



# Tracing the genetic impact of farmed turbot *Scophthalmus maximus* on wild populations

F. D. Prado<sup>1,2</sup>, M. Vera<sup>1</sup>, M. Hermida<sup>1</sup>, A. Blanco<sup>1</sup>, C. Bouza<sup>1</sup>, G. E. Maes<sup>3,4</sup>,  
F. A. M. Volckaert<sup>3</sup>, AquaTrace Consortium<sup>5</sup>, P. Martínez<sup>1,\*</sup>

<sup>1</sup>Department of Zoology, Genetics and Physical Anthropology, Faculty of Veterinary, Universidade de Santiago de Compostela, 27002 Lugo, Spain

<sup>2</sup>CAPES Foundation, Ministry of Education of Brazil, Brasília, Brazil

<sup>3</sup>Laboratory of Biodiversity and Evolutionary Genomics, KU Leuven, 3000 Leuven, Belgium

<sup>4</sup>Laboratory for Cytogenetics and Genome Research, Center for Human Genetics, Genomics Core, KU Leuven, 3000 Leuven, Belgium

<sup>5</sup>European FP7 KBBE project AquaTrace (311920): the development of tools for tracing and evaluating the genetic impact of fish from aquaculture

**ABSTRACT:** The impact of escapees from aquaculture is of general concern for the sustainability of natural resources. Turbot *Scophthalmus maximus* is a marine flatfish of great commercial value whose land-based aquaculture started approx. 40 yr ago; hence, a low impact of escapees is expected on wild populations. However, enhancement of wild stocks using farmed turbot has been carried out along the Northeast Atlantic coasts in the last decades. Recently, a broad panel of single nucleotide polymorphism (SNP) markers (755 SNPs; 1 SNP Mb<sup>-1</sup>) has been used to evaluate the genetic structure of turbot throughout its distribution range, constituting the baseline to evaluate the impact of farmed fish in the wild. Two distinct origins were identified for farmed turbot ( $F_{\text{OR1}}$  and  $F_{\text{OR2}}$ ;  $F_{\text{ST}} = 0.049$ ), which differentiated from wild populations after 5 generations of selection (average  $F_{\text{ST}} = 0.059$ ), and consistent evidence of adaptation to domestication was detected. A notable proportion of fish of farmed ancestry was detected in the wild (15.5%), mainly in the North Sea, where restocking activities have taken place, determining genetic introgression in wild populations. Conversely, effects of land-based aquaculture appear negligible. A simulation exercise supported panels of 40 and 80 SNPs to identify fishes of  $F_{\text{OR1}}$  and  $F_{\text{OR2}}$  ancestry in the wild, respectively. Application to empirical data showed an assignment success (wild/farmed ancestry) of approx. 95% in comparison with the full SNP dataset. The SNP tools will be useful to monitor turbot of farmed ancestry in the wild, which might represent a risk, considering the lower fitness of farmed individuals.

**KEY WORDS:** Aquaculture · Introgression · Restocking · *Scophthalmus maximus* · SNPs · Sustainability · Traceability tool

## INTRODUCTION

Seafood has become one of the most important sources of protein for human consumption (Danancher & Garcia-Vazquez 2011). However, many marine fish stocks have been exploited to depletion (FAO 2016), and natural resources have greatly de-

clined (Worm et al. 2009). Consequently, aquaculture is now emerging as one of the most flourishing areas of animal production (FAO 2016).

Despite the advantage of providing an alternative source of animal protein, negative environmental outcomes, such as land use for feed production (Froehlich et al. 2018) and the escape of domesticated fish

