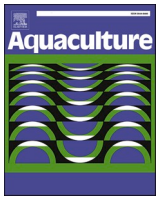
Aquaculture 555 (2022) 738230

Contents lists available at ScienceDirect

Aquaculture

journal homepage: www.elsevier.com/locate/aquaculture



Lack of population genetic structure of lumpfish along the Norwegian coast: A reappraisal based on EST-STRs analyses

Ólöf Dóra Bartels Jónsdóttir [a,1], Davíð Gíslason [b,c,a], Guðbjörg Ólafsdóttir [b], Simo Maduna [d], Snorre B. Hagen [d], Patrick Reynolds [e], Sæmundur Sveinsson [b], Albert K.D. Imsland [a,f,*]

[a] Akvaplan-niva Iceland Office, Akralind 4, 201, Kópavogur, Iceland
[b] Matís, Vínlandsleið 12, 113, Reykjavík, Iceland
[c] Anishinabec Ontario Fisheries Resource Centre, 191 Booth Rd. North Bay, ON P1A 4K3, Canada
[d] Norwegian Institute of Bioeconomy Research (NIBIO), Division of Environment and Natural Resources, P.O.Box 115, N-1431, Ås, Norway
[e] GIFAS AS, Gildeskål, 8140, Inndyr, Norway
[f] Department of Biological Sciences, University of Bergen, High Technology Centre, 5020 Bergen, Norway

ARTICLE INFO

Keywords:
Lumpfish
Population genetics
Aquaculture management
Expressed sequence tags (ESTs) markers

ABSTRACT

Lumpfish is now the single most important cleaner fish species to date and there is an extensive lumpfish translocation along the Norwegian coast. A reliable baseline information about the population genetic structure of lumpfish is a prerequisite for an optimal managing of the species to minimize possible genetic translocation and avoid possible hybridisation and introgression with local populations. The current study is a follow up of the study of Jónsdóttir et al. (2018) using expressed sequence tag-short tandem repeats (EST-STRs) markers. Samples ($N = 291$) were analysed from six sample locations along the Norwegian coastline from south to north, with additional 18 samples of first-generation (from wild fish) reared fish from a fish farm outside Tromsø (North Norway). Present findings show a lack of population differentiation among lumpfish sampling population along the Norwegian coast using EST-STRs, which is in accordance with the findings of Jónsdóttir et al. (2018) where genomic STRs (g-STRs) were analysed. Present findings indicate that should translocated lumpfish escape from salmon sea pens in Norway, this will probably have little impact on the genetic composition of the local lumpfish population.

1. Introduction

The biological control of sea lice in Atlantic salmon farming using cleaner fish has become a feasible alternative due to the increased occurrence of drug resistant lice, the reduced public acceptance of chemotherapeutic use in food production, and the urgent need for an effective and sustainable method of parasite control in Atlantic salmon aquaculture (Denholm et al., 2002; Boxaspen, 2006; Treasurer, 2018). Farmed cleaner fish are preferred due to better biosecurity through vaccination and screening programs, stocking at optimum times and sizes, and in reducing reliance on wild caught fish (Brooker et al., 2018). Cleaner fish are now used as a biological control for sea lice on farmed salmon in Europe and Canada (Imsland et al., 2014a, 2014b, 2014c; Imsland et al., 2021; Skiftesvik et al., 2014; Boyce et al., 2018; Powell et al., 2018). As lumpfish, *Cyclopterus lumpus,* tolerate lower temperatures than wrasse species, their implementation was boosted principally in the northern parts of Norway (Imsland et al., 2014a). In the beginning, the production of juveniles was solely based on wild-caught brood fish, where eggs were stripped, incubated, hatched, and reared to suitable size for transfer to commercial cages (Jonassen et al., 2018). With the demand of lumpfish steadily increasing, this has resulted in gradually moving from wild caught broodfish towards intensive cultivation. Lumpfish production is currently utilizing technology and techniques used in halibut, wrasse and cod aquaculture, and a breeding program for lumpfish was established in 2017 (Imsland et al., 2021). Although wild-caught broodfish still is predominant, nearly all juvenile lumpfish used in salmon sea-pens are from intensive aquaculture production in Norway. In 2019, more than 42 million lumpfish were transferred to cages at commercial marine growth sites for salmonids, which constituted 70% of all cleaner fish used in Norway (Norwegian

* Corresponding author at: Akvaplan-niva Iceland Office, Akralind 4, 201, Kópavogur, Iceland.
E-mail address: albert.imsland@akvaplan.niva.no (A.K.D. Imsland).
[1] **Equal authorship between Jónsdóttir, Gíslason and Imsland**

https://doi.org/10.1016/j.aquaculture.2022.738230
Received 27 February 2022; Accepted 5 April 2022
Available online 12 April 2022

Directorate of Fisheries., 2019). Lumpfish is now the single most important cleaner fish species to date. Given the intensive use of lumpfish along the coast of Norway it is imperative that their use is done with the aim of minimizing possible genetic translocation to avoid possible hybridisation and introgression with local populations. The prerequisite is to have reliable baseline information about the population genetic structure of lumpfish.

Previous studies on the genetic structure of lumpfish (Pampoulie et al., 2014; Garcia-Mayoral et al., 2016; Whittaker et al., 2018) reveal genetic difference on large geographic scale, but less or no genetic difference on small geographic scale, except for two lumpfish subpopulations found West of Greenland. In accordance with those findings, Jónsdóttir et al. (2018) found no indication of significant spatial genetic structuring or positive correlation between geographic and genetic distance among wild lumpfish along the coast of Norway. In all four studies (Pampoulie et al., 2014; Garcia-Mayoral et al., 2016; Whittaker et al., 2018; Jónsdóttir et al., 2018) microsatellites called Short Tandem Repeats (STRs), also known as Simple-Sequence Repeats (SSRs) were used. Microsatellite markers are indeed divided into two types, namely, genomic STRs (g-STRs) and expressed sequence tag (EST) STRs, which originate from arbitrary genome sequencing and RNA transcriptome sequencing, respectively (Wu et al., 2020). Genomic microsatellites (g-STRs) are anonymous markers, whereas EST-STRs are functional molecular markers with the advantages of easier and more efficient development, lower cost and more interspecific transferability (Bouck and Vision, 2010; Wu et al., 2020). Recent research revealed that ESTs are a potentially rich source of STRs that reveal polymorphisms not only within the source taxon, but in related taxa, as well (Ellis et al., 2006). However, EST also have clear potential for use in basic evolutionary applications, such as population genetic analyses (Ellis et al., 2006). In plant studies EST and g-STRs have showed comparable results of population genetic structure in cases of strong genetic differentiation, and genomic STRs have performed slightly better that EST when differentiation is moderate. However, EST had a higher power to detect weak genetic structure compared to g-STRs (Khimoun et al. 2016). The biological function of STR-containing EST unigenes in lumpfish has been assigned into three categories, namely biological process (BP), cellular components (CC) and molecular function (MF) (Maduna et al., 2020). Within the BP category, genes involved in cellular, metabolic and biological regulation comprised the largest portion, while in the CC category the greatest number of genes were found to encode cellular components and cell parts. Likewise, many sequences in the MF category encode proteins with binding and enzymatic activity (Maduna et al., 2020).

EST may be valuable molecular markers for conservation genetic studies in taxa where the development of genomic STRs is impeded by low frequency (Khimoun et al., 2017). Beyond providing more statistical power in paired comparisons, EST also produce cleaner results for scoring as there are fewer null alleles and fewer stutter bands (Khimoun et al., 2017). As EST-STRs are found within transcribed genes, they may not follow neutral expectations and accordingly have higher probability of association to phenotypic effects or mutations, therefore, they may be particularly useful for looking at population differentiation related to local adaptation (Teacher et al., 2012; Postolache et al., 2014; the et al., 2017).

The present study is a follow up study of the Jónsdóttir et al. (2018) study and is done to further examine the population structure of the Norwegian lumpfish using genotypic data derived from EST-STRs, which has shown to have higher power to detect weak genetic structure compared to g-STRs (Khimoun et al., 2017). If in fact, there is a genetic structure to be found within the population of lumpfish along the Norwegian coast, but too weak to be detected with g-STRs alone, we believe that EST-STRs would reveal the structure. To circumvent sampling and temporal variation related issue between the present study and Jónsdóttir et al. (2018), the same samples used in the latter study were analysed in the present study.

In the present study our aim is three-fold. Firstly, to develop EST-STRs genetic markers for lumpfish. Secondly, to use gen-linked markers that may be subject to selection to examine the population structure of lumpfish along the Norwegian coast. Thirdly, to compare the results using gen-linked markers to the results using markers not connected to transcribed genes. The results will provide baseline information to assist in the design of sustainable lumpfish population management and aqua-farming policy development. In relation to the ongoing extensive lumpfish translocation along the Norwegian coast, our study is vital for assessing the risk this practice holds for the wild population and for devising provisions that aim to minimize negative impact and ensure the long-term viability of the species.

