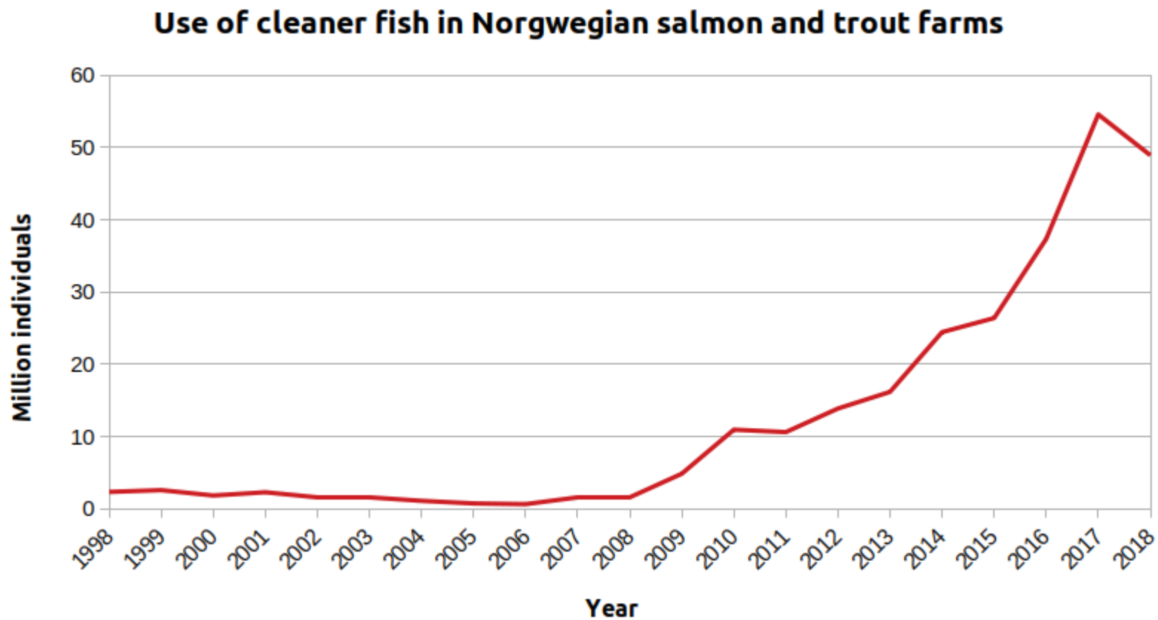
large impact on local populations as escapees have hybridized with local fish, leading to both genetic swamping and reduced fitness (Bolstad et al. 2017; Glover et al. 2017). Salmon farming may also promote inadvertent gene flow of other species such as wrasse, which are used to mitigate sea lice infestations in the farmed salmon (Blanco Gonzalez and de Boer 2017).

Salmonid fish are among the most intensively farmed fish in marine and coastal aquaculture globally. Of all aquaculture species, Atlantic salmon has been ranked #2 in terms of production value, thereby making it the fish species with the highest production value in the world (Cai et al., 2019). Sea lice infestations are a major issue within salmonid aquaculture, in particular the salmon lice (*Lepeophtheirus salmonis*). Salmon lice has been estimated to cost the industry €300-360 million annually and has a greater economic impact than any other parasite (Costello, 2009b; Lafferty et al., 2015). Furthermore, increasing evidence has demonstrated that the lice from aquaculture can cause significant mortality in wild fish populations (Costello, 2009a). Thus, finding a successful treatment, that is both effective as well as safe for the fish and the environment, is of great importance for the salmonid farming industry.

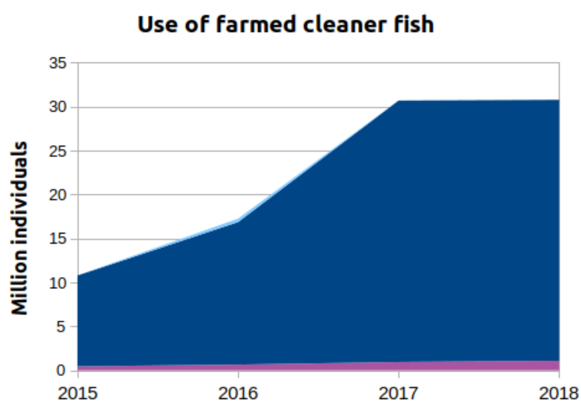
Several species of wrasse exhibit a natural symbiotic cleaning behaviour, removing ectoparasites from larger fish and other organisms (Baliga & Law, 2016). In the late 1980s it was discovered that this natural cleaning behaviour could also be used to reduce infestations of sea-lice (*Lepeophtheirus salmonis* and *Caligus elongatus*) in commercial salmon aquaculture (Bjordal, 1988; Darwall et al., 1992). Since the 1990s a small number of wild-caught wrasse have been used for sea lice control. However, the use of cleaner fish increased dramatically since 2008 (Figure 1), partially due to sea lice developing resistance to widely used pharmaceutical treatments (Besnier et al., 2014; Kaur et al., 2017). The number of cleaner fish used in Norway alone has increased from 1.7 million in 2008 to ~50 million in 2017 and 2018 (Figure 1a).

Currently five fish species cleaner fish are used for parasite control in Norwegian aquaculture: lumpfish (*Cyclopterus lumpus*), ballan wrasse (*Labrus bergylta*), goldsinny wrasse (*Ctenolabrus rupestris),* corkwing wrasse *(Symphodus melops)* and small amounts of rock cook *(Centrolabrus exoletus*) (Norwegian directorate of Fisheries, 2019). Since 2014, when its potential use as a cleaner fish was discovered, lumpfish has become the most commonly used cleaner fish in Norwegian aquaculture (Imsland et al., 2014). The majority of lumpfish are farmed, almost all wrasse are caught in the wild and transported to aquaculture facilities. Currently, the only commercially reared wrasse species is ballan wrasse, although at a very small scale (Figure 1b). Goldsinny and corkwing wrasse are, by far, the most commonly used wild caught cleaner fish (Figure 1c). In 2018, 7.4 million goldsinny and 6.3 million corkwing wrasse were deployed as cleaner fish in Norwegian aquaculture (Norwegian directorate of Fisheries, 2019).

(A)



(B)



(C)

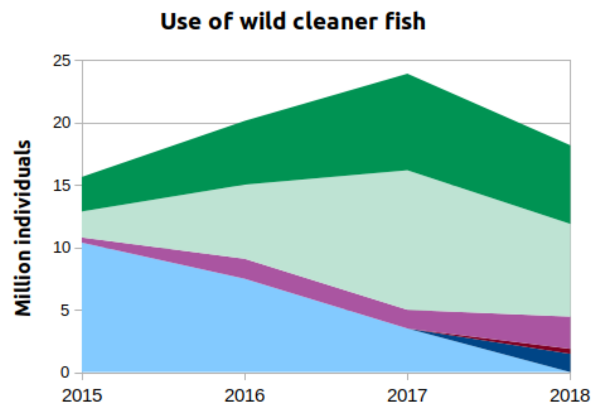


Figure 1. The use of cleaner fish in Norwegian salmon and trout farms (A) between 1998 and 2018. (B) Annual use of farmed cleaner fish by species between 2015 and 2019. (C) Annual use of wild cleaner fish by species between 2015 and 2018. Non-specified refers to wrasse with no species name recorded. Source: Norwegian directorate of Fisheries

The use of cleaner fish as parasite control in other parts of the world is still relatively small but is likely to increase (VKM 2019). While some countries, e.g. Canada, do not allow the use of wild caught cleaner fish in open marine aquaculture (Boyce et al., 2018), others, such as the UK, apply a similar system to Norway with a mix of farmed and wild-caught cleaner fish. Currently, an estimated 1 million wrasse are harvested in southwestern England annually for live transport to salmon farms in Scotland (Devon & Severn, 2017; Riley et al., 2017). Other countries, e.g. Chile, are only starting to investigate the possibility of utilizing cleaner fish for parasite control (Sánchez et al., 2018).

Cleaner fish translocation

Millions of wrasse are used as cleaner fish in Norwegian aquaculture annually, and in many regions the aquaculture demand for cleaner fish exceeds what can be supplied from local stocks. Consequently, large quantities of wild-caught wrasses are imported from other areas often hundreds kilometres away (Figure 2). Since 2010, ballan wrasse, goldsinny wrasse and corkwing wrasse have been targeted by Swedish fisheries and 600 000 to one million wrasse are exported to Norway annually (Andersson, 2019) (Figure 2). Where in Norway wrasses imported from Sweden are deployed was not recorded prior to 2017, when it became mandatory to report source and destination of imported wrasse. Since 2017 we know that the majority of imported wrasse is transported to the Trøndelag region in mid-Norway (Figure 2).

A recent report by the Norwegian Scientific Committee for Food and Environment (VKM) suggests that hybridization between imported cleaner fish and local fish could cause genetic changes with severe negative impact on local populations of corkwing and ballan wrasse and potentially lead to reduced viability and adaptability of local goldsinny wrasse (VKM 2019). They assessed that there is a moderate risk of genetic change in all wrasse species as well as a moderate risk of negative impact from corkwing wrasse spreading beyond the species range. In this report, only wrasse imported from Sweden were addressed, however, much larger numbers of wrasse are being transported long distances within Norway. Southern Norway, adjacent to the Swedish wrasse fisheries, has few fish farms but high densities of wild wrasse (Skiftesvik et al., 2014; VKM 2019). Approximately ~20% of all wild cleaner wrasse are caught in southern Norway annually, but most years less than 1% of all cleaner fish are deployed in that area (Norwegian directorate of Fisheries, 2019). In contrast to imported wrasse, there are currently no requirements to record the source or destination of cleaner fish that are caught in Norway, even though translocation distances can exceed 1000 km.

A Caught wrasse
Million 0 4 8 12 16

B Deployed wrasse
Million 0 5 10 15

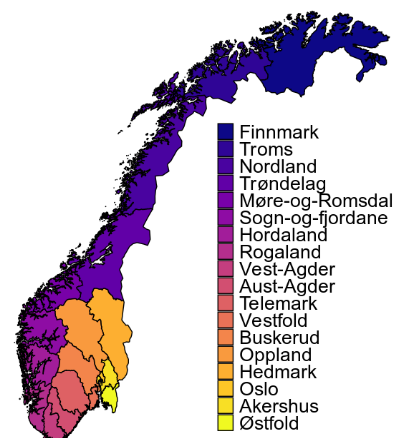
C Imported wrasse
Million 0.0 0.1 0.2 0.3 0.4 0.5 0.6

D Caught corkwing
Million 0 2 4 6 8 10

E Deployed corkwing
Million 0 2 4 6 8

F Imported corkwing
Million 0.0 0.1 0.2 0.3

Figure 2. Map of Norway showing number of (A) caught wrasse, (B) wrasse deployed, (C) destination of imported wrasses from Sweden, (D) caught corkwing, (E) corkwing deployed, (F) destination of imported corkwing from Sweden, in 2017 and 2018 (total) for each county. (G) Map of counties. Catch data and deployment data: Norwegian directorate of Fisheries. Data on imported wrasses from Sweden was provided by the Norwegian Environmental Agency. Disclaimer: The number of actors deploying cleaner fish on the Norwegian south coast are very few. For the sake of anonymity in reported deployment statistics, no species-segregated data for the south coast counties is reported for individual counties and is thus not included in the above map.



Study species

Corkwing wrasse is a marine fish species of the family Labridae native to the eastern Atlantic, with a natural distribution from Morocco to mid-Norway (Figure 3a) (Knutsen et al., 2013; VKM 2019). They can live up to eight to nine years (Darwall et al., 1992; Halvorsen et al., 2016; Uglem et al., 2000), and grow up to 24 cm in length, making it the second largest species of wrasse in Scandinavia (Halvorsen et al., 2016). Similar to other wrasse species, corkwing inhabit rocky shores and reefs along the coast where they can often be found in areas as shallow as at 5 m depth (Skiftesvik et al., 2014). Corkwing wrasse is a territorial and nest building species, with male parental care until eggs have hatched (Halvorsen et al., 2016; Potts, 1985). During the spawning season (May-July) nesting males display bright blue, green and red colours (Figure 3b) in order to attract females to their nests (Potts, 1974). Females are brown/grey in colour and much smaller in body size than the nesting males. A small proportion of males employ female mimicry and do not build nests but rather perform sneak spawning (Figure 3c) (Uglem et al., 2000). The male morphs are believed to be fixed for life and could potentially be genetically determined (Halvorsen et al., 2016). Some concern has been raised that current size limits in the Norwegian wrasse fishery may be sex selective, as nesting males grow faster and mature later than females and sneaker males (Halvorsen et al., 2016, 2017).

Earlier studies of corkwing wrasse have found a reduced genetic diversity in northern Europe aligned with a large genetic break between Atlantic and Scandinavian populations, likely caused by the populations undergoing a bottleneck as it expanded northwards (Knutsen et al., 2013; Robalo et al., 2012). A second genetic break along the Norwegian coast was later discovered by Blanco Gonzalez et al (2016). They found that a long stretch of sandy beaches (<60 km long), which is an unsuitable wrasse habitat, separates southern Skagerrak populations from western North Sea populations. Corkwing wrasse is a non-migratory fish species which lays benthic eggs and is dependent on the planktonic larval stage for dispersal (Darwall et al., 1992). Thus, this large unsuitable habitat might act as an environmental barrier for gene flow. Recent analysis of demographic history by Mattingsdal et al. (2020) shows that the genetic divergence between the populations might be a result of post-glacial recolonization and founder events separating the populations for more than ~10 kya, followed by a secondary contact. Given the low number of hybrids it is likely that the secondary contact is very recent or hybrids are actively selected against (Mattingsdal et al., 2020).

Skagerrak populations south of the genetic break have a much lower genetic diversity than their north-western counterparts, and they also have different life histories (Halvorsen et al., 2016; Mattingsdal et al., 2020). Fish belonging to the southern population grow faster, mature earlier and rarely reach more than four years of age (Halvorsen et al., 2016). Furthermore, the ratio between nesting and sneaker males differs between the two regions, with few sneaker males in the south. However, as Norwegian fisheries only apply a minimum size limit, this could be a result of selective fishery where nesting males are likely to be targeted disproportionately (Halvorsen et al., 2017).

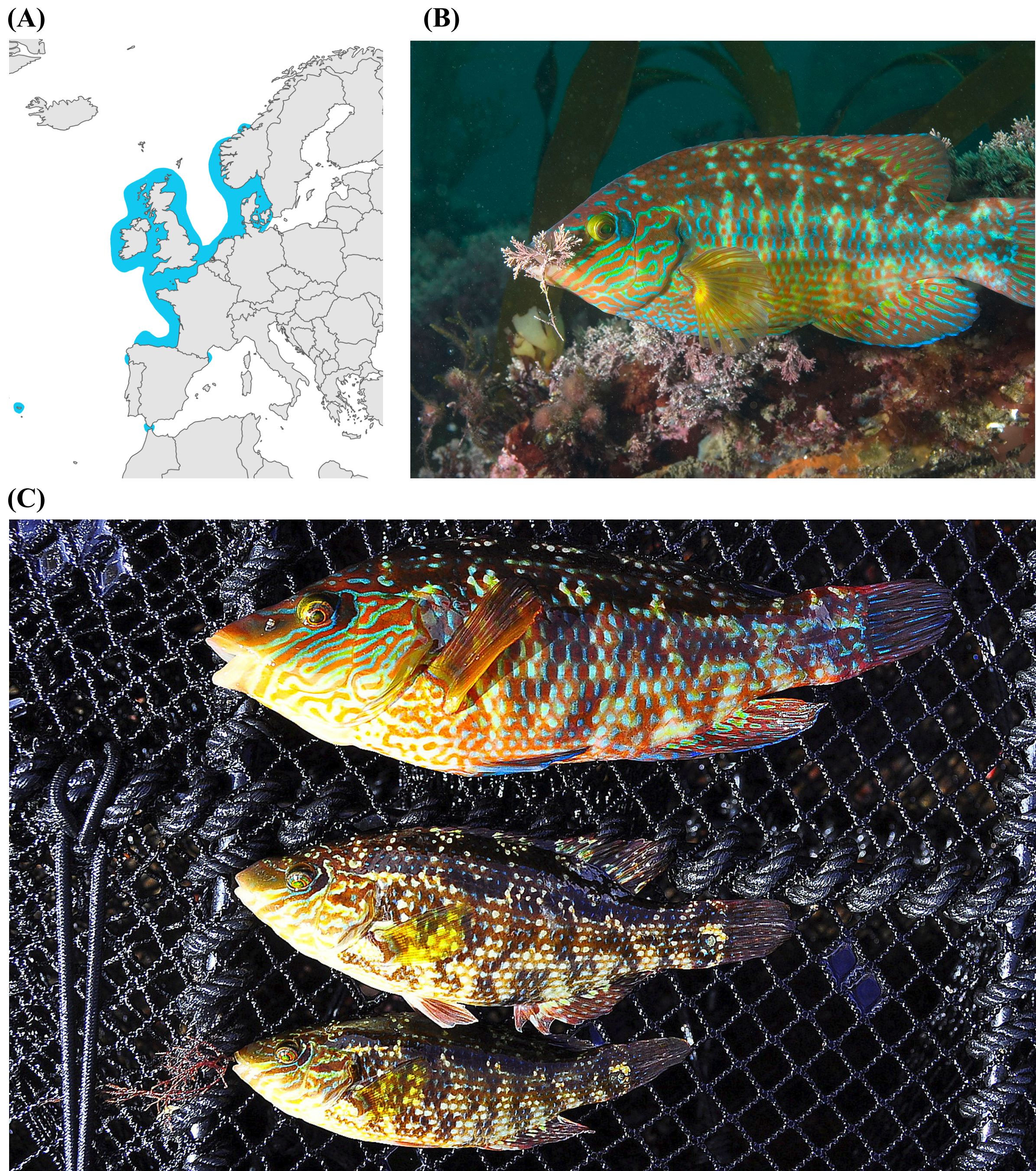


Figure 3. (A) Corkwing wrasse distribution (VKM 2019). (B) A corkwing wrasse nesting male during spawning season, carrying a piece of seaweed. Photo: Paul Naylor at marinephoto.co.uk. (C) Corkwing wrasse sexual reproduction strategies, from top to bottom: nesting male, female and sneaker male. Photo: Tonje K. Sørdalen.

Knowledge gap

Cleaner fish are a low-cost type parasite control and are often considered to be more environmentally friendly than other delousing methods (Liu & Bjelland, 2014). However, both the increasing fishing pressure and the large scale combined with long distance translocation raises concerns of potential overfishing and human-mediated introductions of novel genetic material. A recent study by Jansson et al. (2017) found reduced genetic divergence between wild goldsinny wrasse in aquaculture dense regions in mid-Norway, and populations in southern Norway and Sweden, which indicates past or ongoing gene flow due to translocation.

In recent years an increasing number of observations of corkwing wrasse have been reported in the Flatanger municipality in mid Norway, a region 130 km north of the previously described species range (Maroni & Andersen, 1996). The most natural conclusion would be that the species is expanding its range northwards. However, the Flatanger region is an area densely populated with salmonid aquaculture and is heavily relying on the import of cleaner fish from southern populations. Thus, the question arises whether the newly established population in Flatanger could be a direct effect of imported cleaner fish.

Currently around 50 million cleaner fish are deployed in Norwegian salmonid farms annually. Risks associated with farmed fish escaping aquaculture is a highly debated issue. However, in contrast to salmonids, there are no monitoring programs nor action plans for how to prevent and or deal with escaping cleaner fish. Currently it is unknown how many corkwing cleaner fish have been able to escape, and whether there is a difference between regions in the number of escapees and the extent of genetic admixture with local populations.

Thesis aims

This thesis has three major aims:

1. Investigate whether the newly established population in Flatanger at the northern edge of the corkwing wrasse distribution is a consequence of a northwards range expansion, cleaner fish escaping salmon farms or a mix of both.
2. Investigate the quantity and geographic extent of corkwing wrasse escaping Norwegian salmon farms
3. Develop a tool for management to aid monitoring of escapees mixing with wild populations

Summary of Paper I

In this paper we examined the origin of the recently established population of corkwing wrasse (*Symphodus melops*) in Flatanger, 130 km north of its natural distribution range. Flatanger municipality is an area in Norway with many salmonid farms that rely heavily on

the use and import of cleaner fish such as corkwing wrasse from Skagerrak. Reports have suggested that it is possible for cleaner fish to escape from salmon farms through tears in the net, slipping through the mesh, or even intentional release at the end of the season (Blanco Gonzalez & de Boer, 2017; Svåsand et al., 2017; Woll et al., 2013). However, corkwing wrasse has also increased in abundance in other areas in Scandinavia, suggesting that warmer temperature might allow the species to expand in the north (Knutsen et al. 2013). In this study we aimed to answer the question whether the newly established population in Flatanger was 1.) A direct result of these cleaner fish escaping aquaculture facilities and establishing a feral population, 2.) A result of the species expanding its range northwards, or 3.) Due to a combination of these two processes.

In order to answer this question, we sampled a total of 240 individuals from six different locations, one in Flatanger, two in southwestern Norway, where wrasse is harvested but used locally, and three locations on the Skagerrak–Kattegat coast, where all commercially caught wrasses are transported to salmonid farms in mid- and northern Norway. We used the restriction-site-associated DNA (RAD) sequencing method 2b-RAD (Wang et al., 2012) to identify SNPs and genotype the individuals. Genomic DNA was extracted from fin clips and RAD libraries were prepared according to a protocol modified from Matz & Aglyamova (2014). We pooled all samples with individual barcodes and sequenced as single-read, 50 bp target length sequencing, on an Illumina HiSeq2500 platform. The bioinformatic analysis of the DNA sequences followed a modified de novo pipeline from Pierre de Wit (2016). After removing genotyping errors and uninformative polymorphisms, 4372 SNPs remained.

We estimated population differentiation by calculating pairwise F_{ST}, and used two individual-based clustering methods (STRUCTURE and PCA) to estimate genetic differentiation among individuals. Finally, we investigated the occurrence of hybridization with NEWHYBRIDS in the Flatanger location using 200 highly differentiated SNPs to assign Flatanger individuals to six different hybrid classes (pure western, pure southern, F1, F2, western backcross or southern backcross). We assed accuracy and power to identify individuals of the different hybrid classes with the set of 200 SNPs by simulating and analysing data based of western and southern allele frequencies.

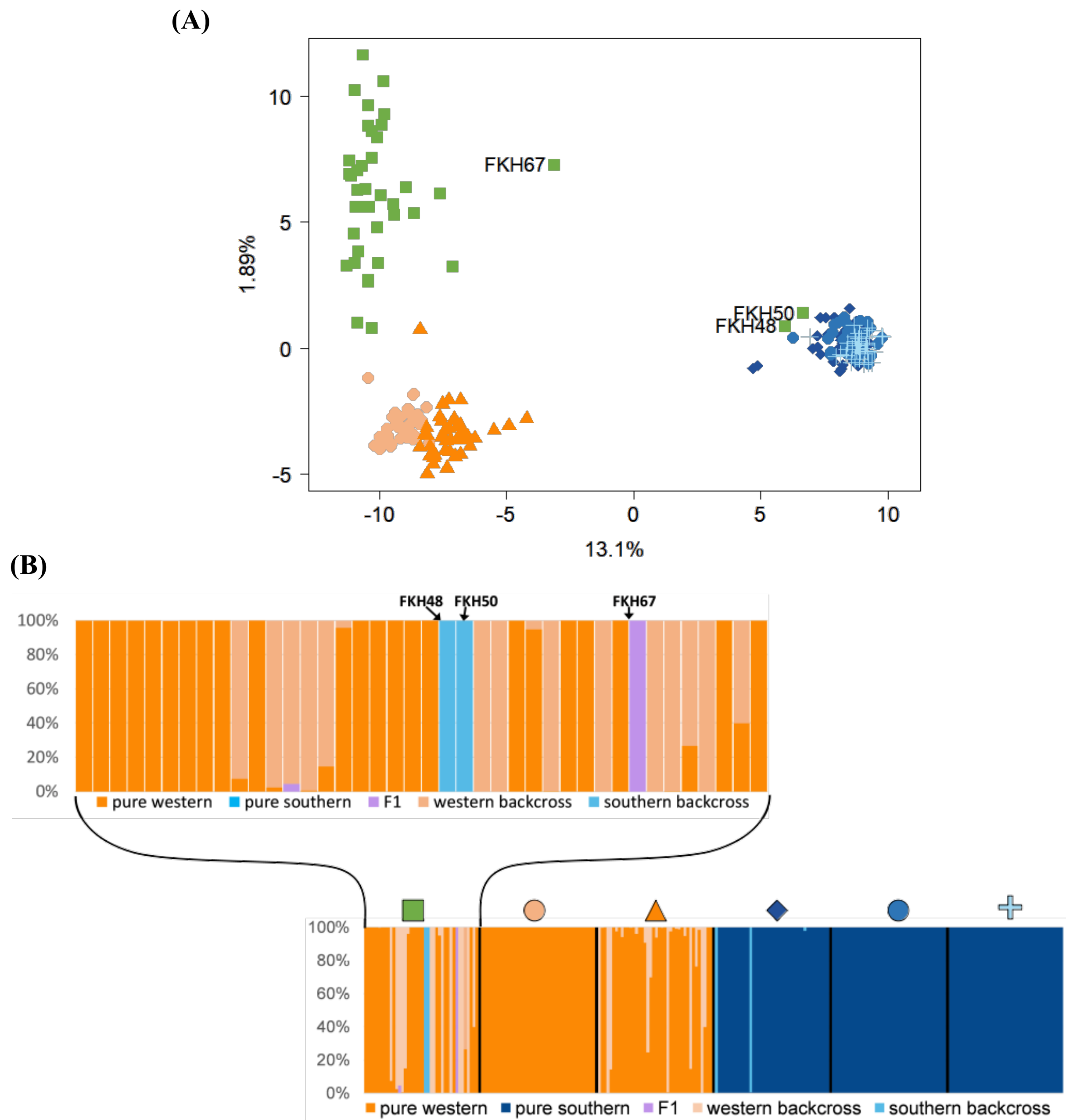


Figure 4. (A) The first (x-axis) and second (y-axis) components of a principal component analysis on 240 corkwing wrasse individuals from 6 locations based on 4357 SNPs. The first component explains 13.1% of the total variation and the second 1.89%. Each point represents one individual, and colour and symbols represent sampling sites. (B) Hybrid analysis of all individuals (bottom) and individuals sampled in Flatanger (top) using the 200 SNPs with highest F_{ST} estimates in NEWHYBRIDS. Each vertical line represents one individual and its probability to belong to one of the six genotype classes, no F2 genotypes were present. Green square = Flatanger in mid Norway. Orange circle and triangle = western Norway. Blue diamond, circle and plus sign = Skagerrak/Kattegat.

We found that Flatanger was overall more genetically similar to the western samples than to southern Skagerrak-Kattegat populations. This suggests that the species is going through a natural range expansion. However, individual based analysis revealed that some individuals were genetically much closer to the Skagerrak-Kattegat populations (Figure 4). Two individuals clustered with the southern population in both STRUCTURE and the PCA, and were identified as southern backcrosses by NEWHYBRIDS (i.e. 75% southern genotype and 25% western genotype). One individual was classified as a F1 hybrid, and an additional 12 individuals from Flatanger had a high probability of being western backcrosses (i.e. 75% western genotype and 25% southern genotype). Thus, there are escapees in Flatanger and they are hybridizing with the local population.

In summary, we found that the Flatanger population is mainly a result of a northward range expansion, but there has also been considerable gene flow from southern populations in Skagerrak and Kattegat. Our results provide the first evidence that corkwing wrasse escape from fish farms and hybridize with local populations. Although more investigation is needed to estimate the magnitude and effects of escapees on local populations and ecosystems, these results provide important information for the future use of translocated cleaner fish.

Summary of Paper II

In **Paper I** we discovered that corkwing wrasse were able to escape and hybridise with local populations at the northern edge of the species distribution, and we could use genetic markers to detect these individuals. However, we only investigated a relatively small number of individuals from a single region. Thus, the geographical extent and magnitude of escapees and introgression is still unknown. To this end we expanded upon our first study by genotyping a large number of wild caught corkwing wrasse along the Norwegian west coast in areas heavily relying on the use of cleaner fish. A second aim was to develop a suite of genetic markers that can be used by management authorities for future monitoring of escapees and hybrids in the wild.

We used 2b-RAD sequences from **Paper I** and mapped them to the genome of *S. melops* (Mattingsdal et al. 2018). We then identified SNP loci with high divergence ($F_{ST} > 0.4$) between western and southern samples, which were used for primer design, amplification and genotype calling, based on the low cost Agena MassARRAY iPLEX Platform (Gabriel et al. (2009). Similarly, to **Paper I**, accuracy, efficiency and power to correctly identify escaping individual hybrids was assessed by simulating data based on western and southern allele frequencies.

In order to cover a large geographic area as possible, samples were collected opportunistically, resulting in varying sample sizes and sample time points. Genomic DNA was extracted from a total of 1955 unique individuals and 105 technical replicates which were then genotyped in four multiplex groups for 106 SNPs. After filtering, the final data set consisted of 1766 unique individuals genotyped for 84 loci with a total of 2.9 % missing data. Genetic differentiation was estimated by calculating pairwise F_{ST} and two individual-based

clustering methods STRUCTURE and PCA. The frequency of escapees and hybrids was estimated with NEWHYBRIDS and accuracy and power was re-assessed with the 84 of SNPs remaining after filtering.

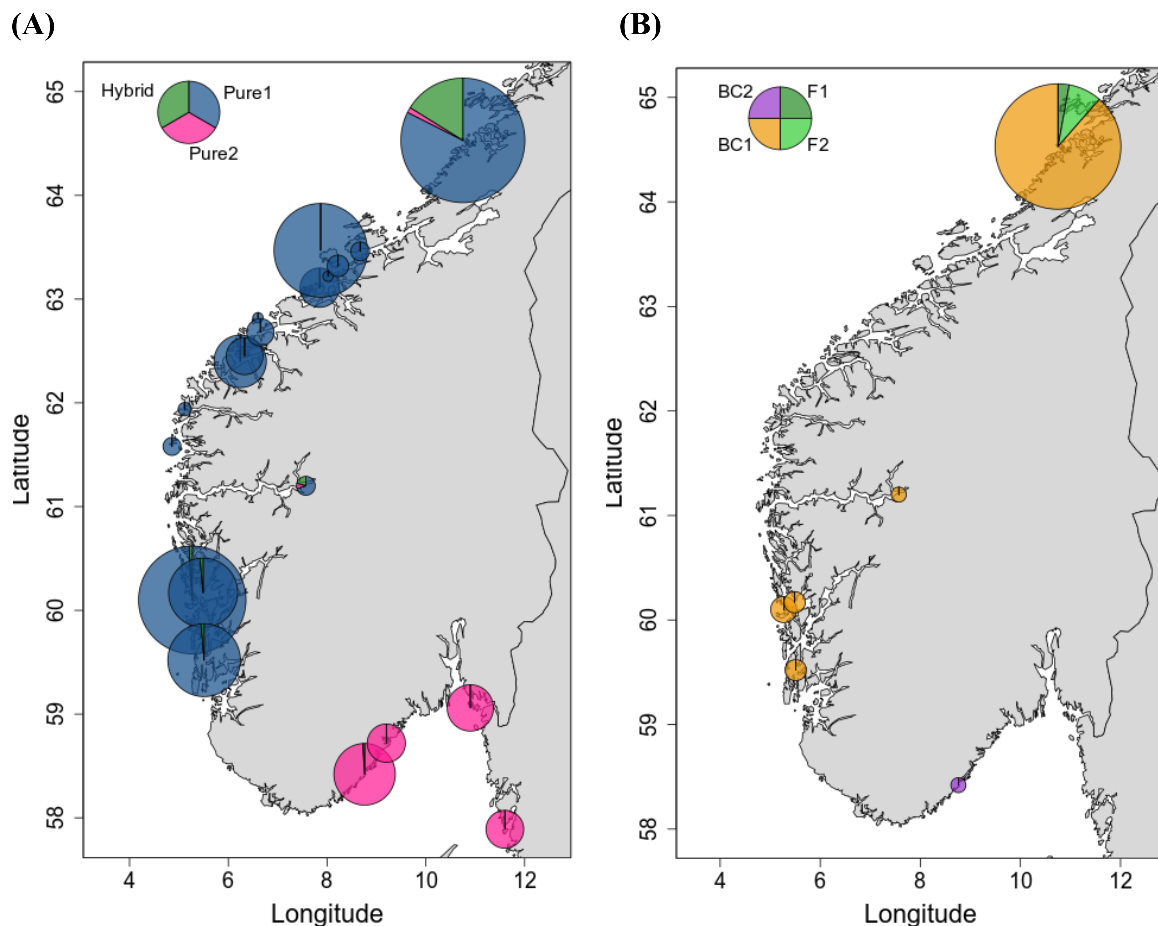


Figure 5. Map displaying proportion of individuals from each sampling site classified by Newhybrid analysis. A) Left map displays individuals classified as pure1 = western genotype, pure2 = south-eastern genotype or hybrid. B) Right map displays the proportion of hybrids assigned to the different hybrid classes F1, F2, backcross with pure1 and backcross with pure2. Sizes reflect the relative number of individuals sampled in a location.

Results show that samples on the Norwegian west coast were similar to each other overall but genetically distinct from Skagerrak samples. However, in addition to the previously known genetic break on the southwest tip of Norway, results from STRUCTURE suggested that there could be a stronger genetic discontinuity along the Norwegian west coast than previously believed. The panel of 84 SNPs had an accuracy above 95% and a power above 95 to correctly classify individuals as western, southern or hybrids. Of the 1519 corkwing wrasse successfully genotyped on the Norwegian west coast, 7 were identified as escapees and 79 as potential hybrids (Figure 5). Almost all of the escapees and hybrids were collected at the northern edge of the population distribution in Flatanger in mid-Norway; the same region as investigated in **Paper I**. We found that escapees and hybrids might constitute up to 20 % of the local population in Flatanger but may be rare elsewhere. Overall these results show that

the relative frequency of escaped and hybridizing individuals is still low in most regions on the Norwegian west coast. However, the introgression of southern genetic material at the northern edge of the species range is likely to alter the local genetic composition and could also obstruct local adaptation, potentially acting as a barrier to further range expansion.

Discussion

Cleaner fish escape and hybridise. These findings raise concerns for how local populations and ecosystems might be affected by the current use of translocated cleaner fish for parasite control. The effects of hybridization between genetically distinct populations are hard to predict and depend on many factors, such as inbreeding, segregating genetic incompatibilities, and locally adapted alleles. Studies of Atlantic salmon have demonstrated significantly lowered fitness in hybrids from domesticated Atlantic salmon and wild populations (Skaala et al., 2012, 2019). Given the known life history differences between southern and western populations of corkwing wrasse, we could expect to see both genetic and phenotypic effects of hybridization. A recent mesocosm study looked at the overall contribution of western and southern individuals to the next generation (F1). Overall, they found that individuals of western origin contributed more to the F1 generation (i.e. produced more offspring) (Blanco Gonzalez et al., 2019). However, in this study western individuals were moved to a southern environment, which is the opposite direction of common cleaner fish translocation. Furthermore, only pure species fitness was assessed, not hybrid fitness which may affect population fitness as a whole. More work is needed to understand how the translocated individuals from southern populations will affect fitness in recipient populations. It is critical to assess phenotypic differences between individuals with native vs. southern origins, and compare fitness between these groups in western Norway in both the field as well as in controlled environments.

As the Flatanger population constitutes the northern boundary of the species distribution, it is likely to play an important role for further northward range expansion. Populations at the periphery of the species distribution often inhabit environmental conditions similar to those just outside the species range, especially if the species exists along an environmental gradient, such as temperature. Thus, edge populations are the most likely populations to carry genotypes that are able to colonize new habitats (Gibson et al., 2009). However, expanding populations will often also experience increased genetic load (Box 1). This is due to many factors such as smaller effective population sizes, population structuring, increased drift, and increased inbreeding and mutational load (Allendorf et al., 2013; Peischl et al., 2013; Sexton et al., 2009). This is often referred to as expansion load (Box 1), which can have long-lasting effects on species, and is believed to be one of the main processes maintaining species boundaries (Peischl et al., 2013). Migration from the source population can benefit the edge population by reducing expansion load by bringing in new alleles and increasing levels of heterozygosity (Allendorf et al., 2013; Bridle et al., 2010). However, gene flow from foreign environments can also disrupt local adaptation and make edge populations more maladapted to the local environment (Gilbert et al., 2017; Kirkpatrick & Barton, 1997), known as migration load. Thus, it is possible that Flatanger populations will benefit from some

migration from some populations, but could quickly become maladapted if introduced individuals come from a very different environment. If western populations are locally adapted to their environment, it is likely that the continued long distance transfer of southern individuals would introduce maladapted alleles into the gene pool and thus work as a barrier to further range expansion.