\*Corresponding author: paulino.martinez@usc.es

(Glover et al. 2017, Arechavala-Lopez et al. 2018), are issues concerning aquaculture production. Genetic differentiation between farmed and wild fish rapidly increases in breeding due to genetic drift and selection (Janssen et al. 2017). In the wild, genetic variability is distributed within and among populations in response to demographic and historical factors but also to environmental adaptation throughout the natural range (Nielsen et al. 2009). Thus, escapees from farming or intentional releases for restocking purposes can impact wild populations in different ways (Glover 2010, Glover et al. 2017), including temporary disturbances, ecological competition, reduction of reproductive performance, population decline, hybridization between wild and farmed lineages, genetic introgression and eventually the genetic extinction of previously adapted populations (Waples 1999). Impact has been evaluated especially for the most important domesticated salmonid species (Arias et al. 1995, Toledo-Guedes et al. 2014, Taranger et al. 2015, Araguas et al. 2017, Glover et al. 2017). However, impact evaluation of all farmed species is crucial for the management of aquaculture considering the ongoing expansion and the difficulty of predicting the ecological and genetic outcomes of escapees.

The evaluation of aquaculture impact requires ecological and genetic knowledge of the target species. Genetic characterization of farmed broodstock is necessary to ascertain differentiation from wild populations for tracking farmed genotypes into the wild and, eventually, for identifying hybridization and introgression (Noble et al. 2014, Habtemariam et al. 2015, Bylemans et al. 2016, Glover et al. 2017). Ideally, one would like to estimate fitness changes in wild populations due to introgression from the correlation between the genetic differences (wild vs. farmed) and the traits involved in adaptation. Various molecular markers have been applied for conservation and management studies of marine fish, including microsatellites (simple sequence repeats [SSRs]) and single nucleotide polymorphism (SNPs); both types of markers have demonstrated their usefulness for discriminating between species and lineages (Seeb et al. 2011, Prado et al. 2017). These markers have enabled researchers to track farmed individuals, identify hybrids and evaluate genetic introgression in wild populations (Karlsson et al. 2014, Bylemans et al. 2016, Pritchard et al. 2016).

The turbot *Scophthalmus maximus* (Scophthalmidae, Pleuronectiformes) is a marine flatfish naturally structured in 4 main regions: the Northeast Atlantic Ocean, Baltic Sea, Mediterranean Sea and Black Sea (Vandamme et al. 2014, Vilas et al. 2015, Prado et al.

2018). Despite low genetic differentiation among Atlantic populations, mainly linked to larval advection, populations adapted to different environmental conditions, mostly driven by salinity and temperature, have been reported (Vilas et al. 2010, 2015, Vandamme et al. 2014, Prado et al. 2018). Turbot is important in the European fishery, mainly in the North Sea (Bouza et al. 2014), but its high commercial value and low supply on the market have promoted its aquaculture, with turbot currently the main worldwide cultured flatfish (Robledo et al. 2017).

Turbot production started in the 1970s in Scotland and France, subsequently expanded to other European countries and more recently developed in PR China (Bouza et al. 2014, Robledo et al. 2017). The autonomous community of Galicia (Northwestern Spain) remains the main European production centre (Martínez et al. 2016). Breeding programs have been carried out for ~5 generations by a few European companies, which have been primarily focused on increasing growth rate, although resistance to pathogens, morphological abnormalities and sex control have also been considered (Martínez et al. 2016, Janssen et al. 2017).

The genome of farmed turbot has been shaped by the specific founder stock at each company, the small effective population size, domestication and selection. This has led to a gradual differentiation from wild stocks (Vilas et al. 2015). Previous studies suggested a loss of genetic diversity of farmed turbot (Coughlan et al. 1998, Estoup et al. 1998, Bouza et al. 2002, Exadactylos et al. 2007, Vilas et al. 2015), but the broodstock representativeness of the samples evaluated has been poorly documented. No monitoring of farmed activities has been carried out, although the risk of escapees is low, given the onshore production (Danancher & Garcia-Vazquez 2011). In addition, programs to restock wild populations using farmed turbot have been carried out with experimental or conservation purposes along the Atlantic coasts of Spain, Denmark, Belgium and Norway but with poor evaluation and documentation of effects (Delbare & Declerck 2000, Ellis et al. 2002, Stottrup et al. 2002, Iglesias et al. 2003, Sparrevojn & Nielsen 2003, Paulsen & Stottrup 2004). For example, the 77 000 ton oil spill of the tanker MV 'Prestige' in Galician waters in 2002 seriously damaged the ecosystem and local fishery, particularly benthic communities and demersal stocks (Viñas et al. 2009). Restocking was promoted mostly with farmed turbot supplied by local companies. Tagging and recapture of released fish in some of the abovementioned areas showed the capability of farmed turbot to survive in the wild. Few dif-

ferences in growth and nutrition in comparison to wild fish were recorded, but nothing has been reported on their subsequent reproductive capacity or fitness (Paulsen & Stottrup 2004, Mariño et al. 2009).