2. Materials and methods

2.1. Sampling areas and protocols

The sampling area extended from Mandal (58°N) in south Norway along the Norwegian cost up to 69°N in the north (Fig. 1). A total of 291 specimens were collected at six fishing grounds, with additional 18 samples of first-generation reared fish from a fish farm (Troms Marin Yngel, Kvaløya, Tromsø, Table 1). All wild individuals were collected by a beach seine and a small tissue sample (fin clip) was collected from each fish and stored in 96% ethanol prior to DNA extraction.

DNA was isolated from all samples using the DNeasy Blood and Tissue Kit following the manufacturer's instructions (Qiagen). DNA quality and quantity were determined with an Epoch Microplate Spectrophotometer (BioTek) prior to genotyping. A total of seventeen EST microsatellite loci were genotyped (*Clu_E03, Clu_E04, Clu_E10, Clu_E11, Clu_E12, Clu_E15, Clu_E20, Clu_E25, Clu_E29, Clu_E34, Clu_E35, Clu_E36, Clu_E37, Clu_E38, Clu_E42, Clu_E45, Clu_E47*. Table 2). Primers were developed following the procedure outlined in Maduna et al. (2020). Five multiplexes PCR reactions were performed in a 10 μl volume containing 1 μl DNA (1–30 ng/μl), 0.8 μl of dNTP (10 mM), 0.75 U Taq polymerase (New England Biolabs Ltd.), 1 μl of 10× Standard Buffer (New England Biolabs Ltd.), 0.08–0.3 μl of a 50:50 ratio of fluorescent dye labelled forward (100 μM) and reverse (100 μM) primer tagged on the 5′-end with a GTTTCTT PIG-tail to enhance PCR amplification (Brownstein et al., 1996). PCR reactions were performed on a Tetrad2 Peltier thermal cycler (BioRad) as follows: initial denaturation step of 4 min at 94 °C followed by 30 cycles of 30 s at 94 °C, 40 s at 58 °C and 40 s at 68 °C, and a final elongation step of 7 min at 68 °C. Samples were analysed on an ABI PRISM 3730 sequencer using the GeneScan-500 LIZ size standard and genotyped with GeneMapper v4.1 (Applied

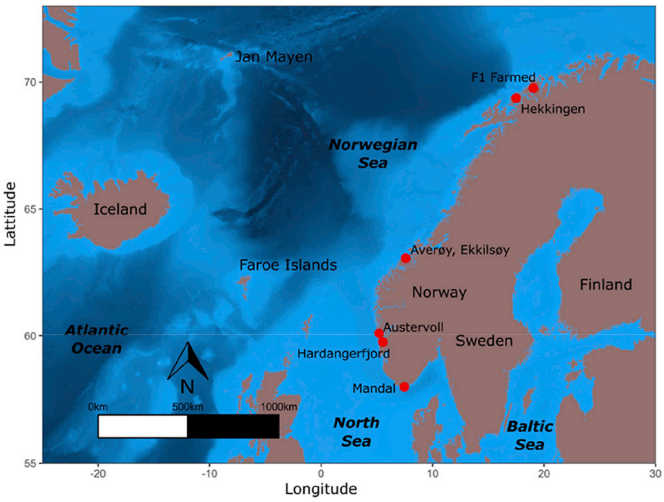


Fig. 1. Sampling location of lumpfish along the Norwegian coast. See Table 1 for details.

Table 1
Sampling locations for lumpfish along the Norwegian coast and from F1 farmed fish.

Sample location	N	L	Latitude	Longitude
Mandal	50	33.8 ± 7.2	N 57.99	E 7.48
Hardangerfjord	42	12.9 ± 9.6	N 59.75	E 5.55
Austevoll	6	29.6 ± 7.0	N 60.10	E 5.19
Averøy (Håholmen)	50	40.2 ± 5.7	N 63.01	E 7.23
Ekkilsøy	50	34.0 ± 5.9	N 63.07	E 7.33
Hekkingen	93	41.3 ± 3.6	N 69.37	E 17.48
F1 farmed fish (Kraknes)	18	23.2 ± 1,2	N 69.76	E 19.05

N = sample size, L = average size (cm) ± standard deviation (SD) of sampled fish.

Biosystems).

2.2. Functional annotation of EST-STR-containing unigenes

Respective EST-STR-containing unigenes were obtained from Maduna et al. (2020) and annotated using BLASTn search against the draft genome of the lumpfish [Entrez query: lumpfish (taxid:8103)] from the NCBI Genome database under BioProject PRJNA625538 (accession number GCF_009769545.1, fCycLum1.pri), the Vertebrate Genomes Project (G10K Consortium) (Table 3).

2.3. Statistical analyses

The EST Microsatellite data set was checked in MICRO-CHECKER v.2.2.3 (Van Oosterhout et al., 2004) for scoring errors due to stuttering and large allele drop out. LOSITAN (Beaumont and Nichols, 1996; Antao et al., 2008) was used to test for neutrality of all EST microsatellite markers. Expected (H_{exp}) and observed (H_{obs}) heterozygosity for samples was estimated using GENETIX 4.02. (Belkhir et al., 2004). Allelic richness (A_r) was estimated using rarefaction methods to correct for unequal sample sizes implemented in ADZE 1.0 (Szpiech et al., 2008). A_r represents expected number of distinct alleles in random sample for standardized sample size g, where g is the number of randomly drawn gene copies. Here g was set to maximum 36 reflecting the lowest sample size for the farmed fish ($N = 18$). It was not possible to estimate allelic richness for Austevoll which only have sample size of 6 individuals. Further it was only possible to estimate A_r for g = 34 for the farmed fish because of missing data. Test for linkage disequilibrium was conducted in GENEPOP 4.7.0 (Rousset, 2008) using the log likelihood ratio statistics (G-test) with 10,000 dememorization, 100 batches and 5000 iterations per batch. GENEPOP was also used for Hardy Weinberg probability test for each locus in each sample using the same number of dememorizations, batches and iterations as for linkage disequilibrium test. Significance of both the linkage disequilibrium and Hardy Weinberg test were corrected using Benjamini and Yekutieli (2001) correction for false discovery rate for multiple tests using p.adjust from package stats in R using false discovering rate (FDR) threshold of 0.05. Test for genotypic differences between samples were done by estimating pairwise F_{ST} (theta, Weir and Cockerham, 1984) value across loci and testing the significance of samples differences using Fisher's exact test implemented in GENEPOP. P value for these tests were corrected using Benjamini and Yekutieli (2001) correction. A package diveRsity in R was used for estimating 95% confidence limits for F_{ST}.

To assess the potential numbers of genetic clusters a Bayesian cluster analysis was performed in STRUCTURE (Pritchard et al., 2000). STRUCTURE was used for only EST-STR data on the one hand, and for combined EST-STR and g-STR data on the other hand. STRUCTURE was run for 10 independent runs of 200,000 iterations for burn-in and 1,000,000 iterations after for $K = 1$ to 10 with 10 repeats. STRUCTURE model assumed an admixed model with correlated allele frequencies. Because relatively no/weak signal of structure was detected, the *LOCPRIOR* model was applied that utilises sampling locations as prior information to assist the clustering and assessed informativeness of the sampling location prior with the $r_{locprior}$ (Hubisz et al., 2009). To

Table 2
Multiplex composition, repeat motifs of alleles, primer dye and concentrations (μM), allele range and numbers, observed (H_{obs}) and expected heterozygosity (H_{exp}) and sequences of primers. A 5' GTTTCTT tail was added to all the reverse primers.