Box 1

Genetic load: the relative difference in fitness between the average genotype and the theoretically fittest genotype in a population. It can also be considered as a measure of the reduction in the mean fitness of a population relative to a population composed entirely of individuals having optimal genotypes. The four primary sources for genetic load are mutation, segregation, drift and migration load.
Mutation load: the decrease in fitness due to the accumulation of deleterious mutations.
Segregation load: is the decrease in fitness caused by heterozygote advantage. This is because two fit heterozygotes will only produce less fit homozygous offspring.
Drift load: accumulation of deleterious alleles due to genetic drift, that are normally retained in the population at low levels by mutation and selection.
Migration load: the reduction in fitness caused by the migration of individuals not adapted to the local environment.
Inbreeding load: the reduction in fitness in inbred populations. This is caused by a combination of increased mutation load and segregation load.
Expansion load: is the reduction in fitness as a result of genetic drift in the front of range expansion which can result in accumulation of deleterious mutations over species range.

Southern corkwing wrasse is also translocated to salmon farms even further north than Flatanger, beyond the current range, where no wild corkwing populations are present. However, it is still unknown if cleaner fish are able to escape and survive in this environment, as well as what potential consequences this could have for local ecosystems. Although escaping cleaner wrasse would have no populations to hybridize with, they may still introduce new diseases or parasites to conspecifics, salmon and other species in the wild (Svåsand et al., 2017; J. W. Treasurer, 2012; Wallace et al., 2015). In addition to the genetic and ecological risks discussed above, some concern has been raised regarding the health and welfare of cleaner fish and other ethical aspects. Many cleaner fish are killed during handling and transportation (up to 40%) or during other delousing procedures, with some estimates as high as 100% mortality (Hjeltnes et al., 2019). In a report by the Norwegian Veterinary Institute it was even stated that this "effectively makes cleaner fish a 'single use' product, which in itself constitutes a welfare challenge for which both the industry and the authorities must find a better solution." (Hjeltnes et al., 2018).

Novelty and significance

This thesis provides the first evidence that translocated wild corkwing wrasse used as cleaner fish in salmon farms are escaping and hybridizing with local populations. With genetic tools, we demonstrate that the recently established Flatanger population is mainly a result of an ongoing northwards range expansion, along with a significant genetic contribution from southern populations. We found that escapees and hybrids may constitute as much as 20 % of the Flatanger population. In other parts along the Norwegian coast, where salmon farming is also common, we found remarkably few escapees and hybrids. This suggests that introgression might be easier, or easier to detect, in smaller edge-populations than in higher-density areas. Finally, we developed a testing suite of 84 SNPs to identify escapees and hybrids, with the purpose to aid future management and monitoring of wild populations of corkwing wrasse.

The use of cleaner fish for parasite control in other parts of the world is likely to increase in the coming years (VKM 2019). This thesis complements previous work on how the use of cleaner fish in aquaculture can affect native populations, and can provide crucial information for the development of a cleaner fish industry globally. Based on the results in this thesis, emphasis should be put on describing existing population structure, to then apply this information in decision making and management. Finally, monitoring should be prioritized in regions with large numbers of imported cleaner fish and/or with small populations, such as at the edge of the species range. Although the evolutionary and ecological significance of escapees warrants further investigation, the results from this thesis should be taken into consideration in the future use of translocated cleaner fish.

Acknowledgments

For many of you, this will be the very first sentence of this thesis you read, and that’s ok (I’m the same). First, I would like to thank my supervisor Carl Andre for his never ending patience, encouragement, and for always taking the time to talk about what’s important, be it dogs or evolutionary theory. Secondly, I wish to thank my co-supervisor Pierre De Wit, who’s been instrumental to my understanding of genomics and bioinformatics, and who (almost) never shies away when I show up at his door with a panicked look on my face. During my PhD, I’ve spent a fair amount of time in both Gothenburg and in Tjärnö. But regardless of where I am, there are always some amazing people around. In addition to all the people, institutions and funding agencies already acknowledged in two papers, I would like to give a special thanks to Kim Tallaksen Halvorsen for his extensive knowledge on wrasse, to Eeva Jansson for good collaboration and discussion and Emma Berdan for her helpful comments, without whom this thesis would not have been possible. Finally, I would like to thank my wife Kristie for putting up with me writing this thesis during an ongoing pandemic.

References

Allendorf, F. W., Luikart, G., & Aitken, S. N. (2013). *Conservation and the genetics of populations* (2nd ed.). Wiley-Blackwell.

Andersson, E. (2019). *Test av selektiva redskap för det svenska fisket efter läppfisk, Aqua reports 2019:1* (p. 54). Sveriges lantbruksuniversitet, Institutionen för akvatiska resurser.

Atalah, J., & Sanchez-Jerez, P. (2020). Global assessment of ecological risks associated with farmed fish escapes. *Global Ecology and Conservation*, *21*, e00842. https://doi.org/10.1016/j.gecco.2019.e00842

Baliga, V. B., & Law, C. J. (2016). Cleaners among wrasses: Phylogenetics and evolutionary patterns of cleaning behavior within *Labridae*. *Molecular Phylogenetics and Evolution*, *94*, 424–435. https://doi.org/10.1016/j.ympev.2015.09.006

Besnier, F., Kent, M., Skern-Mauritzen, R., Lien, S., Malde, K., Edvardsen, R. B., Taylor, S., Ljungfeldt, L. E. R., Nilsen, F., & Glover, K. A. (2014). Human-induced evolution caught in action: SNP-array reveals rapid amphi-atlantic spread of pesticide resistance in the salmon ectoparasite *Lepeophtheirus salmonis*. *BMC Genomics*, *15*(1), 937. https://doi.org/10.1186/1471-2164-15-937

Bjordal, Å. (1988). Cleaning symbiosis between wrasse (*Labridae*) and lice infested salmon (*Salmo salar*) in mariculture. *International Council for the Exploration of the Sea*, *1988*(F:17), 8.

Blakeslee, A. M. H., Manousaki, T., Vasileiadou, K., & Tepolt, C. K. (2020). An evolutionary perspective on marine invasions. *Evolutionary Applications*, *13*(3), 479–485. https://doi.org/10.1111/eva.12906

Blanco Gonzalez, E., & de Boer, F. (2017). The development of the Norwegian wrasse fishery and the use of wrasses as cleaner fish in the salmon aquaculture industry. *Fisheries Science*. https://doi.org/10.1007/s12562-017-1110-4

Blanco Gonzalez, E., Espeland, S. H., Jentoft, S., Hansen, M. M., Robalo, J. I., Stenseth, N. C., & Jorde, P. E. (2019). Interbreeding between local and translocated populations of a cleaner fish in an experimental mesocosm predicts risk of disrupted local adaptation. *Ecology and Evolution*, *0*(0), 1–13. https://doi.org/10.1002/ece3.5246

Blanco Gonzalez, E., Knutsen, H., & Jorde, P. E. (2016). Habitat discontinuities separate genetically divergent populations of a rocky shore marine fish. *PLoS ONE*, *11*(10). https://doi.org/10.1371/journal.pone.0163052

Boyce, D., Ang, K., & Prickett, R. (2018). Cunner and Lumpfish as cleaner fish species in Canada. In J. Treasurer (Ed.), *Cleaner fish biology and aquaculture applications* (pp. 444–467). 5M Publications.

Bridle, J. R., Polechová, J., Kawata, M., & Butlin, R. K. (2010). Why is adaptation prevented at ecological margins? New insights from individual-based simulations. *Ecology Letters*, *13*(4), 485–494. https://doi.org/10.1111/j.1461-0248.2010.01442.x

Cai, J., Zhou, X., Yan, X., Lucente, D., & Lagana, C. (2019). *Top 10 species groups in global*

aquaculture 2017. 12.

Comtet, T., Sandionigi, A., Viard, F., & Casiraghi, M. (2015). DNA (meta)barcoding of biological invasions: A powerful tool to elucidate invasion processes and help managing aliens. *Biological Invasions*, *17*(3), 905–922. https://doi.org/10.1007/s10530-015-0854-y

Costello, M. J. (2009a). How sea lice from salmon farms may cause wild salmonid declines in Europe and North America and be a threat to fishes elsewhere. *Proceedings of the Royal Society B*, *276*(1672), 3385–3394. https://doi.org/10.1098/rspb.2009.0771

Costello, M. J. (2009b). The global economic cost of sea lice to the salmonid farming industry. *Journal of Fish Diseases*, *32*(1), 115–118. https://doi.org/10.1111/j.1365-2761.2008.01011.x

Crooks, J. A. (2002). Characterizing ecosystem-level consequences of biological invasions: The role of ecosystem engineers. *OIKOS*, *97*(2), 153–166. https://doi.org/10.1034/j.1600-0706.2002.970201.x

Darwall, W. R. T., Costello, M. J., Donnelly, R., & Lysaght, S. (1992). Implications of life-history strategies for a new wrasse fishery. *Journal of Fish Biology*, *41*(supplement B), 111–123. https://doi.org/10.1111/j.1095-8649.1992.tb03873.x

De Wit, P. (2016). *IB14_2b-RAD_log_August2016.sh*. https://github.com/DeWitP/BONUS_BAMBI_IDOTEA/blob/master/IB14_2b-RAD_log_August2016.sh

Devon & Severn, I. F. C. A. (2017). Potting Permit Byelaw. *Devon & Severn IFCA, Inshore Fisheries and Conservation Authority*. Devon & Severn IFCA, Inshore Fisheries and Conservation Authority

Faust, E., André, C., Meurling, S., Kochmann, J., Christiansen, H., Jensen, L. F., Charrier, G., Laugen, A. T., & Strand, Å. (2017). Origin and route of establishment of the invasive Pacific oyster Crassostrea gigas in Scandinavia. *Marine Ecology Progress Series*, *575*, 95–105. https://doi.org/10.3354/meps12219

Ficetola, G. F., Bonin, A., & Miaud, C. (2008). Population genetics reveals origin and number of founders in a biological invasion. *Molecular Ecology*, *17*(3), 773–782. https://doi.org/10.1111/j.1365-294X.2007.03622.x

Gabriel, S., Ziaugra, L., & Tabbaa, D. (2009). SNP Genotyping Using the Sequenom MassARRAY iPLEX Platform. *Current Protocols in Human Genetics*, *60*(1), 2.12.1-2.12.18. https://doi.org/10.1002/0471142905.hg0212s60

Gibson, S. Y., Marel, R. C. V. D., & Starzomski, B. M. (2009). Climate Change and Conservation of Leading-Edge Peripheral Populations. *Conservation Biology*, *23*(6), 1369–1373. https://doi.org/10.1111/j.1523-1739.2009.01375.x

Gilbert, K. J., Sharp, N. P., Angert, A. L., Conte, G. L., Draghi, J. A., Guillaume, F., Hargreaves, A. L., Matthey-Doret, R., & Whitlock, M. C. (2017). Local Adaptation Interacts with Expansion Load during Range Expansion: Maladaptation Reduces Expansion Load. *The American Naturalist*, *189*(4), 368–380. https://doi.org/10.1086/690673

Glover, K. A., Quintela, M., Wennevik, V., Besnier, F., Sørvik, A. G. E., & Skaala, Ø. (2012).

Three Decades of Farmed Escapees in the Wild: A Spatio-Temporal Analysis of Atlantic Salmon Population Genetic Structure throughout Norway. *PLoS ONE*, *7*(8). https://doi.org/10.1371/journal.pone.0043129

Glover, K. A., Solberg, M. F., Mcginnity, P., Hindar, K., Verspoor, E., Coulson, M. W., Hansen, M. M., Araki, H., Skaala, Ø., & Svåsand, T. (2017). Half a century of genetic interaction between farmed and wild Atlantic salmon: Status of knowledge and unanswered questions. *Fish and Fisheries*, *June 2016*, 890–927. https://doi.org/10.1111/faf.12214

Halvorsen, K. T., Sørdalen, T. K., Durif, C., Knutsen, H., Olsen, E. M., Skiftesvik, A. B., Rustand, T. E., Bjelland, R. M., & Vøllestad, L. A. (2016). Male-biased sexual size dimorphism in the nest building corkwing wrasse *Symphodus melops*: Implications for a size regulated fishery. *ICES Journal of Marine Science*, *73*(10), 2586–2594. https://doi.org/10.1093/icesjms/fsw135

Halvorsen, K. T., Sørdalen, T. K., Vøllestad, L. A., Skiftesvik, A. B., Espeland, S. H., & Olsen, E. M. (2017). Sex- and size-selective harvesting of corkwing wrasse (*Symphodus melops*)—A cleaner fish used in salmonid aquaculture. *ICES Journal of Marine Science*, *74*(3), 660–669. https://doi.org/10.1093/icesjms/fsw221

Hjeltnes, B., Bang-Jensen, B., Bornø, G., Haukaas, A., & Walde, C. (2018). *The Health Situation in Norwegian Aquaculture 2017* (Norwegian Veterinary Institute).

Hjeltnes, B., Bang-Jensen, B., Haukaas, A., & Walde, C. S. (2019). *The Health Situation in Norwegian Aquaculture 2018* (Norwegian Veterinary Institute).

Imsland, A. K., Reynolds, P., Eliassen, G., Hangstad, T. A., Foss, A., Vikingstad, E., & Elvegård, T. A. (2014). The use of lumpfish (*Cyclopterus lumpus L.*) to control sea lice (*Lepeophtheirus salmonis Krøyer*) infestations in intensively farmed Atlantic salmon (*Salmo salar L.*). *Aquaculture*, *424–425*, 18–23. https://doi.org/10.1016/j.aquaculture.2013.12.033

IPCC. (2019). *Summary for policymakers. In: IPCC Special Report on the Ocean and Cryosphere in a Changing Climate*. https://report.ipcc.ch/srocc/pdf/SROCC_FinalDraft_FullReport.pdf

Jansson, E., Quintela, M., Dahle, G., Albretsen, J., Knutsen, H., André, C., Strand, Å., Mortensen, S., Taggart, J. B., Karlsbakk, E., Kvamme, B. O., & Glover, K. A. (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*(8), 2135–2147. https://doi.org/10.1093/icesjms/fsx046

Kaur, K., Besnier, F., Glover, K. A., Nilsen, F., Aspehaug, V. T., Fjørtoft, H. B., & Horsberg, T. E. (2017). The mechanism (Phe362Tyr mutation) behind resistance in *Lepeophtheirus salmonis* pre-dates organophosphate use in salmon farming. *Scientific Reports*, *7*(1), 1–9. https://doi.org/10.1038/s41598-017-12384-6

Kirkpatrick, M., & Barton, N. H. (1997). Evolution of a species' range. *American Naturalist*, *150*(1), 1–23. https://doi.org/10.1086/286054

Knutsen, H., Jorde, P. E., Gonzalez, E. B., Robalo, J., Albretsen, J., & Almada, V. (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*(6), e67492. https://doi.org/10.1371/journal.pone.0067492

Lafferty, K. D., Harvell, C. D., Conrad, J. M., Friedman, C. S., Kent, M. L., Kuris, A. M., Powell, E. N., Rondeau, D., & Saksida, S. M. (2015). Infectious diseases affect marine fisheries and aquaculture economics. *Annual Review of Marine Science*, *7*, 471–496. https://doi.org/10.1146/annurev-marine-010814-015646

Laikre, L., Schwartz, M. K., Waples, R. S., Ryman, N., & Group, T. G. W. (2010). Compromising genetic diversity in the wild: Unmonitored large-scale release of plants and animals. *Trends in Ecology and Evolution*, *25*(9), 520–529. https://doi.org/10.1016/j.tree.2010.06.013

Liu, Y., & Bjelland, H. vanhauwaer. (2014). Estimating costs of sea lice control strategy in Norway. *Preventive Veterinary Medicine*, *117*(3), 469–477. https://doi.org/10.1016/j.prevetmed.2014.08.018

Maroni, K., & Andersen, P. (1996). Distribution and abundance of wrasse in an area of northern Norway. In M. D. J. Sayer, M. J. Costello, & J. W. Treasurer (Eds.), *Wrasse: Biology and use in Aquaculture.* (pp. 70–73). Fishing News Books.

Mattingsdal, M., Jorde, P. E., Knutsen, H., Jentoft, S., Stenseth, N. C., Sodeland, M., Robalo, J. I., Hansen, M. M., André, C., & Gonzalez, E. B. (2020). Demographic history has shaped the strongly differentiated corkwing wrasse populations in Northern Europe. *Molecular Ecology*, *29*(1), 160–171. https://doi.org/10.1111/mec.15310

Matz, M. V., & Aglyamova, G. (2014). *Protocol for Illumina 2bRAD sample preparation*.

McCallum, H., & Dobson, A. (1995). Detecting disease and parasite threats to endangered species and ecosystems. *Trends in Ecology & Evolution*, *10*(5), 190–194.

Molnar, J. L., Gamboa, R. L., Revenga, C., & Spalding, M. D. (2008). Assessing the global threat of invasive species to marine biodiversity. *Frontiers in Ecology and the Environment*, *6*(9), 485–492. https://doi.org/10.1890/070064

Norwegian directorate of Fisheries. (2019). *Utsett av rensefisk 1998-2018*. Norwegian directorate of Fisheries. https://www.fiskeridir.no/Akvakultur/Tall-og-analyse/Akvakulturstatistikk-tidsserier/Rensefisk

Peischl, S., Dupanloup, I., Kirkpatrick, M., & Excoffier, L. (2013). On the accumulation of deleterious mutations during range expansions. *Molecular Ecology*, *22*(24), 5972–5982. https://doi.org/10.1111/mec.12524

Potts, G. W. (1974). The colouration and its behavioural significance in the corkwing wrasse, *Crenilabrus melops*. *Journal of the Marine Biological Association of the United Kingdom*, *54*(04), 925–938. https://doi.org/10.1017/S0025315400057659

Potts, G. W. (1985). The Nest Structure of the Corkwing Wrasse, *Crenilabrus Melops* (*Labridae: Teleostei*). *Journal of the Marine Biological Association of the United Kingdom*, *65*(02), 531–546. https://doi.org/10.1017/S002531540005058X

Riley, A., Jeffery, K., Cochrane-dyet, T., White, P., & Ellis, J. (2017). *Northern European*

Wrasse – Summary of commercial use, fisheries and implications for management. Cefas, Centre for Environment, Fisheries and Aquaculture Science.

Rius, M., Bourne, S., Hornsby, H. G., & Chapman, M. A. (2015). Applications of next-generation sequencing to the study of biological invasions. *Current Zoology*, *61*(3), 488–504. https://doi.org/10.1093/czoolo/61.3.488

Robalo, J. I., Castilho, R., Francisco, S. M., Almada, F., Knutsen, H., Jorde, P. E., Pereira, A. M., & Almada, V. (2012). Northern refugia and recent expansion in the North Sea: The case of the wrasse *Symphodus melops* (Linnaeus, 1758). *Ecology and Evolution*, *2*(1), 153–164. https://doi.org/10.1002/ece3.77

Sánchez, J. C., Mancilla, J., Hevia, M., & Saez, P. J. (2018). The Patagonian blenny (Eleginops maclovinus): A Chilean native fish with potential to control sea lice (Caligus rogercresseyi) infestations in salmonids. In J. W. Treasurer (Ed.), *Cleaner fish biology and aquaculture applications.* 5M Publishing.

Sexton, J. P., McIntyre, P. J., Angert, A. L., & Rice, K. J. (2009). Evolution and Ecology of Species Range Limits. *Annual Review of Ecology, Evolution, and Systematics*, *40*(1), 415–436. https://doi.org/10.1146/annurev.ecolsys.110308.120317

Skaala, Ø., Besnier, F., Borgstrøm, R., Barlaup, B., Sørvik, A. G., Normann, E., Østebø, B. I., Hansen, M. M., & Glover, K. A. (2019). An extensive common-garden study with domesticated and wild Atlantic salmon in the wild reveals impact on smolt production and shifts in fitness traits. *Evolutionary Applications*, *12*(5), 1001–1016. https://doi.org/10.1111/eva.12777

Skaala, Ø., Glover, K. A., Barlaup, B. T., Svåsand, T., Besnier, F., Hansen, M. M., & Borgstrøm, R. (2012). Performance of farmed, hybrid, and wild Atlantic salmon (*Salmo salar*) families in a natural river environment. *Canadian Journal of Fisheries and Aquatic Sciences*, *69*(12), 1994–2006. https://doi.org/10.1139/f2012-118

Skiftesvik, A. B., Durif, C. M. F., Bjelland, R. M., & Browman, H. I. (2014). *Distribution and habitat preferences of five species of wrasse (Family Labridae) in a Norwegian fjord*. *72*(October), 890–899.

Svåsand, T., Grefsrud, E. S., Karlsen, Ø., Kvamme, B. O., Glover, K. S., Husa, V., & Kristiansen, T. S. (red.). (2017). Risikorapport norsk fiskeoppdrett. *Fisken Og Havet*, *2*.

Tepolt, C. K., Darling, J. A., Blakeslee, A. M. H., Fowler, A. E., Torchin, M. E., Miller, A. W., & Ruiz, G. M. (2020). Recent introductions reveal differential susceptibility to parasitism across an evolutionary mosaic. *Evolutionary Applications*, *13*(3), 545–558. https://doi.org/10.1111/eva.12865

Treasurer, J. W. (2012). Diseases of north European wrasse (*Labridae*) and possible interactions with cohabited farmed salmon, *Salmo salar L*. *Journal of Fish Diseases*, *35*(8), 555–562. https://doi.org/10.1111/j.1365-2761.2012.01389.x

Troost, K. (2010). Causes and effects of a highly successful marine invasion: Case-study of the introduced Pacific oyster *Crassostrea gigas* in continental NW European estuaries. *Journal of Sea Research*, *64*(3), 145–165. https://doi.org/10.1016/j.seares.2010.02.004

Uglem, I., Rosenqvist, G., & Wasslavik, H. S. (2000). Phenotypic variation between dimorphic males in corkwing wrasse. *Journal of Fish Biology*, *57*(1), 1–14. https://doi.org/10.1006/jfbi.2000.1283

Viard, F., & Comtet, T. (2015). 18. Applications of DNA-based Methods for the Study of Biological Invasions. In *Biological Invasions in Changing Ecosystems* (pp. 411–435). De Gruyter. https://doi.org/10.1515/9783110438666-025

Viard, F., David, P., & Darling, J. A. (2016). Marine invasions enter the genomic era: Three lessons from the past, and the way forward. *Current Zoology*, *62*(6), 629–642. https://doi.org/10.1093/cz/zow053

VKM, Rueness, E. K., Berg, P. R., Gulla, S., Halvorsen, K. A. T., Järnegren, J., Malmstrøm, M., Mo, T. A., Rimstad, E., de Boer, H., Eldegard, K., Hindar, K., Hole, L. R., Kausrud, K., Kirkendall, L., Måren, I., Nilsen, E. B., Thorstad, E. B., Nielsen, A., & Velle, G. (2019). *Assessment of the risk to Norwegian biodiversity from import of wrasses and other cleaner fish for use in aquaculture. Opinion of the Panel on Alien Organisms and Trade in Endangered Species of the Norwegian Scientific Committee for Food and Environment* (No. 15; VKM Report 2019). Norwegian Scientific Committee for Food and Environment (VKM).

Wallace, I. S., Donald, K., Munro, L. A., Murray, W., Pert, C. C., Stagg, H., Hall, M., & Bain, N. (2015). A survey of wild marine fish identifies a potential origin of an outbreak of viral haemorrhagic septicaemia in wrasse, *Labridae*, used as cleaner fish on marine Atlantic salmon, *Salmo salar* L., farms. *Journal of Fish Diseases*, *38*(6), 515–521. https://doi.org/10.1111/jfd.12259

Wang, S., Meyer, E., McKay, J. K., & Matz, M. V. (2012). 2b-RAD: a simple and flexible method for genome-wide genotyping. *Nature Methods*, *9*(8), 808–810. https://doi.org/10.1038/nmeth.2023

Woll, A. K., Bakke, S., Aas, G. H., Solevåg, S. E., Skiftesvik, A. B., & Bjelland, R. M. (2013). *Velferd leppefisk i merd* (p. 30). Møreforsking MARIN.

PAPER I

ROYAL SOCIETY OPEN SCIENCE

rsos.royalsocietypublishing.org

Research

Cite this article: Faust E, Halvorsen KT, Andersen P, Knutsen H, André C. 2018 Cleaner fish escape salmon farms and hybridize with local wrasse populations. *R. Soc. open sci.* **5**: 171752.
http://dx.doi.org/10.1098/rsos.171752

Received: 27 October 2017
Accepted: 13 February 2018

Subject Category:
Biology (whole organism)

Subject Areas:
genomics/ecology/evolution

Keywords:
aquaculture, wrasse, sea lice, hybrid, RAD, salmon

Author for correspondence:
Ellika Faust
e-mail: ellika.faust@gmail.com

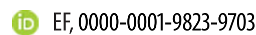
Electronic supplementary material is available online at https://dx.doi.org/10.6084/m9.figshare.c.4019425.

THE ROYAL SOCIETY PUBLISHING

Cleaner fish escape salmon farms and hybridize with local wrasse populations

Ellika Faust[1], Kim Tallaksen Halvorsen[2], Per Andersen[3], Halvor Knutsen[4,5,6] and Carl André[1]

[1]Department of Marine Sciences - Tjärnö, University of Gothenburg, 45296 Strömstad, Sweden
[2]Institute of Marine Research, Austevoll Research Station, Storebø, Norway
[3]Marine senior advisor Nord-Trøndelag, 7770 Flatanger, Norway
[4]Institute of Marine Research, Flødevigen, Norway
[5]Centre for Ecological and Evolutionary Synthesis, Department of Biosciences, University of Oslo, Oslo, Norway
[6]Centre for Coastal Research, University of Agder, Kristiansand, Norway


EF, 0000-0001-9823-9703

The genetic impact of farmed fish escaping aquaculture is a highly debated issue. However, non-target species, such as cleaner fish used to remove sea lice from farmed fish, are rarely considered. Here, we report that wild corkwing wrasse (*Symphodus melops*), which are transported long distances to be used as cleaner fish in salmon farms, escape and hybridize with local populations. Recently, increasing numbers of corkwing wrasse have been reported in Flatanger in Norway, north of its described distribution range, an area heavily relying on the import of cleaner fish from Skagerrak. Using genetic markers identified with 2bRAD sequencing, we show that, although the Flatanger population largely is a result of a northward range expansion, there is also evidence of considerable gene flow from southern populations in Skagerrak and Kattegat. Of the 40 corkwing wrasses sampled in Flatanger, we discovered two individuals with clear southern genotypes, one first-generation hybrid, and 12 potential second-generation hybrids. In summary, we provide evidence that corkwing wrasse escape from fish farms and hybridize with local populations at the leading edge of an ongoing range expansion. Although the magnitude and significance of escapees warrant further investigation, these results should be taken into consideration in the use of translocated cleaner fish.

1. Introduction

Marine species display a range of levels of genetic divergence among populations, from panmictic species to species with marked genetic structure, as a consequence of reduced gene

© 2018 The Authors. Published by the Royal Society under the terms of the Creative Commons Attribution License http://creativecommons.org/licenses/by/4.0/, which permits unrestricted use, provided the original author and source are credited.

flow, genetic drift and/or local adaptation [1]. Transferring individuals between spatially separated populations that are genetically distinct is likely to result in genetic changes to native populations. Such changes could involve shifts in allelic composition, loss of genetic variation, erosion of local adaptation and/or breakdown of population structure [2]. Human-mediated releases of genetically different individuals to native populations are increasingly common. Farmed fish escaping aquaculture is a serious threat to wild fish populations, through competition, transfer of diseases and pathogens, and gene flow through interbreeding [3]. There are many examples from open-pen farming of salmonids, where escapees have hybridized with local river populations, leading to genetic swamping and reduced fitness [4,5].

Salmon farming may also promote inadvertent gene flow in populations of species of wrasses (Labridae) in Norway and the UK, where wild wrasses are caught and used as cleaner fish to mitigate the increasing problems of sea lice infestations in the farmed salmon [6,7]. These wrasses are relatively small predatory fish, typically abundant at shallow depths on rocky coastlines in northern Europe. They had little to no commercial value until their function as cleaner fish in captivity was discovered and applied in the late 1980s [8–10]. The use of cleaner fish increased drastically in 2010 as a result of sea lice evolving resistance to the most widely used pharmaceutical treatments [11]. In Norway, the national landings of wrasse have now surpassed 20 million fish annually [12]. However, in mid-Norway, the demand for cleaner fish exceeds the supply from local stocks, and wild-caught wrasses are imported from southern Norway and western Sweden, areas where salmon farming is absent [13]. Similarly, in the UK, most salmon farms are situated in Scotland, but due to local supply not meeting the demand, an estimated 1 million wrasse are harvested in southwestern England annually for live transport to Scotland [14,15]. Furthermore, the UK wrasse fishery is largely undocumented, and the records of landed wrasse are rarely specified by species, only under a generic wrasse code. The lack of data on species composition and landings makes it difficult to assess the impact of the wrasse fishery. This is a concern that has received increasing attention in recent years, resulting in restrictions on wrasse fisheries in southwestern UK by regional Inshore Fisheries and Conservation Authorities (IFCA) [15–17].

In Norway, two species, the goldsinny wrasse (*Ctenolabrus rupestris*) and the corkwing wrasse (*Symphodus melops*), are the most commonly used wild cleaner fish, with 39% and 52% of the total Norwegian official landings 2016, respectively (Norwegian directorate of Fisheries; https://www.fiskeridir.no/Yrkesfiske/Tema/Leppefiske/Registrert-uttak-av-leppefisk-i-2017). A recent study found relatively low genetic divergence between wild goldsinny populations in farming areas in mid-Norway and populations in southern Norway and Sweden, suggesting inadvertent gene flow [18]. In contrast to the goldsinny, which generally shows weak population structure, the corkwing has highly differentiated populations in Scandinavia with a strong genetic break between southern and western Norway and overall lower genetic diversity in the southern area [19]. The difference in population structure between the two species could be related to differences in population connectivity caused by distinct reproductive strategies: the goldsinny is a broadcast spawner with a fraction of the eggs being pelagic, while the corkwing lay benthic eggs in seaweed nests [6,20,21]. Furthermore, southern corkwing populations have been found to grow faster and mature earlier than the populations further north [22], which aligns with the genetic break [19]. Thus, if corkwing with southern origin escapes and hybridizes with local populations further north, we could expect to see changes in genotype composition with possible phenotypic effects.

The corkwing's northern distribution range was earlier reported to extend to the Trondheims Fjord in mid-Norway. In the Flatanger municipality, North Trøndelag county, 130 km further north, no corkwing was found during extensive field surveys of wrasse in the 1990s [23]. However, in recent years, occasional observations of corkwing have been reported in North Trøndelag (but not further north; Norwegian Fishermen's Sales Organization 2016, personal communication), indicating a recent northward range expansion. Knutsen *et al*. [24] proposed that the current increase in abundance in southern Scandinavia is a result of population growth due to rising temperatures, and that the predicted rise in sea temperature could facilitate a northward expansion. The other possibility would be that this northward expansion is a direct result of wrasse escaping from the salmon pens through tears in the net, small fish slipping through the mesh [13,25] or intentional release at the end of the season [26].

Here, we investigate the origin of wild corkwing wrasse captured in Flatanger, amid salmon farms where wrasses are currently used as cleaner fish and rely heavily on the import of wrasse from southern Norway and Sweden. We used the restriction-site-associated DNA (RAD) sequencing method 2b-RAD [27] to simultaneously discover and genotype thousands of SNPs (single nucleotide polymorphisms) across the entire genome [24]. Our objective is to investigate whether the wild corkwing in Flatanger

rsos.royalsocietypublishing.org R. Soc. open sci. 5: 171752

Figure 1. Map of sampling locations. Kristiansand, Strömstad and Kungsbacka are referred to as 'southern population', Austevoll and Stavanger as 'western population' and Flatanger as 'mid-Norwegian population'.

represents: (i) the leading edge of an ongoing northward range expansion [24], (ii) escaped wrasse from aquaculture with origin from Skagerrak and Kattegat or (iii) a mix of both. To answer these questions, we compare SNPs from corkwing wrasse collected in Flatanger with wrasse collected: (i) in western Norway, where wrasse is harvested but used locally, and (ii) further south on the Skagerrak–Kattegat coast, where all wrasses are harvested for live transport to salmon farms in mid- and northern Norway.

2. Material and methods

2.1. Sampling and DNA extraction

With the help of commercial fishermen and local researchers, we collected corkwing wrasse from Flatanger in mid-Norway; from two locations in western Norway: Austevoll and Stavanger (western population); and from three locations at the Skagerrak–Kattegat coast: Kristiansand, Strömstad and Kungsbacka (southern population) (figure 1). Fin clips from forty individuals per location were taken in June–October 2016 and stored in 96% ethanol until further analysis. For fish sampled in Flatanger, we dissected otoliths and aged them by counting annual growth increments following Halvorsen *et al*. [22]. Additional sampling information, such as coordinates and sampling location in relation to salmon farms, can be found in electronic supplementary material, S1 and S2.

Genomic DNA was extracted using DNeasy® Blood & Tissue Kit (Qiagen) with optional RNAse treatment (200 mg RNAse), and purified and concentrated with standard ethanol/isopropanol precipitation. DNA quantity and quality (i.e. presence of contaminants, degradation etc.) were assessed using Qubit® ds DNA BR AssayKit (Invitrogen–ThermoFisher Scientific) and on a 1% agarose gel. 2b-RAD libraries were prepared following a protocol modified from Matz & Aglyamova [28], available in a dedicated GitHub repository (https://github.com/ellikafaust/S.melopsPopGen). All individual DNA samples were tagged with unique barcodes and then pooled in sets of 24 per sequencing lane, including technical replicates of four individuals to control for methodical artefacts. Pooling was done by sampling site, where each sample (40 individuals) was divided in two independent pools that were sequenced in separate lanes. This was done to minimize the risk of mixing up samples during library preparation, while having two independent pools to account for any lane bias. Single-read, 50 bp target length sequencing on Illumina HiSeq2500 platform was conducted at the SNP&SEQ Technology Platform in Uppsala, Uppsala University.

2.2. Bioinformatics

The bioinformatic analysis of the DNA sequences followed a modified de novo pipeline from Pierre de Wit [29] using scripts developed by Mikhail Matz (scripts and manual available at https://github.com/z0on/2bRAD_denovo). First, low-quality reads and redundant sequences (i.e. restriction sites and duplicates) were removed. Remaining fragments were then clustered into rad tags, allowing up to three mismatches among reads (identity threshold 91%) and with a minimum depth of 20 reads. Individual genotypes were called, following the criteria of Mindp = 5 (min depth for calling a homozygote), hetero = 0.8 (max fraction of heterozygotes allowed), aobs = 20 (min number of times allele has to be observed across all samples) and strbias = 20 (strand bias cut-off). Four technical replicates per lane were used to control for methodical artefacts using the recalibrateSNPs.pl script. Variants that had been identically genotyped between the replicates were used as reference for non-parametric quality recalibration of all variants, estimating their probability of being 'true' SNPs. Loci with recalibrated quality below 20 and alleles with quality below 20 were removed. Only variants with less than 75% heterozygotes and less than 50% missing data were kept for thinning (removing) of the dataset. SNPs occurring on the same RAD-tag were removed, leaving only the SNP with the highest minor allele frequency (MAF) in each RAD-tag. Technical replicates and poorly sequenced individuals (individuals with more than 50% missing data) were removed. Finally, we removed loci that were missing in more than 30% of the individuals or with a global MAF below 1%. Initially, different levels of minor allele frequency (maf 0%, 1% and 5%) were tested. As the different datasets did not change the outcome of the analyses (data not shown), we only present results from loci with maf > 1%, maintaining the most number of loci, while still removing genotyping errors and uninformative polymorphisms [30]. Data conversions between different software technologies were done using PGD spider [31].

2.3. Statistical analysis

2.3.1. Population diversity and differentiation

We used the R package diveRsity [32] in R v. 3.3.2 [33] to calculate observed and expected heterozygosity for each locus in the different samples. Whether observed heterozygosity (*H*o) values deviated from expected heterozygosity (*H*e) was assessed by calculating F_{IS} according to Weir & Cockerham [34]. Deviations from Hardy–Weinberg (HW) proportions were estimated with exact tests, with *p*-values calculated according to the complete enumeration method [35] and adjusted for multiple testing using false discovery rate (FDR) correction [36]. Loci that deviated ($q < 0.05$) from HW proportions in more than one of the samples were subsequently removed. Weir & Cockerham's F_{ST} was estimated for each population pair and over all samples using diveRsity. Statistical significance of F_{ST} values was assessed using Fisher's exact probability test with 5000 Monte Carlo replicates, followed by FDR correction.