In this study, we aimed to identify farmed turbot either from escapees or restocking and evaluate their impact on wild populations. More specifically, a previously reported set of SNPs developed through double digestionRAD (ddRAD) technology (Prado et al. 2018) was used to (1) assess genetic differentiation between wild and farmed populations and identify differentiated genomic regions resulting from domestication and selection processes at farms; (2) ascertain the presence of fish of farmed ancestry in the wild and the degree of genetic introgression of wild populations; and (3) develop and validate a practical molecular tool useful to monitor farmed individuals in the wild.

## MATERIALS AND METHODS

### Farm broodstock sampling

Three Spanish and one French companies, all located in the Atlantic area, are the main European producers of turbot *Scophthalmus maximus*. Fry are either reared locally on the farm or sold to other companies along the Atlantic coast from Portugal to Norway. Since the 1990s, all 4 companies maintain breeding programs, mostly focussing on enhanced growth rate. They each use a broodstock of ~400 to ~500 fish and select for 5 generations (Janssen et al. 2017). Accordingly, our analysis of the impact of aquaculture and restocking was focused on the broodstock of those 4 companies. For confidentiality reasons, the 4 companies anonymously supplied a representative sample of their broodstock, including ~45 individuals per company and totalling 174 individuals.

Considering that turbot farming occurs mainly in the Northeast Atlantic, only native populations from the Atlantic and Baltic were considered. Wild samples comprised 686 individuals from 17 sampling sites in 9 different areas: Baltic Sea (BAS); Skagerrak, a transition area between the Atlantic and the Baltic Sea (T); Norway Sea (NOR); North Sea (NS); Iceland (ICE); Ireland (IR); English Channel (ECH); Biscay Bay (BB); and Spain coast (SP-W). These sites were previously studied by Prado et al. (2018) and details can be found in Table S1 (Supplement 1 at [www.int-res.com/articles/suppl/q010p447\\_supp1.pdf](http://www.int-res.com/articles/suppl/q010p447_supp1.pdf)). This sampling is representative of turbot in the Atlantic, especially considering the very low genetic

differentiation reported ( $F_{ST} < 0.003$  excluding the Baltic Sea; Prado et al. 2018). Additionally, a temporal sample collected along the Spanish coast after the *Prestige* oil spill (SP.new,  $N = 48$ ) was included for comparison with an older one (SP.old) prior to the *Prestige* oil spill (Table S1).

### SNP genotyping

Genomic DNA was extracted from fin clips of each individual and used to construct individual genomic libraries for SNP genotyping with the ddRAD technology (Peterson et al. 2012). A panel of 755 SNPs was genotyped following the modified ddRAD method reported by Prado et al. (2018) in the newly sampled turbot (174 farmed and 48 SP.new) and combined with the previous dataset of wild populations ( $n = 686$ ) for a total of 908 individuals.