Locus	Repeat motif	Fluorescent dye/label	Concentration (μM)	Allele range	No Alleles	H_{obs}	H_{exp}	Forward sequence	Reverse sequence
Multiplex 1									
Clu-E03	CTTT	FAM	1,0	128–204	18	0.73	0.75	AGATACTTAGCAATACTCACACG	CAAAATTCCACTTGAACAGAA
Clu-E10	GGAA (CG)n	PET	3,0	128–200	18	0.79	0.8	GAGAGAAGAACATCCACGAT	CTCCTTGCATTCTTGTTTTT
Clu-E12	(*CA*)n	NED	1,5	100–110	6	0.52	0.54	TTGAAAATCCCTGAAAGTACA	GCAACCATGAGTGATTATTGT
Clu-E34	ACA	NED	2,0	138–171	12	0.65	0.64	TTAAACATACCGACAGTAGCC	CGGATTTAGTCATGAAACGTA
Multiplex 2									
Clu-E04	GT	FAM	1,0	156–172	8	0.7	0.69	TTCTCAATGGTAAGAGCATGT	GACCTCAGCCAGATTACTGT
Clu-E36	AC (AAC)n	PET	2,0	145–197	22	0.93	0.92	CCCCTACAGAGTGATGAGC	GCTGGAAACATGTATAAGTACG
Clu-E47	(*AGC*)n	VIC	0,8	111–126	6	0.61	0.59	GAAGCAGAGGTCCAGTCTAC	CAAATCTGACAGCCCATC
Multiplex 3									
Clu-E11	CCG	PET	2,0	244–265	7	0.49	0.62	TCCATGTCGTGTTAATTTCTT	AAATGAATGGCGGTGGAG
Clu-E15	TGT (GGA)n	FAM	1,5	123–160	6	0.14	0.15	ACGACCTCGTCTTGAGTCTTA	TGACGAATATTTTCAGTGCAT
Clu-E25	(*AGA*)n	PET	2,0	108–186	21	0.76	0.81	CAGGACGTACATGAGGAGATA	AGCTTCACGTTCAGAGCTT
Multiplex 4									
Clu-E20	TTACA	PET	1,5	147–162	4	0.06	0.06	TGTTGTATTTCCTTCTGTTCC	GATAAAAGCGTCAGCTAAATG
Clu-E37	CAA	VIC	1,0	133–154	8	0.66	0.7	CATGCTTTGCTAGTTTGTTTT	GAAAGAAAATCAGGAATGGTC
Clu-E42	GTAG	NED	1,5	164–168	2	0.02	0.04	TAGTTGGAATTCTTGCTTCAA	CCTACCTACCTACCTACCTACG
Clu-E45	AC	FAM	0,5	235–255	11	0.71	0.72	TGATCAAGCACACTAAAGACTG	TACTTAGGGCCTAGGCTACAT
Multiplex 5									
Clu-E29	TAT	PET	2,0	194–215	8	0.68	0.69	CACAACGTTTCTGTCAGCTAT	TCATGACCATGTGGTCTTATT
Clu-E35	GT	NED	2,0	146–150	3	0.46	0.5	CAAAAATAGCAGCGTTAAAGA	TTACTCAACTCGCCTAAAAGA
Mean					9.82				

Table 3
Annotation of EST-STR-containing unigenes.

Loci	GenBank Acc.	BLASTn Hit			
		Chromosome GenBank Acc.	Chromosome	Unigene position (bp)	Candidate gene or closest genes to the EST-STR
Clu_E03	MW962126	NC_046977.1	12	3,904,711–3,905,398	kynurenine–oxoglutarate transaminase 1; n-lysine methyltransferase kmt5a-a
Clu_E04	MW962119	NC_046974.1	9	12,978,235–12,978,992	inositol polyphosphate-5-phosphatase a
Clu_E10	MW962127	NC_046988.1	23	15,442,442–15,442,608	cugbp elav-like family member 2 isoform x1; scm-like with four mbt domains protein 2 isoform x1
Clu_E11	MW962113	NC_046987.1	22	13,434,951–13,435,533	fanconi anemia group m protein; heme-binding protein soul3
Clu_E12	MW962114	NC_046967.1	2	8,598,459–8,599,209	e3 ubiquitin-protein ligase midline-1-like
Clu_E15	MW962120	NC_046987.1	22	21,422,626–21,423,031	sphingosine-1-phosphate phosphatase 1; wd repeat-containing protein 89
Clu_E20	MW962128	NC_046972.1	7	3,742,381–3,742,792	slit-robo rho gtpase-activating protein
Clu_E25	MW962115	NC_046969.1	4	27,622,803–27,623,503	scl-interrupting locus protein homolog
Clu_E29	MW962121	NC_046968.1	3	9,330,611–9,331,067	bromodomain adjacent-zinc finger domain protein 2b
Clu_E34	MW962117	NC_046973.1	8	2,740,031–2,740,392	myosin phosphatase rho-interacting protein-like; retinoic acid-induced protein 1
Clu_E35	MW962122	NC_046972.1	7	19,080,325–19,080,950	zinc finger protein 568-like
Clu_E36	MW962123	NC_046986.1	21	21,109,040–21,109,332	ceramide kinase-like protein
Clu_E37	MW962124	NC_046967.1	2	4,430,568–4,431,110	transmembrane protein 182-like; actin-related protein 3
Clu_E42	MW962125	NC_046979.1	14	1,795,125–1,795,580	solute carrier family 22 member 5-like isoform x1; interferon regulatory factor 1b
Clu_E45	MW962118	NC_046981.1	16	11,015,486–11,015,872	low quality protein: run and sh3 domain-containing protein 1; histone-lysine *n*-methyltransferase ash1l-like isoform x1
Clu_E47	MW962116	NC_046969.1	4	25,395,315–25,395,631	sun domain-containing ossification factor isoform x3; sun domain-containing ossification factor isoform x2

determine the number of populations results from STRUCTURE runs were scrutinized in STRUCTURESELECTOR web server to estimate the most likely number of populations (*K*) using the Puechmaille method (Puechmaille, 2016). Data was then visualized using the CLUMPAK server (Kopelman et al., 2015).

The R package ADEGENET v.2.1.3 (R core team, 2014; Jombart, 2008) was used to perform DAPC on clusters pre-defined by sampling location using the *dapc()* function. For DAPCs without spatial prior information, we inferred the number of discrete genetic clusters using the *find.clusters()* function in ADEGENET, which runs successive *K*-means clustering with an increasing number of clusters (*k*). Ten independent runs of the *find.clusters* function was performed with the *diffNgroup* option selected to identify the sharp changes in fit of models with different number of clusters based on the Bayesian information criterion (BIC) score. Then, selection of optimal *k* was done by applying the BIC score as recommended by Jombart et al. (2010). For all DAPC analyses, we determined the number of principal components to retain using the cross-validation approach implemented by the function *xvalDapc()* with 100 repetition in ADEGENET.

The relative importance of geographical distance (isolation-by-distance) and genetic distance assessed as F_{ST} $(1\text{-}F_{ST})^{-1}$ was examined using the Mantel test (Mantel, 1967) implemented in library ape v5.4–1 in R (Paradis and Schliep, 2019), setting the number of permutations to 999. Pairwise geographic distances between sampling sites were calculated following the coastline (range of distances between 63 and 1850 km).

3. Results

3.1. Genetic diversity

Test for scoring errors of the microsatellite markers in MICRO-CHECKER indicated that one locus, Clu_E38, had scoring error and lack of heterozygotes, thus this locus was removed before further analyses. Test for loci under selection using LOSITAN indicated that one locus, Clu_E11, might be under selection, however we kept this locus in all calculations. All EST microsatellite markers were highly polymorphic with number of alleles ranging from 2 for Clu_E42 to 22 for Clu_E36, and with mean number of alleles over all markers of 10.0 (Table 2). Observed (H_{obs}) and expected heterozygosity (H_{exp}) was generally high but variable with lowest heterozygosity of 0.02 for loci Clu_E42 and highest of 0.93 for loci Clu_E36 (Table 2). For all samples mean number of alleles ranged from 3.81 in Austevoll, which has very low sample size, to 8.63 for Hekkingen (Table 4). Observed and expected heterozygosity was even over all samples, where both observed and expected heterozygosity was lower for g-STR dataset, and there was significant difference in mean observed and expected heterozygosity between g-STR and EST-STR datasets (TukeyHDS test in R). Allelic richness (A_r) was also even over samples ranging from 4.71 for the farmed fish to 5.81 for Hardangerfjord (Table 4).

Test for linkage disequilibrium over all samples was significant in 7 tests of 120 (5.8%), no test was significant after correction of *P* values for multiple tests. Test for HWE for each population and loci were never significant after correction of P values for multiple tests. Based on these results all samples and loci were used for further analyses of genetic structure.

3.2. Genetic structure

Samples genotypic differentiation estimated as pairwise F_{ST} between samples over all loci was low but significant for few samples mostly involving the F1 farmed sample. The F1 farmed samples were significantly different from all wild samples (Table 5). Only in one case were wild samples significantly different as the samples from Mandal and Hekkingen were significantly different (Table 5).