2.3.2. Individual-based clustering

Missing genotypes can induce patterns of similarity or differentiation that are easily confused with genetic structure. To detect such biases, we clustered individuals based on their identity-by-missingness in PLINK v. 1.9 [37,38] where pairwise distances between individuals are calculated from the proportion of missing sites which are not shared between individual pairs. Pairwise distances were visualized with a multidimensional scaling plot.

To estimate and visualize genetic differentiation among individuals, we applied two individual-based clustering methods, STRUCTURE v. 2.3.4 [39] and principal component analysis (PCA) in the R package ade4 [40–42]. STRUCTURE uses model-based Bayesian clustering to find the most probable number of population clusters *K*. Once *K* is defined, it estimates the posterior probability of each individual's genotype to originate in each cluster. STRUCTURE analyses were performed assuming uncorrelated allele frequencies, allowing admixture and with no locprior. The burn-in period was set to 10 000 and the number of Markov chain Monte Carlo (MCMC) repetitions to 50 000. Clusters *K* from 1 to 7 were run three times per *K*. The different runs were merged for visual analysis with CLUMPAK [43]. Calculations of the most probable number of population clusters (*K*) were estimated using STRUCTURE HARVESTER [44] by calculating the posterior probability for each value of *K* (mean lnP(*K*)) and the modal value of Delta *K*. The second individual-based clustering method (PCA) uses a multivariate exploratory approach that makes no prior assumptions about how many populations exist or boundaries between populations. Allele frequencies were centred but not scaled and missing data were replaced by mean allele frequencies

with the function scaleGen in ADEGENET [45,46]. PCA was performed using the function dudi.pca in ade4.

2.3.3. Hybridization

To remove potential bias in hybrid analysis, 200 SNPs with the highest overall F_{ST} were tested for linkage disequilibrium (LD) in Genepop on the web [47] using 10 000 dememorizations, 100 batches and 5000 iterations per batch. SNPs with significant LD after FDR corrections were removed and replaced with new SNPs until no significant comparisons remained. To assess the accuracy, efficiency and power to correctly identify individuals belonging to different hybrid classes, we used the R package HYBRIDDETECTIVE [48]. We used the function freqbasedsim_AlleleSample to generate three replicates of three simulated data sets with pure parents (Pure_A and Pure_B), first- and second-generation hybrids (F1 and F2) and backcrosses between F1 and pure parents (BC_A and BC_B). The datasets contained 720 individuals and were based on the genotype frequencies from the 200 loci in the western (PureA = Austevoll and Stavanger) and southern (PureB = Kristiansand, Strömstad and Kungsbacka) samples. Simulations were analysed in NEWHYBRIDS v. 1.1 [49] which estimates the posterior probability of each individual to belong to one of the six hybrid classes. The analysis was done using the uniform prior option and default genotype proportions with a burn-in period of 50 000 iteration and 300 000 MCMC sweeps. Power was estimated as the product of efficiency (correctly assigned individuals over the known individuals per class) and accuracy (correctly assigned individuals over individuals assigned to that class) as described in HYBRIDDETECTIVE [48].

Finally, we investigated the occurrence of hybridization in the northern-most location Flatanger in mid-Norway with the software NEWHYBRIDS. Individuals from Skagerrak (Kristiansand, Strömstad and Kungsbacka) and western Norway (Austevoll and Stavanger) were included in the runs as the pure parent genotypes using the 'z' and 's' options. The analysis was performed using the same 200 loci as for the simulated data, displaying the highest overall F_{ST} estimates and no LD. The data were analysed using the uniform prior option, default genotype proportions and the burn-in period was set to 50 000 and the number of MCMC sweeps after burn-in to 300 000.

3. Results

3.1. Genetic diversity and population differentiation

From the total 48 technical replicates (four for each pool of 20), we called 237 090 SNPs (average 4939 ± 126 s.d. per replicate pair). Of these, 9% ± 0.05% s.d. were inconsistent between technical replicates. Data filtering resulted in a total of 4372 polymorphic SNPs, and none of the 240 individuals had to be removed due to missing data. Of the 57 600 missing data comparisons, only 479 pairwise comparisons have an identity of missingness higher than 20% (max 47%), and no obvious patterns of identity by missingness can be observed (S3). F_{IS} estimates indicate heterozygote excess in all samples (mean F_{IS} ranging from −0.344 to −0.052). Fifteen loci deviated significantly ($q < 0.05$) from HW proportions in more than one sample and were subsequently removed, leaving 4357 SNPs for final analysis. No more than eight loci deviate significantly from HW proportions in any of the western or southern samples. However, a much higher number of loci deviate from HW proportions in the Flatanger sample. Almost all of the loci display negative F_{IS} values, indicating heterozygosity excess. Furthermore, wrasse from the western population display an overall higher genetic diversity (mean $Ho = 0.30$, mean $He = 0.32$, polymorphic loci = 95.2%) compared with wrasse from the southern population (mean $Ho = 0.26$, mean $He = 0.24$, polymorphic loci = 82.3%). The Flatanger population shows the highest genetic diversity (mean $Ho = 0.50$, mean $He = 0.35$, polymorphic loci = 97%). Global genetic differentiation, estimated as $F_{ST} = 0.0789$, is significantly ($p < 0.05$) different from zero. Pairwise F_{ST} estimates (S4) demonstrate higher genetic differentiation between the western and southern populations ($F_{ST} = 0.101$–0.1312) than among the southern samples ($F_{ST} = 0.0023$–0.0030) or between the western samples ($F_{ST} = 0.0065$). Overall, Flatanger is genetically more similar to the western population ($F_{ST} = 0.0243$–0.0277) than the southern population ($F_{ST} = 0.1163$–0.1258).

3.2. Individual-based clustering

STRUCTURE analyses suggest the existence of two, potentially three, genetically differentiated clusters (figure 2; electronic supplementary material, figure S5). The first two clusters correspond to the divide

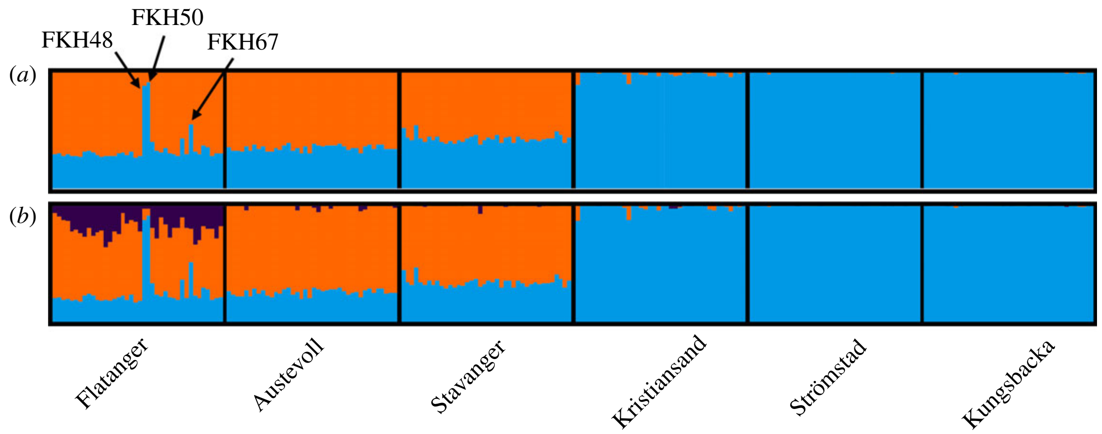


Figure 2. STRUCTURE cluster assignment of corkwing wrasse based on 4357 SNPs for $K = 2$ (*a*) and $K = 3$ (*b*). Each vertical line represents one individual and the colour shows the proportion of each individual assigned to the *K* different genetic clusters. Individuals from Skagerrak/Kattegat cluster together (blue) and individuals from western Norway cluster together (orange), visualizing the genetic break between southern and western populations. Majority of individuals in mid-Norway (Flatanger) cluster with the western population, with the exception of FKH48 and FKH50, which cluster with southern population, and FKH67, which does not group to either cluster.

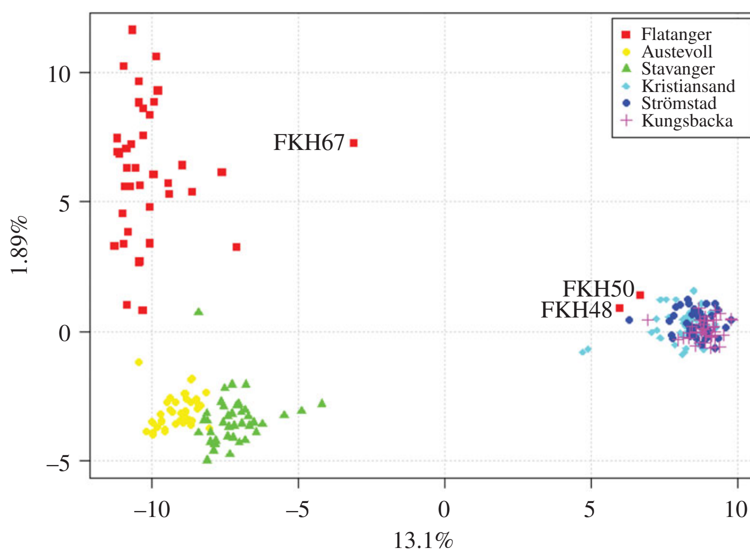


Figure 3. The first (*x*-axis) and second (*y*-axis) components of a principal component analysis on 240 corkwing wrasse individuals from 6 locations based on 4357 SNPs. The first component explains 13.1% of the total variation and the second 1.89%. Additional components explain less than 1% of the total variance each, and are not shown. Each point represents one individual, which is colour coded by sampling site. On the first axis, majority of individuals from Flatanger cluster with individuals from western Norway (left), but two individuals from Flatanger (FKH48 and FKH50) cluster together individuals from Skagerrak/Kattegat (right) and one individual (FKH67) separates from both clusters. On the second axis, individuals from Flatanger are more separated, but overall closer to Skagerrak/Kattegat than western Norway.

between southern and western populations (blue and orange, respectively), in concordance with pairwise F_{ST} estimates and previous studies [19]. Most individuals from Flatanger were assigned to the western population for $K = 2$, and partially to a third cluster (purple) for $K = 3$. However, two individuals from Flatanger (FKH48 and FKH50) were assigned to the southern population (blue). Another individual from the Flatanger sample (FKH67) was assigned equally to both populations, suggesting admixture.

To estimate and visualize genetic differentiation among individuals without prior assumptions about the population model, we conducted a PCA (figure 3). The first principal component separates data into two main clusters, which correspond closely to southern and western clusters observed in the

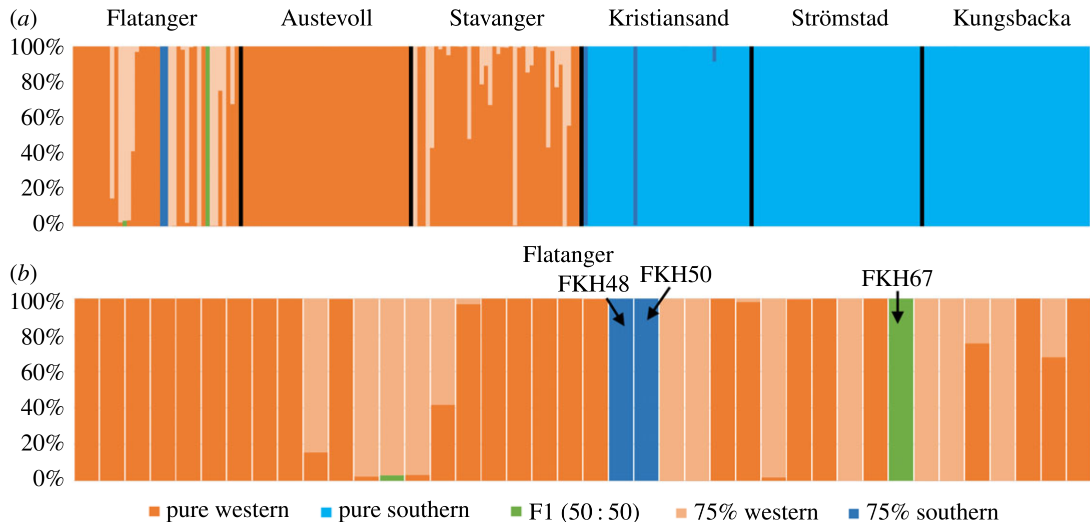


Figure 4. Hybrid analysis of all individuals (*a*) and individuals sampled in Flatanger (*b*) using the 200 SNPs with highest F_{ST} estimates and no LD. Each vertical line represents one individual and its probability belongs to one of the six genotype classes: pure western, pure southern, F1 hybrid (50 : 50 western:southern), F2 hybrid (none present) or backcrosses (75% western or southern) between F1 and pure western or pure southern. Hybrids are only detected in Flatanger and the two samples next to the genetic break, Stavanger and Kristiansand. Out of the 40 individuals from Flatanger, we discovered two individuals with clear southern genotypes (FKH48 and FKH50), one first-generation hybrid (FKH67), and twelve potential second-generation hybrids (light orange).

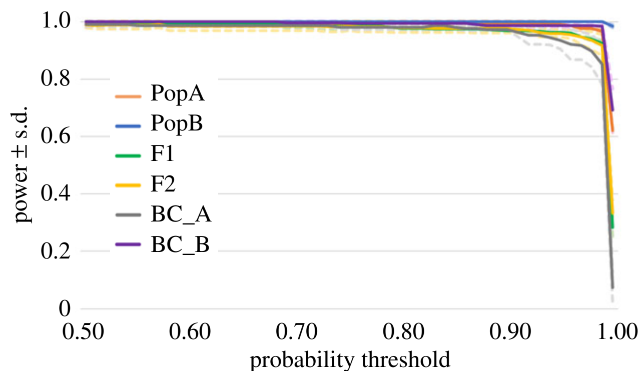


Figure 5. Hybrid detection power at different probability thresholds based on three sets of simulated genotype data from 200 SNPs with the highest overall F_{ST} and no LD. Solid lines represent the six genotype classes, pure parents (PopA = western population and PopB = southern population), first- and second-generation hybrids (F1 and F2) and backcrosses (BC_A and BC_B). The dashed lines represent the standard deviation among the simulations for each class.

STRUCTURE analysis. The second principal component (*y*-axis) splits the Flatanger population from the western population, placing Flatanger closer to the southern population than the western. Succeeding components explain less than 1% of the total variance each, and are not shown. The same two individuals from Flatanger (FKH48 and FKH50) which were assigned to the southern cluster in the STRUCTURE analyses group with the southern cluster in the PCA. The Flatanger individual which was assigned equally to both clusters in the STRUCTURE analyses (FKH67) is closer to the southern cluster than any of the other individuals from the Flatanger or western population.

3.3. Hybridization

We used the software NEWHYBRIDS to identify potential hybrids in Flatanger (figure 4). The two individuals, assigned to the southern cluster in both STRUCTURE and the PCA (FKH48 and FKH50), were identified as southern backcrosses, i.e. 75% southern genotype and 25% western genotype. FKH67 was detected as a F1 hybrid, carrying 50% of both the southern and western genotype. Furthermore, another 12 individuals from Flatanger have a high probability (greater than 50%) of being western

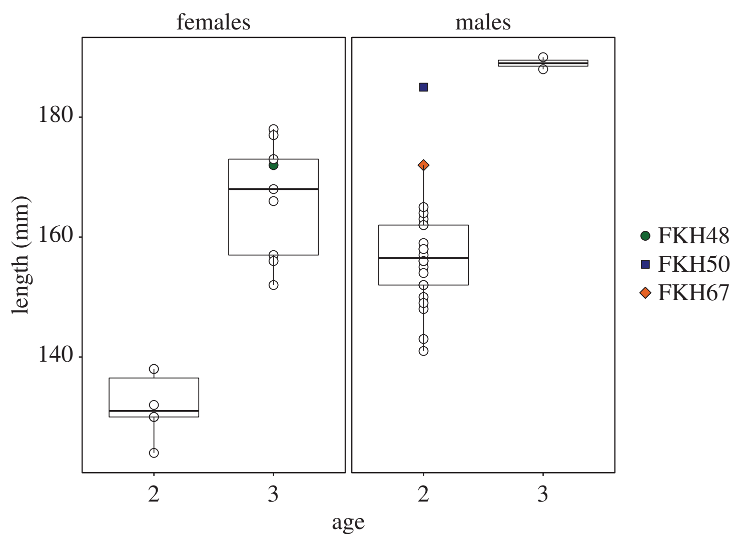


Figure 6. Boxplots showing length at age for corkwing wrasse sampled in Flatanger. FKH48, FKH50 and FKH67 are individuals with genotypes closely resembling southern populations.

backcrosses, i.e. 75% western genotype and 25% southern genotype. Some fish from the 'pure' southern and western samples closest to the southern/western genetic break (two individuals in Kristiansand and eight in Stavanger) are also distinguished as genetic backcrosses (figure 4*a*), indicating gene flow across the break. Simulated data demonstrated high efficiency, accuracy and power to detect individuals from all of the six hybrid classes given the battery of 200 loci used (figure 5; electronic supplementary material, figure S6). The battery of SNPs is able to call individuals as pure western, pure southern, F1, F2, western or southern backcross with a power above 95% at a probability threshold of 90%.

Comparison of length measurements for individuals of the same sex and age (figure 6) shows that the F1 individual (FKH67) and one of the individuals with southern genotype (FKH48) are the largest 2-year-old males in the sample. The second individual with a southern genotype (FKH50) is a 3-year-old female above median length.

4. Discussion

Here, we provide the first evidence that translocated corkwing wrasse escape salmon farms and hybridize with local populations. Our results support previous studies by finding marked genetic differentiation between southern Skagerrak corkwing wrasse populations and those in western Norway. We expand on current knowledge by discovering that almost half of the individuals sampled at the northern limit of the species distribution range have partial southern genotypes. Three of these individuals carry 50% or more of the southern genotype. We discuss the potential consequences of human-mediated gene flow and the concerns with the current practice of large-scale translocation of wrasse.

4.1. Population diversity and differentiation

As expected under isolation by distance, pairwise F_{ST} estimates (S2) demonstrate that the Flatanger as a whole is genetically most similar to Austevoll, followed by Stavanger, while almost 10-fold more differentiated from the southern sampling locations. We observe similar patterns of genetic differentiation in the individual-based clustering methods for a majority of individuals from Flatanger (figures 2 and 3). This suggests that the Flatanger population is largely a result of an ongoing northward range expansion, as suggested by Knutsen *et al.* [24]. It is possible that a more continuous sampling along the west coast of Norway would have improved upon these results by adding samples closer to Flatanger, and hence more likely to have contributed to a range expansion.

While we find a clear western/southern genetic break and an overall lower genetic diversity in the southern, Skagerrak region [19,24], the highest diversity can be seen in Flatanger, which is rather surprising, considering that this area has been colonized recently [19,23,24] and is on the leading edge of a range expansion [19,24]. Typically, a reduction in genetic diversity is to be expected when a species

colonizes a new area [50]. The high genetic diversity in Flatanger is, therefore, likely to be a result of multiple sources of origin and recent interbreeding [50], as indicated by the fact that roughly 40% of all loci demonstrated a significant heterozygosity excess in the Flatanger sample.

4.2. Hybridization

Two individuals (FKH48 and FKH50) exhibit high similarity to the southern population while differentiating from all western and Flatanger individuals. They clearly cluster with southern individuals in STRUCTURE and PCA, suggesting a southern genotype. A third individual (FKH67) did not cluster with either southern or western populations, and was classified as a F1 hybrid (50 : 50 western:southern) by NEWHYBRIDS (figure 4). Furthermore, NEWHYBRIDS found twelve Flatanger individuals to have more than 50% probability of being western backcrosses. This strongly supports ongoing hybridization between the southern and western genotypes in the wild, which has previously only been documented in captivity [26]. We also detected two potential backcrosses in Kristiansand and seven in Stavanger (figure 4*b*) in addition to the hybrids discovered in Flatanger. Stavanger and Kristiansand are the two samples collected closest to the genetic break on the western and the southern side, respectively. Except for Flatanger, we did not detect any indication of hybrids in any of the other samples further from the genetic break, indicating the existence of isolated populations [19].

The relatively high number of southern–western hybrids in Flatanger is, therefore, convincing evidence of escapement and hybridization of cleaner fish sourced from Skagerrak and/or Kattegat. Recently, Jansson *et al*. [18] showed there to be much lower differentiation than expected in goldsinny wrasse between Flatanger and Skagerrak populations indicating escapees and possibly hybridization. Unfortunately, there are no official records on the locations of source and destination of wrasses used as cleaner fish, which could have facilitated further interpretation of these results. Upon consulting with the four wrasse transport companies, they confirmed that the clear majority of wrasse being translocated in Norway are exported from Skagerrak–Kattegat coast to farms in mid- and northern Norway. Furthermore, translocations of wrasse from western Norway to mid-Norway have been strongly discouraged by food-safety authorities due to the possibility of wrasse being a carrier of pancreas disease which affects farmed salmon and is endemic in western Norway south of Hustadvika [51]. Combined, this supports the conclusion that western backcrosses in Flatanger must have been the result of hybridization with southern genotypes from Skagerrak and/or Kattegat. We did not find any western backcrosses east of Kristiansand in the Skagerrak. Consequently, the western backcross genotypes we found in Flatanger are likely a result of second-generation hybridization that occurred after translocation. Two of the companies reported to also have transported wrasse from Skagerrak to farms in western Norway. Thus, it is presently unclear whether the occurrence of western backcrosses in the Stavanger area is a result of human-mediated translocation, or if it is due to occasional natural gene-flow across the genetic break between the southern and western populations.

The onset of gene flow between previously isolated populations may have genetic, physiological and ecological consequences. The corkwing wrasse in Flatanger most likely colonized the area within the last two decades [23]. This and low catch rates attest to a very low abundance in the Flatanger area compared to regions further south (Per Andersen 2016, personal observation), rendering this population more vulnerable to hybridization events. Presently, fishing for wrasse in Sweden is allowed from 15 May, and occurs during their spawning period in May and June [52]. Hence, there is a possibility that ready-to-spawn corkwing are escaping during the spawning season, increasing the probability of hybridization. In Norway, the wrasse fishery is closed until the end of the spawning season [22], which reduces the chances of hybridization. In the UK, wrasse fishery has no temporal restrictions nationally, but in 2017 three southwestern IFCAs implemented byelaws that restrict wrasse fishery to certain periods of the year in specified areas [15–17].

4.3. Implications

The effects of hybridization between genetically distinct populations are hard to predict and depend on many factors. Fitness can increase as a result of introducing favourable alleles and genotypes (overdominance), or because of deleterious alleles being sheltered (heterosis) [50]. The three individuals with more than 50% southern genotype tended to be larger than the native fish at the same age. Although a conclusion cannot be reached without a larger sample size of hybrids, this is consistent with earlier findings of southern corkwing growing faster than western [22]. If the faster growth and larger body size for southern populations have a genetic basis, hybrids may have a fitness advantage in reproduction,

either through sexual selection for large males or higher fecundity of large females. Alternatively, a reduction in fitness can occur due to genetic incompatibilities (intrinsic outbreeding depression) or reduced adaptation to the local environment (extrinsic outbreeding depression) [50]. The life history differences between southern and western populations have been suggested to reflect temperature differences between these regions [22]. If there is local adaptation, it is likely that the continued transfer of unfit individuals would cause the loss of locally adapted alleles and genotypes, known as genetic swamping [53]. However, introgression and admixture of the southern genotype into the Flatanger population are likely to continue, whether there is an increase of fitness or not. This is because all of the hybrids' progeny will also be hybrids [50].

Populations on the boundary of a species range exist in conditions similar to the habitats just outside the distribution range, making them more likely to carry genotypes that are able to colonize new habitats [54]. As the Flatanger population constitutes the northern boundary of the species distribution, it is likely to play an important role for future adaptation potential, and range expansion. However, the asymmetric gene flow to the edges of a species range can obstruct this adaptation [55]. Admixing with southern genotypes might, therefore, work as a barrier to further range expansion. Furthermore, southern corkwing is also translocated to salmon farms even further north, to the Nordland county (Jacob Meland, Lovundlaks 2017, personal communication), where no wild corkwing populations are present. This could facilitate further spreading of southern genotypes beyond the current natural range. In addition to the genetic and ecological risks discussed above, escaping wrasse may introduce new diseases or parasites to conspecifics, salmon and other species in the wild [13,56,57]. Murray [58] argues that the risks of disease transfer from cleaner fish to salmon are small compared to the risk posed by sea lice, but disease transfer to the local populations of wrasse and other species was not considered. With ongoing hybridization, the risk of disease transfer may be an even greater threat to local populations, because hybrids may be more susceptible to diseases and parasites, as seen in other fish species [59,60].

In the face of climate-induced changing environments, conservation of populations on the leading edge should be prioritized to maximize future adaptive potential [54,61,62]. We argue that any evaluation of the risks with the translocation of wrasse needs to include effects on wild populations and ecosystems. However, prohibiting long-distance transport and sourcing wrasse locally might also pose a problem as local stocks are prone to overexploitation [12,22,63]. An obstacle for effective management is that the current practice of cleaner fish use is poorly documented and regulated. Norwegian law states that aquacultures are obligated to report all escaping fish from aquaculture installations, but presently only the target species cultured are recorded. Moreover, Norwegian and UK transporters are not required to log and report the source or the destination of cleaner fish, which complicates the possibilities to assess and address the problem of escapees.

5. Conclusion

We provide the first evidence that translocated wild corkwing wrasse used as cleaner fish in salmon farms escape and hybridize with local populations at the northern limit of its distribution. These findings provide important information for aquaculture management and conservation of wild populations of non-target species, and have implications for the increasing use of cleaner fish as parasite control in fish farms, which is both poorly documented and regulated. Moving genetic material between isolated populations could drastically alter the genetic composition, erode population structure and potentially result in loss of local adaptation, hampering the species expansion. The geographical extent and magnitude of introgression and the ecological consequences remain unknown for this and other wrasse species. It is urgent to address these gaps of knowledge, as there is no immediate sign of reduction of the current practice in Norway, and wrasse are increasingly being deployed in other areas such as the UK.

Ethics. Fish sampling was conducted in compliance with the Norwegian Animal Welfare Act (LOV 2009-06-19 nr 97) and the Swedish Board of Agriculture (Dnr 35-2016, addition to 59-2015). We strived to minimize handling time and stress imposed on the fish.

Data accessibility. Data are deposited on Dryad: http://dx.doi.org/10.5061/dryad.tv553 [64]. Raw data are available on NCBIs Sequence Read Archive (BioProject PRJNA415388). Scripts used for bioinformatic analysis can be found at https://github.com/z0on/2bRAD_denovo. Log of the pipeline used can be found at https://github.com/ellikafaust/S.melopsPopGen/blob/master/2bRAD_log_escapee.

Authors' contributions. K.T.H. and C.A. conceived the study. E.F., K.T.H., H.K. and C.A. planned the research. E.F., K.T.H., P.A. and H.K. conducted and coordinated sample collection. E.F. conducted the molecular laboratory and bioinformatic work. K.T.H. performed morphometrics and age analysis. E.F. and C.A. performed statistical analyses.

E.F. drafted the manuscript with the help from K.T.H. and C.A. All authors contributed to the final writing and approval of the publication.

Competing interests. We declare we have no competing interests.

Funding. E.F. was funded by EU Interreg project MarGen; Swedish Research Council FORMAS. H.K. was funded by EU Interreg project MarGen; Norwegian Research Council (grant number 234328/E40). C.A. was funded by EU Interreg project MarGen; Swedish Research Council FORMAS.

Acknowledgements. We thank local fishermen, fisherwomen and research colleagues for assisting with sampling, especially Åsa Strand for help with Swedish sampling, Lars Mårvik for collecting corkwing in Flatanger, and Reine Andreasson for collecting corkwing in Strömstad. Torkel Larsen assisted with the otolith work and age-reading. Pierre de Wit gave valuable advice on the 2bRAD method and Mats Töpel on SNP data storage and analysis. Sequencing was performed by the SNP&SEQ Technology Platform in Uppsala. The facility is part of the National Genomics Infrastructure (NGI) Sweden and Science for Life Laboratory. The SNP&SEQ Platform is also supported by the Swedish Research Council and the Knut and Alice Wallenberg Foundation. All bioinformatics analyses were run on the Albiorix computer cluster (http://albiorix.bioenv.gu.se/) at the Department of Marine Sciences, University of Gothenburg. This study was conducted within the Linnaeus Centre for Marine Evolutionary Biology (CeMEB).

References

1. Hauser L, Carvalho GR. 2008 Paradigm shifts in marine fisheries genetics: ugly hypotheses slain by beautiful facts. *Fish.* **9**, 333–362. (doi:10.1111/j.1467-2979.2008.00299.x)
2. Laikre L, Schwartz MK, Waples RS, Ryman N, Group TGW. 2010 Compromising genetic diversity in the wild: unmonitored large-scale release of plants and animals. *Trends Ecol. Evol.* **25**, 520–529. (doi:10.1016/j.tree.2010.06.013)
3. Jensen DT, Thorstad EB, Uglem I, Fredheim A. 2010 Escapes of fishes from Norwegian sea-cage aquaculture: causes, consequences and prevention. *Aquac. Environ. Interact.* **1**, 71–83. (doi:10.3354/aei00008)
4. Glover KA *et al.* 2017 Half a century of genetic interaction between farmed and wild Atlantic salmon: status of knowledge and unanswered questions. *Fish Fish.* **18**, 890–927. (doi:10.1111/faf.12214)
5. Bolstad GH *et al.* 2017 Gene flow from domesticated escapes alters the life history of wild Atlantic salmon. *Nat. Ecol. Evol.* **1**, 124. (doi:10.1038/s41559-017-0124)
6. Darwall W, Costello M, Donelly R, Lysaght S. 1992 Implications of life-history strategies for a new wrasse fishery. *J. Fish Biol.* **41**, 111–123. (doi:10.1111/j.1095-8649.1992.tb03873.x)
7. Skiftesvik AB *et al.* 2014 Wrasse (*Labridae*) as cleaner fish in salmonid aquaculture—the Hardangerfjord as a case study. *Mar. Biol. Res.* **10**, 289–300. (doi:10.1017/CBO9781107415324.004)
8. Bjordal Å. 1988 Cleaning symbiosis between wrasse (Labridae) and lice infested salmon (Salmo salar) in mariculture. International Council for the Exploration of the Sea, 1988/F:17.
9. Bjordal Å. 1991 Wrasse as cleaner fish for farmed salmon. *Prog. Underw. Sci.* **16**, 17–28.
10. Deady S, Varian SJA, Fives JM. 1995 The use of cleaner-fish to control sea lice on two Irish salmon (*Salmo salar*) farms with particular reference to wrasse behaviour in salmon cages. *Aquaculture* **131**, 73–90. (doi:10.1016/0044-8486(94)00331-H)
11. Besnier F *et al.* 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**, 937. (doi:10.1186/1471-2164-15-937)
12. Halvorsen KT, Larsen T, Sørdalen TK, Vøllestad LA, Knutsen H, Olsen EM. 2017 Impact of harvesting cleaner fish for salmonid aquaculture assessed from replicated coastal marine protected areas. *Mar. Biol. Res.* **13**, 1–11. (doi:10.1080/17451000.2016.1262042)
13. Svåsand T, Grefsrud ES, Karlsen Ø, Kvamme BO, Glover K, Husa V, Kristiansen TS. 2017 Risikorapport norsk fiskeoppdrett 2017. Fisk. og havet. SN 2-2017.
14. Riley A, Jeffery K, Cochrane-dyet T, White P, Ellis J. 2017 Northern European wrasse—summary of commercial use, fisheries and implications for management.
15. Devon & Severn IFCA. 2017 Potting Permit Byelaw. Management of the 'live' wrasse pot fishery. Devon & Severn Inshore Fisheries and Conservation Authority. See https://secure.toolkitfiles.co.uk/clients/15340/sitedata/4F/Byelaw_development_reports/Wrasse/Final-Wrasse-v-3-new-cover-Aug16th-2017.pdf.
16. Cornwall IFCA. 2017 Live wrasse fishery guidance 2017–18. Cornwall Inshore Fisheries and Conservation Authority. See https://secure.toolkitfiles.co.uk/clients/17099/sitedata/Code_of_practice/live-wrasse-fishery-guidnece.pdf.
17. Southern IFCA. 2017 Wrasse fishery guidance. Southern Inshore Fisheries and Conservation Authority. See https://secure.toolkitfiles.co.uk/clients/25364/sitedata/files/Wrasse-Guidance.pdf.
18. Jansson E *et al.* 2017 Genetic analysis of goldsinny wrasse reveals evolutionary insights into population connectivity and potential evidence of inadvertent translocation via aquaculture. *ICES J. Mar. Sci.* **74**, 2135–2147. (doi:10.1093/icesjms/fsx046)
19. Blanco GE, Knutsen H, Jorde PE. 2016 Habitat discontinuities separate genetically divergent populations of a rocky shore marine fish. *PLoS ONE* **11**, e0163052. (doi:10.1371/journal.pone.0163052)
20. Potts GW. 1985 The nest structure of the corkwing wrasse, crenilabrus melops (Labridae: Teleostei). *J. Mar. Biol. Assoc. UK* **65**, 531–546. (doi:10.1017/S002531540005058X)
21. Hilldén N-O. 1984 Behavioural ecology of the labrid fishes (Teleostei, Labridae) at Tjärnö on the Swedish west coast.
22. Halvorsen KT, Sørdalen TK, Durif C, Knutsen H, Olsen EM, Skiftesvik AB, Rustand TE, Bjelland RM, Vøllestad LA. 2016 Male-biased sexual size dimorphism in the nest building corkwing wrasse (*Symphodus melops*): implications for a size regulated fishery. *ICES J. Mar. Sci.* **73**, 2586–2594. (doi:10.1093/icesjms/fsw135)
23. Maroni K, Andersen P. 1996 Distribution and abundance of wrasse in an area of northern Norway. In *Wrasse: biology and use in aquaculture* (eds MDJ Sayer, MJ Costello, JW Treasurer), pp. 70–73. Oxford, UK: Fishing News Books.
24. Knutsen H, Jorde PE, Blanco Gonzalez E, Robalo JI, Albretsen J, Almada V. 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. (doi:10.1371/journal.pone.0067492)
25. Woll AK, Bakke S, Aas GH, Solevåg SE, Skiftesvik AB, Bjelland RM. 2013 Velferd leppefisk i merd.
26. Blanco GE, de Boer F. 2017 The development of the Norwegian wrasse fishery and the use of wrasses as cleaner fish in the salmon aquaculture industry. *Fish. Sci.* **83**, 661–670. (doi:10.1007/s12562-017-1110-4)
27. Wang S, Meyer E, McKay JK, Matz MV. 2012 2b-RAD: a simple and flexible method for genome-wide genotyping. *Nat. Methods* **9**, 808–810. (doi:10.1038/nmeth.2023)
28. Matz MV, Aglyamova G. 2014 Protocol for Illumina 2bRAD sample preparation. See http://www.bio.utexas.edu/research/matz_lab/matzlab/Methods_files/2bRAD_protocol-1.pdf (accessed 26 May 2017).
29. De Wit P. 2016 IB14_2b-RAD_log_August2016.sh. See https://github.com/DeWitP/BONUS_BAMBI_IDOTEA/blob/master/IB14_2b-RAD_log_August2016.sh (accessed 26 May 2017).
30. Roesti M, Salzburger W, Berner D. 2012 Uninformative polymorphisms bias genome scans for signatures of selection. *BMC Evol. Biol.* **12**, 94. (doi:10.1186/1471-2148-12-94)
31. Lischer HEL, Excoffier L. 2012 PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics* **28**, 298–299. (doi:10.1093/bioinformatics/btr642)
32. Keenan K, Mcginnity P, Cross TF, Crozier WW, Prodöhl PA. 2013 diveRsity: an R package for the estimation and exploration of population genetics parameters and their associated errors. *Methods*

Ecol. Evol. **4**, 782–788. (doi:10.1111/2041-210X.12067)