### Identification and genetic characterization of farmed and wild samples

A Bayesian clustering method implemented in STRUCTURE 2.3.4 (Pritchard et al. 2000) was used to evaluate the number of population units ( $K$ ) in the farmed samples. We opted for an admixture ancestry model, correlated allele frequencies, 10000 iterations of burn-in, 50000 Markov Chain Monte-Carlo steps and 5 independent runs for each  $K$  tested ( $K = 1$  to 5). Results were processed with STRUCTURE HARVESTER 0.3 (Earl & von Holdt 2012) to estimate the number of  $K$  that better fitted with the data based on the  $\Delta K$  method described by Evanno et al. (2005). CLUMPP 1.1.2 (Jakobsson & Rosenberg 2007) was applied to estimate the average membership coefficients ( $q$ ) among the different tested runs. Two different clusters were identified in farmed turbot and thus taken as a reference for further analyses ( $F_{OR1}$  and  $F_{OR2}$ ; see 'Results: Identification and impact of farmed fish on wild populations'). A second STRUCTURE analysis identified individuals of farmed ancestry in the wild populations. In this case, all wild samples were included along with  $F_{OR1}$  and  $F_{OR2}$ , testing  $K = 1$  to 6 and using the same parameters applied before.

Wild samples from the Atlantic Ocean and the Baltic Sea were characterized genetically by removing all individuals considered of farmed ancestry ( $q_w$ -values  $< 0.9$ , where  $w =$  wild cluster) (see 'Results'). Genetic diversity was estimated by the expected heterozygosity ( $H_e$ ) and the percentage of polymorphic loci

at 99% ( $P_{99}$ ) using ARLEQUIN 3.5.2 (Excoffier & Lischer 2010). Pairwise  $F_{ST}$  between sampling sites and groups of samples was estimated with ARLEQUIN 3.5.2 (Excoffier & Lischer 2010) using 10 000 permutations for significance at  $p < 0.05$  and also after sequential Bonferroni correction (Rice 1989). Discriminant analyses of principal components (DAPC) for groups of samples were performed with ADEGENET 2.0.0 (Jombart & Ahmed 2011) using an R platform (R Development Core Team 2014). PGDSPIDER 2.0.8.3 program (Lischer & Excoffier 2012) was used for file conversion in each statistics analysis.

To evaluate genetic differentiation between wild and farmed turbot resulting from domestication, only wild samples from the Atlantic area were used (hereafter  $W_{ATL}$ ) since they were the source for founding the farmed broodstock. For that, 2 statistical approaches implemented in BAYESCAN and LOSITAN (Narum & Hess 2011, Shimada et al. 2011, Souche et al. 2015) were applied, comparing  $W_{ATL}$  vs.  $F_{OR1}$ ,  $W_{ATL}$  vs.  $F_{OR2}$  and  $F_{OR1}$  vs.  $F_{OR2}$ . BAYESCAN 2.01 (Foll & Gaggiotti 2008) was run using 20 pilot runs, 5000 iterations, 5000 burn-in steps and a sample size of 5000. Outliers were considered significant when showing a logarithm of the probability of departure from neutrality, the Bayes factor  $\log_{10}(BF)$ ,  $> 1.3$  ( $p > 0.95$ ) or  $> 2$  ( $p > 0.99$ ). The FDISt method implemented in LOSITAN software (Antao et al. 2008) was used through 100 000 simulations, a population size of 50 according to  $N_e$  (effective population size) estimates of farm samples, infinite allele mutation model and 'neutral mean  $F_{ST}$ ', applying a confidence interval (CI) of 0.99 and false discovery rate (FDR) of 0.01. Previous studies have suggested that the BAYESCAN approach is the most conservative (Narum & Hess 2011), and thus, SNPs were considered as confident outliers when detected with BAYESCAN and suggestive when only identified with LOSITAN. The R platform was used to plot figures and graphics.

BLASTn was used to localize the genomic position of outliers in the turbot genome ( $e$ -value  $< 1e-5$ ). Their relationship with either previously described outliers (Vilas et al. 2010, 2015, Prado et al. 2018) or quantitative trait loci (QTLs) related to productive traits (Martínez et al. 2016) was investigated by anchoring their sequences to the turbot genetic map. We identified candidate genes putatively associated with domestication or selection close to selected outliers ( $\pm 250$  kb) through gene mining with the turbot genome browser (<http://denovo.cnag.cat/genomes/turbot/>, Figueras et al. 2016) according to the moderate linkage disequilibrium detected in turbot (Saura et al. 2017). Comparison of outlier genomic positions

between both farmed stocks was investigated to search for signals of convergent selection. Functional enrichment of the genomic windows close to outliers was also investigated with Blast2GO (FDR  $< 5\%$ ) and using the turbot genes as background (Figueras et al. 2016).