Test for population structure using both location of samples (LOCPRIOR model) and not using the location of samples, show that there is no population structure in lumpfish along the Norwegian coast (Fig. 2). The Puechmaille method implemented in STRUCTURESELECTOR estimated that the most likely number of populations were two. All estimators MedMed K, MedMean K, MaxMed K and MaxMean K showed K = 2 as most likely. However, the structure graph showed that all individuals had ~0.5% probability of belonging to the two clusters. When there is no population structure typically the proportion of the sample assigned to each population is roughly symmetric (~ 1/K in each population), and most individuals will be fairly admixed. By using median value of Ln(Pr Data) (Pritchard et al., 2000) the k for which Pr(K = k) is 1 had the highest probability whereas K = 2 to 10 had zero probability. Same results were obtained when using combined data for EST-STR and g-STR. Discriminant analysis of principal components (DAPC) were used, at first we determined the number of clusters de novo using function find.clusters() which determined that 5 clusters were most likely for the EST-STR and 8 for the g-STR data. However, when 70 PC and 6 discriminant factors were retained for DAPC, all samples largely overlapped indicating no population structure (Fig. 3). Similar results were obtained for g-STR data set (Jónsdóttir et al., 2018) where 80 PC and 6 discriminant factors were retained (Fig. 3). In accordance with low

Table 4
Genetic variation for lumpfish samples, showing results for both g-STRs (Jónsdóttir et al., 2018) and EST-STRs (present study).

Population	G-STRs (Jónsdóttir et al., 2018)					EST-STRs (This study)					G-STRs+EST-STRs				
	Na	H_{obs}	H_{exp}	A_r	F_{IS}	Na	H_{obs}	H_{exp}	A_r	F_{IS}	Na	H_{obs}	H_{exp}	A_r	F_{IS}
Mandal	8.0	0.53	0.51	5.7 ± 1.1	−0.02	6.88	0.595	0.583	5.51 ± 0.81	−0.01	7.4	0.56	0.55	5.55 ± 0.63	−0.016
Hardangerfjord	6.6	0.46	0.49	5.2 ± 1.1	0.06	7.19	0.554	0.581	5.81 ± 0.92	0.06	6.9	0.51	0.54	5.55 ± 0.68	0.063
Austevoll	5.3	0.48	0.42	NA	−0.06	3.81	0.51	0.534	NA	0.134	3.6	0.50	0.48	NA	0.054
Averøy (Håholmen)	7.5	0.53	0.54	5.7 ± 1.1	0.02	7.38	0.543	0.57	5.67 ± 0.87	0.057	7.4	0.54	0.55	5.66 ± 0.68	0.039
Ekkilsøy	6.9	0.5	0.51	5.2 ± 0.9	0.03	7.06	0.538	0.557	5.60 ± 0.86	0.044	7.0	0.52	0.54	5.42 ± 0.62	0.038
Hekkingen	8.6	0.51	0.50	5.4 ± 1.1	−0.02	8.63	0.559	0.574	5.67 ± 0.82	0.031	8.6	0.54	0.54	5.54 ± 0.64	0.011
F1 Farmed fish	4.7	0.48	0.49	4.7 ± 0.7*	0.02	4.75	0.555	0.525	4.71 ± 0.74*	−0.03	4.7	0.52	0.50	4.69 ± 0.51*	−0.007

Average number of alleles (N_a), observed (H_{obs}) and expected (H_{exp}) heterozygosity, allelic richness (A_r) based on rarefaction method using a minimum sample size (i.e. number of alleles) of 36, F_{IS}, the within-population inbreeding coefficient.
* Sample size (i.e. number of alleles) of 34 for F1 farmed fish used for calculation of allelic richness because of missing data.

values of differentiation observed using above statistical tests, there was no correlation between geographical and genetic distance ($R^2 = 0.0015$, $P = 0.89$) and thus no indication of Isolation-by-distance. (Fig. 4).

4. Discussion

The high conserved microsatellites that have recently been developed from expressed sequence tags (EST) in lumpfish were used in the present study to attain a better understanding of population structure of lumpfish along the coast of Norway, especially because of its role as a lice-eating cleaner fish in salmon aquaculture. In accordance with previous results (Jónsdóttir et al., 2018) using anonymous microsatellites (g-STRs), the present findings, using EST-STRs, show a lack of population divergence among lumpfish along the Norwegian coast ranging from Mandal in the south (58°N) to Hekkingen (69°N) in the north (Fig. 1). The exact same specimens of lumpfish were used in both

Table 5
F_{ST}, *P* values (GENEPOP, Weir and Cockerham, 1984) and 95% CI (diveRsity í R) for EST-STR data (current study) and g-STR (Jónsdóttir et al., 2018, for comparison). 95% CI show differences between samples for both markers. P values were adjusted for multiple tests.

Samples Pairwise tests	EST-STR				g-STR			
	F_{ST}	Lower 95% CI	Upper 95% CI	*P*	F_{ST}	Lower 95% CI	Upper 95% CI	*P*
Mandal vs. Hardangerfjord	−0.0007	−0.0078	0.0094	1	0.0007	−0.0073	0.0098	0.5602
Mandal vs. Austevoll	−0.0046	−0.0384	0.0568	1	0.0047	−0.0296	0.0616	1
Mandal vs. Averøy (Håholmen)	0.0041	−0.0036	0.0144	0.3738	0.001	−0.0062	0.01	0.8629
Mandal vs. Ekkilsøy	0.0005	−0.0062	0.0091	0.7042	0.0039	−0.0034	0.0138	**0.0052**
Mandal vs. Hekkingen	0.002	−0.0031	0.009	**0.0331**	−0.0004	−0.0054	0.0062	1
Mandal vs. F1 Farmed fish	0.0147	5e-04	0.0349	**0.0001**	0.0186	0.0029	0.0396	**0.0004**
Hardangerfjord vs. Austevoll	−0.0157	−0.0501	0.0477	1	−0.0076	−0.0414	0.0461	1
Hardangerfjord vs. Averøy (H.)	0.0013	−0.0051	0.0133	1	0.0044	−0.0051	0.016	0.2282
Hardangerfjord vs. Ekkilsøy	−0.0013	−0.0091	0.0092	1	−0.0003	−0.0091	0.0104	1
Hardangerfjord vs. Hekkingen	0.0025	−0.0053	0.0129	1	0.0016	−0.0045	0.0098	0.1661
Hardangerfjord vs. F1 Farmed f.	0.0216	0.0046	0.0444	**0.0015**	0.0185	0.0013	0.04	**0.0013**
Austevoll vs. Averøy (H.)	0.0066	−0.0294	0.0682	1	0.0115	−0.0216	0.0617	1
Austevoll vs. Ekkilsøy	−0.0073	−0.0431	0.0573	1	0.0133	−0.0226	0.0634	1
Austevoll vs. Hekkingen	−0.0037	−0.0384	0.0624	1	0.0086	−0.0227	0.0558	1
Austevoll vs. F1 Farmed fish	0.0172	−0.0266	0.0973	**0.0332**	0.0047	−0.0365	0.0632	1
Averøy (H.) vs. Ekkilsøy	0.002	−0.006	0.0123	1	0.0014	−0.0058	0.011	0.1470
Averøy (H.) vs. Hekkingen	0.0076	1e-04	0.0175	1	0.0017	−0.0042	0.0098	1
Averøy (H.) vs. F1 Farmed fish	0.0314	0.011	0.0571	**0.0000**	0.0204	0.0032	0.0421	**0.0013**
Ekkilsøy vs. Hekkingen	0.0019	−0.004	0.0098	1	0.0000	−0.0053	0.0065	0.6242
Ekkilsøy vs. F1 Farmed fish	0.0191	0.0028	0.0412	**0.0002**	0.0185	0.0003	0.0438	**0.0007**
Hekkingen vs. F1 Farmed fish	0.0149	9e-04	0.0332	**0.0002**	0.0133	−0.0011	0.0331	**0.0022**

Numbers in bold indicate statistically significant tests.

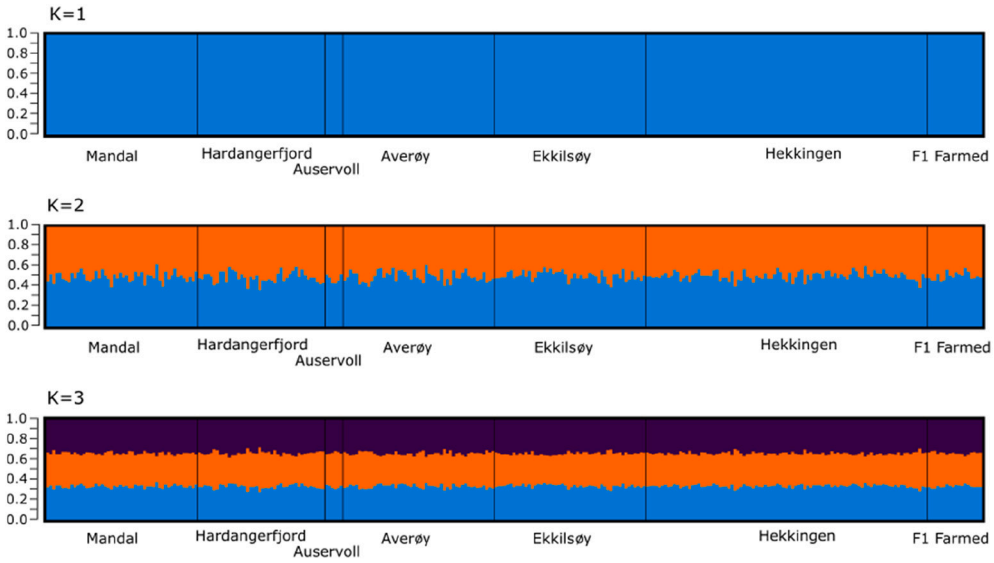


Fig. 2. Bayesian cluster analysis conducted in STRUCTURE (Pritchard et al., 2000). Shown are clustering for K = 1 to 3 for seven samples of lumpfish collected in costal Norway. Within each plot, vertical bars represent individuals while colours indicate the different clusters detected. K = 1 was the most likely number of populations.