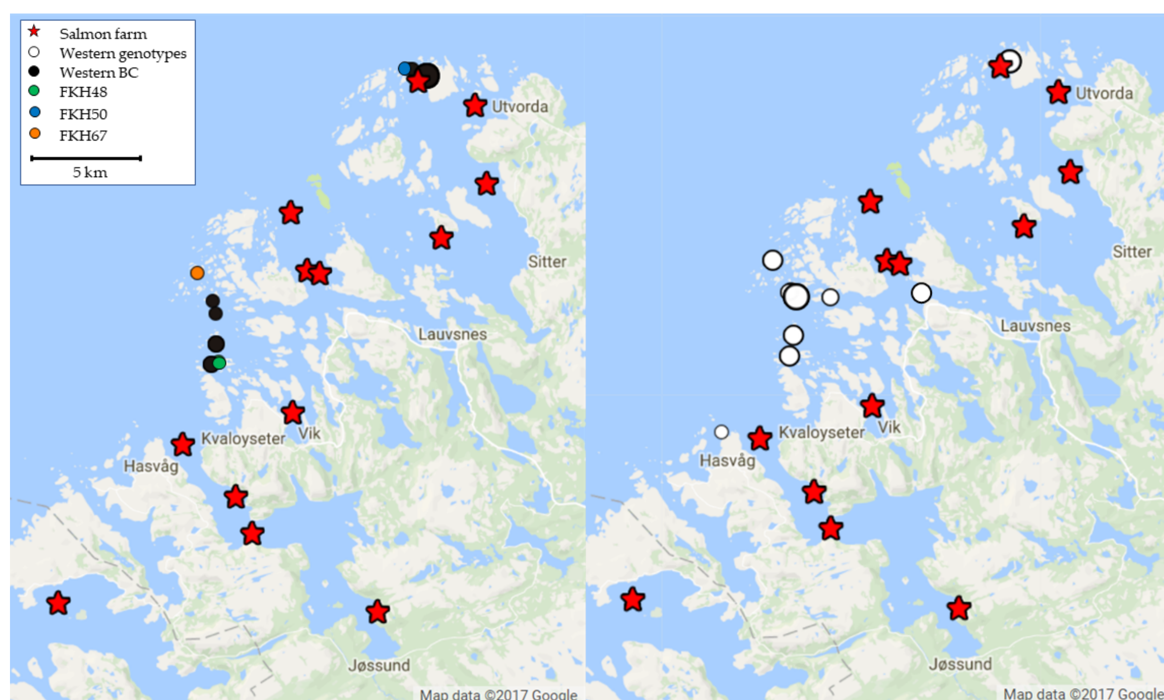
33. R Core Team. 2015 *R: a language and environment for statistical computing*. Vienna, Austria: R foundation for statistical computing. See http://www.r-project.org.
34. Weir BS, Cockerham CC. 1984 Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358–1370.
35. Louis EJ, Dempster ER. 1987 An exact test for Hardy-Weinberg and multiple alleles. *Int. Biometric Soc.* **43**, 805–811. (doi:10.2307/2531534)
36. Benjamini Y, Hochberg Y. 1995 Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* **57**, 289–300. (doi:10.2307/2346101)
37. Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ. 2015 Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 1–16. (doi:10.1186/s13742-015-0047-8)
38. Purcell S, Chang C. 2017 PLINK. 9.
39. Pritchard JK, Stephens M, Donnelly P. 2000 Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959. (doi:10.1111/j.1471-8286.2007.01758.x)
40. Chessel D, Dufour AB, Thioulouse J. 2004 The ade4 package—I: one-table methods. *R News* **4**, 5–10. (doi:10.2307/3780087)
41. Dray S, Dufour AB, Chessel D. 2007 The ade4 package—II: two-table and K-table methods. *R News* **7**, 47–52. (doi:10.1159/000323281)
42. Dray S, Dufour AB. 2007 The ade4 package: implementing the duality diagram for ecologists. *J. Stat. Softw.* **22**, 1–20. (doi:10.1.1.177.8850)
43. Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I. 2015 CLUMPAK: a program for identifying clustering modes and packaging population structure inferences across K. *Mol. Ecol. Resour.* **15**, 1179–1191. (doi:10.1111/1755-0998.12387)
44. Earl DA, Von Holdt BM. 2012 STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* **4**, 359–361. (doi:10.1007/s12686-011-9548-7)
45. Jombart T. 2008 Adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403–1405. (doi:10.1093/bioinformatics/btn129)
46. Jombart T, Ahmed I. 2011 adegenet 1.3–1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* **27**, 3070–3071. (doi:10.1093/bioinformatics/btr521)
47. Rousset F. 2008 GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Mol. Ecol. Resour.* **8**, 103–106. (doi:10.1111/j.1471-8286.2007.01931.x)
48. Wringe BF, Stanley RRE, Jeffery NW, Anderson EC, Bradbury IR. 2017 HYBRIDDETECTIVE: a workflow and package to facilitate the detection of hybridization using genomic data in R. *Mol. Ecol. Resour.* **17**, e275–e284. (doi:10.1111/1755-0998.12704)
49. Anderson EC, Thompson EA. 2002 A model-based method for identifying species hybrids using multilocus data. *Genetics* **160**, 1217–1229.
50. Allendorf FW, Luikar G, Aitken SN. 2013 *Conservation and the genetics of populations*, 2nd edn. Hoboken, NJ: Wiley-Blackwell.
51. Olsen AB, Jensen BB, Nilsen H, Grøntvedt RN, Gjerset B, Taksdal T, Høgåsen HR. 2011 Risikovurdering for spredning av pancreas disease virus (PD-virus) ved bruk av leppefisk i norsk laksefiskoppdrett.
52. Swedish Agency for Marin and Water Management. 2017 Swedish exempted fishing permits.
53. Lenormand T. 2002 Gene flow and the limits to natural selection. *Trends Ecol. Evol.* **17**, 183–189. (doi:10.1016/S0169-5347(02)02497-7)
54. Gibson SY, Marel RC, Starzomski BM. 2009 Climate change and conservation of leading-edge peripheral populations. *Conserv. Biol.* **23**, 1369–1373. (doi:10.1111/j.1523-1739.2009.01375.x)
55. Kirkpatrick M, Barton NH. 1997 Evolution of a species' range. *Am. Nat.* **150**, 1–23. (doi:10.1086/286054)
56. Treasurer JW. 2012 Diseases of north European wrasse (Labridae) and possible interactions with cohabited farmed salmon, *Salmo salar* L. *J. Fish Dis.* **35**, 555–562. (doi:10.1111/j.1365-2761.2012.01389.x)
57. Wallace IS, Donald K, Munro LA, Murray W, Pert CC, Stagg H, Hall M, Bain N. 2015 A survey of wild marine fish identifies a potential origin of an outbreak of viral haemorrhagic septicaemia in wrasse, Labridae, used as cleaner fish on marine Atlantic salmon, *Salmo salar* L., farms. *J. Fish Dis.* **38**, 515–521. (doi:10.1111/jfd.12259)
58. Murray AG. 2017 A model of the process of spillover and adaption leading to potential emergence of disease in salmon held with cleaner fish used to control lice. *Aquaculture* **473**, 283–290. (doi:10.1016/j.aquaculture.2017.02.028)
59. Currens KP, Hemmingsen AR, French RA, Buchanan DV, Schreck CB, Li HW. 1997 Introgression and susceptibility to disease in a wild population of rainbow trout. *North Am. J. Fish. Manag.* **17**, 1065–1078. (doi:10.1577/1548-8675(1997)017<1065:IASTDI>2.3.CO;2)
60. Goldberg TL, Grant EC, Inendino KR, Kassler TW, Claussen JE, Philipp DP. 2005 Increased infectious disease susceptibility resulting from outbreeding depression. *Conserv. Biol.* **19**, 455–462. (doi:10.1111/j.1523-1739.2005.00091.x)
61. Haak AL, Williams JE, Neville HM, Dauwalter DC, Colyer WT. 2010 Conserving peripheral trout populations: the values and risks of life on the edge. *Fisheries* **35**, 530–549. (doi:10.1577/1548-8446-35.11.530)
62. Rehm EM, Olivas P, Stroud J, Feeley KJ. 2015 Losing your edge: climate change and the conservation value of range-edge populations. *Ecol. Evol.* **14**, 4315–4326. (doi:10.1002/ece3.1645)
63. Halvorsen KT. 2017 Selective harvesting and life history variability of corkwing and goldsinny wrasse in Norway: implications for management and conservation. Doctoral thesis, University of Oslo.
64. Faust E, Halvorsen KT, Andersen P, Knutsen H, André C. 2018 Data from: Cleaner fish escape salmon farms and hybridize with local wrasse populations. Dryad Digital Repository. (http://dx.doi.org/10.5061/dryad.tv553)

SUPPLEMENTARY

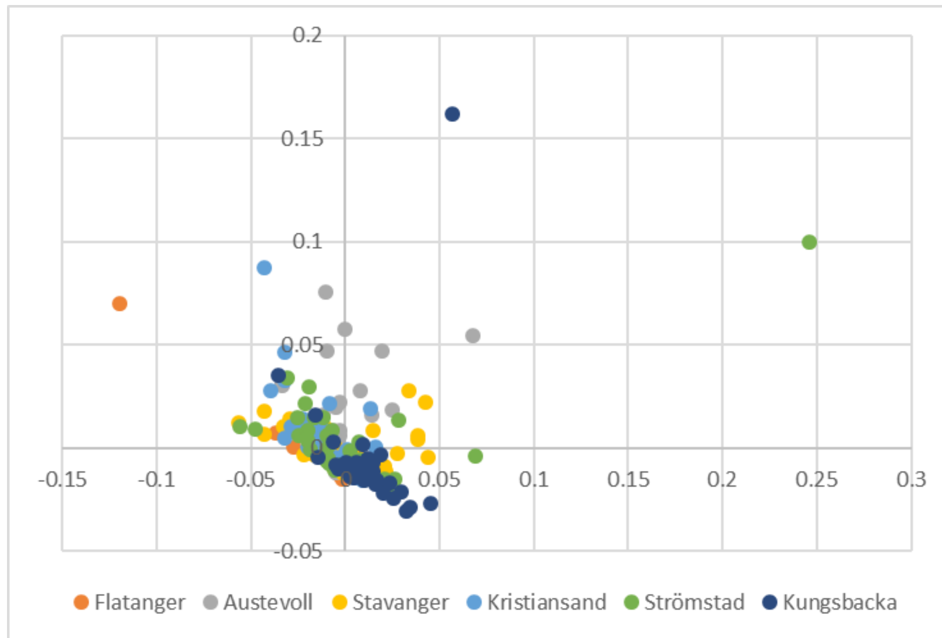
S1 Sampling information of corkwing wrasse (*Symphodus melops)*. Sample size indicates the number of individuals collected at each site.

Sample name	Area	Sample size	Year	Area Coordinates
Flatanger	Norwegian Sea	40	2016	64.513943, 10.679272
Austevoll	North Sea N	40	2016	60.096642, 5.269618
Stavanger	North Sea S	40	2016	58.963804, 5.942488
Kristiansand	Skagerrak W	40	2016	58.187534, 8.048447
Strömstad	Skagerrak E	40	2016	58.947222, 11.000000
Kungsbacka	Kattegat	40	2016	57.403142, 11.907473

S2 Map of sampling locations in Flatanger and the nearby open pen Salmon farms (data from https://kart.fiskeridir.no/akva). Left: Sampling locations of Western backcrosses (black) and Southern backcrosses (coloured). Right: Sampling locations of Western genotypes (open circles). The size of the circles corresponds to the number of individuals at each of the locations 1-5.) Visualisation done on http://www.copypastemap.com.



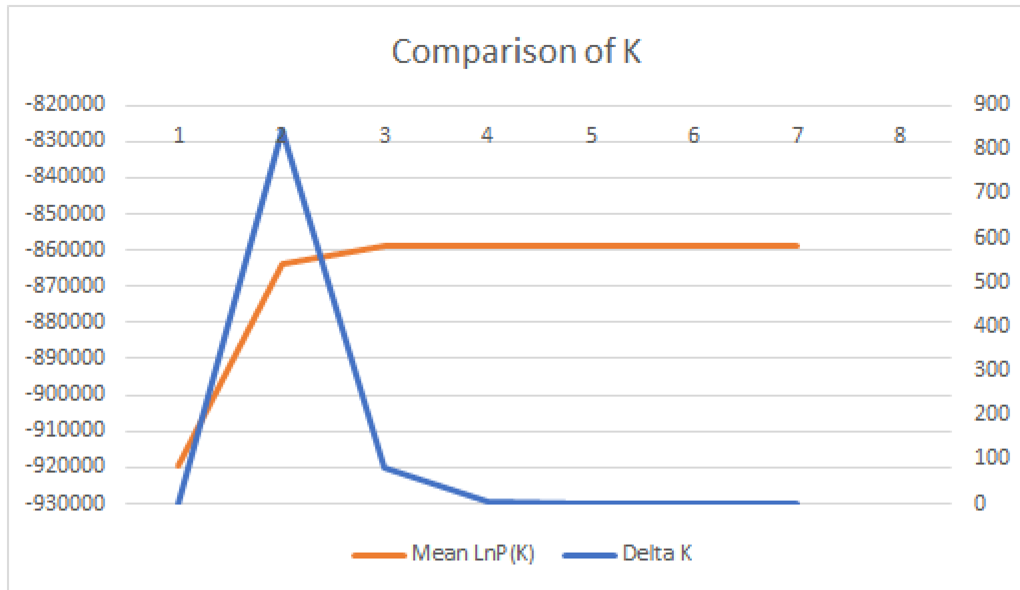
S3 Multidimensional scaling plot showing identity by missingness calculated in PLINK [1]. First (x-axis) and second (y-axis) dimensions of 240 corkwing wrasse individuals from 6 locations based on identity by missingeness in 4357 SNPs. Each point represents one individuals, which are colour coded by sampling site. No clear structure or pattern of missingness can be seen.



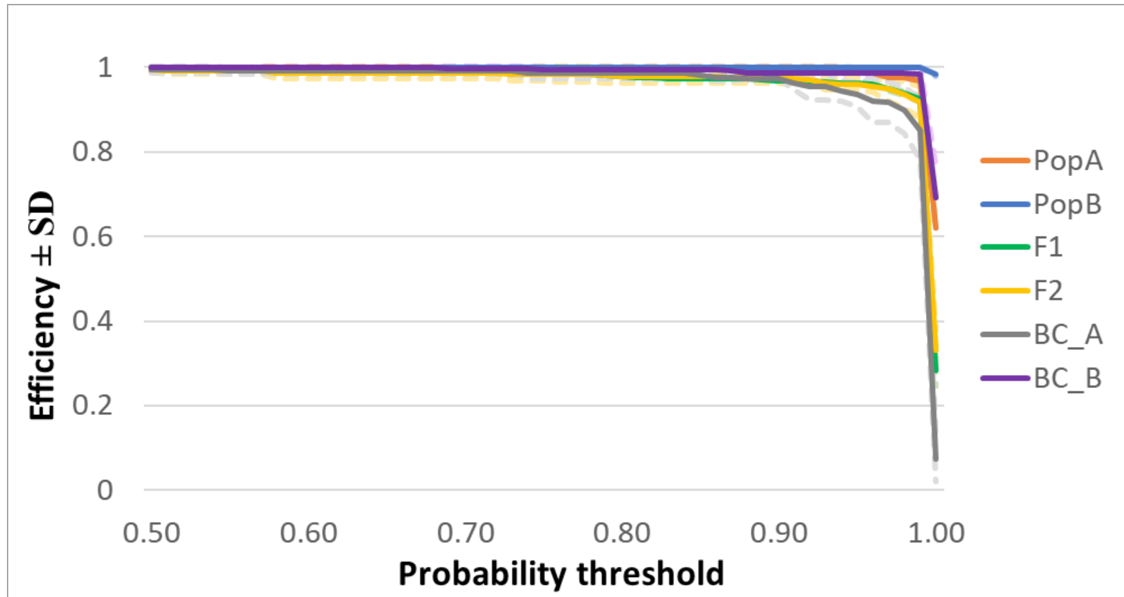
S4 Matrix of pairwise F_{ST} according to Weir and Cockerham (1984) calculated with the R package diveRsity [2]. Significance of the estimates was tested using Fisher's Exact tests with 10000 MC reps. * q < 0.001

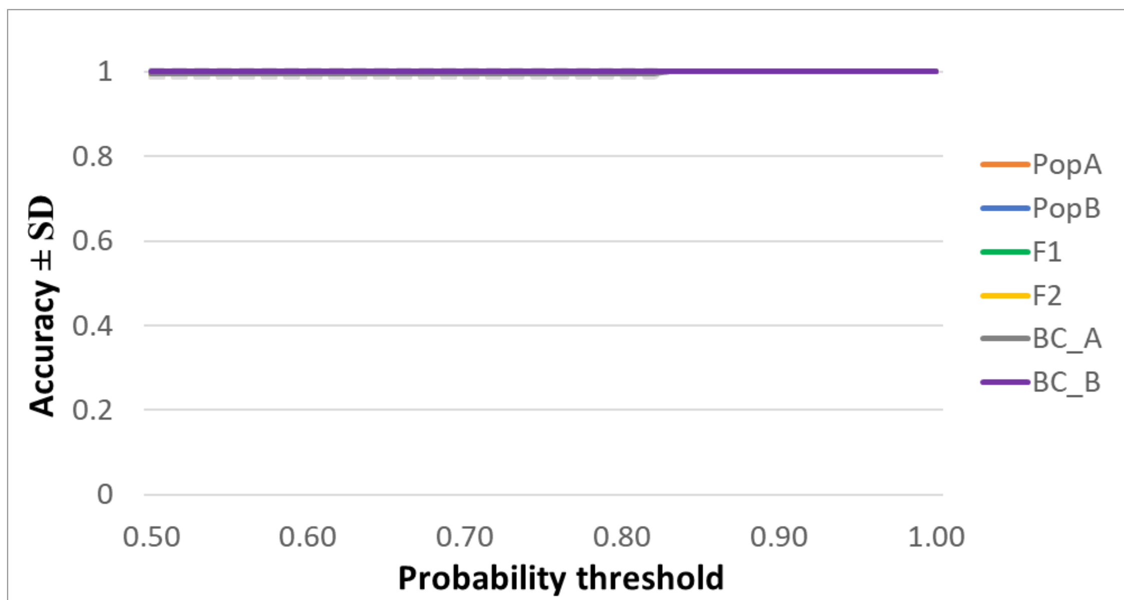
	Flatanger	Austevoll	Stavanger	Kristiansand	Strömstad	Kungsbacka
Flatanger						
Austevoll	0.0243*					
Stavanger	0.0277*	0.0065*				
Kristiansand	0.1163*	0.1213*	0.101*			
Strömstad	0.1218*	0.1281*	0.1072*	0.0023		
Kungsbacka	0.1258*	0.1312*	0.1102*	0.0030	0.0029	

S5 A visualisation of mean LnP(K) ± SD and delta K for K 1-7 clusters run 3 times in STRUCTURE [3]. Values calculated using Structure Harvester, http://taylor0.biology.ucla.edu/structureHarvester [4].



S6 Assignment efficiency and accuracy at different probability thresholds based on three sets of simulated genotype data [5] from 200 SNPs with the highest over all F_{ST} and no LD. Solid lines represent the 6 genotype classes, pure parents (PopA = western population and PopB = southern population), first and second-generation hybrids (F1 and F2) and backcrosses (BC_A and BC_B). The dashed lines stand for the standard deviation among the simulations for each class. Efficiency = correctly assigned individuals over the known individuals per class. Accuracy = correctly assigned individuals over individuals assigned to that class.





1. Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ. 2015 Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* **4**, 1–16. (doi:10.1186/s13742-015-0047-8)

2. Keenan K, Mcginnity P, Cross TF, Crozier WW, Prodöhl PA. 2013 diveRsity: An R package for the estimation and exploration of population genetics parameters and their associated errors. *Methods Ecol. Evol.* **4**, 782–788. (doi:10.1111/2041-210X.12067)

3. Pritchard JK, Stephens M, Donnelly P. 2000 Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959. (doi:10.1111/j.1471-8286.2007.01758.x)

4. Earl DA, Von Holdt BM. 2012 STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* **4**, 359–361. (doi:10.1007/s12686-011-9548-7)

5. Wringe BF, Stanley RRE, Jeffery NW, Anderson EC, Bradbury IR. 2017 HYBRIDDETECTIVE: A workflow and package to facilitate the detection of hybridization using genomic data in R. *Mol. Ecol. Resour.* **17**, e275–e284. (doi:10.1111/1755-0998.12704)

PAPER II

Large scale survey of escape and hybridization of cleaner fish in aquaculture

Ellika Faust[1#], Eeva Jansson[2#*], Carl André[1], Kim Tallaksen Halvorsen[3], Geir Dahle[2], Halvor Knutsen[4,5], María Quintela[2], Kevin A. Glover[2,6]

[1]*Department of Marine Sciences – Tjärnö, University of Gothenburg, 45296 Strömstad, Sweden,*
[2]*Institute of Marine Research, P. O. Box 1870 Nordnes, N-5817 Bergen, Norway*
[3]*Institute of Marine Research, Austevoll Research Station, Storebø, N-5392, Norway*
[4]*Institute of Marine Research, Flødevigen, Nye Flødevigveien 20, N-4817 His, Norway*
[5]*Centre of Coastal Research, University of Agder. P.O.Box 422, N-4604 Kristiansand, Norway*
[6]*Institute of Biology, University of Bergen, P. O. Box 7803, N-5020 Bergen, Norway*

[#]*These authors contributed equally to this work*
**Corresponding author: email: eeva.jansson@hi.no*

Abstract

Translocation and introduction of new organisms can have considerable negative impact on local populations and ecosystems. Despite this it is still common practise in agri- and aquaculture. Every year millions of wild caught wrasse are transported hundreds of kilometres to be used as cleaner fish for parasite control in Norwegian salmon farms. It was recently discovered that translocated cleaner fish are able to escape, survive and reproduce. Here we have developed a panel of 84 SNPs that can be used to detect escaping corkwing wrasse (*Symphodus melops*) and their first and second generation hybrids. Applying these markers to ~2000 individuals, we found that escapees and hybrids may constitute up to 20 % of the local population at the northern edge of the species distribution. The introgression of southern genetic material at the northern edge of the species range alters the local genetic composition, but could also obstruct local adaptation, and act as a potential barrier to further range expansion. Surprisingly, in other parts of the species distribution, where salmon farming is also common, we found no escapees and few hybrids. A possible explanation is that smaller, marginal and newly established populations are more prone to introgression and random drift effects. However, the current lack of reporting makes it difficult to evaluate possible causes for why we only see few escapees and hybrids in other aquaculture-dense areas. Reporting escapees at the end of a season and the source and destination of translocated cleaner fish would improve the ability to assess current and future risks associated with the use of cleaner fish for parasite control.

Introduction

Moving organisms outside their natural boundaries comes with many potential problems and can have many and diverse effects on the ecosystems (Atalah & Sanchez-Jerez, 2020). Introductions can for instance affect some species through ecological competition, either by becoming their prey or predator, or by competing for available resources (Evangelista, Cucherousset, & Lecerf, 2019). Introduced individuals can also carry pathogens, that being unknown to the local population, can quickly spread into a novel environment, which has not been able to develop any form of resistance (Tepolt et al., 2020). Furthermore, if the introduced populations are genetically distant from the local ones, introgression and admixture can lead to altered population structure (Glover et al., 2012), lower effective population size, and reduced fitness through outbreeding depression (Blakeslee, Manousaki, Vasileiadou, & Tepolt, 2020; Glover et al., 2017; Laikre, Schwartz, Waples, Ryman, & Group, 2010). Donor populations and ecosystems can also be negatively affected if harvest leads to disruption in species interactions and ecosystem function (Halvorsen, Larsen, et al., 2017), and adverse genetic effects such as loss of diversity due to dwindling population size (Allendorf, England, Luikart, Ritchie, & Ryman, 2008). Despite the known problems, introduction of species into new areas and translocation of individuals from foreign populations are common practice in aquaculture. These actions aim to increase catches, mitigate loss of wild stocks, and restore or even create new fisheries. Likewise, many species are harvested in large numbers in the wild to provide food or other services to cultured species such as cleaner fish to delouse salmonids.

Wrasses (*Labridae*) are a large and diverse family of marine fish with over 600 described species worldwide. Many wrasses show natural cleaning behaviour, i.e. they feed on ectoparasites from other fish species' skin. In Norway alone, millions of wrasse are utilized as cleaner fish and translocated hundreds of kilometres every year to be used for parasite control in salmon farms (Norwegian directorate of Fisheries, 2019). Although often considered as an environment-friendly form of parasite control (Liu & Bjelland, 2014), the increasing fishing pressure and large-scale translocation of cleaner fish raise concerns about potential overfishing and human-mediated gene flow from translocated individuals to wild populations. There are many examples of salmonids escaping open-pen aquaculture and hybridising with local populations, leading to genetic swamping and reduced fitness (Bolstad et al., 2017; Glover et al., 2017). Recently, several studies have collectively demonstrated that also wrasses are able to escape and likely hybridise and introgress with local populations (Blanco González et al., 2019; Faust, Halvorsen, Andersen, Knutsen, & André, 2018; Jansson et al., 2017). However, the geographical extent, magnitude of introgression and the ecological consequences are largely unknown. In contrast to regulations for salmonid farming, there are currently no requirements for preventing escape of cleaner fish from sea-cages, nor reporting escapes when they occur.

The use of wrasse as cleaner fish for sea lice control in commercial aquaculture was first implemented in the late 1980s (Bjordal, 1988), and modest numbers of wrasse have been used as cleaner fish annually ever since. However, the use of cleaner fish increased

dramatically in the last decade a result of sea lice developing resistance to widely used pharmaceutical treatments (Besnier et al., 2014; Kaur et al. 2017). The number of cleaner fish used in Norway alone has increased from 1.7 million in 2008 to ~50 million in 2017 and 2018 (Norwegian directorate of Fisheries, 2019). Currently there are five different species used as cleaner fish in Norwegian aquaculture, lumpfish (*Cyclopterus lumpus*), ballan wrasse (*Labrus bergylta*), goldsinny wrasse (*Ctenolabrus rupestris*), corkwing wrasse *(Symphodus melops*) and small amounts of rock cook (*Centrolabrus exoletus*). Lumpfish, whose potential use as a cleaner fish was discovered in 2014, has since become the most commonly used cleaner fish (Imsland et al., 2014). The majority of lumpfish are farmed whilst almost all wrasses are caught wild and transported to aquaculture facilities. Currently, ballan wrasse is the only commercially reared wrasse species, albeit at a very small scale (Norwegian directorate of Fisheries 2019). Goldsinny and corkwing wrasse are, by far, the most commonly used wild caught cleaner fish. In 2018, 7.4 million goldsinny and 6.3 million corkwing wrasse were deployed as cleaner fish in Norwegian aquaculture.

The four wrasse species inhabit shallow rocky areas along the coast from the Mediterranean Sea in the south, to the Norwegian coast in the north. In recent years, their abundance has shifted northwards and diminished in the south, which has been suggested to be due to increased sea water temperatures (Knutsen et al., 2013). These species differ in their ecology and life history characteristics in several ways, but they are all believed to be territorial and non-migratory, thus almost exclusively dependent on the planktonic early life stages for dispersal (Darwall, Costello, Donnelly, & Lysaght, 1992; Skiftesvik, Durif, Bjelland, & Browman, 2014). Previous studies of genetic population structure have shown large differences between populations in the northern and southern part of the distribution, as well as patterns of isolation by distance along the Scandinavian coastline (D'Arcy, Mirimin, & FitzGerald, 2013; Jansson et al., 2017; Knutsen et al., 2013; Robalo et al., 2012; Seljestad et al., 2020). However, the most striking divergence is the genetic break (F_{ST} ~ 0.1) in corkwing wrasse which is located in south-western Norway (Blanco González et al. 2016; Mattingsdal et al., 2020). The break only spans <60 km and has been suggested to be a result of post-glacial recolonization and founder events separating the populations for more than ~10 kya (Mattingsdal et al., 2020).

Corkwing wrasse is a nest building species and spawns benthic eggs, which are dependent on paternal care until hatching. Nesting males are brightly coloured and significantly larger than females or sneaker males, which mimic the females' brown colour and smaller body size (Halvorsen et al., 2016). Currently, nesting males are disproportionately targeted by Norwegian fisheries, which are regulated by a minimum size limit (Halvorsen, Sørdalen, et al., 2017). However, size, maturity and proportion of nesting males to sneaker males do not seem to be consistent across populations. Recent studies suggest that populations south of the genetic break are growing faster, maturing earlier, having a shorter life span and a lower proportion of sneaker males to nesting males (Halvorsen et al., 2016).

The strong differentiation found over the genetic break south and west coast of Norway allowed for the development of genomic tools to identify escaping individuals as well as first- and second-generation hybrids between escaping southern individuals and local populations

(Faust et al., 2018). Faust and colleagues showed in their study (2018) that translocated corkwing wrasses can escape and hybridize with local populations at the northern edge of the species distribution limit in Flatanger, Norway. Of the 40 corkwing wrasse they sampled, two were identified as southern escapees and 13 as potential first or second generation hybrids. However, the results were limited in geographic scope, and more samples are needed to quantify the extent of escape and introgression of corkwing wrasse inadvertently translocated from southern to northern areas of Norway in association with aquaculture of salmon. In the present study, we aimed to investigate the quantity and geographic extent of escaping and hybrid individuals on the Norwegian west coast. In order to achieve this, we developed a panel of genome-wide SNPs, and analysed ~2000 corkwing wrasse from aquaculture-dense regions in western Norway and potential source populations in Skagerrak.

Material & Methods

SNP selection and Bioinformatics

In order to find discriminant and divergent SNPs, we used published 2b-RAD sequences from Faust et al. (2018) available at NCBI (Bioproject PRJNA415388) together with additional unpublished sequences. The additional sequences were sampled and processed in the same way as the published ones using a modified version of 2b-RAD (Wang, Meyer, McKay, & Matz, 2012) full procedure (Faust et al., 2018). Sequences were mapped using bowtie2 (Langmead & Salzberg, 2012) to the published *Symphodus melops* genome (Mattingsdal et al., 2017). Variants calling was done following the GATK pipeline (McKenna et al., 2010) using UnifiedGenotyper after realigning sequences around indels and recalibrating base quality (BQSR). Variant score quality was recalibrated (VQSR) using site identity across technical replicates as a training set. To ensure high confidence in genotype and SNPs, we used vcftools (Danecek et al., 2011) filtering on quality by depth (QD < 2.0), strand bias (FS > 60, SOR > 2) and mapping quality (MQ < 40). Sites with more than 10% missing data and with a fraction of heterozygotes above 0.5 (possible lumped paralogs) were removed, leaving a total of 10 747 SNPs.

To select the most divergent SNPs between import and export populations, a pairwise comparison was conducted between one sample from western Norway (Austevoll) and three from the exporting region in southern Norway and western Sweden (Risör, Sandefjord and Kungsbacka, respectively). Each sample consisted of 40 individuals. A total of 387 SNPs, distributed over 270 contigs, were identified among the 500 highest F_{ST} values in all three pairwise comparisons. SNPs displaying F_{ST} values > 0.4 were then used to design the final panel for genotyping. Reading and converting between file formats was done using VcfR radiator (Knaus & Grünwald, 2016, 2017) and Radiator (Gosselin, 2019), and the package diveRsity (Keenan, McGinnity, Cross, Crozier, & Prodöhl, 2013) was used to calculate pairwise F_{ST}.

SNP locus primer design, amplification and genotype calling was based on the Agena MassARRAY iPLEX Platform, as described by Gabriel et al. (2009). Selected SNP loci were analyzed in four assay groups (Supplementary material, Table S1). Accuracy, efficiency and

power of the four assays to correctly identify escaping individuals from the two populations and their potential offspring was estimated using R package HYBRIDDETECTIVE (Wringe, Stanley, Jeffery, Anderson, & Bradbury, 2017a). Genotype frequencies from the reference samples in Austevoll and Risör with 40 individuals each were used to simulate three replicates of three independent data sets with pure parents (Pure1 and Pure2), first and second-generation hybrids (F1 and F2), and backcrosses between F1 and pure parents (BC1 and BC2). The simulated data sets contained 288 individuals and were analysed using the R package parallelnewhybrid (Wringe, Stanley, Jeffery, Anderson, & Bradbury, 2017b) and NEWHYBRIDS v. 1.1 (Anderson & Thompson, 2002), which estimates the posterior probability of each individual to belong to one of the six hybrid classes. The analysis was done using default priors and genotype proportions with a burn-in period of 50 000 iteration and 300 000 MCMC sweeps. In case of non-convergent MCMC chains, simulations were re-analyzed. Power was estimated as the product of efficiency (correctly assigned individuals over the known individuals per class) and accuracy (correctly assigned individuals over individuals assigned to that class) as described in Wringe et al. (2017a). Simulations demonstrated a high efficiency (> 94%), accuracy (> 98%), and power (> 94) to detect individuals from all of the six hybrid classes (Supplementary material, Figure S2).

Data collection and processing

Sampling

In total, 1954 corkwing wrasse were collected from 22 localities in western and middle Norway which represent the primary region that cleaner fish originating from southern Norway and Sweden are translocated to for to delouse salmon on commercial farms (Table 1; Figure 1). As the aim was to cover a wide area and as many locations as possible, an opportunistic sampling scheme was introduced leading to very uneven sample sizes per location (range 1-365) and a time span of six years (from 2013 to 2018). Collection emphasis was focused in mid-Norway (counties of Trøndelag and Møre og Romsdal), which is the primary recipient area of translocated corkwing wrasses, and where the hybridization between local and translocated fish had already been proven. Five hundred fish were collected in three consecutive years (2016-2018) in Flatanger (FLA16-18 in Fig. 1), which roughly represents the species' current northernmost distribution limit. Of those, 105 fish collected in 2016 were already used in Faust et al. (2018), whereas samples from 2017 (N=365) and 2018 (N=30) were collected for the current study. Smøla is an island municipality ~200 kilometres south from Flatanger with a high density of fish farms. In 2017-2018, 271 fish were collected there (SMO 17-18 in Fig. 1) to increase the sampling effort in mid-western Norway. Additional 126 corkwing wrasses from 8 locations from mid-Norway were obtained as bycatch from a research cruise conducted in 2017 (Table 1) and included. Dense sampling in mid-Norway was complemented with 83 fish collected in Sula in 2013 (SUL13 in Fig. 1). A total of 974 fish from southwestern and south-eastern parts of the study region were collected during summer months (June-September) in 2013-2018 (Fig. 1; Table 1). All fish were caught by trained research personnel or professional fishermen using fyke nets and pots, killed upon catch and samples taken immediately. Alternatively,

killed whole fish were stored frozen until sampling in laboratory facilities. From each fish, a fin clip sample (~1 x 1cm) was taken for genetic analysis and stored in absolute ethanol. When possible, biological data (length, weight and sex) were collected.

Genotyping

Genomic DNA was extracted from fin clips using the Qiagen DNeasy Blood & Tissue Kit in 96-well plates following the manufacturer's instructions. A total of 1954 unique individuals and 105 technical replicates were genotyped in four multiplex groups for 106 SNPs. Loci that did not produce reliable clustering patterns were removed (N=17). Loci and individuals with more than 20% missing data were removed from the data leaving 1766 individuals and 85 SNPs. Genotyping robustness was evaluated by calculating concordance between 79 successfully genotyped technical replicates, removing any loci with more than 2 discordant genotypes. One locus showed several discrepancies between genotypes (Supplementary material, Figure S1) and was removed. The final data set consisted of 1766 unique individuals genotyped for 84 loci with a total of 2.9 % missing data.

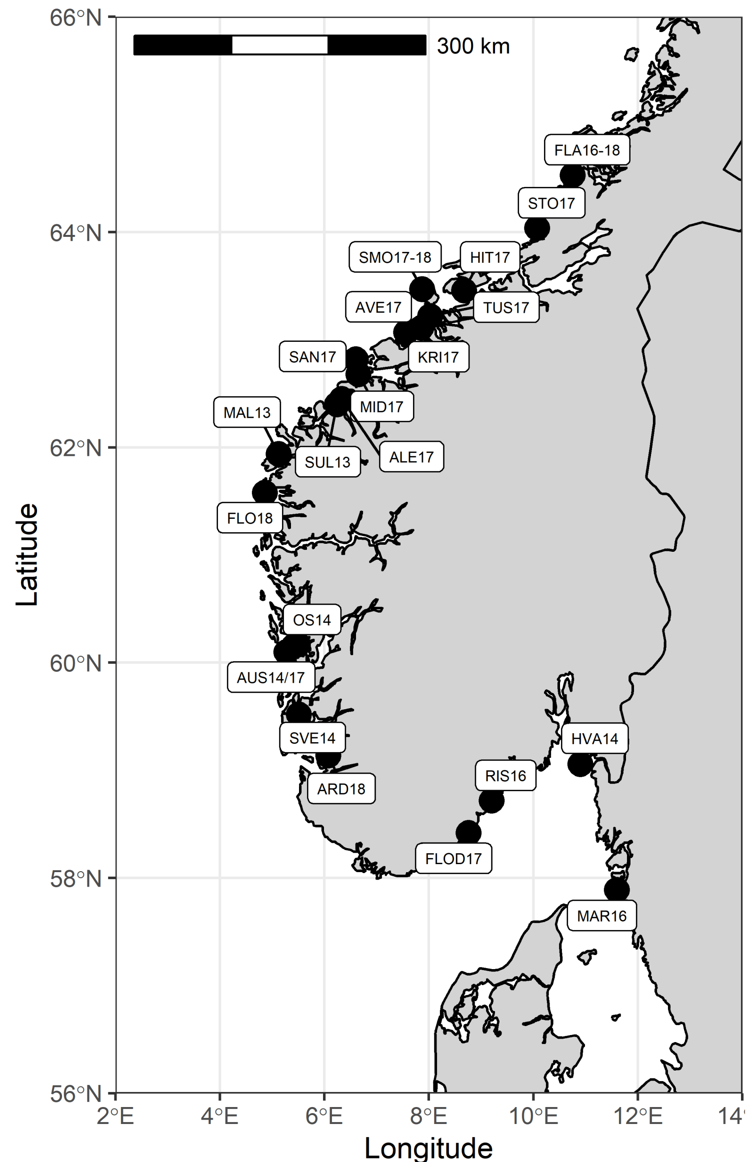


Figure 1. Map of corkwing wrasse sampling locations with respective abbreviations. For details see Table 1.