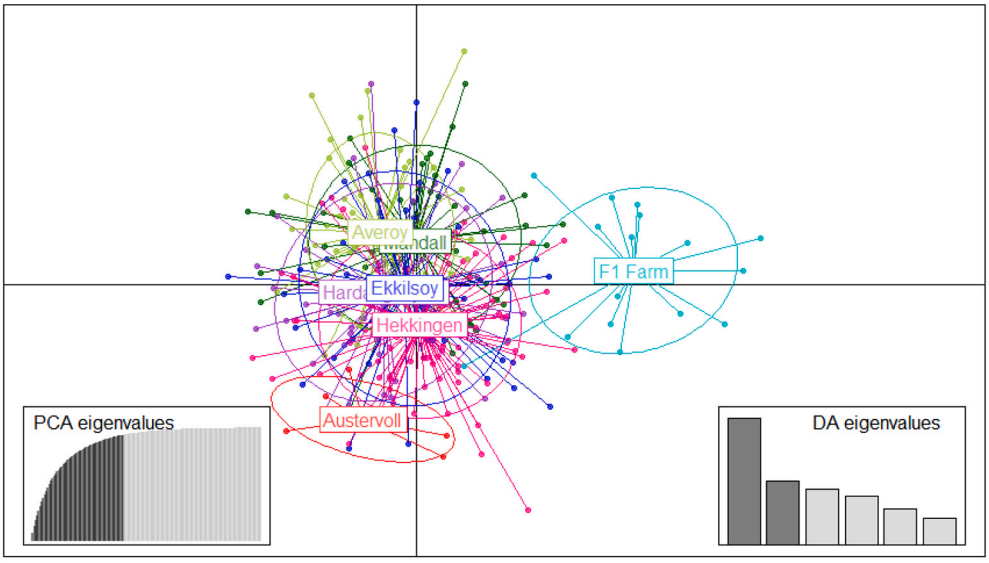
studies, but it should be pointed out that since the 100 specimens collected within Averøy area (analysed as one sample site in Jónsdóttir et al., 2018) were collected from two sampling sites several kilometres apart, 50 specimen each, the samples were analysed as two sampling sites (Averøy Håholmen and Ekkilsøy) in the present study (Table 1).

Only few studies directly compare relative performances of EST-STRs and genetic STRs to detect genetic differentiation in populations. As EST-STRs are located in coding regions, the higher stability of their flanking regions might reduce the frequency of null alleles, thus producing cleaner results for scoring (Leigh et al., 2003; Rungis et al., 2004) and fewer stutter bands (Leigh et al., 2003; Woodhead et al., 2003; Eujayl et al., 2004; Pashley et al., 2006), making them ideal for genetic studies in which genotypic errors should be strictly avoided, e.g. in fine-scale population genetic studies. (Kim et al., 2008). Despite these advantages, however, EST are not without their drawbacks (Ellis and Burke, 2007). Because of higher conservation of DNA sequences in coding compared to noncoding sequences, EST have generally less allelic variability than g-STRs, potentially leading to differences in estimates of population genetic parameters such as genetic differentiation (Khimoun er al., 2017). Despite less variability of EST-STR alleles, these markers still reveal sufficient levels of variation for the vast majority of population genetic applications (Ellis and Burke, 2007). Khimoun et al. (2017) argue that the higher allelic variability of g-STRs may provide them an advantage over EST-STRs in cases of moderate genetic differentiation, however, both EST-STRs and g-STRs showed to have equivalent ability to detect strong genetic differentiation. Moreover, EST-STRs might perform better than g-STRs to infer population structure in cases of weak genetic differentiation, thus making it interesting to employ EST-STR analysis on the lumpfish samples of Norway. Moreover, Postolache et al. (2014) argue that EST-STRs generally display lower polymorphism compared to g-STRs and lower frequency of null alleles, but higher genetic differentiations among populations. Other studies report either lower performances (Coulibaly et al., 2005) of EST-STRs compared with g-STRs or equivalent performances (Woodhead et al., 2005; Yatabe et al., 2007; Kim et al., 2008).

A comparison of genetic variability between EST-STRs (this study) and g-STRs (Jónsdóttir et al., 2018), using the same lumpfish samples, shows that the average number of alleles (Na) was lower in EST-STRs than in g-STRs in three samples (Mandal, Austevoll and Averøy), the allelic richness (A_r) was lower in Mandal and Averøy, however, the EST-STRs heterozygosity was higher in all samples (Table 4).

Population genetic structure of lumpfish in coastal areas using g-STRs has previously been investigated around Iceland by Pampoulie et al. (2014) and along West Greenland by Garcia-Mayoral et al. (2016) showing contrasting results. Pampoulie et al. (2014) investigated the population structure of lumpfish on both large and small geographic scale and found no population divergence among lumpfish collected at several locations around Iceland, whereas a significant genetic divergence on large geographic scale in the North Atlantic was revealed. In a more recent study, Garcia-Mayoral et al. (2016) investigated the population structure of lumpfish along the West Greenland coast using g-STRs and found significant structuring with isolation by distance in the West Greenland samples, moreover, two major subpopulations, north and south, were suggested. Whittaker et al. (2018) investigated the genetic (using g-STRs) and phenotypic differentiation of lumpfish across the North Atlantic and found five genetically distinct groups located in the West Atlantic (USA and Canada), Mid Atlantic (Iceland), East Atlantic (Faroe Islands, Ireland, Scotland, Norway and Denmark), English Channel (England) and Baltic Sea (Sweden). Significant phenotypic differences were also found, with Baltic lumpfish growing more slowly, attaining a higher condition factor and maturing at a smaller size than North Atlantic lumpfish. In contrast Jónsdóttir et al. (2018) found no indication (using g-STRs) of significant spatial genetic structuring or of positive correlation between geographic and genetic distance among lumpfish sampled along the Norwegian coast, a conclusion supported by the present study using EST-STRs. Even though one statistical test (F_{ST}, *P*-values, Table 5) showed genetic differentiation (in the present study) between two sample sites (Mandal and Hekkingen), the F_{ST} was low and the results was not supported by other statistical tests, such as Bayesian cluster analysis conducted in STRUCTURE (Fig. 2) and discriminant analysis of principal components (DAPC, Fig. 3), thus, it is concluded that even though there might be weak genetic difference, it is not statistically significant. The indication of genetic difference using EST-STRs when genetic structure of a species is weak is in accordance with the findings of Khimoun et al. (2017). The results of studies using both g-STRs (Jónsdóttir et al., 2018) and EST-STRs (present study) in lumpfish suggest that translocated individuals that escape from aquaculture in Norway will apparently have a similar STR-allele distribution as the local fish. However, as the authors pointed out (Jónsdóttir et al., 2018), lower number of g-STRs alleles per loci observed in farmed fish (F1 generation) compared to their wild counterparts might suggest either a bias due to low sample size (18 samples) or a possible selection affecting

A) g-STRs (Jónsdóttir et al.2018) for all samples.



B) EST-STRs (this study) for all samples.