Statistical analysis

To ease analysing and discussion phases, samples were ordered from north to south along the coastline. From now on, samples are referred to either with their sampling location name, corresponding abbreviation, or according to larger geographic groups. Larger geographic groups are defined as "western" (Norwegian west coast), "southern" (Norwegian south coast and Swedish west coast) or as "mid-western" (>62° N), "south-western" (<62° N, <8°E) and "south-eastern" (<60°N, >8°E) samples (Table 1). Unless otherwise stated, data manipulation and visualisation of results was done using R v3.6.1 (R Core Team, 2019) and Rstudio v1.2.5019 (RStudio Team, 2019), mainly with Tidyverse packages (Wickham et al., 2019).

Table 1. Corkwing wrasse sample information. Samples are arranged from north to south following the Scandinavian coastline.

Sample location	Abbreviation**	County	Geographic group	Collection year	Geographic location* Lat (N)	Lon (E)	Sample size	Number of genotypes
Flatanger	FLA16***	Trøndelag	Mid-Western	2016	64.53	10.75	105	95
Flatanger	FLA17	Trøndelag	Mid-Western	2017	64.53	10.75	365	307
Flatanger	FLA18	Trøndelag	Mid-Western	2018	64.53	10.75	30	30
Hitra	HIT17#	Trøndelag	Mid-Western	2017	63.46	8.67	10	10
Edøya/Smøla	SMO17#	Møre og Romsdal	Mid-Western	2017	63.32	8.22	13	13
Smøla	SMO18	Møre og Romsdal	Mid-Western	2018	63.47	7.87	258	245
Tustna	TUS17#	Møre og Romsdal	Mid-Western	2017	63.22	8.02	3	3
Kristiansund	KRI17#	Møre og Romsdal	Mid-Western	2017	63.11	7.85	44	43
Averøy	AVE17#	Møre og Romsdal	Mid-Western	2017	63.07	7.56	3	3
Sandøy	SAN17#	Møre og Romsdal	Mid-Western	2017	62.82	6.60	3	3
Midsund	MID17#	Møre og Romsdal	Mid-Western	2017	62.68	6.65	22	21
Ålesund	ALE17#	Møre og Romsdal	Mid-Western	2017	62.45	6.33	40	38
Sula	SUL13	Møre og Romsdal	Mid-Western	2013-2014	62.40	6.24	83	77
Måløy	MAL13	Vestland	South-Western	2013-2014	61.94	5.12	5	5
Flora	FLO18	Vestland	South-Western	2018	61.58	4.86	9	9
Os	OS14	Vestland	South-Western	2013-2014	60.17	5.49	156	134
Austevoll	AUS14	Vestland	South-Western	2013-2014	60.10	5.27	108	91
Austevoll	AUS17	Vestland	South-Western	2016-2017	60.10	5.27	249	233
Sveio	SVE14	Vestland	South-Western	2013-2014	59.52	5.51	182	148
Årdalsfjorden	ARD18	Rogaland	South-Western	2018	59.14	6.07	14	10
Flødevigen	FLOD17	Arendal	South-Eastern	2016-2017	58.42	8.76	110	106
Risør	RIS16	Agder	South-Eastern	2016	58.72	9.20	41	41
Hvaler	HVA14	Østfold	South-Eastern	2014	59.06	10.90	60	60
Marstrand	MAR16	Västra Götaland (Sweden)	South-Eastern	2016	57.89	11.60	40	40

* Given geographic location is an approximate midpoint for several sampling locations

** Samples marked with "#" were received as bycatch during research cruise in South Trøndelag and Møre og Romsdal counties.

***These samples are the same ones as used in study by Faust et al. 2018

Genetic diversity and divergence

Observed and expected heterozygosity for each locus across samples was calculated using the R package diveRsity (Keenan et al., 2013). Deviations from expected heterozygosity (He) were assessed by calculating F_{IS} according to Weir & Cockerham (1984). Deviations from expected Hardy–Weinberg proportions (HWE) were estimated with Exact test, and p-values calculated according to the complete enumeration method and adjusted for multiple testing using Bonferroni correction (Louis & Dempster, 1987). Loci that deviated from HW proportions in more than half of the samples were subsequently removed. Weir & Cockerham's pairwise F_{ST} was estimated for each population pair as well as global F_{ST} across all samples. Statistical significance of F_{ST} values was assessed using Fisher's exact probability test with 5000 Monte Carlo replicates, followed by Bonferroni correction. Sample from Stoksund (STO17; see Table 1) was excluded from all genetic diversity and divergence analysis due to the sample size of one individual not being sufficient to make any estimates.

Individual-based clustering and cline models

To estimate and visualise genetic differentiation among individuals we applied two individual-based clustering methods, STRUCTURE v.2.3.4 (Pritchard, Stephens, & Donnelly, 2000) and Principal Component Analysis (PCA) in the R package ade4 (Chessel, Dufour, & Thioulouse, 2004, p. 4; Dray & Dufour, 2007, p. 4; Dray, Dufour, & Chessel, 2007, p. 4). STRUCTURE is a model-based Bayesian clustering method that uses a predefined number of K clusters to estimate the posterior probability of each individual's genotype to originate from each cluster. STRUCTURE analyses were performed for the dataset including all samples using the default admixture model with correlated allele frequencies. To test the performance of different clustering algorithms, simulations were run with and without *a priori* location information (Hubisz, Falush, Stephens, & Pritchard, 2009). A total of 70 000 MCMC (Markov Chain Monte Carlo) repetitions were run and the first 20 000 were discarded as burn-in. K was set from 1 to 6, and the number of iterations was set to 5. To determine the optimal solution for K, the StructureSelector software (Li & Liu, 2018) was utilized. The software summarizes results as the optimal Ln Pr(X|K) given by the STRUCTURE software and the ad hoc summary statistic ΔK by Evanno et al. (2005), which identifies the uppermost level of population hierarchy. Moreover, StructureSelector software produces and visualizes four alternative statistics (MedMed, MedMean, MaxMed and MaxMean) described by Puechmaille (2016). Results from the runs for the different values of K were averaged with CLUMPAK (Kopelman, Mayzel, Jakobsson, Rosenberg, & Mayrose, 2015) using the LargeKGreedy algorithm and 2000 repeats. The second individual-based clustering method (PCA) uses a multivariate exploratory approach that makes no prior assumptions about how many populations exist or boundaries between them. Allele frequencies were centred but not scaled and missing data were replaced by mean allele frequencies with the function scaleGen in ADEGENET (Jombart, 2008; Jombart & Ahmed, 2011).

Cline analysis are used to estimate the shape, centre and width of the sigmoid curves generated by molecular, phenotypic or environmental markers, and to test for concordance

and coincidence in these parameters between markers (Gay et al., 2008). Geographic cline analyses over a 1200 km transect between Flatanger (Norway) and Marstrand (Sweden) were conducted with the R package HZAR (Derryberry et al., 2014). The fifteen models implemented in HZAR were fitted to the allele frequency of every individual locus to determine the position, width and shape of clines over the geographic distance. A reference cline was built using STRUCTURE Q-score for the total dataset and 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 centre did not overlap with the STRUCTURE Q-score ($Qb = 1 - Qs$). Temporal replicates were pooled and sampled populations with small sample size (<10) were removed.

Hybridization

In order to ensure high efficiency, accuracy, and consequently power to detect true escapees and hybrids with a filtered dataset of 84 markers, a second round of simulations was performed. The same procedure was used for both simulation and analysis as described above for the full panel of 106 SNPs. After simulations, the occurrence of escapees and hybridization along the Norwegian coast was investigated with the software NEWHYBRIDS. Analyses were done using the uniform prior option, default genotype proportions, and the burn-in period was set to 50 000 and the number of MCMC sweeps after burn-in to 300 000. Map visualisation was done using the R packages shapefiles (Stabler, 2013) and mapplots (Gerritsen, 2018).

Result

Genotype validation and power estimation

Individual genotyping was evaluated by comparing concordance between technical replicates. A total of 79 individuals were successfully genotyped twice with less than 20% missing data. Genotyping concordance was 93.2% across markers and 93.6% across individuals. Discordant genotypes were few (total of 16 mismatches) and the majority of mismatches were due to missing data in one or both genotypes. Discordant genotypes were present in only two markers, one with 2 discordant genotypes and one with 15. Locus SYMME_00004618_13817, with 15 discordant genotypes was removed from further analysis, which resulted in a final dataset of 84 SNPs. Simulated hybrid data showed that the final panel of 84 SNPs maintained a high accuracy (> 92%), efficiency (> 83%) and power (> 81) to assign all six hybrid classes (pure western, pure southern, F1, F2, western and southern backcross) at probability thresholds between 0.5 and 0.9. Furthermore, when pooling the F1, F2, western and southern backcrosses as a single hybrid class these numbers increased to > 97% accuracy, > 95% efficiency and > 95% in power (Supplementary material, Figure S3).

Genetic diversity

The overall diversity showed a similar pattern to what has been observed in previous studies, with much lower diversity south of the genetic break (Supplementary material, Table S2). The mean observed heterozygosity ranged between 0.184-0.187 in south-eastern samples and

0.315-0.413 in all western samples ($p = 0.002$). Similarly, allelic richness was significantly lower ($p = 0.002$) in south-eastern samples (1.42-1.43) compared to the western samples (1.78-1.90). Differences in these diversity indices were statistically significant also when both western samples were compared separately with south-eastern samples ($p < 0.05$ in all cases). Moreover, mean allelic richness was higher ($p = 0.015$) in south-western samples ($A_R = 1.83$) than in mid-western samples ($A_R = 1.80$), but no difference was observed for mean observed heterozygosity ($p = 0.284$). The majority of markers showed no deviation from HWE in any of the sampled locations and only two markers deviated significantly from HWE in more than six locations. Initial comparisons showed little to no difference in results when removing these two markers, and consequently, all markers were kept for further analysis. Overall, nine out of all sample populations deviated significantly from HWE. Eight of which showed heterozygosity deficiency (F_{IS} 0.017-0.06) and one heterozygosity excess (F_{IS} -0.012); all were observed in western Norway.

Pairwise F_{ST} estimates between sampled populations showed an overall lower genetic differentiation within each of the three geographic groups than between them (Supplementary material, Table S3). Within group differentiation was lowest in south-eastern samples (mean F_{ST} of 0.0005 ± 0.0011), followed by the mid-western samples ($F_{ST} = 0.0054 \pm 0.0052$) and highest in south-western samples ($F_{ST} = 0.0120 \pm 0.0146$). When comparing divergence within and between the three geographic areas, the genetic differentiation within the western samples were order of magnitude lower (mean-$F_{ST} = 0.0216 \pm 0.0119$) than between the western and the south-eastern samples ($F_{ST\ (mid\text{-}west_vs_south\text{-}east)} = 0.5155 \pm 0.0699$ and $F_{ST\ (south\text{-}west_vs_south\text{-}east)} = 0.4757 \pm 0.0549$). Of the western samples, Flatanger17 showed clearly lower differentiation toward the south-eastern samples (mean-$F_{ST} = 0.3704 \pm 0.0089$) with all other pairwise comparisons ranging between 0.4106 - 0.6070.

Population structure and individual assignment

In concordance with pairwise F_{ST} measurements, individual-based clustering using STRUCTURE differentiated the south-eastern cluster (pink) from the western samples (blue) (K=2 in Fig. 2). K=2 was clearly supported as the highest level of population hierarchy by the Evanno method (Supplementary material, Figure S4b). Support for additional substructure was also evident: Adding one additional cluster (i.e. K=3) splits western samples into two distinct clusters between Sula and Måløy implying an additional genetic break (green and blue in Fig. 2; note that these clusters correspond to our Mid-Western and South-Western geographic groups, Table 1). Sampling location given as *a priori* clearly increased resolution power between the two western groups on an individual level for K=3 (Fig. 2; Supplementary material, Figure S5a), but had little to no effect on the estimated admixture proportions with K=2. Despite STRUCTURE gave clear clustering solutions with these two levels (K=2 and 3) of population division, additional methods that were utilized favoured solutions for even higher levels of Ks (4-5; Supplementary material, Figure S4b and S4a). However, visual inspection of the corresponding bar plots (Figures S5a-b) show that instead of creating new (vertical) separations between those well-supported groupings of two or three, these clusters would merely build up additional layer(s) of difference, and are thus likely technical artefacts depending on the model assumptions.

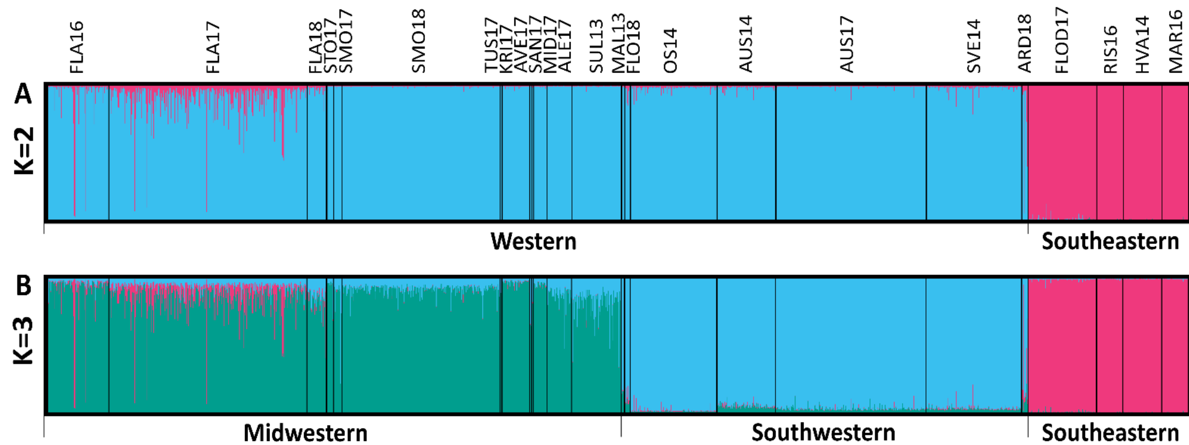


Figure 2. STRUCTURE cluster assignment of 1766 corkwing wrasse individuals based on 84 SNPs for K = 2 (A) and 3 (B) with sampling location given as a priori. Each vertical bar represents one individual and the colour the proportion of that individual assigned to the different genetic clusters. Individuals are sorted from North (left) to South (right).

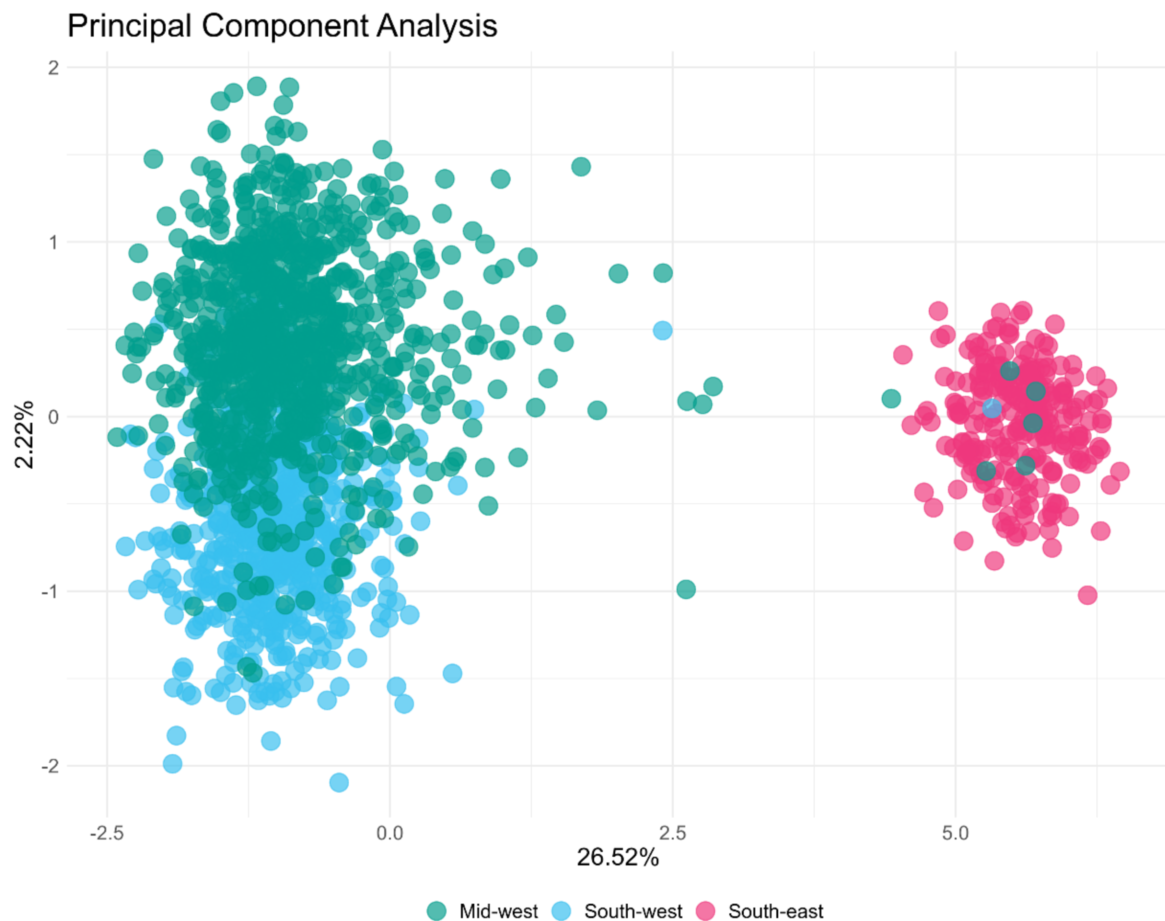


Figure 3. First (x-axis) and second (y-axis) component of a principal component analysis (PCA) on 1766 corkwing wrasse individuals based on 84 SNPs. The first component explains 26.5% of the total variation and the second 2.2%. Each point represents one individual and colours represent the three geographic regions.

Assignment of individuals into genetic clusters with K=2 was straightforward. When investigating individual membership coefficients (q), the vast majority of all fish (94.6%) had a q of 0.95 or higher corresponding to their “own” geographic group (western or south-eastern). However, six individuals from Flatanger and one from Årdalsfjorden in western Norway were assigned with a high proportion (q > 0.98) to the south-eastern cluster. Moreover, nine individuals had roughly equally admixed genotypes (q=0.4-0.6 to both clusters), 38 had moderate representation (q=0.2-0.4) of the south-eastern cluster in their genomes, and 40 rather low but still notable portions (q=0.1-0.2), all suggesting varying degree of admixture between the clusters. When fish were assigned into three clusters (K=3) instead of two, they were still highly concordant with their geographic origin: Mid-western individuals had a mean assignment of 0.899 (±0.058), south-western 0.857 (±0.105) and south-eastern 0.997 (±0.004).

To estimate and visualize genetic differentiation among individuals without prior assumptions about the population model, we conducted a Principal Coordinate Analysis (PCA). The PCA demonstrated a similar pattern as seen in the STRUCTURE cluster analysis (Fig. 3, Supplementary material, Figure S6). The first axis (x-axis, accounting for 26.5% of the total variation) clearly separates south-eastern samples (pink) from western samples (blue and green). The second axis (y-axis, explaining 2.2% of the variation) separates the mid-western samples (green) from south-western samples (blue), but with a degree of overlap between the clusters. The seven individuals previously identified in the STRUCTURE analysis clearly cluster together with individuals from the south-eastern cluster also in the PCA. Individuals previously identified as possible admixed in STRUCTURE analysis are also in the PCA located between the western and south-eastern clusters.

The reference cline based on the STRUCTURE Q-score fitted an optN model, with the centre situated at 799 km (787-1087) (Supplementary material, Figure S7a). All the 84 loci fitted cline models with centres ranging between 706 and 1062 km (Supplementary Table S4) and none of them was significantly displaced from the STRUCTURE reference cline (Supplementary material, Figure S7b). This means that all loci showed a similar pattern of divergence. The cline centre is located close to the habitat break on the southwest tip of Norway.

Hybridization

Samples were screened for potential hybrids using the software NEWHYBRIDS which estimates each individual's probability of belonging to predefined classes (pure western, pure south-eastern, F1, F2, western backcross and south-eastern backcross). Of the 1766 individuals analysed, all of them could be assigned with a probability > 50% to be either pure western (blue) or pure southern (pink) or hybrid (green) (Figure 4a). When distinguishing between the different hybrid classes (F1, F2, backcross 1 and backcross 2), all but one individual could be assigned with a probability > 50% (Figure 4b and Supplementary material Figure S8a). When increasing the probability threshold to > 80 %, 1715 individuals could still be assigned to the different hybrid classes. Among the western samples, seven individuals had a very high probability (> 90%) to be of pure-eastern origin, six in Flatanger and one in Årdalsfjorden. The majority of all potential hybrids could also be found in

Flatanger where 70 individuals had more than a 50% probability to be F1, F2, western backcross or south-eastern backcrosses. In all other western samples only nine individuals could be identified as potential hybrids, and all of them as western backcrosses.

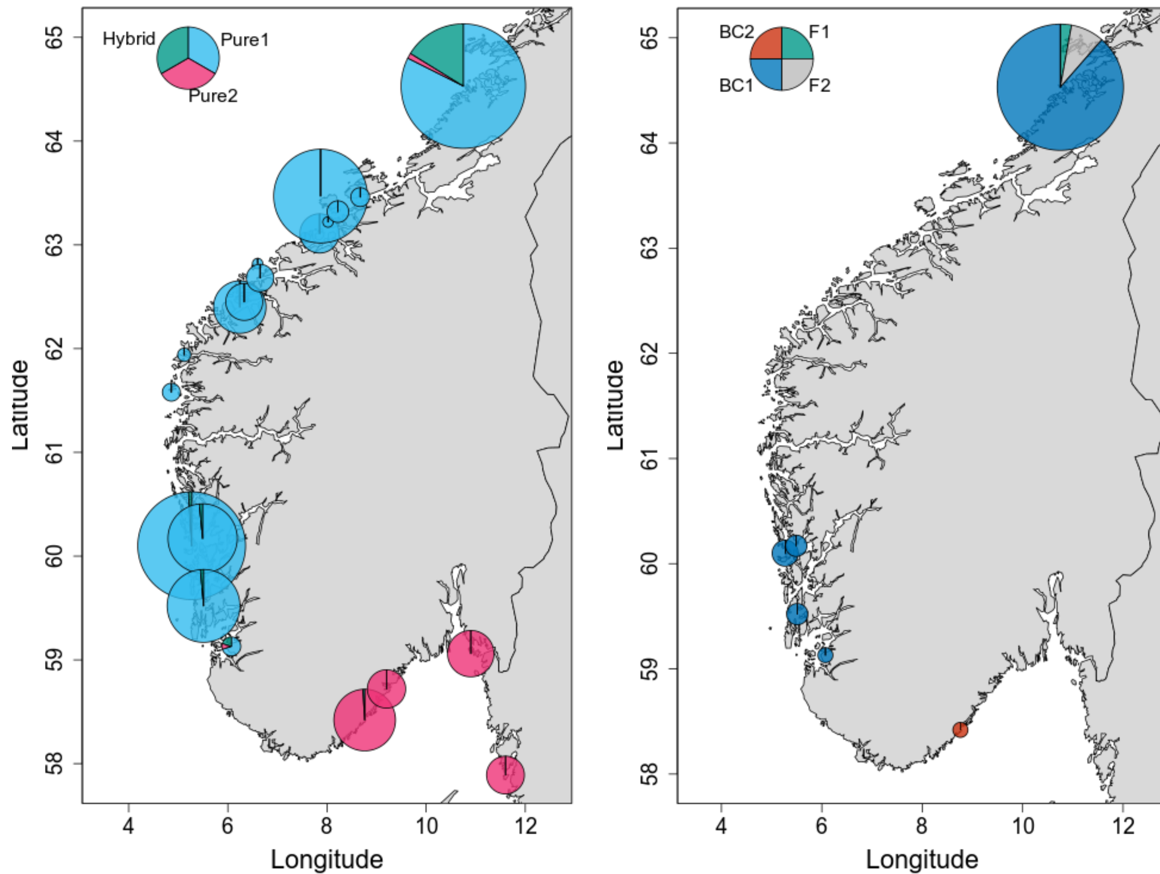


Figure 4. Proportion of individuals from each sampling site classified by Newhybrid analysis. A) Left map displays individuals classified as pure1 = western genotype, pure2 = south-eastern genotype or hybrid with a > 50% probability. B) Right map displays the proportion of hybrids assigned to the different hybrid classes F1, F2, backcross with pure1 and backcross with pure2. Sizes reflect the relative number of samples.

Discussion

In this study, we developed and implemented a panel of diagnostic SNPs to quantify the proportion of escaped and hybridised corkwing wrasse with a southern origin in middle and western Norway, where translocated cleaner fish are used for parasite control. The panel of 84 SNPs, which can detect escapees and hybrids with a power > 0.95, identified a total of 7 escapees and 79 potential hybrids and back-crossed individuals on the Norwegian west coast. Most of these were identified in samples from the northern part of the species distribution in mid-west Norway, which also represents the main area of import from southern latitudes.

Genetic differentiation among wrasse populations

The panel of diagnostic SNPs were developed to identify genetic differences between source populations in southern Norway and Sweden versus local wild wrasse populations in receiving areas in western Norway. In concordance with previous studies (Blanco González et al., 2016; Faust et al., 2018; Mattingsdal et al., 2020), we found strong genetic divergence between corkwing wrasse on the west coast and south coast of Norway. All SNPs showed a similar pattern of divergence with a cline centre located close to the habitat break on the south-western tip of Norway (Figure 2, Supplementary material Figure S7).

In addition to the known genetic break, the STRUCTURE clustering analysis indicated the existence of a second break located 62° N on the Norwegian west coast. Blanco González et al. (2016) described a pattern of moderate isolation by distance along the west coast of Norway and found their northern most samples (Vestnes 62.65 ° and Smøla N 63.32° N) to be distinct from south-western samples. However, few studies compare samples north of 60.2° N and none has included sampled the areas between 60° N and 62.4° N. Despite that relatively few individuals (N=14) were available from this region in this study, all of them were clearly clustered within the south-western group, indicating that there could be a stronger genetic discontinuity than previously suggested. However, given that the markers were chosen to be able to distinguish south-eastern samples from western samples, they might not be ideal for genetic population structure inference in this specific region. It is therefore not possible to disentangle the nature of this break, i.e. the degree of divergence or whether selection or neutral processes are at play.

Extent of escapees and hybridization

The hybrid analysis identified a total of 7 individuals as potential escapees and 79 as potential hybrids on the Norwegian west coast. The majority of these individuals were caught in Flatanger in Trøndelag (6 potential escapees and 70 hybrids) in the northern part of the species distribution. The only other pure south-eastern individual was found in Årdalsfjorden less than 60 km from the sandy beaches in Jæren and the genetic break (Blanco González et al. 2016). Out of the 10 individuals successfully genotyped in Årdalsfjorden, one was of south-eastern origin and two were hybrids. In all other south-western samples, we found no more than one or two potential hybrids. However, given the proximity to the genetic break, it is not possible to say whether these individuals are the direct result of escapees or natural gene flow across the break (Mattingdal et al. 2020). Besides in the Flatanger area, we did not detect any potential escapees or hybrids in other parts of the Trøndelag county or its neighbour county Møre og Romsdal, despite the relatively large number of fish sampled. Below we discuss possible causes for why we only see few escapees and hybrids in aquaculture dense areas other than Flatanger.

The lack of escapees or hybrids reported in Møre og Romsdal compared to Trøndelag could be explained by a combination of different factors: 1.) Corkwing wrasse only expanded into Trøndelag recently, population size is small and thus escapees and hybrids are easier to detect, 2.) Smaller populations make it easier for escapees to establish due to less competition (Rhymer & Simberloff, 1996), 3.) There is less import from the south-eastern population to

Møre og Romsdal and/or less individuals are escaping. The abundance of corkwing in mid-Norway (i.e. Trøndelag and Møre og Romsdal counties) has only recently increased, suggestively indicated by the catch-per-unit effort (CPUE) data from fishermen in this region (Figure 5). In Smøla, the catch rates increased in 2015 and have levelled out after 2017. The population in Flatanger appears to still be in an early phase of establishment and was virtually absent from catches until 2018.

Corkwing wrasse is a territorial species where nest densities are not dependant of availability but are determined by the territorial behaviour of nesting males (Potts, 1985). In denser populations, the aggressive territorial behaviour can result in fewer successful courtships and mating encounters (Myhre, Forsgren, & Amundsen, 2013). A recent study investigated hybridization potential between western and southern corkwing in an experimental mesocosm setting (Blanco González et al., 2019). The authors found that individuals of western origin had a significantly larger contribution of breeders compared to individuals of southern origin, and suggested a potential fitness advantage in western populations. However, individuals were moved in the opposite direction of common cleaner fish translocation, with western individuals being introduced to a southern environment, rather than southern individuals to a western environment. If western individuals have indeed higher fitness in general, this could effectively prevent gene flow. Even weak negative selection against translocated genotypes would be effective in reducing the frequency of escapees and hybrids in a large population, such as in Smøla. However, in a smaller population as in Flatanger, selection would be less effective and survival and reproduction success would be more dependent on chance and genetic drift (Allendorf, Luikart, & Aitken, 2013; Bridle & Vines, 2007).

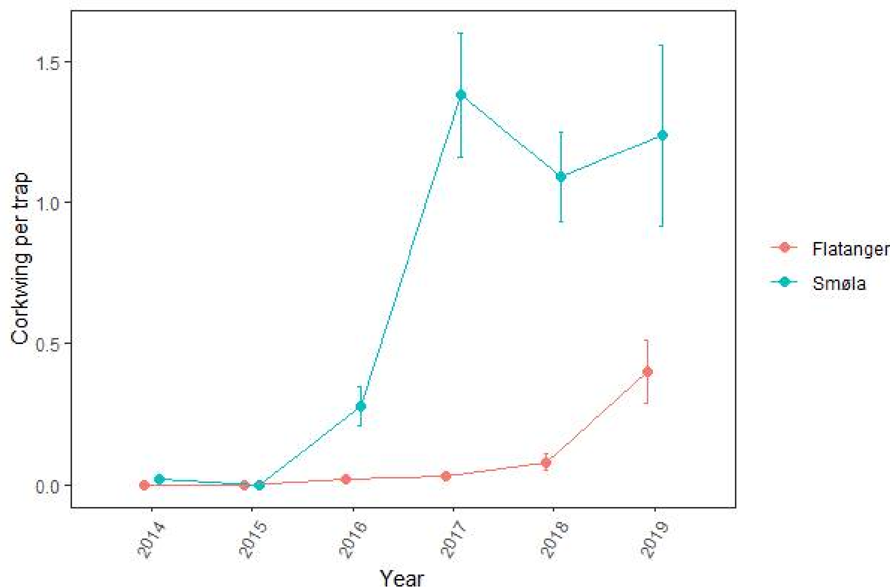


Figure 5. Development in raw catch-per-unit effort (CPUE) for corkwing caught in commercial trap fishery (one fisher per location). CPUE is calculated as the total N corkwing caught, divided on the total number of traps sampled in each year. Error bars show ±SE of the mean.

Overall, more cleaner fish are used in Trøndelag than in Møre og Romsdal, but species-specific data shows that this is not always the case for corkwing wrasse. Species segregated data of cleaner fish use has only been collected since 2015 and it is not possible to know how much corkwing was used prior to this. However, in two of the four years for which data is available (2016 and 2017), almost twice the amount of corkwing wrasse were used in Møre og Romsdal compared to Trøndelag (Norwegian directorate of Fisheries, 2019). Given the higher densities as indicated by the higher CPUE in Møre og Romsdal, it is likely that more fish is sourced locally than in Trøndelag. This is also corroborated by 2017 and 2018 import records from Sweden. Since reporting started in 2017, more than 3 times as many corkwing wrasse have been transported to Trøndelag compared to Møre og Romsdal (data on imported wrasses from Sweden 2017–2018 was provided by the Norwegian Environmental Agency).

During 2017 and 2018, an average of less than 0.4 million corkwing wrasse were imported from Sweden per year. During the same years, an average of 7 million wild corkwing wrasse were used in Norwegian aquaculture. Thus, Swedish imports constitute less than 6% of corkwing wrasse used as cleaner fish in Norwegian aquaculture. However, source and destination of corkwing wrasse caught in Norway is not reported. This makes it difficult to estimate how much of the corkwing wrasse used in commercial salmon farming originates from the southern coast of Norway, as opposed to local sources. Catch numbers suggest that on average 20% of wild-caught cleaner fish are caught of the southern coast of Norway annually, but most years less than 1% of all cleaner fish is being deployed in the region (Norwegian directorate of Fisheries, 2019). Given the current lack or reporting it is not possible to estimate where southern corkwing are transported to. The lack of reporting also complicates potential estimation of the number of escapees. Although all Norwegian fish farms are obligated to report escaping fish, currently this is only applied to the target species being farmed.

Implications

The effects of hybridization between genetically distinct populations are hard to predict and depend on many factors. First, the prevalence, that is the number of escapees vs the local population size, will be important. Direct escapees can cause ecological effects and transmit diseases and pathogens. If hybridization occurs, genetic effects can also be anticipated. Several escapees and backcrossed individuals were identified in the northernmost localities. In addition, the Structure analysis indicates that in Flatanger, a majority of the investigated individuals show admixture (Fig. 2). This means that a notable fraction of the population gene pool has a southern origin. In contrast, we did not detect such introgression in e.g. Smøla despite frequent and abundant translocation of fish from south to this region. Although we do not know if admixture creates consistent pattern across the genome, our results indicate a clear alteration of the genetic composition of the Flatanger population.

The ecological consequences of hybridization in the northern edge population are unknown but given the considerable difference in important abiotic factors between this region and southern Norway and Sweden, inadvertently translocated individuals are likely to be maladapted and have lower fitness in the recipient populations. For example, the onset of the reproduction is affected by photo-period and temperature (Stone 1996), which implies the

possibility that hybrids might initiate spawning at an unfavourable time-of-year, resulting in reduced survival of offspring. Furthermore, genetic differences may include life history, physiological and morphological traits that negatively affect fitness, thus reducing the overall population viability, as well as the capacity to naturally expand further north as the environment changes. Future work in this direction should assess phenotypic differences between individuals with native and southern origin, and ideally do field studies comparing fitness between these groups (e.g. tagging experiments, field observations of reproduction) and/or controlled common garden experiments to assess differences in phenotypic plasticity and physiology. Such studies have unequivocally demonstrated lower fitness of domesticated Atlantic salmon offspring in wild populations (Skaala et al. 2012 and 2019).

The recently established Flatanger population is on the leading edge of the current species range, and is thus likely to carry favorable genetic material also for future range expansion northwards (Gibson, Marel, & Starzomski, 2009). However, the ongoing asymmetric gene flow from southern genotypes could obstruct further adaptation and range expansion (Kirkpatrick & Barton, 1997). Investigating if local adaptation of the admixed populations in the northern part of the species distribution is affected would require experimental studies. However, given predicted climate-changes of warmer sea temperatures, populations at the northern edge of species distributions should be prioritised. These are likely the populations with the best adaptive potential to lead the species range expansion in a future environment of global warming. We argue that any evaluation of the risk of translocation should not only include wrasse imported from Sweden but also the existing knowledge of genetically distinct populations within Norway. An obstacle for effective management is that the current practice of cleaner fish use is poorly documented and regulated. Although all Norwegian fish farms are obligated to report escaping fish, currently only the target species being farmed (i.e. Salmonids) are recorded. Moreover, the lack of knowledge regarding the source and destination of cleaner fish transported within Norway is a big obstacle to assess and address the challenge of escapees.

Conclusion

We have developed a SNP panel with high power to detect corkwing wrasse translocated from Skagerrak-Kattegat to the Norwegian west coast as well as first and second generation hybrids. Using these markers, we found that the geographical extent of escapees and potential hybrids is largely limited to areas at the northern edge of the species distribution where the number of escapees and potential hybrids may constitute up to 20% of the population. These results provide an important knowledge, a baseline of the geographical extent and magnitude of hybridization, and a tool for management and monitoring of the future use of corkwing wrasse as a cleaner fish for parasite control. Moving genetic material between distant populations could drastically alter the genetic composition, erode population structure and potentially result in loss of local adaptation, hampering the species expansion. The translocation and number of escaping cleaner fish is today poorly documented and regulated. The ecological consequences of escapees and hybrids remain unknown for this and other wrasse species.

Acknowledgements

We thank Lars Måvik, Nils Marius Holm, Eva Farestveit, Lisbeth Sælemyr, Torkel Larsen, and Gunnar Didriksen for their great contribution in the field. We acknowledge Fredrik Staven and Per Andersen for their assistance in sampling. We also thank Kjell Nedreaas, Askeir Aglen and Håkon Otterå who kindly provided samples from research cruises; and Reidun Bjelland and Anne Berit Skiftesvik for contributing old samples. Håkan Berg-Rolness and Dawit Berhe Ghebretnsae helped us to take samples from frozen fish, for which we want to express our gratitude. Gaute Seljestad, Laila Unneland, and Sofie Knutar are thanked for their assistance in the laboratory. Pierre de Wit gave valuable advice on the 2bRAD method and Mats Töpel on SNP data storage and analysis. Bioinformatics analyses for SNP selection were run on the Albiorix computer cluster (http://albiorix.bioenv.gu.se/) at the Department of Marine Sciences, University of Gothenburg. This study was funded by the Norwegian Ministry for Trade, Industry and Fisheries, by the Swedish Research Council FORMAS, by the EU Interreg projects MarGen and MarGenII, and by the Linnaeus Centre for Marine Evolutionary Biology (CeMEB).

References

Allendorf, F. W., England, P. R., Luikart, G., Ritchie, P. A., & Ryman, N. (2008). Genetic effects of harvest on wild animal populations. *Trends in Ecology & Evolution*, *23*(6), 327–337. https://doi.org/10.1016/j.tree.2008.02.008

Allendorf, F. W., Luikar, G., & Aitken, S. N. (2013). *Conservation and the genetics of populations* (2nd ed.). Wiley-Blackwell.

Anderson, E. C., & Thompson, E. a. (2002). A model-based method for identifying species hybrids using multilocus data. *Genetics*, *160*(3), 1217–1229.

Atalah, J., & Sanchez-Jerez, P. (2020). Global assessment of ecological risks associated with farmed fish escapes. *Global Ecology and Conservation*, *21*, e00842. https://doi.org/10.1016/j.gecco.2019.e00842

Besnier, F., Kent, M., Skern-Mauritzen, R., Lien, S., Malde, K., Edvardsen, R. B., … Glover, K. A. (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*(1), 937. https://doi.org/10.1186/1471-2164-15-937

Bjordal, Å. (1988). Cleaning symbiosis between wrasse (Labridae) and lice infested salmon (*Salmo salar*) in mariculture. *International Council for the Exploration of the Sea*, *1988*(F:17), 8.

Blakeslee, A. M. H., Manousaki, T., Vasileiadou, K., & Tepolt, C. K. (2020). An evolutionary perspective on marine invasions. *Evolutionary Applications*, *13*(3), 479–485. https://doi.org/10.1111/eva.12906

Blanco Gonzalez, E., Espeland, S. H., Jentoft, S., Hansen, M. M., Robalo, J. I., Stenseth, N. C., & Jorde, P. E. (2019). Interbreeding between local and translocated populations of a cleaner fish in an experimental mesocosm predicts risk of disrupted local adaptation. *Ecology and Evolution*, *0*(0), 1–13. https://doi.org/10.1002/ece3.5246

Blanco González, E., Knutsen, H., & Jorde, P. E. (2016). Habitat discontinuities separate genetically divergent populations of a rocky shore marine fish. *PLoS ONE*, *11*(10). https://doi.org/10.1371/journal.pone.0163052

Bolstad, G. H., Hindar, K., Robertsen, G., Jonsson, B., Sægrov, H., Diserud, O. H., … Karlsson, S. (2017). Gene flow from domesticated escapes alters the life history of wild Atlantic salmon. *Nature Ecology & Evolution*, *1*(5), 0124. https://doi.org/10.1038/s41559-017-0124

Bridle, J.R., & Vines, T.H. (2007) Limits to evolution at range margins: when and why does adaptation fail? *Trends in Ecology and Evolution, 22*(3), 140–147.

Chessel, D., Dufour, A. B., & Thioulouse, J. (2004). The ade4 package - I: One-table methods. *R News*, *4*(1), 5–10. https://doi.org/10.2307/3780087

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., … 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

D'Arcy, J., Mirimin, L., & FitzGerald, R. (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*(3), 685–693. https://doi.org/10.1093/icesjms/fst018

Darwall, W. R. T., Costello, M. J., Donnelly, R., & Lysaght, S. (1992). Implications of life-history strategies for a new wrasse fishery. *Journal of Fish Biology*, *41*(supplement B), 111–123. https://doi.org/10.1111/j.1095-8649.1992.tb03873.x

Dray, S., & Dufour, A. B. (2007). The ade4 Package: Implementing the Duality Diagram for Ecologists. *Journal of Statistical Software*, *22*(4), 1–20. https://doi.org/10.1.1.177.8850

Dray, S., Dufour, A. B., & Chessel, D. (2007). The ade4 package-II: Two-table and K-table methods. *R News*, *7*(2), 47–52. https://doi.org/10.1159/000323281

Evangelista, C., Cucherousset, J., & Lecerf, A. (2019). Contrasting ecological impacts of geographically close invasive populations. *Oecologia*, *189*(2), 529–536. https://doi.org/10.1007/s00442-018-04333-5

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology*, *14*(8), 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Faust, E., Halvorsen, K. T., Andersen, P., Knutsen, H., & André, C. (2018). Cleaner fish escape salmon farms and hybridize with local wrasse populations. *Royal Society Open Science*, *5*(3), 171752. https://doi.org/10.1098/rsos.171752

Gabriel, S., Ziaugra, L., & Tabbaa, D. (2009). SNP Genotyping Using the Sequenom MassARRAY iPLEX Platform. *Current Protocols in Human Genetics*, *60*(1), 2.12.1-2.12.18. https://doi.org/10.1002/0471142905.hg0212s60

Gerritsen, H. (2018). *mapplots: Data Visualisation on Maps*. Retrieved from https://CRAN.R-project.org/package=mapplots

Gibson, S. Y., Marel, R. C. V. D., & Starzomski, B. M. (2009). Climate Change and Conservation of Leading-Edge Peripheral Populations. *Conservation Biology*, *23*(6), 1369–1373. https://doi.org/10.1111/j.1523-1739.2009.01375.x

Glover, K. A., Solberg, M. F., Mcginnity, P., Hindar, K., Verspoor, E., Coulson, M. W., … Svåsand, T. (2017). Half a century of genetic interaction between farmed and wild Atlantic salmon: Status of knowledge and unanswered questions. *Fish and Fisheries*, (June 2016), 890–927. https://doi.org/10.1111/faf.12214

Glover, K.A., Quintela, M., Wennevik, V., Besnier, F., Sørvik, AG.E., Skaala, O. (2012). Three decades of farmed escapees in the wild: A spatio-temporal analysis of population genetic structure throughout Norway. *PLOS One,* 7(8), e:43129. https://doi.org/10.1371/journal.pone.0043129

Gosselin, T. (2019). *radiator: RADseq Data Exploration, Manipulation and Visualization using R*. https://doi.org/10.5281/zenodo.1475182

Halvorsen, K. T., Larsen, T., Sørdalen, T. K., Vøllestad, L. A., Knutsen, H., & Olsen, E. M. (2017). Impact of harvesting cleaner fish for salmonid aquaculture assessed from replicated coastal marine protected areas. *Marine Biology Research*, 1–11. https://doi.org/10.1080/17451000.2016.1262042

Halvorsen, K. T., Sørdalen, T. K., Durif, C., Knutsen, H., Olsen, E. M., Skiftesvik, A. B., … Vøllestad, L. A. (2016). Male-biased sexual size dimorphism in the nest building corkwing wrasse *Symphodus melops*: implications for a size regulated fishery. *ICES*

Journal of Marine Science, *73*(10), 2586–2594. https://doi.org/10.1093/icesjms/fsw135

Halvorsen, K. T., Sørdalen, T. K., Vøllestad, L. A., Skiftesvik, A. B., Espeland, S. H., & Olsen, E. M. (2017). Sex- and size-selective harvesting of corkwing wrasse (*Symphodus melops*)—a cleaner fish used in salmonid aquaculture. *ICES Journal of Marine Science*, *74*(3), 660–669. https://doi.org/10.1093/icesjms/fsw221

Hubisz, M. J., Falush, D., Stephens, M., & Pritchard, J. K. (2009). Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, *9*(5), 1322–1332. https://doi.org/10.1111/j.1755-0998.2009.02591.x

Imsland, A. K., Reynolds, P., Eliassen, G., Hangstad, T. A., Foss, A., Vikingstad, E., & Elvegård, T. A. (2014). The use of lumpfish (*Cyclopterus lumpus* L.) to control sea lice (*Lepeophtheirus salmonis* Krøyer) infestations in intensively farmed Atlantic salmon (*Salmo salar* L.). *Aquaculture*, *424–425*, 18–23. https://doi.org/10.1016/j.aquaculture.2013.12.033

Jansson, E., Quintela, M., Dahle, G., Albretsen, J., Knutsen, H., André, C., … Glover, K. A. (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*(8), 2135–2147. https://doi.org/10.1093/icesjms/fsx046

Jombart, T. (2008). Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, *27*(21), 3070–3071. https://doi.org/10.1093/bioinformatics/btr521

Kaur, K., Besnier, F., Glover, K.A., Nilsen, F., Aspehaug, V.T., Fjortoft, H.B., Horsberg, T.E. (2017) The mechanism (Phe362Tyr mutation) behind resistance in *Lepeophtheirus salmonis* pre-dates organophosphate use in salmon farming. *Scientific reports*, *7*(1), 12349. https://doi.org/10.1038/s41598-017-12384-6

Keenan, K., McGinnity, P., Cross, T. F., Crozier, W. W., & Prodöhl, P. A. (2013). diveRsity: An R package for the estimation of population genetics parameters and their associated errors. *Methods in Ecology and Evolution*, *4*(8), 782–788. https://doi.org/10.1111/2041-210X.12067

Kirkpatrick, M., & Barton, N. H. (1997). Evolution of a species' range. *American Naturalist*, *150*(1), 1–23. https://doi.org/10.1086/286054

Knaus, B. J., & Grünwald, N. J. (2016). VcfR: an R package to manipulate and visualize VCF format data. *BioRxiv*. Retrieved from http://dx.doi.org/10.1101/041277

Knaus, B. J., & Grünwald, N. J. (2017). VCFR: a package to manipulate and visualize variant call format data in R. *Molecular Ecology Resources*, *17*(1), 44–53.

Knutsen, H., Jorde, P. E., Gonzalez, E. B., Robalo, J., Albretsen, J., & Almada, V. (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*(6), e67492. https://doi.org/10.1371/journal.pone.0067492

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). CLUMPAK: a program for identifying clustering modes and packaging population

structure inferences across K. *Molecular Ecology Resources*, *15*(5), 1179–1191. https://doi.org/10.1111/1755-0998.12387

Laikre, L., Schwartz, M. K., Waples, R. S., Ryman, N., & Group, T. G. W. (2010). Compromising genetic diversity in the wild: Unmonitored large-scale release of plants and animals. *Trends in Ecology and Evolution*, *25*(9), 520–529. https://doi.org/10.1016/j.tree.2010.06.013

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Li, Y.-L., & Liu, J.-X. (2018). StructureSelector: A web-based software to select and visualize the optimal number of clusters using multiple methods. *Molecular Ecology Resources*, *18*(1), 176–177.

Liu, Y., & Bjelland, H. vanhauwaer. (2014). Estimating costs of sea lice control strategy in Norway. *Preventive Veterinary Medicine*, *117*(3), 469–477. https://doi.org/10.1016/j.prevetmed.2014.08.018

Louis, E. J., & Dempster, E. R. (1987). An Exact Test for Hardy-Weinberg and Multiple Alleles. *International Biometric Society*, *43*(4), 805–811.

Mattingsdal, M. (2017, November 10). Symphodus_melops.fasta. https://doi.org/10.6084/m9.figshare.5590003.v1

Mattingsdal, M., Jorde, P. E., Knutsen, H., Jentoft, S., Stenseth, N. C., Sodeland, M., … Gonzalez, E. B. (2020). Demographic history has shaped the strongly differentiated corkwing wrasse populations in Northern Europe. *Molecular Ecology*, *29*(1), 160–171. https://doi.org/10.1111/mec.15310

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., … DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303. https://doi.org/10.1101/gr.107524.110

Myhre, L. C., Forsgren, E., & Amundsen, T. (2013). Effects of habitat complexity on mating behavior and mating success in a marine fish. *Behavioral Ecology*, *24*(2), 553–563. https://doi.org/10.1093/beheco/ars197

Norwegian directorate of Fisheries. (2019). Utsett av rensefisk 1998-2018. Retrieved 22 April 2020, from Norwegian directorate of Fisheries website: https://www.fiskeridir.no/Akvakultur/Tall-og-analyse/Akvakulturstatistikk-tidsserier/Rensefisk

Potts, G. W. (1985). The Nest Structure of the Corkwing Wrasse, Crenilabrus Melops (Labridae: Teleostei). *Journal of the Marine Biological Association of the United Kingdom*, *65*(02), 531–546. https://doi.org/10.1017/S002531540005058X

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945–959. https://doi.org/10.1111/j.1471-8286.2007.01758.x

Puechmaille, S. J. (2016). The program structure does not reliably recover the correct population structure when sampling is uneven: subsampling and new estimators alleviate the problem. *Molecular Ecology Resources*, *16*(3), 608–627.

R Core Team. (2019). *R: A Language and Environment for Statistical Computing*. Retrieved from https://www.R-project.org/

Rhymer, J. M., & Simberloff, D. (1996). Extinction by Hybridization and Introgression. *Annual Review of Ecology and Systematics*, *27*(1), 83–109. https://doi.org/10.1146/annurev.ecolsys.27.1.83

Robalo, J. I., Castilho, R., Francisco, S. M., Almada, F., Knutsen, H., Jorde, P. E., … Almada, V. (2012). Northern refugia and recent expansion in the North Sea: the case of the wrasse *Symphodus melops* (Linnaeus, 1758). *Ecology and Evolution*, *2*(1), 153–164. https://doi.org/10.1002/ece3.77

RStudio Team. (2019). *RStudio: Integrated Development Environment for R*. Retrieved from http://www.rstudio.com/

Seljestad, G. W., Quintela, M., Faust, E., Halvorsen, K. T., Besnier, F., Jansson, E., … Glover, K. A. (Accepted to be published in Ecology and Evolution). "A cleaner-break": Genetic divergence between geographic groups and sympatric phenotypes revealed in ballan wrasse *(Labrus bergylta)* (p. 66).

Skaala, Ø., Besnier, F., Borgstrom, R., Barlaup, B., Sorvik, AG., Normann, E., Ostebo, B.I., Hansen, M.M., Glover, K.A. (2019) An extensive common-garden study with domesticated and wild Atlantic salmon in the wild reveals impact on smolt production and shifts in fitness traits. *Evolutionary Applications*, 12(5), 1001–1016. https://doi.org/10.1111/eva.12777

Skaala, Ø., Glover, K.A., Barlaup, B.T., Svåsand, T., Besnier, F., Hansen, M.M., Borgstrøm, R. (2012) Performance of farmed, hybrid, and wild Atlantic salmon (*Salmo salar*) families in a natural river environment. *Canadian Journal of Fisheries and Aquatic Sciences*, 69(12), 1994–2006. https://doi.org/10.1139/f2012-118

Skiftesvik, A. B., Durif, C. M. F., Bjelland, R. M., & Browman, H. I. (2014). *Distribution and habitat preferences of five species of wrasse (Family Labridae) in a Norwegian fjord*. *72*(October), 890–899.

Stabler, B. (2013). *shapefiles: Read and Write ESRI Shapefiles*. Retrieved from https://CRAN.R-project.org/package=shapefiles.

Stone, J. 1996. Preliminary trials on the culture of goldsinny and corkwing wrasse. Pages 142–167 *in* M. D. J. Sayer, M. J. Costello, and J. W. Treasurer, editors. Wrasse: Biology and use in aquaculture. Oxford: Fishing News Books.

Tepolt, C. K., Darling, J. A., Blakeslee, A. M. H., Fowler, A. E., Torchin, M. E., Miller, A. W., & Ruiz, G. M. (2020). Recent introductions reveal differential susceptibility to parasitism across an evolutionary mosaic. *Evolutionary Applications*, *13*(3), 545–558. https://doi.org/10.1111/eva.12865

Wang, S., Meyer, E., McKay, J. K., & Matz, M. V. (2012). 2b-RAD: a simple and flexible method for genome-wide genotyping. *Nature Methods*, *9*(8), 808–810. https://doi.org/10.1038/nmeth.2023

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-Statistics for the analysis of population structure. *Evolution*, *38*(6), 1358–1370.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wringe, B. F., Stanley, R. R. E., Jeffery, N. W., Anderson, E. C., & Bradbury, I. R. (2017a). HYBRIDDETECTIVE: A workflow and package to facilitate the detection of

hybridization using genomic data in R. *Molecular Ecology Resources*, *17*(6), e275–e284. https://doi.org/10.1111/1755-0998.12704

Wringe, B. F., Stanley, R. R. E., Jeffery, N. W., Anderson, E. C., & Bradbury, I. R. (2017b). parallelnewhybrid: an R package for the parallelization of hybrid detection using newhybrids. *Molecular Ecology Resources*, *17*(1), 91–95. https://doi.org/10.1111/1755-0998.12597

Supplementary

Supplementary tables

Table S1. Assay information for all 106 SNP markers. The 84 markers included in the final analysis after filtering are marked in bold.

Assay	SNP_ID	2nd-PCRP	1st-PCRP	UEP_SEQ
W1	**SYMME_00001019_259056**	ACGTTGGATGTTGGGAGAACTACGACTCAC	ACGTTGGATGTACTGAGGGTTTGCCTCTTG	CTCACATGCCAACACAC
W1	**SYMME_00000351_687296**	ACGTTGGATGGGTAAAAGATTTAAAAGGCCC	ACGTTGGATGCCGGCTAAAAATAAAGTCTG	AAAAGGCCCTTGTATCA
W1	**SYMME_00023818_156483**	ACGTTGGATGGGCAAGTCCTGTCAGGAAAC	ACGTTGGATGAGAGAATGGATCAGTCGAGG	ATTGACAGGACGAGGAG
W1	**SYMME_00006679_62104**	ACGTTGGATGACACCAACGTGGGACTTATG	ACGTTGGATGTCACTTGGCTTGGCCAAAAC	GGGACTTATGTCACATGC
W1	**SYMME_00007129_47437**	ACGTTGGATGGAGATGAGCCCAGAGATTAC	ACGTTGGATGGAAAAAGAGGACGACCTTGC	CTTTAACCGACCTTCCAAA
W1	**SYMME_00023738_421399**	ACGTTGGATGACTGGGACGAGAAACCTTTG	ACGTTGGATGTCAACCGCTGATTGTGCAAC	CTGATCACGCTCGACTACT
W1	**SYMME_00004764_285875**	ACGTTGGATGTCAGCACTTTGAGCAGAGAC	ACGTTGGATGTCCGTCTGGATCACGATGAG	TGGTGAACGTTTTGTAGAC
W1	**SYMME_00004729_285551**	ACGTTGGATGAGGTGTAGTTATGAGGGAGC	ACGTTGGATGATGACTGCCCCGTGTTCCTG	caAGACAAGCTGTCCTCGAT
W1	**SYMME_00000109_1247032**	ACGTTGGATGATTCTCTCCACTGCAGACAC	ACGTTGGATGTTTTGTGCAGACAGAAGGTG	ggaAGACACGAAAATCATGC
W1	**SYMME_00023453_96630**	ACGTTGGATGTAAAACGCCTCTACATGCCC	ACGTTGGATGTTCGCCACGTTAATGTGAGG	tCCTCTACATGCCCCGCACGA
W1	**SYMME_00001278_977669**	ACGTTGGATGATTACAGATCAGTGGGTGCG	ACGTTGGATGGAGCCTGTTAGCTTAGCTTC	atATCAGTGGGTGCGACCTCA
W1	**SYMME_00023225_405765**	ACGTTGGATGGCTTGAAGAGATTGTGACTG	ACGTTGGATGGCTACTCAAACAAAGTGACTC	ACTGAAGTATTGATTGGGAAA
W1	**SYMME_00006025_40517**	ACGTTGGATGAAGGAAGACAACTGCTCAGG	ACGTTGGATGAGACAGTGCTATCAGACAGG	cgctAAATTGCTGCCTACCTAA
W1	**SYMME_00023574_386442**	ACGTTGGATGTCTACCGTCATCGGTGAAAC	ACGTTGGATGGCACTTGGACATTAGCACTG	cTATCTGTTGGTTACAGCAGAC
W1	**SYMME_00002115_94621**	ACGTTGGATGGCCCGTCAGCAAGTTTTTCG	ACGTTGGATGGCTGCGTCAGTCAATTTATC	ccCAGCAAGTTTTTCGAGGTAA
W1	SYMME_00000376_965994	ACGTTGGATGCAGAGACCCCGAGTATTTAG	ACGTTGGATGGTTGGCTCGAGGTGAATTCT	gttgAAGTTAAGTCAGAGTGGG
W1	**SYMME_00023529_208893**	ACGTTGGATGTGAACTGCAAATGGCTTTGG	ACGTTGGATGTACCTTTCTACGGCAGCATC	GGAAATAAAGCAAAGACAAACAC
W1	**SYMME_00001181_204509**	ACGTTGGATGCAGAGTATTTACTTGATGGC	ACGTTGGATGGAACCCGAGAAGAAATGACC	tcTTTACTTGATGGCACCCGATTC
W1	**SYMME_00002662_67836**	ACGTTGGATGGGAAAGCAGTAGCTTCTTAC	ACGTTGGATGTACACAGAGATCACGAGATG	agtcCACTCACGTTGTTGATCATC
W1	**SYMME_00003318_375819**	ACGTTGGATGCGTCACTCGTGATTCTCTTC	ACGTTGGATGCTGGTTTCATGGCAAGCCTC	tttCTTCCCTTTTTTTTTCTCGCT
W1	**SYMME_00023380_3836**	ACGTTGGATGAGGCGATGTCGAGAGAGAAG	ACGTTGGATGTGGCGCGGATGTAAATTCTC	gggaGACGACGAACGACGTGCAAA
W1	**SYMME_00002116_350960**	ACGTTGGATGTTTGTGTGCGATAGAGAGCC	ACGTTGGATGATGGATGCAGTGGATGTAAG	tcCTGGCAGATGTGTCGATCAACGC
W1	**SYMME_00023319_1170712**	ACGTTGGATGCAGTGCATTGATTCACGCAG	ACGTTGGATGATTGACAGCAAGGTGACTTC	tcagTGATTCACGCAGGAAATTCGC
W1	**SYMME_00010158_75346**	ACGTTGGATGGGCCTCCCGTATAATAATGG	ACGTTGGATGAGTGAAGTTTGCGATCGAAG	taaCTTTTGTGTCTTTAGCCTCTCTC
W1	**SYMME_00023197_121823**	ACGTTGGATGCACTTCTTGTTGCAGAGTTC	ACGTTGGATGTTCGTTTGGAATCCCGATGC	gaacCAGTTGACGGCTTTAAGATAAA
W1	**SYMME_00023145_2392136**	ACGTTGGATGCTGGTCGATCATAAAACGTTG	ACGTTGGATGAGCCCAACCGATATGTTTGC	ggggTCAGGACACTACTGGTGTATCT
W1	**SYMME_00013619_22969**	ACGTTGGATGTAGAGTCGGTCTCCTCTAAC	ACGTTGGATGCACAGGAAGTGACATCACTC	gtcgCCTCTAACGAGGTCACTGCCCCC

W1	**SYMME_00012017_14751**	ACGTTGGATGGGTGGGATTTGAACCCGCAT	ACGTTGGATGCTGCACGAAAAATCAATCTG	ccttTACGCACAGACCTCGAGACTACT
W2	**SYMME_00001242_101541**	ACGTTGGATGTCCTCACCTATCCATACACC	ACGTTGGATGGCAGGTCACAACAACCAATG	TCCACACCTCTGCACAC
W2	**SYMME_00002850_26023**	ACGTTGGATGCTTTAGGTCGTCTGCAGAAC	ACGTTGGATGAACCTGGAATTTAGCGCCTG	TCTGCAGAACGCGCTAC
W2	SYMME_00023206_68332	ACGTTGGATGAAGAAGCAGGTGCAGCATTG	ACGTTGGATGGTTATTGTGGACATGTCCCG	GTGCAGCATTGCAGAAA
W2	**SYMME_00023953_207211**	ACGTTGGATGAATCGTGACAGCAGCTTCTC	ACGTTGGATGACTTTGCTGTGGAGACCTTG	CTGCAATCGCCATCCCTC
W2	**SYMME_00023262_816209**	ACGTTGGATGACCATCAGATGAGAAAGGCG	ACGTTGGATGACAAAGCTGTGCGTATAGCC	CTATATATGCGCAAGCTG
W2	**SYMME_00007078_108905**	ACGTTGGATGTTGCAGAAGCTCTCTCCATC	ACGTTGGATGTACTGTATTAACCCGGCCTC	GCTCTCTCCATCCTCCGCA
W2	**SYMME_00008787_21327**	ACGTTGGATGAGTGGTTGCGTGCACAGTTG	ACGTTGGATGGAGACGGATGAAAAGGTCAG	ccTGACCTTTCGATGACCT
W2	SYMME_00023260_787080	ACGTTGGATGAAGAGAGTGATCCATGTGGC	ACGTTGGATGGAAAGTAAAGGATGTGACCC	CATAATGCTCGTGTCTGAT
W2	**SYMME_00000097_411069**	ACGTTGGATGGTTCATCTGTACTCTGTACC	ACGTTGGATGCCCAGAATCTGACATTTGTG	TGTACTCTGTACCTGCATCA
W2	**SYMME_00000375_42132**	ACGTTGGATGAAAATTCCTCAGGTGCAGGG	ACGTTGGATGTCACTTCAATGACCTTTAC	AAAGCAGAGCAAGCTAATCG
W2	SYMME_00003351_308412	ACGTTGGATGGTGTGTGTGAAATGATGGG	ACGTTGGATGTTTATCGGCTGGCAGATTA	ccAAAATGACAACTAGACACC
W2	**SYMME_00001599_151719**	ACGTTGGATGGCCTAATCACATTGAGAAGC	ACGTTGGATGTGGATGTCGTCTTAGTGTCG	AGACAAAAATGCTCCAAGTGA
W2	**SYMME_00007721_148340**	ACGTTGGATGGTCAGCGAGGCGAGTAAGAT	ACGTTGGATGCCCACAGGGTAACAACAATT	CGAGGCGAGTAAGATGCAAAT
W2	**SYMME_00003452_329369**	ACGTTGGATGCACTATGATGCCTGCTCAAG	ACGTTGGATGCACAAATAACTGTCTTCGGC	ccCATCCATATGTCAACGGTAA
W2	**SYMME_00011099_16235**	ACGTTGGATGACTGAGGGCCAGTTGTAAAG	ACGTTGGATGAAACCCGGGAGACGAATGAG	aaaTCGATAAAACTGCCTCTCA
W2	**SYMME_00005187_201142**	ACGTTGGATGGATGGCAGGACTTTTTCAGA	ACGTTGGATGTGCATTCAGCGTTGGCATTT	ggGCTTAACTGTGACGAATATG
W2	**SYMME_00001871_333256**	ACGTTGGATGTTGTCCCCAGCATGTGATAC	ACGTTGGATGTGATTTCACATGTTGGGCAG	tcATCAAAACAGTTGTGCTAAAT
W2	**SYMME_00002972_463186**	ACGTTGGATGGCCGACGGCAAATCAAAAAC	ACGTTGGATGACCAGAATCAGCAGCGATAC	ttACGAAGCAGAGAAGTCGTCCG
W2	**SYMME_00006969_69288**	ACGTTGGATGCAGGCAGATTCAATCTGACG	ACGTTGGATGCTGTCCACTTTTTGGTTTGC	gaaAGAACAGATCTGAGTGCTAC
W2	**SYMME_00003129_299972**	ACGTTGGATGGATCTGGGCTCAGTTCATTG	ACGTTGGATGTCCAAGTGCTGATCTAGGAG	ggGGTTACGGTAAAGATGCTGAA
W2	**SYMME_00023242_777958**	ACGTTGGATGTGTGAGAATTTACTACGTGC	ACGTTGGATGCTTTTGCTTCGAGGTCTGTG	ccTGCCAAATACTGCATTTTTTTT
W2	SYMME_00003544_13076	ACGTTGGATGGAGAAGACAGTTTTCACCCG	ACGTTGGATGTTAGTGGCGACATTCTCCTC	cttCTCGTATTTTAGAGTCTTAACG
W2	**SYMME_00024144_124879**	ACGTTGGATGTCCGAGAGTGTTTGTGACAG	ACGTTGGATGTCAGTGCTGCTAAATAGACC	cGTTAGCAATGCAATCGAATTTTT
W2	**SYMME_00024110_27315**	ACGTTGGATGTTCTCCTTTTGGCGTCACTG	ACGTTGGATGAGCAACGTTAAGACCAGCAC	gTAGTTGACGCGAGATTAGTGCAAA
W2	SYMME_00002174_27173	ACGTTGGATGAAACAAAGAGGAAAGCAGGG	ACGTTGGATGCAGACTCAGAAGAAGTCAGC	ggGAGGAAAGCAGGGGATTCGTCTT
W2	**SYMME_00024098_237080**	ACGTTGGATGATAAGGGCGAGCTATTCGTG	ACGTTGGATGGAGATTTCAGGCTTGGTCAG	gTAAGGGCGAGCTATTCGTGGGCGGC
W2	SYMME_00023208_352029	ACGTTGGATGACGCAGCGTGCAAATCTCTC	ACGTTGGATGCCCTTTCCATTCTTCTCCTC	cCGCAGCGTGCAAATCTCTCGCCTGAA
W3	**SYMME_00023308_56447**	ACGTTGGATGGAGCAAGCCGACAGAATAAG	ACGTTGGATGCTCTACTGGTTGGCTGAAAG	CGACCCCCATGCACTCT
W3	**SYMME_00001073_633173**	ACGTTGGATGCCTTCGCATCACTCTTCGAC	ACGTTGGATGTGGCAGAAGCAGCATTTTCG	CCAGACAAGCTGCCCAG
W3	**SYMME_00002674_685013**	ACGTTGGATGCATTTCCATCTGCAGGCAGG	ACGTTGGATGGTCTTAGGTCTTAGCATGTG	CTGCAGGCAGGTCGACA
W3	**SYMME_00003351_50140**	ACGTTGGATGGAGGCAATAAACAAGCAGAC	ACGTTGGATGCTCAGTTTGATCACTCGTCC	ACAAGCAGACAAGTCGG
W3	**SYMME_00002315_632632**	ACGTTGGATGGCAGTCCACGTTCTTATGTC	ACGTTGGATGCGATTACACAAAGCACTGGG	TCAGGAGAATGGCGATGG
W3	**SYMME_00001760_**	ACGTTGGATGAGTCCACTGT	ACGTTGGATGAGGAAATGG	aTAATTGCACCATTCTC

	736838	GTTTGTGCTG	AGGCATGGAAC	GG
W3	SYMME_00004618_13817	ACGTTGGATGTACCCACTGCAGTTGTCTCG	ACGTTGGATGAATCCAGCCCTCATTTTGCC	ctAGTTGTCTCGTCTGATG
W3	SYMME_00000376_1475703	ACGTTGGATGAGTCGAGTCTGGTTCATCAG	ACGTTGGATGGTTAGTCCTGATGCAGTGTG	CTTACCATTTGAGGTTGAG
W3	**SYMME_00010876_30588**	ACGTTGGATGGATTCCGTCGACTCGTCATC	ACGTTGGATGTAAGGAAACTACGGAAGCCC	GGGGAGAACGAAGGCGAGC
W3	SYMME_00000230_1023038	ACGTTGGATGAGGAAGGTAGTCACAGGTTG	ACGTTGGATGAATAGCGGTCTCTTGTCAGC	AGTCACAGGTTGGACTTGAG
W3	**SYMME_00004807_79507**	ACGTTGGATGCAAGCCGAAGGCTTGAATTG	ACGTTGGATGAGCTGATTGGCAGGTTGTTC	cccATTTTATTGCTGCACGTA
W3	**SYMME_00001692_108228**	ACGTTGGATGTTTGGGTCAAACTGGTCGAG	ACGTTGGATGAAGTGTGTAAGTGCACAGAG	ACTGGTCGAGTATTTTGCACA
W3	**SYMME_00001977_272813**	ACGTTGGATGAAGCAGGAGATCATGACCAG	ACGTTGGATGTAGCTTCTTTCAGCCTCCAG	gaGGCAGTACAAAAAGGAGGC
W3	SYMME_00023242_770809	ACGTTGGATGGCCAAACGAATCTCAAAGTG	ACGTTGGATGGTTGTAACTTCTCATTGCACG	TCGAACTAGACCTTCACACTTA
W3	**SYMME_00024209_145735**	ACGTTGGATGACAACAAGACGAGATGCGTG	ACGTTGGATGTCTCGGACCGTGATGAACTG	aACGAGATGCGTGCGATTAATT
W3	**SYMME_00003452_457206**	ACGTTGGATGTAGCTGTGATACAGCATTAC	ACGTTGGATGAAACAACACGTCACGAGCAC	cTAATCCATGAGATTCTTTCAGG
W3	SYMME_00024389_110000	ACGTTGGATGAGGAAGACTCCCACTTTGCC	ACGTTGGATGTGCCTGTGGCTGAAGAACG	gGAGGCAGACGGGTCGAGTCCAG
W3	SYMME_00001278_696676	ACGTTGGATGGTCAATTATTGTCTTTCTGC	ACGTTGGATGACAGTGACAGTTTGTAGCTC	cccTCTTTCTGCATATTTTTCGTG
W3	**SYMME_00000053_1922641**	ACGTTGGATGTGGACATGAAGGAGTCACAC	ACGTTGGATGATATCCCTCCGCTTGCAAAC	accAAAGAAAGGCCCACTAATCTT
W3	SYMME_00002916_300935	ACGTTGGATGTTCTGAAGATGACACGACGC	ACGTTGGATGTTAAGAATGCTGCGAACGAG	gaACACTGTTTGCAGGTTACTCGG
W3	**SYMME_00000230_2293915**	ACGTTGGATGGCTTTCCACTTACAGATACC	ACGTTGGATGGAATATCCCATCCAACCACG	ACAGATACCTAAATGATCTGTACAC
W3	**SYMME_00023188_229499**	ACGTTGGATGATCGCGACGTGAAATGATGC	ACGTTGGATGTATTATCTGCCGTTAGCGCC	aGCGACGTGAAATGATGCGACTGAT
W3	**SYMME_00002548_480157**	ACGTTGGATGGTACTCGTCTATGAGTCTAC	ACGTTGGATGAGCCATTAACATGCTAACAG	cccTTGACATTCTTAACAGCGATCAC
W3	**SYMME_00008867_17068**	ACGTTGGATGTCCTGCCAGCGATTGCAATG	ACGTTGGATGGACTTTGAAGGAGCTTGGTG	gCCTGCCAGCGATTGCAATGCAACAA
W3	**SYMME_00010761_23994**	ACGTTGGATGTGAGTATACCAGAGGAGAGG	ACGTTGGATGCAGCTCAACCGATGCATGTG	cAGAAAAACGGCTGCAGCTTCCAGTT
W3	SYMME_00023181_495503	ACGTTGGATGGCCGGATGGAGAAACAGCAA	ACGTTGGATGTTAACCGTCAATGGGCCTTG	aCCGGATGGAGAAACAGCAACGAACA
W4	**SYMME_00023385_679926**	ACGTTGGATGTGCTCTTCCTCTCCATTCAC	ACGTTGGATGATAGCCCCAGAATTCACAGC	TCTCCATTCACTGCATG
W4	**SYMME_00002806_257690**	ACGTTGGATGAAATCCCCCACAAAACCCTG	ACGTTGGATGGCGCAGAGTTATGATACGTG	CCCTGCAGAGGACGAAC
W4	**SYMME_00003210_166008**	ACGTTGGATGCAGAGACTATGTCCTACCTG	ACGTTGGATGTGTCAGTGCTTTTTCTGTGG	CACTGGGCAGCAGCGAT
W4	**SYMME_00002054_508210**	ACGTTGGATGAAAATGTCTGCATTTTCCGC	ACGTTGGATGGCACATTTTGCTTGGTGAGG	TGCATTTTCCGCATCACT
W4	**SYMME_00009579_72077**	ACGTTGGATGATTATGGAAGGGAAGCGGAC	ACGTTGGATGTTGGTGCTTTAGACAGACGC	GTGAGGGGACGTTCACTG
W4	**SYMME_00004127_40394**	ACGTTGGATGAGTGTACAGGCCTGAAAGAC	ACGTTGGATGCTTAAAGGAAACTCGGCGTG	ccTGCCAAATAAACCCGAG
W4	**SYMME_00024054_274460**	ACGTTGGATGGAAGGCACCTACACTAAGAG	ACGTTGGATGGAACGCTGACCTTGACTTTG	ACTAAGAGTAACGAGCTCC
W4	**SYMME_00003115_60176**	ACGTTGGATGGGCAATATTCAATAGTGTGTG	ACGTTGGATGCTCTTCCTCTCCATGTTTGC	ATAGTGTGTGTGAGGAAGT
W4	**SYMME_00011044_1109**	ACGTTGGATGATTCAGTGTCTTCACCGTTG	ACGTTGGATGATATTCAGTCCCTCCTGAGC	GACGATGTGTCTGCAGAGTC
W4	**SYMME_00000436_144204**	ACGTTGGATGCACAGACTTAGATTTAAGGC	ACGTTGGATGAATTGGTTCAGTGACATTGG	cCTTAGATTTAAGGCAACTCC
W4	**SYMME_00023191_105880**	ACGTTGGATGTAATGTCCTCTTTCAGGACG	ACGTTGGATGGACAGATTCATGGACCTGAG	CTTTCAGGACGTAATAGACGA
W4	**SYMME_00024072_36827**	ACGTTGGATGGGAAGATGGATGACAGAGAG	ACGTTGGATGCTGGTGTCAGTCAGGTCTTT	ggcTGACAGAGAGACGAGAGA
W4	SYMME_00000367_402845	ACGTTGGATGGTGAGAAGATAACAGACAGG	ACGTTGGATGGTGCTGTCGATTAAATGCTG	cccAACTGCAAATATTTCGAGA