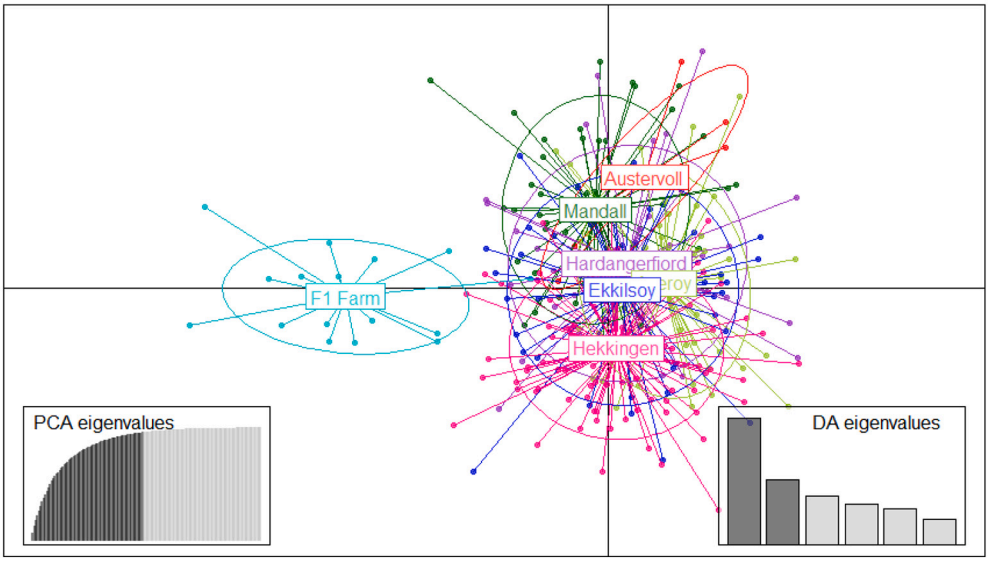


Fig. 3. Discriminant analysis of principal components (DAPC) for a) g-STR (Jónsdóttir et al., 2018) and b) EST-STR (this study) for all lumpfish samples. The number of PCs retained was determine by cross validation in DAPC. Each dot represent individual with centroid denoting the mean of the samples (colour) and ellipses are drawn around 2/3 of the points and represent the distribution of the points.

the farmed fish. Indeed, lower number of alleles per loci for farmed fish compared to wild fish was also observed in the present study using EST-STRs (Table 4). Individuals bred for their desirable characteristics differ from their origin in adaptive traits and the authors therefore suggested that the potential effects of escapes should be investigated further using not only neutral but also selective markers.

In the present study the same sampling material was analysed using functional genetic markers (EST-STRs) giving similar results as found for non-functional genetic markers (g-STRs), where low/no genetic structure was found for lumpfish along the Norwegian coast using STRs. Based solely on the results for both g-STRs and EST-STRs analyses, the former advise of non-negative genetic effect of translocation of farmed juvenile lumpfish along the Norwegian coast is, therefore, still valid, given that wild lumpfish are used for juvenile production to prevent possible inbreeding, however, cautionary approach is always recommended. In contrast, translocation of lumpfish between countries can represent a potential threat to local populations as studies have shown genetic structuring of lumpfish across the North Atlantic (Whittaker et al., 2018). For example, over 85% of all lumpfish deployed in Scotland during 2017 originated from eggs imported from Iceland and Norway, and none came from local sources (Treasurer, 2018). In Ireland, 70% of lumpfish deployed during 2015–2016 were derived from eggs imported from Iceland and Norway (Bolton-Warberg et al., 2018), while in the Faroe Islands nearly all lumpfish used during 2014–2016 were of Icelandic origin (Johannesen et al., 2018; Steinarsson and Árnason, 2018). Whether escapes have a genetic impact on local lumpfish populations

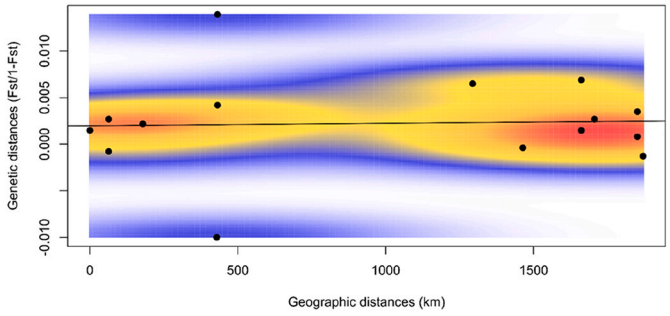


Fig. 4. Relationship between geographical (km shortest waterway) and genetic distance (estimated as $F_{ST}/(1\text{-}F_{ST})$) among pairs of lumpfish samples along the Norwegian coastline. The line shows the results of non-significant ($R^2 = 0.0015$, $P = 0.89$) regression between geographical and genetic distance. Mantels permutation test $P = 0.451$. Colours show the density of the points in the graph. Highest density as red, then yellow and blue. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

will depend on the number of escapees, their reproductive success, and the extent of genetic differentiation between local and introduced fish, but none of these parameters are currently known (Powell et al., 2018; Whittaker et al., 2018).

Lumpfish may undertake long annual distance migrations from spawning areas to winter grounds and vice versa (Schopka, 1974). In fact, tagging data revealed extensive movements during this time with the majority of tagged lumpfish exhibiting a displacement distance (shortest distance between tagging and recapture location) up to 350 km, however one individual was recaptured 587 km from its tagging location. Holst (1993) investigated the geographic distribution of lumpfish in the Norwegian Sea and found that lumpfish is widely distributed throughout the Norwegian Sea, with the largest concentrations occurring in the areas close to the polar front i.e. north of 72°N. This wide distribution might indicate large population mixing that can counteract genetic structuring of local stocks. In general, based on the results of no observed genetic structure in the wild lumpfish population along the Norwegian coast using both g-STRs and EST-STRs (Jónsdóttir et al., 2018; present study), current farming practises, including widespread transport of juveniles along the Norwegian coast will probably not have a negative impact on the wild stock of lumpfish. This is an important finding in regard to the species' use in aquaculture, as recent studies on two other cleaning fish species, ballan and corkwing wrasse (Gonzales et al., 2016; Quintela et al., 2016; Seljestad et al., 2020), commonly used in Norwegian aquaculture (Skiftesvik et al., 2014; Anon, 2016), have revealed genetic divergence among wild populations questioning the sustainability of the current practise (capture and long-distance translocation) of these wrasse species. Seljestad et al. (2020) pointed out that the management of wrasse in Scandinavia is currently facing two major challenges, both linked to their extensive use as cleaner fish to delouse farmed salmon in sea-cages. These challenges include the sustainable harvest of wild populations to provide wrasse as cleaner fish for the aquaculture industry, and the potential aquaculture-driven inadvertent translocation of wrasse from southern Norway and Sweden to western and middle Norway. Studies of other species of wrasse subjected to the same aquaculture-driven harvest and translocation regimes have also indicated escapes and potential hybridization between translocated individuals and local populations (Jansson et al., 2017; Faust et al., 2018). In contrast the management situation for lumpfish is very different as all lumpfish used in salmon sea-cages are from intensive aquaculture production. Secondly, both current study and our previous study (Jónsdóttir et al., 2018) have not revealed a genetic structure of the species along the Norwegian coast using g-STRs and EST-STRs, indicating low negative effect of translocation of lumpfish in Norway. It should though be kept in mind that other genetic markers (e.g. SNP) might add to our knowledge and understanding of the genetic structure of lumpfish in the Norwegian waters, and therefore it is always necessary to take precautions in translocation of animals.

Discriminant analysis of principal components and pairwise F_{ST} estimates indicate genetic differentiation between farmed F1 fish and the wild samples from the Norwegian coast. Possibly, this differentiation exhibited by the farmed strain is due to random genetic drift of allele frequencies in the hatcheries. This has been well documented in other species (Danzmann et al., 1989; Coughlan et al., 1998; Sanchez et al., 2012). However, some element of domestication or hatchery/husbandry (deliberate or inadvertent) selection may have led also to changes in the genetic composition of farmed strains compared with source populations (Lorenzen et al., 2012). Cultured fish are being released on a very large scale, both accidentally and intentionally (McGinnity et al., 2003; Lorenzen et al., 2012; Glover et al., 2017). Such interactions may be problematic or positive (Lorenzen et al., 2012). Problematic interactions include displacement of wild fish through ecological interactions, reductions in fitness and genetic diversity of populations subject to interbreeding between cultured and wild fish (Van Poorten et al., 2011; Glover et al., 2017). Positive interactions include increases in total population abundance that can support fisheries or counteract ecological and genetic risks in endangered populations (Hilderbrand, 2002; Lorenzen et al., 2012). The risk of loss of diversity is greatest when fish of cultured origin contribute substantially to the mixed population, but have a much lower effective population size than the wild population. This situation can arise relatively easily because high fecundity of fish combined with high survival of early life stages in culture makes it possible to produce very large numbers of offspring from very few parents. This may be helped by factorial mating designs that can increase effective population size to about twice the census size (Lorenzen et al., 2012). Current production of lumpfish as cleanerfish is often based on a few parents (Jónsdóttir et al., 2018; Jonassen et al., 2018; Imsland et al., 2021), but based on current findings and published knowledge on this topic we strongly advise for genetic monitoring of hatchery populations of lumpfish by increasing the number of effective breeders whenever possible.