W4	**SYMME_00006358_220367**	ACGTTGGATGTCACTCTCTACGTAAACCCC	ACGTTGGATGAGCCACAGCTCGAATTGAAC	gAGAATGGGCTTCAGAAGTCAT
W4	**SYMME_00000564_900003**	ACGTTGGATGTTTGTAACCACAGACTGGGC	ACGTTGGATGATCTGAAACACCGCCTTCAC	tACCGGGTTTCTTCTATAGCTGT
W4	SYMME_00002174_10525	ACGTTGGATGAAGCATTCCTGCTGCAAGAC	ACGTTGGATGATCATGTGTACGGCACGTTG	tgaAAGACCAGAGACGATAAGGT
W4	**SYMME_00023722_192560**	ACGTTGGATGGTGTCATTCTTCGTCTGTCC	ACGTTGGATGGATTAATGACTCCTTGGCCC	ccTCTGTCCTTCAGTAACAAGCAA
W4	SYMME_00003544_104554	ACGTTGGATGGTGGTTTCAAGGCGCATCAG	ACGTTGGATGATCATGAGGGTAACGTCAGC	tCATCAGGCTCAGCAGGCTGGTCG
W4	SYMME_00001033_46636	ACGTTGGATGTGGGTTTGTGCAAGTACTGG	ACGTTGGATGTGATATTTTGAACATATCAC	ggcATGAGGAATGGTCCGATACAT
W4	**SYMME_00023145_4178055**	ACGTTGGATGGGGAAGAAACTCTTCCACAC	ACGTTGGATGACACCAGTGTTGTCAGCTTC	TTCCACACGAGTTAATGTCCAATCA
W4	SYMME_00000564_674855	ACGTTGGATGTGGCGTTTTTTGCTGGTCTC	ACGTTGGATGTACACAGAAGCAAAGTGCCG	gCTTTATTTGTGACCCTGTCGGGAA
W4	**SYMME_00005187_163571**	ACGTTGGATGTAAAGGCACTGCTGTTCAAG	ACGTTGGATGTCTGAATGCAGCAGGCTTAC	aaTGTAGTGTGAAAAAAAAGGGCAA
W4	**SYMME_00023225_397158**	ACGTTGGATGTCGTGTTTATTCACGGCGTC	ACGTTGGATGCTTTAGTGAGCAGGACCATC	tCTGACACGGTGTCACATTCTCAAAA
W4	SYMME_00001635_333554	ACGTTGGATGTACACACGTGTATGCCACTG	ACGTTGGATGAGGAGCTTTGAGAGCAAGTC	gCGTGTATGCCACTGCAGGCTGCGAT
W4	SYMME_00001467_618978	ACGTTGGATGGCCTGCACCATTTACACTTG	ACGTTGGATGTGCCAAATTACACACTCTGC	cctACTTTCCTAATTTCTTCAGTAATT

Table S2. Sample size, allelic richness, mean values of unbiased expected heterozygosity, observed heterozygosity, F_{IS} and p-values for global Hardy Weinberg exact test averaged across 84 loci.

Sample	N	A_R	H_{obs}	H_{exp}	F_{IS}	HWE
FLA16	95	1.824	0.376	0.396	0.055	0.001
FLA17	307	1.849	0.396	0.413	0.045	0
FLA18	30	1.823	0.385	0.397	0.020	0.724
HIT17	10	1.797	0.382	0.382	-0.064	1
SMO17	13	1.804	0.388	0.388	-0.034	0.994
SMO18	245	1.797	0.370	0.383	0.031	0
TUS17	3	1.792	0.381	0.384	-0.181	1
KRI17	43	1.792	0.367	0.381	0.021	0.019
AVE17	3	1.793	0.391	0.386	-0.214	1
SAN17	3	1.813	0.367	0.379	-0.173	1
MID17	21	1.775	0.351	0.372	0.017	0.996
ALE17	38	1.787	0.375	0.378	0.003	0.099
SUL13	77	1.785	0.354	0.378	0.060	0
ARD18	10	1.895	0.411	0.443	0.034	1
MAL13	5	1.817	0.360	0.396	-0.039	1
FLO18	9	1.825	0.353	0.398	0.072	0.999
OS14	134	1.822	0.405	0.398	-0.012	0
AUS14	91	1.818	0.391	0.396	0.017	0
AUS17	233	1.822	0.379	0.400	0.054	0
SVE14	148	1.824	0.394	0.400	0.018	0
FLOD17	106	1.433	0.184	0.191	0.022	1
RIS16	41	1.423	0.186	0.189	0.005	1
HVA14	60	1.427	0.185	0.189	0.019	1
MAR16	40	1.428	0.187	0.189	-0.014	1

Table S3. Lower left trimatrix display W&C F_{ST} estimates and upper right trimatrix display respective p-values. Bold cells have a p-value < 0.05 and grey are significant values after Bonferroni correction.

	FLA16	FLA17	FLA18	HIT17	SMO17	SMO18	TUS17	KRI17	AVE17	SAN17	MID17	ALE17	SUL13	ARD18	MAL13	FLO18	OS14	AUS14	AUS17	SVE14	FLOD17	RIS16	HVA14	MAR16
FLA16		**0**	0.099	0.792	0.092	**0**	0.868	**0**	0.997	0.997	**0.014**	**0**	**0**	**0**	0.492	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
FLA17	**0.004**		0.162	0.957	**0.014**	**0**	0.974	**0**	0.999	0.999	**0.007**	**0**	**0**	**0.010**	0.422	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
FLA18	0.006	0.003		0.996	**0.033**	**0.033**	0.991	**0.008**	0.999	1	0.059	0.109	**0.001**	**0.017**	0.591	**0.004**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
HIT17	0.003	0.004	0		0.981	0.998	1	0.906	1	1	0.942	0.981	0.730	0.549	0.904	0.354	**0.001**	0.276	0.315	**0.047**	**0**	**0**	**0**	**0**
SMO17	0.007	**0.012**	**0.011**	0.001		0.109	1	0.126	1	1	0.089	0.146	0.084	**0.009**	0.357	**0.020**	**0**	**0.004**	**0.000**	**0.006**	**0**	**0**	**0**	**0**
SMO18	**0.007**	**0.010**	**0.004**	0	0.004		0.940	**0.038**	0.999	0.999	0.311	**0.010**	**0.002**	**0**	0.219	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
TUS17	0.010	0	0	0.006	0	0.000		0.919	1	1	0.977	0.966	0.976	0.974	0.996	0.922	0.599	0.789	0.817	0.825	**0**	**0**	**0**	**0**
KRI17	**0.007**	**0.011**	**0.009**	0.007	0.008	0.004	0.003		0.999	0.944	0.148	**0.010**	**0**	**0**	0.188	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
AVE17	0	0	0	0.001	0.005	0	0.016	0		1	1	0.998	0.991	1	1	1	0.999	1	0.999	1	**0**	**0**	**0**	**0**
SAN17	0.011	0.010	0	0.010	0.003	0.003	0	0.024	0.007		1	0.999	1	1	1	1	0.959	1	0.996	0.997	**0**	**0**	**0**	**0**
MID17	**0.008**	**0.009**	0.011	0	0.005	0.003	0	0.006	0	0.012		0.455	0.146	**0.001**	0.568	0.069	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
ALE17	**0.012**	**0.012**	0.008	0	0.006	**0.004**	0.012	**0.007**	0	0.022	0.002		**0.001**	**0**	0.126	**0.009**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
SUL13	**0.010**	**0.012**	**0.007**	0.001	0.006	**0.003**	0.000	**0.010**	0	0	0.004	**0.005**		**0**	0.502	**0.001**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
ARD18	**0.033**	**0.023**	**0.036**	0.052	**0.044**	**0.050**	0.046	**0.053**	0.001	0.026	**0.038**	**0.056**	**0.047**		0.948	0.851	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
MAL13	0.012	0.012	0.014	0.025	0.027	0.016	0.022	0.024	0	0.004	0.016	0.019	0.005	0.011		1	0.564	0.619	0.703	0.643	**0**	**0**	**0**	**0**
FLO18	**0.028**	**0.027**	**0.028**	0.023	**0.035**	**0.030**	0.036	**0.046**	0.011	0.010	0.031	**0.022**	**0.023**	0.028	0		**0.004**	**0.008**	**0.011**	**0.025**	**0**	**0**	**0**	**0**
OS14	**0.024**	**0.027**	**0.020**	**0.025**	**0.021**	**0.021**	0.018	**0.027**	0	0.016	**0.025**	**0.020**	**0.019**	**0.039**	0	**0.014**		**0.021**	**0**	**0.050**	**0**	**0**	**0**	**0**
AUS14	**0.020**	**0.022**	**0.016**	0.015	**0.018**	**0.017**	0.013	**0.023**	0	0.009	**0.020**	**0.018**	**0.014**	**0.038**	0.002	**0.015**	**0.001**		**0.020**	**0.044**	**0**	**0**	**0**	**0**
AUS17	**0.025**	**0.025**	**0.017**	0.018	**0.017**	**0.020**	0.010	**0.026**	0	0.011	**0.023**	**0.015**	**0.016**	**0.041**	0	**0.012**	**0.001**	**0.001**		0.280	**0**	**0**	**0**	**0**
SVE14	**0.022**	**0.024**	**0.017**	**0.019**	**0.017**	**0.020**	0.015	**0.025**	0	0.012	**0.021**	**0.016**	**0.016**	**0.034**	0.003	**0.012**	**0.001**	**0.001**	0		**0**	**0**	**0**	**0**
FLOD17	**0.453**	**0.383**	**0.511**	**0.573**	**0.568**	**0.459**	**0.607**	**0.527**	**0.595**	**0.604**	**0.555**	**0.546**	**0.512**	**0.438**	**0.573**	**0.558**	**0.471**	**0.488**	**0.452**	**0.463**		0.963	0.740	0.997
RIS16	**0.418**	**0.364**	**0.473**	**0.554**	**0.543**	**0.438**	**0.603**	**0.492**	**0.592**	**0.602**	**0.527**	**0.510**	**0.478**	**0.413**	**0.563**	**0.535**	**0.439**	**0.451**	**0.428**	**0.432**	0.001		0.994	0.983
HVA14	**0.432**	**0.371**	**0.492**	**0.565**	**0.557**	**0.445**	**0.607**	**0.507**	**0.598**	**0.606**	**0.541**	**0.527**	**0.492**	**0.428**	**0.572**	**0.550**	**0.453**	**0.467**	**0.438**	**0.445**	0	0		0.999
MAR16	**0.418**	**0.364**	**0.474**	**0.554**	**0.543**	**0.437**	**0.603**	**0.489**	**0.592**	**0.601**	**0.524**	**0.509**	**0.476**	**0.411**	**0.563**	**0.536**	**0.440**	**0.451**	**0.429**	**0.433**	0	0.003	0	

Table S4. Model-fitting for the different markers and parameter estimates for the geographic cline ranging from Flatanger to Marstrand. 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.

Marker	Model	AICc	Centre (km)	Width (km)	δ_M	τ_M	δ_L	τ_L	p_{min}	p_{max}	loglike
STRUCTURE Q-score	optN	627.292	799.4 (786.7, 1087.2)	23.3 (0.0, 168.9)					0.00409	0.89123	-309.6292
SYMME_00000230_2293915	fixL	80.234	941.1 (768.6, 1044.9)	224.7 (93.0, 541.4)			49.40234	0.00004	0.56500	1.00000	-36.09707
SYMME_00000351_687296	fixL	137.388	817.9 (770.7, 1052.2)	544.2 (149.3, 831.3)			94.14454	0.00001	0.45900	0.96700	-64.67724
SYMME_00001019_259056	fixL	116.056	767.9 (736.7, 1058.6)	152.8 (13.7, 362.0)			36.67244	0.00011	0.50000	1.00000	-54.01118
SYMME_00001278_977669	fixL	126.596	884.6 (832.1, 1056.2)	384.1 (100.0, 529.2)			146.75380	0.00011	0.44800	0.98800	-59.28127
SYMME_00001692_108228	fixL	203.006	927.6 (843.2, 1064.3)	448.2 (83.6, 601.8)			210.63860	0.00005	0.37500	1.00000	-97.48641
SYMME_00001977_272813	fixL	125.701	818.3 (854.0, 1065.1)	674.6 (81.5, 535.7)			86.70604	0.00024	0.00000	1.00000	-58.822
SYMME_00002548_480157	fixL	177.925	875.6 (773.8, 1080.6)	312.2 (22.0, 530.8)			118.25190	0.00005	0.15700	0.84200	-84.94583
SYMME_00002806_257690	fixL	68.882	912.8 (757.4, 1077.1)	420.2 (30.2, 717.0)			151.04970	0.00002	0.47300	0.86800	-30.42424
SYMME_00002850_26023	fixL	324.532	810.8 (762.7, 1048.8)	331.5 (48.0, 430.7)			92.20118	0.00005	0.01300	0.97600	-158.2491
SYMME_00003452_457206	fixL	129.891	814.7 (752.9, 1050.1)	161.4 (5.7, 334.7)			54.49966	0.00014	0.43900	1.00000	-60.92881
SYMME_00005187_163571	fixL	141.773	883.9 (776.5, 1066.6)	474.6 (45.4, 669.6)			169.39900	0.00031	0.37500	0.95000	-66.86962
SYMME_00006969_69288	fixL	226.462	906.1 (825.2, 1058.4)	445.8 (83.1, 672.5)			136.31460	0.00013	0.12700	0.92100	-109.2142
SYMME_00009579_72077	fixL	83.769	818.5 (747.4, 1061.4)	364.5 (23.1, 604.8)			88.22731	0.00003	0.05000	0.59500	-37.86781
SYMME_00013619_22969	fixL	112.525	757.2 (732.2, 1049.1)	285.4 (47.1, 463.1)			35.72526	0.00001	0.40000	1.00000	-52.24568
SYMME_00023191_105880	fixL	125.618	778.1 (733.4, 1045.9)	195.4 (15.3, 419.4)			34.41165	0.00007	0.55900	1.00000	-58.79201
SYMME_00023242_777958	fixL	102.751	774.8 (725.7, 1037.6)	161.9 (9.9, 394.7)			28.26784	0.00010	0.40700	1.00000	-47.35829
SYMME_00024072_36827	fixL	116.062	911.3 (835.6, 1060.8)	497.8 (78.4, 655.0)			192.49440	0.00019	0.21300	0.85000	-54.01428
SYMME_00000053_1922641	optN	72.140	810.4 (780.8, 1085.3)	3.2 (0.0, 227.9)					0.00002	0.51431	-32.04971
SYMME_00000097_411069	optN	59.272	862.5 (775.2, 1089.4)	1.4 (0.0, 319.0)					0.79811	0.99647	-25.61927
SYMME_00000109_1247032	optN	267.679	834.9 (787.0, 1088.0)	11.4 (0.0, 180.3)					0.00005	0.69776	-129.8228
SYMME_00000375_42132	optN	108.158	784.9 (723.9, 1086.0)	17.3 (0.0, 332.7)					0.14571	0.65653	-50.06171
SYMME_00000436_144204	optN	95.703	799.1 (730.7, 966.2)	304.3 (117.5, 645.4)					0.06925	0.51413	-43.83465
SYMME_00000564_900003	optN	81.569	993.3 (786.6, 1088.9)	82.2 (0.0, 344.7)					0.47619	0.88151	-36.76738
SYMME_00001073_633173	optN	72.192	706.6 (685.5, 790.0)	154.8 (25.0, 493.1)					0.75541	0.98683	-32.079

SYMME_00001181_204509	optN	97.772	786.5 (771.8, 1087.7)	5.2 (0.0, 249.4)	0.67596	0.99681	-44.8694
SYMME_00001242_101541	optN	109.745	781.3 (717.2, 1086.8)	26.6 (0.0, 247.7)	0.61193	0.98244	-50.85559
SYMME_00001599_151719	optN	118.294	811.2 (777.7, 1088.6)	38.8 (0.0, 260.5)	0.62308	0.99653	-55.13
SYMME_00001760_736838	optN	172.044	788.4 (720.9, 1090.2)	15.2 (0.0, 271.4)	0.45939	0.88288	-82.00522
SYMME_00002054_508210	optN	170.245	811.3 (782.6, 1088.4)	41.8 (0.0, 237.5)	0.22122	0.78816	-81.10554
SYMME_00002115_94621	optN	137.893	803.5 (764.0, 1087.3)	49.1 (0.0, 317.1)	0.08890	0.57843	-64.92947
SYMME_00002116_350960	optN	207.230	985.1 (785.4, 1089.8)	38.2 (0.0, 305.3)	0.12829	0.67438	-99.59824
SYMME_00002315_632632	optN	51.997	791.0 (729.4, 1175.8)	10.7 (0.1, 623.2)	0.36571	0.75265	-21.97763
SYMME_00002662_67836	optN	144.101	1056.2 (785.9, 1105.3)	90.6 (0.0, 317.6)	0.12573	0.61083	-68.03384
SYMME_00002972_463186	optN	122.806	850.6 (753.0, 1090.4)	186.2 (0.0, 540.4)	0.17075	0.57603	-57.38504
SYMME_00003115_60176	optN	100.931	800.1 (773.0, 1087.6)	6.8 (0.0, 401.7)	0.28786	0.70577	-46.44891
SYMME_00003129_299972	optN	211.385	791.9 (775.2, 1088.7)	27.9 (0.0, 217.6)	0.07304	0.64361	-101.6756
SYMME_00003318_375819	optN	174.911	787.6 (786.0, 1088.6)	3.6 (0.0, 244.3)	0.08229	0.67692	-83.43841
SYMME_00003351_50140	optN	328.985	1036.3 (786.4, 1090.5)	57.6 (0.0, 223.0)	0.06836	0.81013	-160.4756
SYMME_00003452_329369	optN	126.138	957.6 (781.9, 1089.3)	53.4 (0.0, 218.0)	0.58967	0.99997	-59.0514
SYMME_00004127_40394	optN	211.269	954.9 (785.2, 1088.9)	249.5 (0.0, 502.2)	0.11255	0.69296	-101.6174
SYMME_00004729_285551	optN	276.066	799.3 (782.7, 1090.5)	30.7 (0.0, 207.2)	0.02135	0.74185	-134.0164
SYMME_00004764_285875	optN	169.743	785.8 (724.5, 1089.4)	13.6 (0.0, 332.7)	0.26258	0.74755	-80.85494
SYMME_00004807_79507	optN	100.326	779.8 (722.5, 1089.2)	44.1 (0.1, 351.6)	0.56357	0.96702	-46.14467
SYMME_00005187_201142	optN	192.819	838.3 (771.9, 1070.2)	304.9 (2.0, 423.4)	0.00029	0.52992	-92.39211
SYMME_00006358_220367	optN	214.920	807.9 (780.5, 1088.0)	32.2 (0.0, 279.7)	0.30455	0.86892	-103.4423
SYMME_00006679_62104	optN	191.850	895.4 (782.5, 1087.8)	15.4 (0.0, 212.7)	0.00003	0.53221	-91.90804
SYMME_00007078_108905	optN	204.570	806.0 (776.1, 1090.7)	57.8 (0.0, 275.3)	0.12339	0.70402	-98.26834
SYMME_00007129_47437	optN	134.095	741.4 (716.5, 810.2)	2.9 (0.0, 107.9)	0.20898	0.70077	-63.03017
SYMME_00007721_148340	optN	208.301	787.7 (717.8, 1084.4)	32.3 (0.1, 182.7)	0.32865	0.85839	-100.1331
SYMME_00008787_21327	optN	207.892	788.2 (776.6, 1086.5)	12.2 (0.0, 265.9)	0.07351	0.62437	-99.92902
SYMME_00008867_17068	optN	142.177	806.4 (774.7, 1087.7)	29.0 (0.0, 224.9)	0.57553	0.99992	-67.07149
SYMME_00010158_75346	optN	105.307	857.8 (720.0, 1153.1)	327.8 (0.2, 836.1)	0.26277	0.67423	-48.63661
SYMME_00010876_30588	optN	121.158	784.9 (723.8, 1089.5)	20.0 (0.0, 275.3)	0.52326	0.95012	-56.562
SYMME_00011044_1109	optN	149.536	875.5 (775.8, 1087.5)	53.4 (0.0, 382.9)	0.17492	0.63277	-70.75142
SYMME_00011099_16235	optN	148.623	782.4 (718.0, 870.9)	29.1 (0.0, 137.7)	0.29368	0.76379	-70.29453
SYMME_00012017_14751	optN	139.430	1016.2 (779.0, 1087.8)	9.0 (0.0, 237.1)	0.51388	0.98933	-65.69758
SYMME_00023145_2392136	optN	129.217	829.9 (757.8, 1089.7)	202.5 (0.0, 381.0)	0.68376	0.99980	-60.5915
SYMME_00023145_4178055	optN	142.693	812.3 (771.9, 1087.4)	49.7 (0.0, 243.8)	0.56730	0.97083	-67.32976
SYMME_00023188_229499	optN	132.348	1054.5 (784.6, 1156.4)	99.4 (0.1, 382.2)	0.31053	0.76573	-62.15717

SYMME_00023197_121823	optN	121.933	784.5 (750.6, 1086.7)	67.5 (0.0, 246.9)					0.62361	0.99994	-56.94974
SYMME_00023225_397158	optN	123.050	823.7 (779.4, 1089.8)	70.1 (0.0, 311.8)					0.13794	0.70899	-57.50637
SYMME_00023225_405765	optN	250.733	798.5 (783.1, 1089.7)	16.8 (0.0, 242.3)					0.18547	0.82220	-121.3499
SYMME_00023262_816209	optN	128.969	786.2 (776.9, 1086.6)	1.1 (0.0, 268.5)					0.54069	0.94564	-60.46736
SYMME_00023308_56447	optN	131.185	805.9 (775.0, 1088.6)	39.8 (0.0, 262.4)					0.05107	0.53964	-61.57515
SYMME_00023319_1170712	optN	158.398	1062.1 (785.4, 1089.3)	21.8 (0.0, 277.1)					0.09444	0.60554	-75.18232
SYMME_00023380_3836	optN	122.873	785.1 (738.9, 1085.6)	9.2 (0.1, 309.6)					0.11138	0.54200	-57.41965
SYMME_00023385_679926	optN	87.845	1034.5 (776.4, 1118.0)	119.3 (0.0, 480.5)					0.15007	0.55965	-39.90547
SYMME_00023453_96630	optN	320.818	806.8 (783.7, 1090.1)	38.3 (0.0, 237.0)					0.31280	0.90341	-156.3922
SYMME_00023529_208893	optN	99.745	790.6 (763.5, 1087.2)	14.1 (0.0, 293.0)					0.20324	0.64053	-45.85568
SYMME_00023574_386442	optN	198.977	811.0 (782.9, 1088.8)	42.6 (0.0, 284.4)					0.31359	0.82960	-95.47196
SYMME_00023722_192560	optN	242.737	793.6 (785.0, 1088.9)	21.3 (0.0, 205.6)					0.00388	0.63842	-117.3514
SYMME_00023738_421399	optN	159.557	1016.7 (785.6, 1087.3)	81.3 (0.0, 335.3)					0.21937	0.74524	-75.76167
SYMME_00023818_156483	optN	90.835	965.0 (783.0, 1087.7)	34.8 (0.1, 238.1)					0.69060	0.99999	-41.40093
SYMME_00023953_207211	optN	153.121	788.4 (768.2, 1090.0)	27.2 (0.0, 246.5)					0.53889	0.91627	-72.54345
SYMME_00024054_274460	optN	159.753	945.7 (786.5, 1085.5)	9.5 (0.0, 195.3)					0.52945	0.99995	-75.8596
SYMME_00024110_27315	optN	74.054	866.3 (780.9, 1088.9)	36.6 (0.0, 361.3)					0.25477	0.63648	-33.00966
SYMME_00024144_124879	optN	126.689	989.9 (785.1, 1088.4)	68.6 (0.0, 275.8)					0.50817	0.92971	-59.32702
SYMME_00024209_145735	optN	111.076	1005.4 (773.1, 1087.4)	34.4 (0.0, 225.4)					0.70144	0.99998	-51.52109
SYMME_00002674_685013	typL	171.673	798.0 (723.9, 1012.1)	366.5 (129.4, 534.9)			14.12170	0.00002	0.00000	1.00000	-81.81944
SYMME_00003210_166008	typL	192.638	826.0 (770.2, 1014.1)	466.4 (196.1, 651.6)			63.11648	0.00001	0.00000	1.00000	-92.30223
SYMME_00006025_40517	typL	100.855	731.0 (698.4, 1015.5)	718.8 (243.7, 932.8)			14.36357	0.00005	0.00000	1.00000	-46.41094
SYMME_00010761_23994	typL	87.767	1030.2 (975.0, 1108.9)	737.5 (356.1, 986.5)			278.61970	0.00027	0.00000	1.00000	-39.8663
SYMME_00001871_333256	typM	194.425	944.0 (791.9, 1073.6)	660.2 (33.2, 867.2)	210.28310	0.00006			0.00000	1.00000	-93.19481
SYMME_00024098_237080	typM	171.621	944.5 (782.6, 1075.8)	760.8 (57.7, 980.7)	224.89700	0.00005			0.00000	1.00000	-81.79336

Supplementary figures

Figure S1. Genotyping robustness was evaluated by calculating concordance between 79 successfully genotyped technical replicates across 85 loci. Figure displays A) genotype concordance by locus and B) genotype concordance by individual for 79 technical replicates across 85 loci. Each bar represents the proportion of genotypes that were concordant (i.e. identical), missing (i.e. one or both replicates could not be genotyped) or discordant (i.e. replicates had different genotypes).

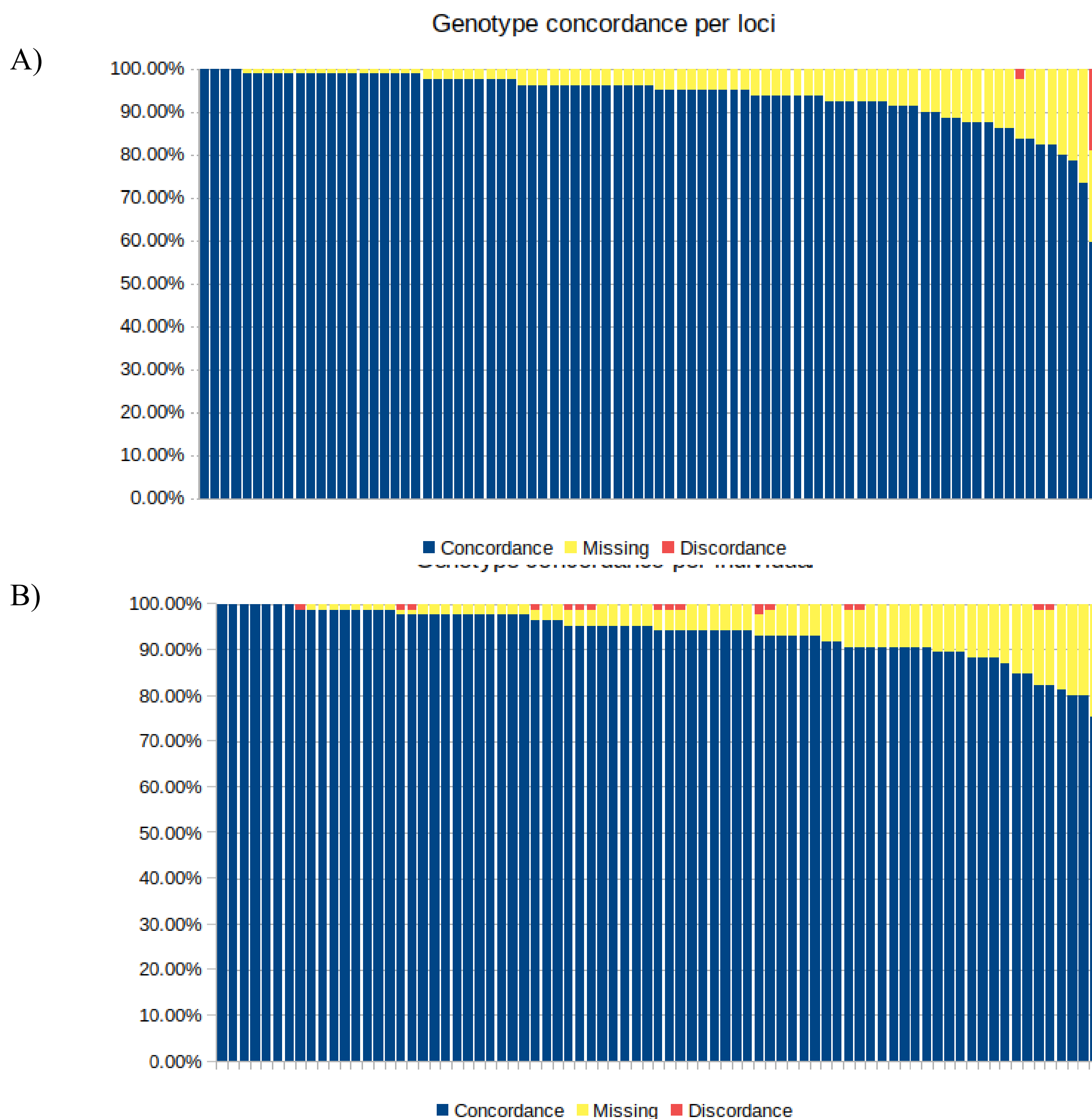


Figure S2. Hybrid detection accuracy, efficiency and power at different critical posterior probability thresholds. Solid lines are averages of three replicates of three simulated genotype data sets for 106 SNPs (A-C) and 84 SNPs (D-F). The dashed lines is the standard deviation among the simulations. Colours represent the 6 genotype classes, Pure1 = western population, Pure2 = south-eastern population, F1, F2 BC1 = F1 backcrosses with western population, and BC2 = F1 backcrosses with south-eastern populations. Accuracy = correctly assigned individuals over total individuals assigned to that class. Efficiency = correctly assigned individuals over the known individuals per class. Power = Accuracy * Efficiency. A) Accuracy shows that at critical posterior probability thresholds between 0.5 and 1.0, of the individuals assigned to a given class, > 98% of them will have been assigned correctly. B) Efficiency indicates that > 94% of individuals in each class will be identified at critical posterior probability thresholds between 0.5 and 0.9. C) Which results in power > 0.94 at critical posterior probability threshold between 05 and 0.9 D) Accuracy shows that at critical posterior probability thresholds between 0.5 and 1.0, of the individuals assigned to a given class, > 92% of them will have been assigned correctly. E) Efficiency indicates that > 83% of individuals in each class will be identified at critical posterior probability thresholds between 0.5 and 0.9. F) Which results in power > 0.81 at critical posterior probability threshold between 05 and 0.9.