5. Conclusion

Present findings show a lack of population divergence among lumpfish along the Norwegian coast using EST-STRs. Based on these results, should translocated offspring of wild lumpfish individuals escape from salmon sea pens in Norway, this will probably have little impact on the genetic composition of the local lumpfish population. Present data indicate low number of parental fish being used in aquaculture production of lumpfish and it is advised to increase the numbers of parental fish.

Declaration of Competing Interest

There is no conflict of interest in relation to this study.

Acknowledgements

We are grateful to Mr. Thor Jonassen, Mr. Thor Arne Hangstad and Mrs. Mette Remen at Akvaplan-niva, Norway and Mrs. Caroline Durif and Mr. Nola Velimir at Marine Research Institute, Bergen, Norway, for their assistance in sampling the biological samples of lumpfish. Special thanks are addressed to Ágústa Sigfúsdóttir and Steinunn Magnúsdóttir for laboratory assistance at Matís. The financial assistance Lerøy Seafood Group, Research Council of Norway (RFFNord project 282460) and the Icelandic Research Council (Rannís, 186971-0611) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the authors and are not necessarily to be

attributed to the funding bodies.

References

Anon, 2016. Orientering om regulering av fisket etter leppefisk (orientation on the regulation of wrasse fisheries in Norway). In: Norwegian Directorate of Fisheries, Bergen, Norway, nr. 19/2016, p. 13 (in Norwegian).

Antao, T., Lopes, A., Lopes, R.J., Beja-Pereira, A., Luikart, G., 2008. LOSITAN: a workbench to detect molecular adaptation based on a Fst-outlier method. BMC Bioinform. 9, 1–5.

Beaumont, M.A., Nichols, R.A., 1996. Evaluating loci for use in the genetic analysis of population structure. Proc. R. Soc. B Biol. Sci. 263, 1619–1626.

Belkhir, K., Borsa, P., Chikhi, L., Raufaste, N., Bonhomme, F., 2004. Genetix 4.05, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier.

Benjamini, Y., Yekutieli, D., 2001. The control of the false discovery rate in multiple testing under dependency. Ann. Stat. 29, 1165–1188.

Bolton-Warberg, M., O'Sullivan, S.M., Power, A.M., Moore, A.I., Wilson, L., Sproll, F., Fitzpatrick, R., 2018. Cleaner fish use in Ireland. In: Treasurer, J.W. (Ed.), Cleaner Fish Biology and Aquaculture Applications. 5M Publishing Ltd., Sheffield, pp. 386–413.

Bouck, A., Vision, T., 2010. The molecular ecologist's guide to expressed sequence tags. Mol. Ecol. 16, 907–924.

Boxaspen, K., 2006. A review of the biology and genetics of sea lice. ICES J. Sea Res. 63, 1304–1316.

Boyce, D., Ang, K.P., Prickett, R., 2018. Cunner and lumpfish as cleaner fish species in Canada. In: Treasurer, J.W. (Ed.), Cleaner Fish Biology and Aquaculture Applications. 5M Publishing Ltd., Sheffield, pp. 444–467.

Brooker, A.J., Papadopoulou, A., Gutierrez, C., Rey, S., Davie, A., Migaud, H., 2018. Sustainable production and use of cleaner fish for the biological control of sea lice: recent advances and current challenges. Vet. Rec. 183, 383.

Brownstein, M.J., Carpten, J.D., Smith, J.R., 1996. Modulation of non-templated nucleotide addition by Taq DNA polymerase: primer modifications that facilitate genotyping. Biotechniques 20, 1004–1010.

Coughlan, J.P., Imsland, A.K., Galvin, P.T., FitzGerald, R.D., Nævdal, G., Cross, T.F., 1998. Microsatellite DNA variation in wild populations and farmed strains of turbot *Scophthalmus maximus* from Ireland and Norway. J. Fish Biol. 52, 916–922.

Coulibaly, I., Gharbi, K., Danzmann, R.G., Yao, J., Rexroad, C.E., 2005. Characterization and comparison of microsatellites derived from repeat-enriched libraries and expressed sequence tags. Anim. Genet. 36, 309–315.

Danzmann, R.G., Ferguson, M.M., Allendorf, F.W., 1989. Genetic variability and components of fitness in hatchery strains of rainbow trout. J. Fish Biol. 35, 313–319.

Denholm, I., Devine, G.J., Horsberg, T.E., Sevatdal, S., Fallang, A., Nolan, D.V., Powell, R., 2002. Analysis and management of resistance to chemotherapeutants in salmon lice *Lepeophtheirus salmonis* (Krøyer) (Copepoda: Caligidae). Pest Manag. Sci. 58, 528–536.

Ellis, J.R., Burke, J.M., 2007. EST-SSRs as a resource for population genetic analyses. Heredity 99, 125–132.

Ellis, J.R., Pashley, C.H., Burke, J.M., McCauley, D.E., 2006. High genetic diversity in a rare and endangered sunflower as compared to a common congener. Mol. Ecol. 15, 2345–2355.

Eujayl, I., Sledge, M.K., Wang, L., May, G.D., Chekhovskiy, K., Zwonitzer, J.C., Mian, M., 2004. *Medicago truncatula* EST-SSRs reveal cross-species genetic markers for *Medicago* spp. Theor. Appl. Genet. 108, 414–422.

Faust, E., Halvorsen, K.T., Andersen, P., Knutsen, H., André, C., 2018. Cleaner fish escape salmon farms and hybridize with local wrasse populations. R. Soc. Open Sci. 5, 1–12.

Garcia-Mayoral, E., Olsen, E., Hedeholm, R., Post, S., Nielsen, E.E., Bekkevold, D., 2016. Genetic structure of West Greenland populations of lumpfish *Cyclopterus lumpus*. J. Fish Biol. 89, 2625–2642.

Glover, K.A., Solberg, M.F., McGinnity, P., Hindar, K., Verspoor, E., Coulson, M.W., Svasand, T., 2017. Half a century of genetic interaction between farmed and wild Atlantic salmon: status of knowledge and unanswered questions. Fish Fish. 18, 890–927.

Gonzales, E.B., Knutsen, H., Jorde, P.E., 2016. Habitat discontinuities separate genetically divergent populations of a rocky shore marine fish. PLoS One 11, 1–17.

Hilderbrand, R.H., 2002. Simulating supplementation strategies for restoring and maintaining stream resident cutthroat trout populations. North Am. J. Fish. Manag. 22, 879–887.

Holst, J.C., 1993. Observations on the distribution of lumpsucker (*Cyclopterus lumpus* L.) in the Norwegian Sea. Fish. Res. 17, 369–372.

Hubisz, M.J., Falush, D., Stephens, M., Pritchard, J.K., 2009. Inferring weak population structure with the assistance of sample group information. Mol. Ecol. Resour. 9, 1322–1332.

Imsland, A.K., Reynolds, P., Eliassen, G., Hangstad, T.A., Foss, A., Vikingstad, E., Elvegård, T.A., 2014a. The use of lumpfish (*Cyclopterus lumpus* L.) to control sea lice (*Lepeophtheirus salmonis* Krøyer) infestations in intensively farmed Atlantic salmon (*Salmo salar* L.). Aquaculture 425-426, 18–23.

Imsland, A.K., Reynolds, P., Eliassen, G., Hangstad, T.A., Nytrø, A.V., Foss, A., Elvegård, T.A., 2014b. Assessment of growth and sea lice infection levels in Atlantic salmon stocked in small-scale cages with lumpfish. Aquaculture 433, 137–142.

Imsland, A.K., Reynolds, P., Eliassen, G., Hangstad, T.A., Nytrø, A.V., Foss, A., Elvegård, T.A., 2014c. Notes on behaviour of lumpfish in sea pens with and without Atlantic salmon. J. Ethol. 32, 117–122.

Imsland, A.K.D., Reynolds, P., Hangstad, T.A., Madura, S., Hagen, S., Jónsdóttir, Ó.D.B., Lindberg, K.S., 2021. Quantification of grazing efficacy, growth and health score of different lumpfish (*Cyclopterus lumpus* L.) families: possible size and gender effects. Aquaculture 530, 735925.

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 J. Mar. Sci. 74, 2135–2147.

Johannesen, Á., Arge, R., Eliasen, K., 2018. Rearing, farm application and behaviour of lumpfish (*Cyclopterus lumpus*) in the Faroes. In: Treasurer, J. (Ed.), Cleaner Fish Biology and Aquaculture Applications. 5M publications, Sheffield, UK, pp. 435–449.

Jombart, T., 2008. ADEGENET: a R package for the multivariate analysis of genetic markers. Bioinform. 24, 1403–1405.

Jombart, T., Devillard, S., Balloux, F., 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet. 11, 1–15.

Jonassen, T.M., Lein, I., Nytrø, A.V., 2018. Hatchery management of lumpfish. In: Treasurer, J. (Ed.), Cleaner Fish Biology and Aquaculture Applications. 5M publications, Sheffield, UK, pp. 114–136.