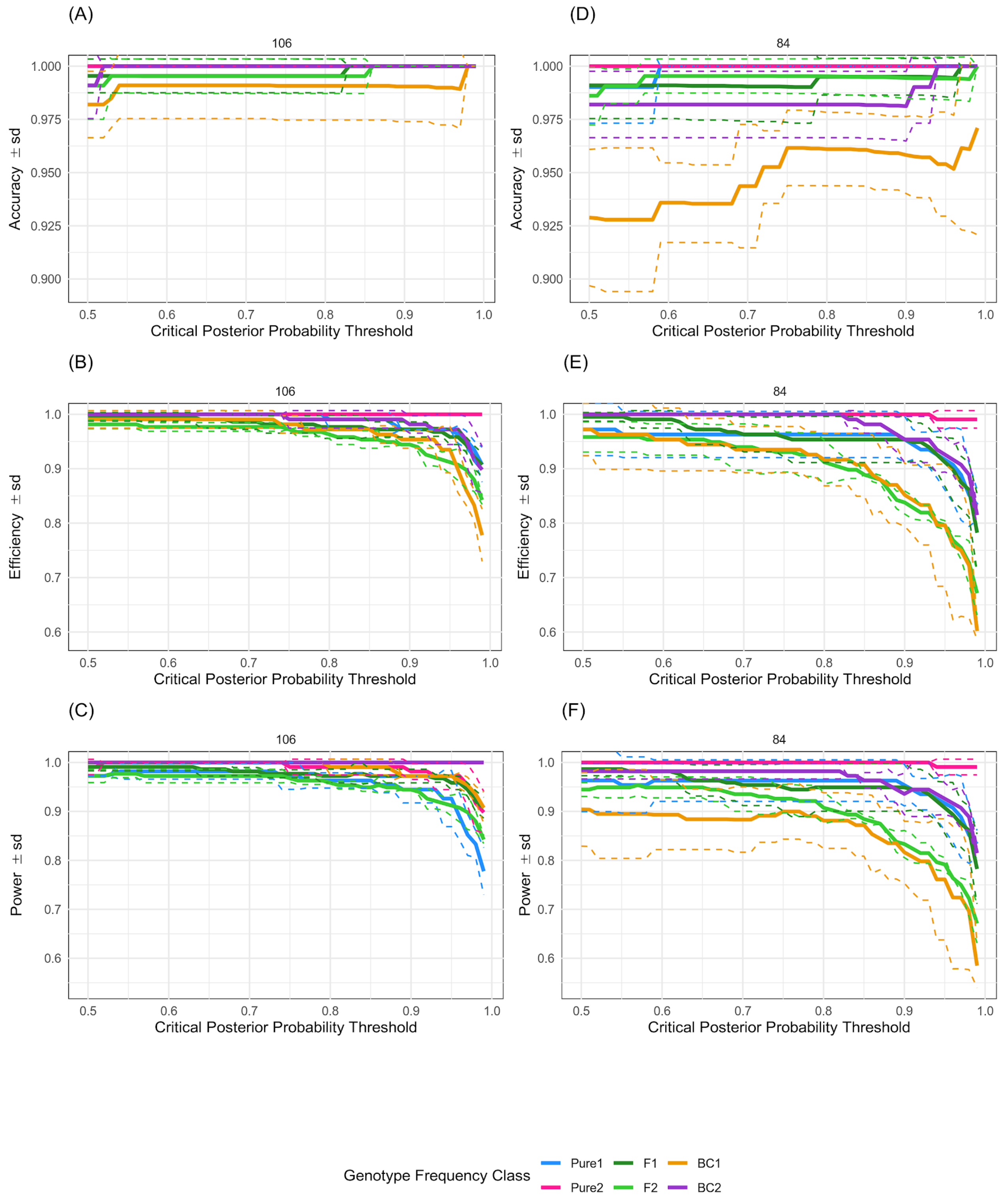
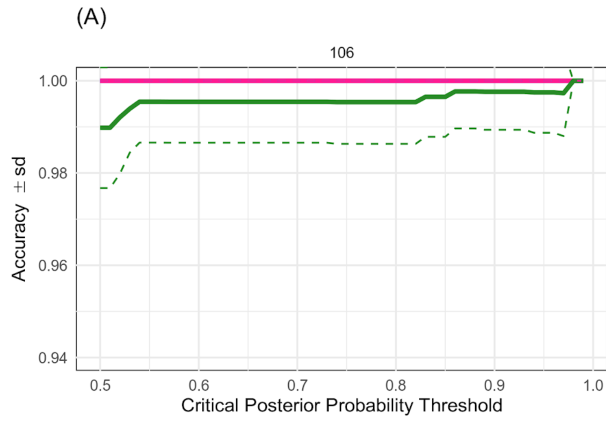
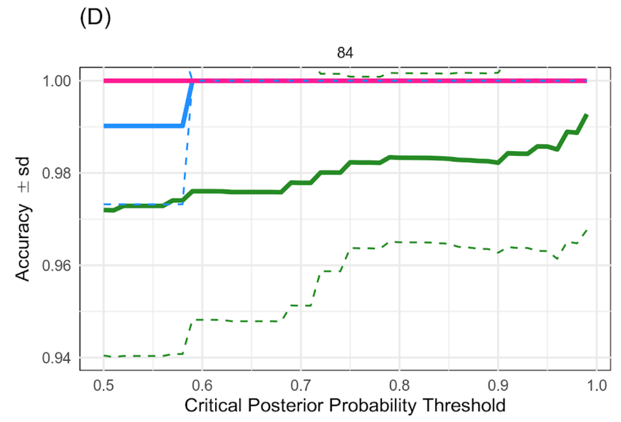
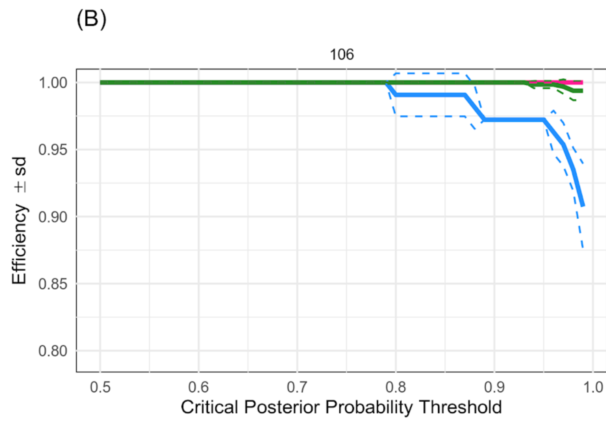


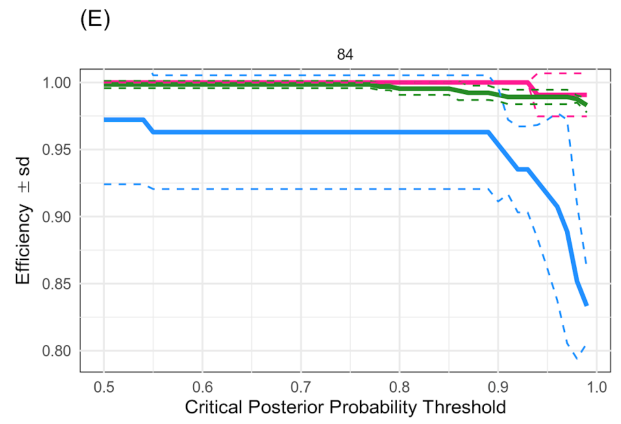
Figure S3. Hybrid detection accuracy, efficiency and power at different critical posterior probability thresholds. Solid lines are averages of three replicates of three simulated genotype data sets for 106 SNPs (A-C) and 84 SNPs (D-F). The dashed lines is the standard deviation

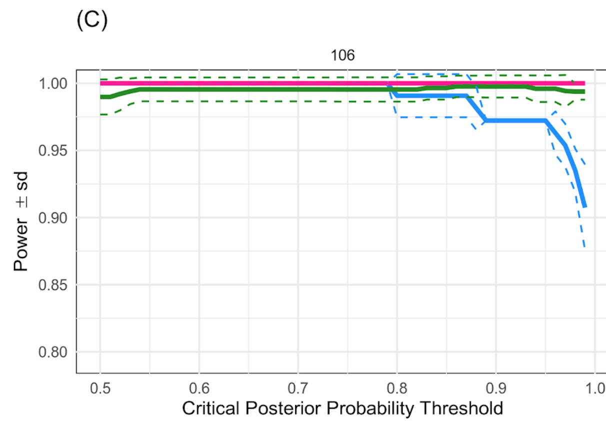
among the simulations. Colours represent the 3 genotypes, Pure1 = western population, Pure2 = south-eastern population, Hybrid = first or second generation hybrid. Accuracy = correctly assigned individuals over total individuals assigned to that class. Efficiency = correctly assigned individuals over the known individuals per class. Power = Accuracy * Efficiency. A) Accuracy shows that at critical posterior probability thresholds between 0.5 and 1.0, of the individuals assigned to a given class, > 98% of them will have been assigned correctly. B) Efficiency indicates that > 97% of individuals in each class will be identified at critical posterior probability thresholds between 0.5 and 0.9. C) Which results in power > 0.97 at critical posterior probability threshold between 05 and 0.9 D) Accuracy shows that at critical posterior probability thresholds between 0.5 and 1.0, of the individuals assigned to a given class, > 97% of them will have been assigned correctly. E) Efficiency indicates that > 95% of individuals in each class will be identified at critical posterior probability thresholds between 0.5 and 0.9. F) Which results in power > 0.95 at critical posterior probability threshold between 05 and 0.9.

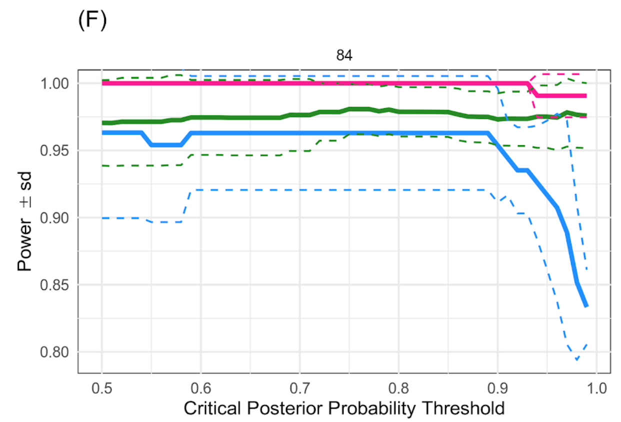








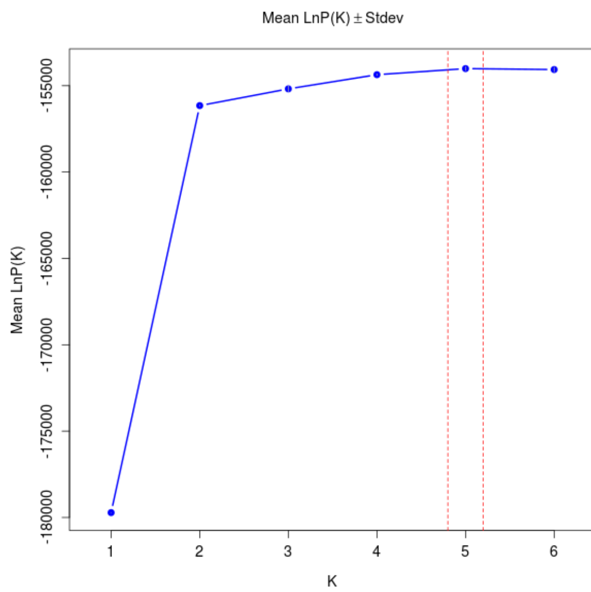




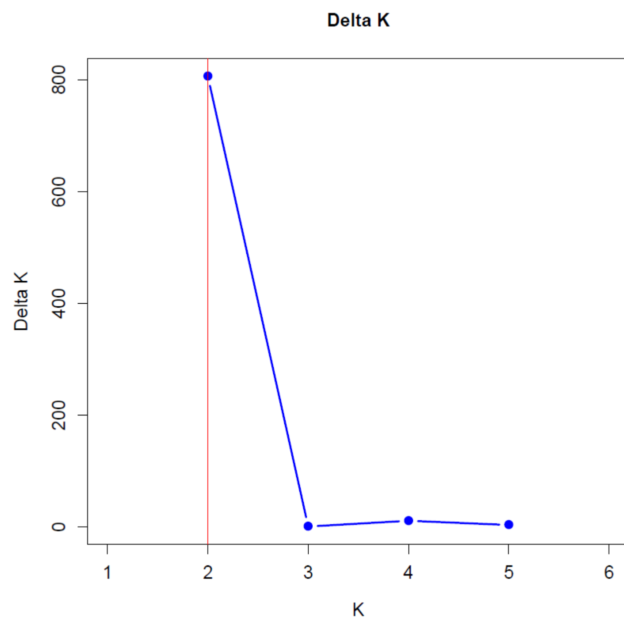
Genotype Frequency Class — Pure1 — Pure2 — Hybrid

Figure S4a-c. Results from Structure cluster analysis for optimal K using different approaches. Red lines in each figure shows the most supported solution for that particular approach.

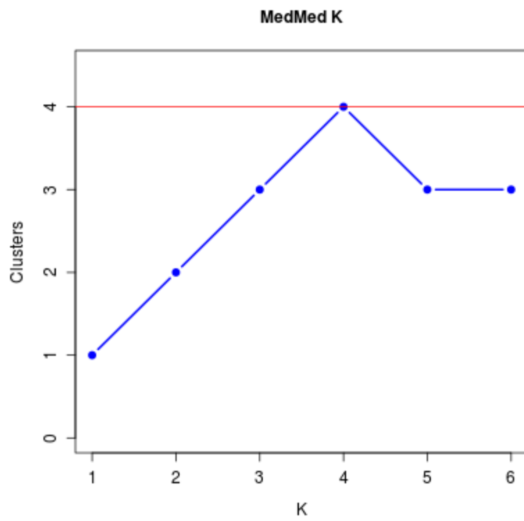
S4a. The most likely number of K (5) based on the log probability of data (lnPrX|K).

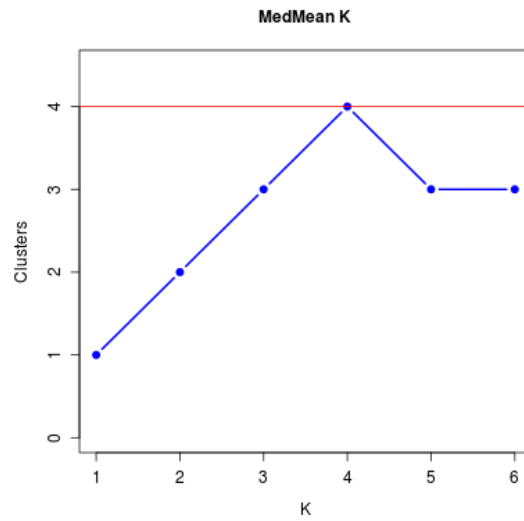


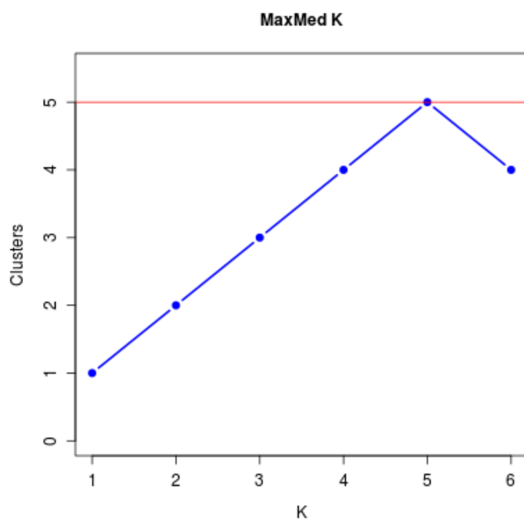
S4b. The most likely number of K (2) based on ΔK.



S4c. The most likely number of K (4 or 5) based on four different Puechmaille method calculations.







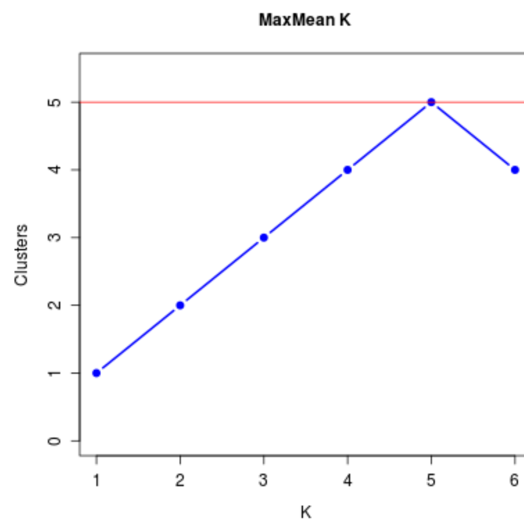
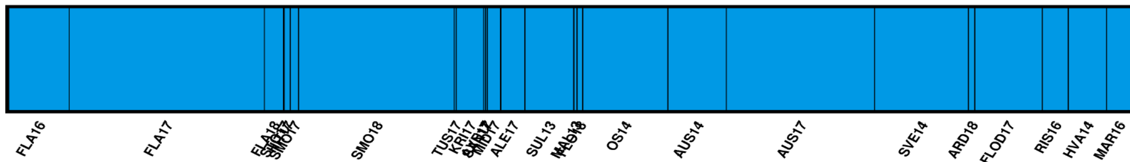
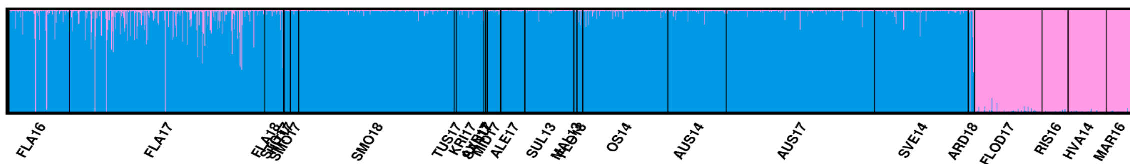


Figure S5a. Structure bar plots for K values from 1 to 6 without a priori.

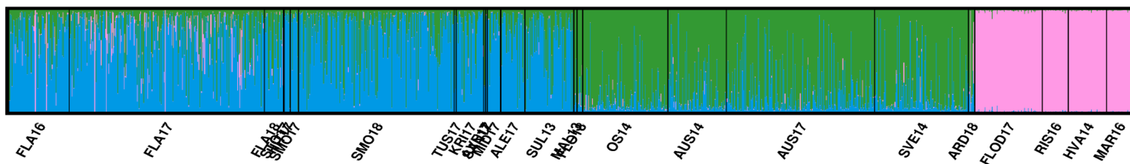
K=1



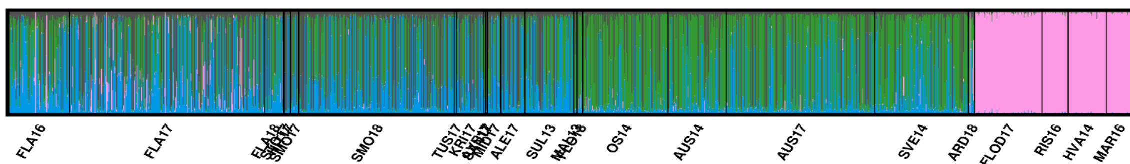
K=2



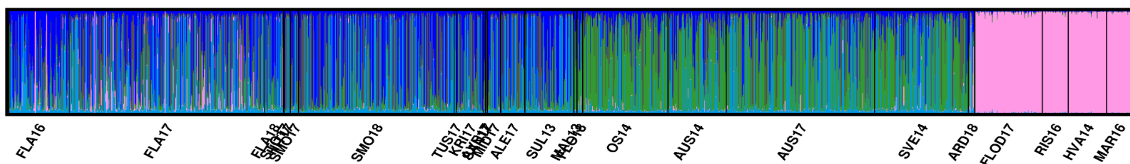
K=3



K=4



K=5



K=6

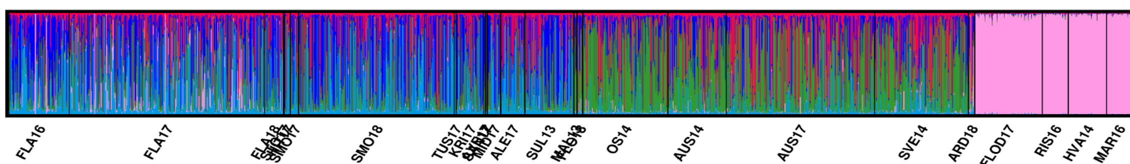
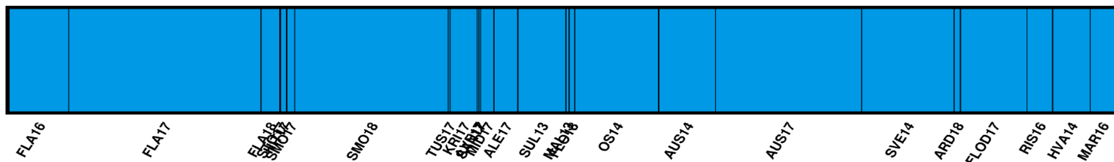
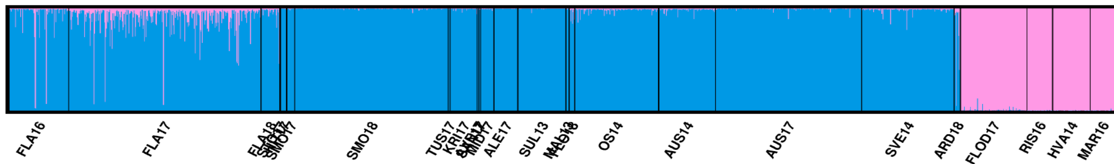


Figure S5b. Structure bar plots for K values from 1 to 6 with sampling location given as a priori.

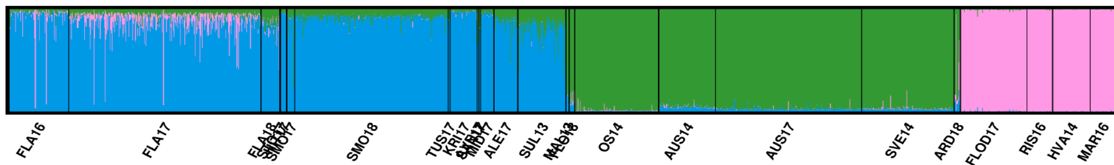
K=1



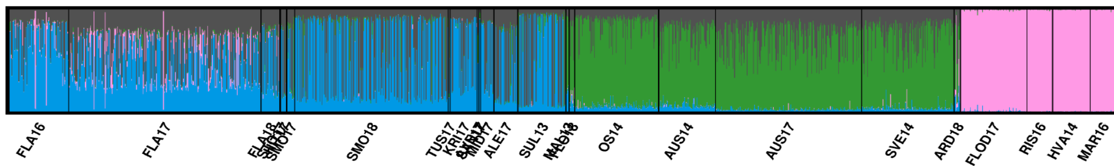
K=2



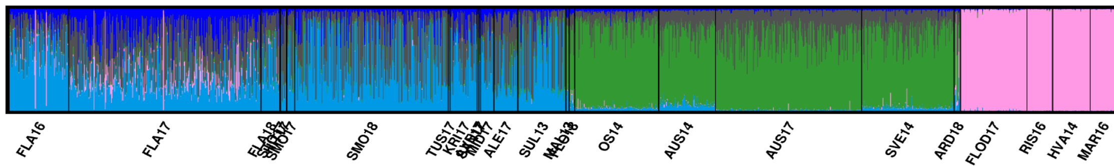
K=3



K=4



K=5



K=6

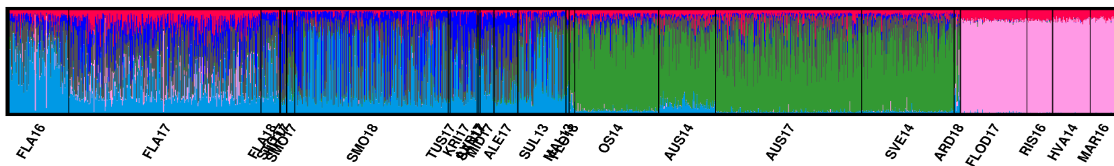


Figure S6. First (x-axis) and second (y-axis) component of a principal component analysis (PCA) on 1766 corkwing wrasse individuals based on 84 SNPs. The first component explains 26.5% of the total variation and the second 2.2%. Each point represents one individual and colours represent the different samples. For reference see Table 1

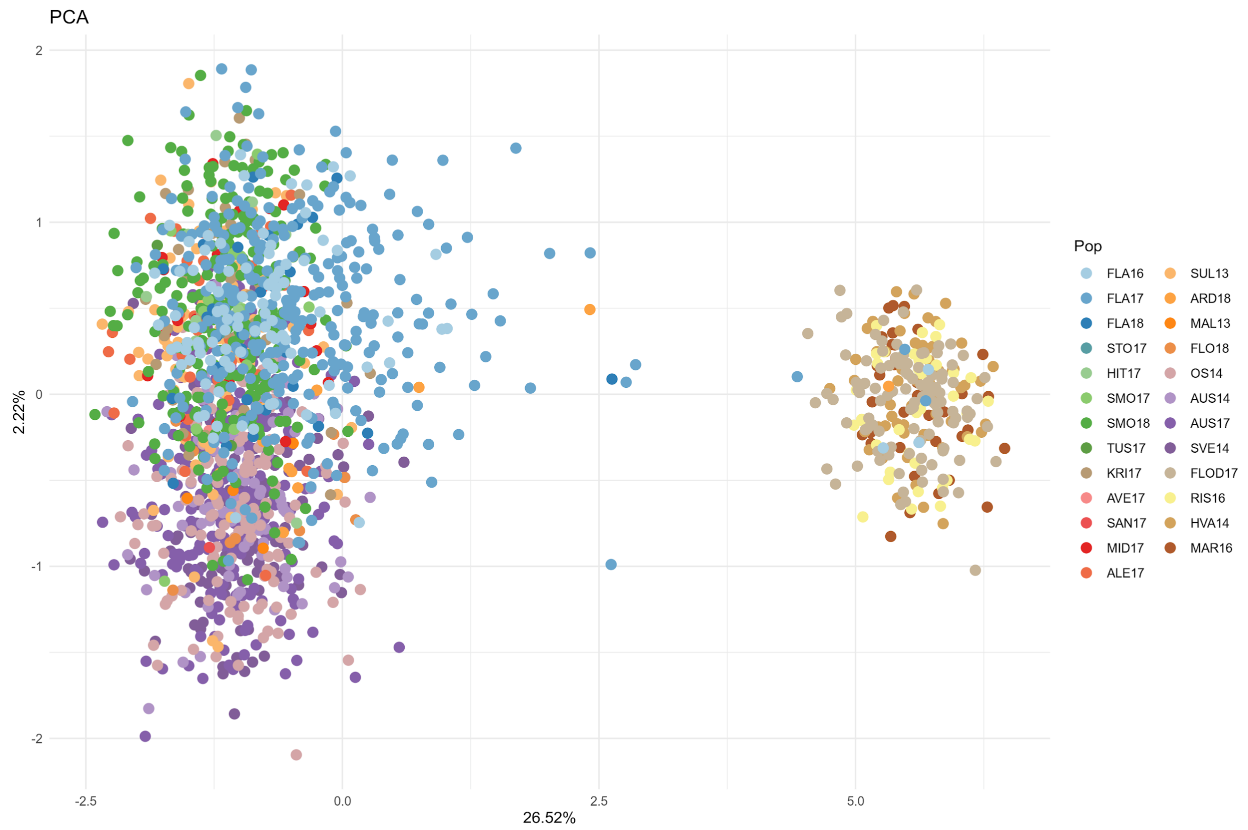
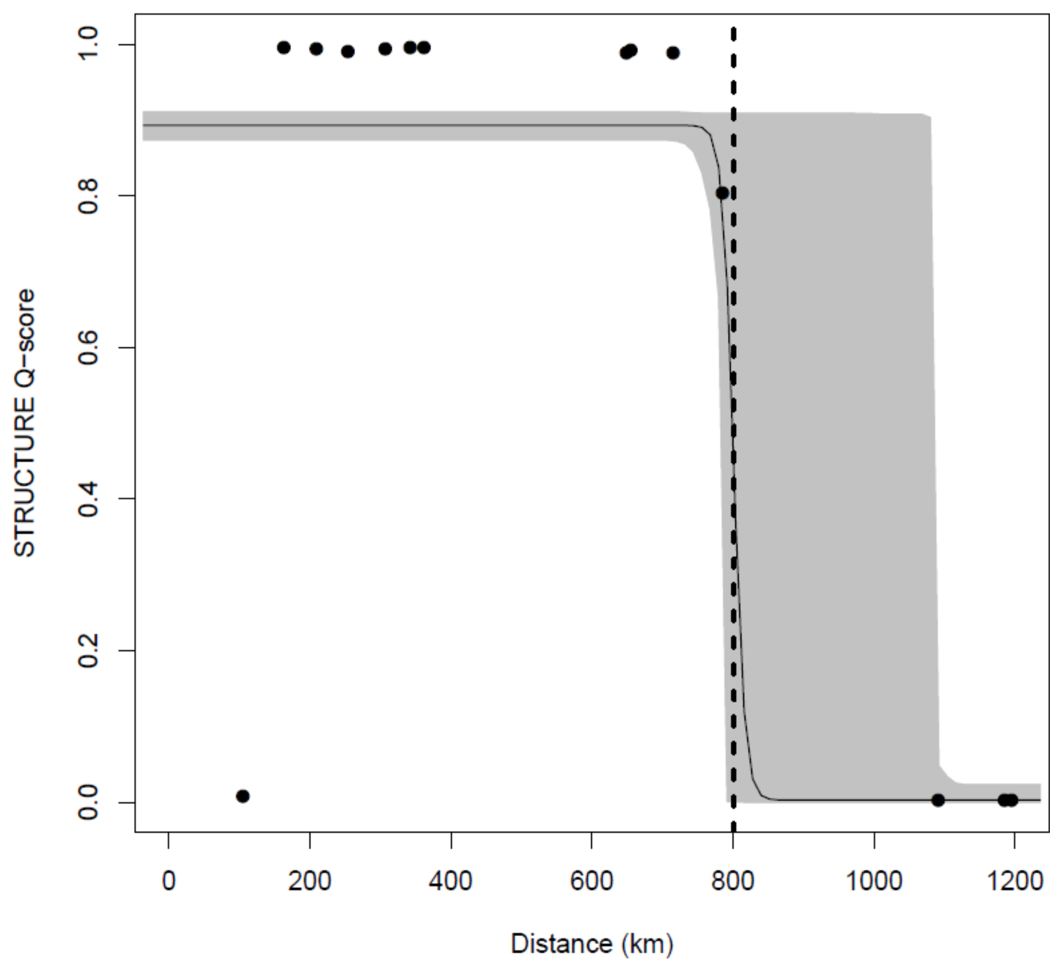


Figure S7. Geographical cline analysis for corkwing wrasse: a) Reference cline based on the STRUCTURE Q-score, and b) cline centres (and their support limits based on two log-likelihood units, in km) obtained by fitting curves for every SNP locus. Cline centres are measured as the distance along the 1200 km long transect ranging from Flatanger to Marstrand. The red dashed lines depict upper and lower values for STRUCTURE Q-score reference cline.

A)



B)

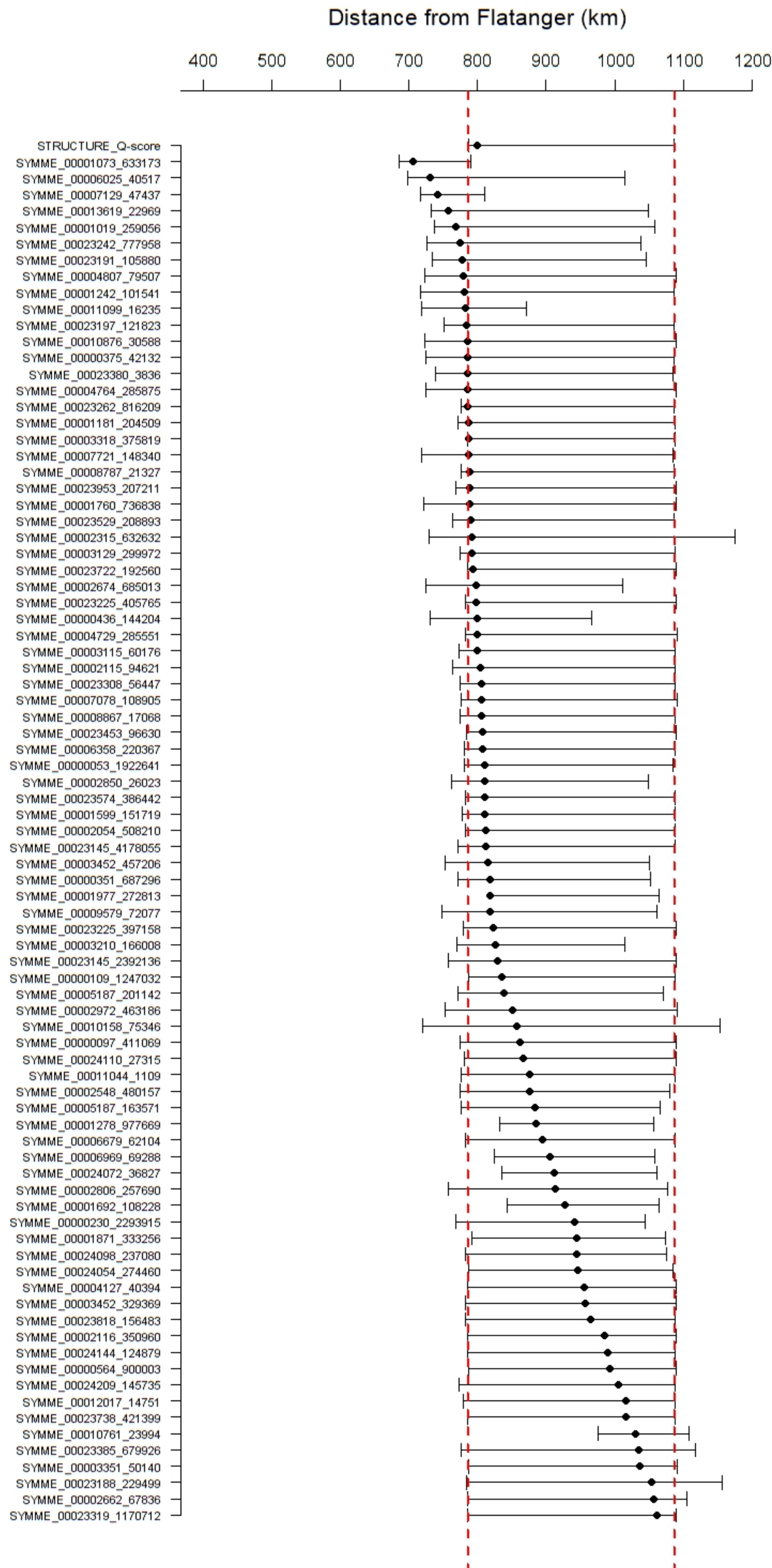
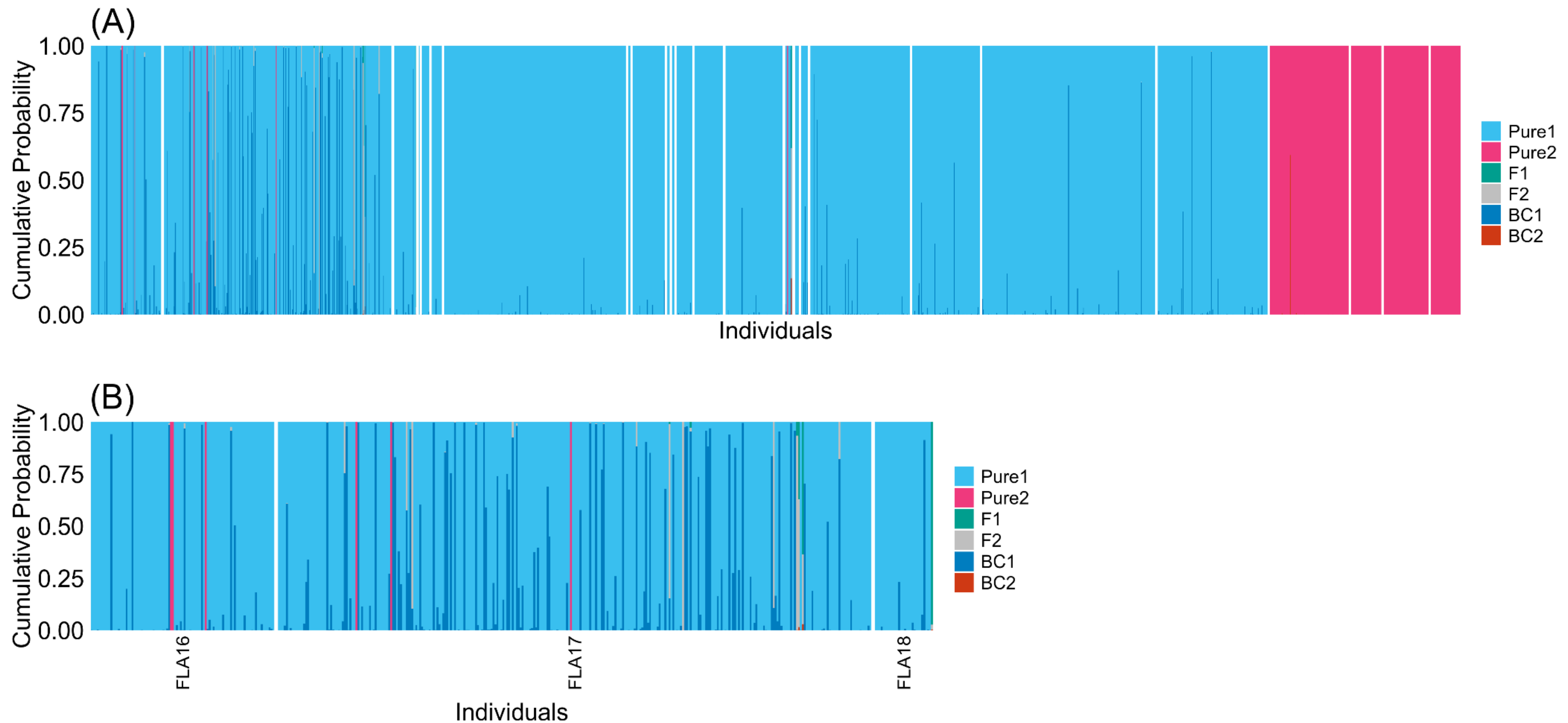


Figure S8. Hybrid analysis of (A) all 1766 individuals using 84 SNPs and (B) zoomed in on individuals sampled in Flatanger 2016, 2017, 2018 and Årdalsfjorden 2018. Each line represents one individual and its probability to belong to one of the six genotype classes: pure 1, pure 2, F1 hybrid, F2 hybrid, or backcrosses between F1 and pure 1 or pure 2. Out of the 432 individuals from Flatanger, we discovered six individuals with clear south-eastern genotypes, two first generation hybrids, and 70 potential second-generation hybrids. In Årdalsfjorden ten individuals were genotyped and one individual was classified as a south-eastern genotype, one as a potential second generation hybrid and one could not be assigned with > 50% probability.



Genetic Identification of Corkwing Wrasse Cleaner Fish Escaping from Norwegian Aquaculture

The use of wrasses as cleaner fish in salmon farms has increased exponentially over the last decade. Wild-caught fish are transported long distances to be used where local stocks of wrasses do not exist or cannot meet the demand. This thesis provides evidence that corkwing wrasse captured in Skagerrak and translocated to salmon farms off Trondheim escape the farms and hybridize with local populations at the leading edge of a natural ongoing range expansion. A significant fraction of the northern edge population are escapees or hybrids, but only few hybrids can be found in other areas along the Norwegian coast. A set of genetic markers was developed to aid future monitoring of wild populations and detection of escaping and hybridising individuals. Overall, these findings provide important information both for aquaculture management and conservation of wild populations, and have implications for the increasing use of cleaner fish as parasite control in fish farms.

Ellika Faust received her MSc in biology from the University of Gothenburg in 2017 and started her PhD in 2018.

ISBN978-91-7833-934-1 (TRYCKT)
ISBN 978-91-7833-935-8 (PDF)