Jónsdóttir, Ó.D.B., Schregel, J., Hagen, S.B., Tobiassen, C., Aarnes, S.G., Imsland, A.K.D., 2018. Population genetic structure of lumpfish along the Norwegian coast: aquaculture implications. Aquac. Int. 26, 49–60.

Khimoun, A., Ollivier, A., Faivre, B., Garnier, S., 2017. Level of genetic differentiation affects relative performances of expressed sequence tag and genomic SSRs. Mol. Ecol. Resour. 17, 893–903.

Kim, K.S., Ratcliffe, S.T., French, B.W., Liu, L., Sappington, T.W., 2008. Utility of EST-derived SSRs as population genetics markers in a beetle. J. Hered 99, 112–124.

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. Mol. Ecol. Resour. 15, 1179–1191.

Leigh, F., Lea, V., Law, J., Wolters, P., Powell, W., Donini, P., 2003. Assessment of EST- and genomic microsatellite markers for variety discrimination and genetic diversity studies in wheat. Euphytica 133, 359–366.

Lorenzen, K., Beveridge, M.C.M., Mangel, M., 2012. Cultured fish: integrative biology and management of domestication and interactions with wild fish. Biol. Rev. 87, 639–660.

Maduna, S., Vivian-Smith, A., Jónsdóttir, Ó.D.B., Imsland, A.K.D., Klutsch, C.F.C., Nyman, T., Hagen, S.B., 2020. Genome and transcriptome-derived microsatellite loci in lumpfish *Cyclopterus lumpus*: molecular tools for aquaculture, conservation and fisheries management. Sci. Rep. 10 (article 559).

Mantel, N., 1967. The detection of disease clustering and a generalized regression approach. Cancer Res. 27, 209–220.

McGinnity, P., Prodöhl, P., Ferguson, A., Hynes, R., Maoiléidigh, Ó., Baker, N., Cross, T., 2003. Fitness reduction and potential extinction of wild populations of Atlantic salmon *Salmo salar*, as a result of interactions with escaped farm salmon. Proc. R. Soc. Lond. Ser. B 270, 2443–2450.

Norwegian Directorate of Fisheries, 2019. Sale of farmed cleaner fish 2012-2019. Available from URL. https://www.fiskeridir.no/Akvakultur/Tall-og-analyse/Akvakulturstatistikk-tidsserier/Rensefisk.

Pampoulie, C., Skírnisdottir, S., Ólafsdóttir, G., Helyar, S.J., Thorsteinsson, V., Jónsson, S.Þ., Kasper, J.M., 2014. Genetic structure of the lumpfish *Cyclopterus lumpus* across the North Atlantic. ICES J. Mar. Sci. 71, 2390–2397.

Paradis, E., Schliep, K., 2019. Ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35, 526–528.

Pashley, C.H., Ellis, J.R., McCauley, D.E., Burke, J.M., 2006. EST databases as a source for molecular markers: lessons from Helianthus. J. Hered. 97, 381–388.

Postolache, D., Leonarduzzi, C., Piotti, A., Spanu, I., Roig, A., Fady, B., Vendramin, G.G., 2014. Transcriptome versus genomic microsatellite markers: highly informative multiplexes for genotyping Abies alba mill. and congeneric species. Plant Mol. Biol. Report. 32, 750–760.

Powell, A., Treasurer, J.W., Pooley, C.L., Keay, A.J., Lloyd, R., Imsland, A.K., Garcia de Leaniz, C., 2018. Cleaner fish for sea-lice control in salmon farming: challenges and opportunities using lumpfish. Rev. Aquac. 10, 683–702.

Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of population structure using multilocus genotype data. Genetics 155, 945–959.

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. Mol. Ecol. Resour. 16, 608–627.

Quintela, M., Danielsen, E.A., Lopez, L., Barreiro, R., Svåsand, T., Knutsen, H., Glover, K. A., 2016. Is the ballan wrasse (*Labrus bergylt*) two species? Genetic analysis reveals within-species divergence associated with plain and spotted morphotype frequencies. Int. Zool. 11, 162–172.

R Core Team, 2014. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. http://www.r-project.org/.

Rousset, F., 2008. GENEPOP′007: a complete re-implementation of the GENEPOP software for Windows and Linux. Mol. Ecol. Resour. 8, 103–106.

Rungis, D., Bérubé, Y., Zhang, J., Ralph, S., Ritland, C., Ellis, B., Ritland, K., 2004. Robust simple sequence tags. Theor. Appl. Genet. 109, 1283–1294.

Sanchez, P., Vinas, J., Bremer, J.R.A., Ambrosio, P.P., Flos, R., 2012. Loss of genetic variability in a hatchery strain of Senegalese sole (*Solea senegalensis*) revealed by sequence data of the mitochondrial DNA control region and microsatellite markers. Sco. Mar. 76, 225–235.

Schopka, S.A., 1974. Preliminary results from tagging of lumpsucker (*Cyclopterus lumpus*), in Icelandic waters 1971-1974. In: ICES CM.1974/F:18. Demersal Fish Northern Committee, 6p.

Seljestad, G.W., Quintela, M., Faust, E., Halvordsen, K.T., Besnier, F., Jansson, E., Glover, K.A., 2020. A cleaner break: genetic divergence between geographic groups and sympatric phenotypes revealed in ballan wrasse (*Labrus bergylta*). Ecol. Evol. 10, 6120–6135.

Skiftesvik, A.B., Blom, G., Agnalt, A.L., Durif, C.M.F., Browman, H., Bjelland, R.M., Mortensen, S., 2014. Wrasse (Labridae) as cleaner fish in salmonid aquaculture – the Hardangerfjord as a case study. Mar. Biol. Res. 10, 289–300.

Steinarsson, A., Árnason, T., 2018. Rearing of cleaner fish use in Iceland. In: Treasurer, J. W. (Ed.), Cleaner Fish Biology and Aquaculture Applications. 5M Publishing Ltd., Sheffield, pp. 414–428.

Szpiech, Z.A., Jakobsson, M., Rosenberg, N.A., 2008. ADZE: a rarefaction approach for counting alleles private to combinations of populations. Bioinformatics 24, 2498–2504.

Teacher, A., Kähkönen, K., Merilä, J., 2012. Development of 61 new transcriptome-derived microsatellites for the Atlantic herring (*Clupea harengus*). Conserv. Genet. Resour. 4, 71–74.

Treasurer, J.W., 2018. An introduction to sea lice and the rise of cleaner fish. In: Treasurer, J.W. (Ed.), Cleaner Fish Biology and Aquaculture Applications, Cleaner Fish Biology and Aquaculture Applications. 5M Publishing Ltd., Sheffield, pp. 3–24.

Van Oosterhout, C., Hutchinson, W.F., Wills, D.P.M., Shipley, P., 2004. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Mol. Ecol. Notes 4, 535–538.

Van Poorten, B.T., Arlinghaus, R., Daedlow, K., Heartel-Borer, S.S., 2011. Social-ecological interactions, management panaceas, and the future of wild fish populations. Proc. Natl. Acad. Sci. 108, 12554–12599.

Weir, B.S., Cockerham, C.C., 1984. Estimating *F*-statistics for the analysis of population structure. Evolution 38, 1358–1370.

Whittaker, B.A., Consuegra, S., de Leaniz, C.G., 2018. Genetic and phenotypic differentiation of lumpfish (*Cyclopterus lumpus*) across the North Atlantic: implications for conservation and aquaculture. PeerJ 6, e5974.

Woodhead, M., Russek, J., Squirrell, J., Hollingsworth, P.M., Cardle, L., Ramsay, L., Powell, W., 2003. Development of EST-SRRs from the alpine lady-fern, Athyrium distentifolium. Mol. Ecol. Notes 3, 287–290.

Woodhead, M., Russell, J., Squirrell, J., Hollingsworth, P.M., Mackenzie, K., Gibby, M., Powell, W., 2005. Comparative analysis of population genetic structure in *Athyrium distentifolium* (Pteridophyta) using AFLPs and SSRs from anonymous and transcribed gene regions. Mol. Ecol. 14, 1681–1695.

Wu, Q., Zang, F., Xie, X., Ma, Y., Zheng, Y., Zang, D., 2020. Full-length transcriptome sequencing analysis and development of EST-SSR markers for the endangered species *Populus wulianensis*. Sci. Rep. 10 article 16249.

Yatabe, Y., Kane, N.C., Scotti-Saintagne, C., Rieseberg, L.H., 2007. Rampant gene exchange across a strong reproductive barrier between the annual sunflowers, *Helianthus annuus* and *H. petiolaris*. Genetics 175, 1883–1893.