### Identification of farmed individuals in the wild

The USEPOPINFO option implemented in STRUCTURE, recommended to identify hybridization when reference populations are well defined (Pritchard et al. 2000), was used to identify individuals of farmed ancestry in the native range of turbot in the Baltic Sea and the Atlantic Ocean. For that, wild Atlantic specimens, after excluding individuals of farmed ancestry, were pooled to constitute the pure wild reference sample ( $W_{ATL}$ ), and  $F_{OR1}$  and  $F_{OR2}$  were used as farmed reference samples. These 3 samples were flagged as populations of known origin (POPFLAG = 1) to trace the genetic ancestry of all wild fish from the Atlantic Ocean and Baltic Sea (POPFLAG = 0). Three distinct genetic clusters with no migration were assumed ( $K = 3$ ;  $W_{ATL}$ ,  $F_{OR1}$  and  $F_{OR2}$ ) (migprior option = 0), using the same aforementioned parameters. The membership coefficient ( $q$ ) was used to classify individuals as wild or farmed ancestry considering the self-assignment test of wild individuals to the wild reference sample  $W_{ATL}$ . The  $q$ -values  $< 0.9$  were considered as 'admixed genotypes' of farmed ancestry (either pure, F1, backcrosses [BC], or more advanced hybrids).

The relative impact of farming and restocking was also evaluated in Galicia (NW Spain), a suitable area where aquaculture production occurs since the 1980s and the restocking took place between 2006 and 2011 after the *Prestige* oil spill. The 2 temporal samples outlined before (SP.old and SP.new) were analysed using STRUCTURE (USEPOP-INFO), following the same methodology and reference samples to detect fishes of farmed ancestry in both samples.

### Simulation of a molecular tool for monitoring escapees

Three lists of the 755 SNPs ranked by decreasing values of  $F_{ST}$  were obtained by comparing  $W_{ATL}$  vs.  $F_{OR1}$ ,  $W_{ATL}$  vs.  $F_{OR2}$  and  $F_{OR1}$  vs.  $F_{OR2}$  using GENEPOP 4.0 (Rousset 2008). Starting from allelic frequencies estimated for reference populations ( $W_{ATL}$ ,  $F_{OR1}$  and  $F_{OR2}$ ), 2 simulated populations,

one for training and one for testing, were generated for each of the 3 reference populations using HYBRIDLAB 1.0 (Nielsen et al. 2006). The testing populations were used to obtain the various hybrid categories: F1 ( $W_{ATL} \times F_{ORI1}$  or  $W_{ATL} \times F_{ORI2}$ ) and backcrosses BC<sub>F</sub> ( $F1 \times F_{ORI1}$  or  $F1 \times F_{ORI2}$ ) and BC<sub>W</sub> ( $F1 \times W_{ATL}$ ). A total of 500 individuals were generated for each simulated population, either pure or hybrid, and the analysis of  $W_{ATL}$  vs. farmed samples was done both for  $F_{ORI1}$  and  $F_{ORI2}$  independently.

The program STRUCTURE 2.3.4 was used again to assign simulated individuals to each category in different scenarios (see below). The program was sequentially executed 100 times using default parameters and the USEPOPINFO option defining the farmed (either  $F_{ORI1}$  or  $F_{ORI2}$ ) and  $W_{ATL}$  reference samples. Assignment was tested by progressively adding the most divergent SNPs one by one up to 100 SNPs (from highest to lowest  $F_{ST}$ ). Three different scenarios of progressive complexity were considered. In the simplest scenario, pure categories were tested (farmed vs. wild). In a second scenario, F1 hybrids were also included, such that 3 categories were evaluated (farmed, wild and F1). Finally, in the third and more complex scenario, backcrosses were also included, which led to the assignment of individuals to 5 categories (farmed, wild, F1, BC<sub>W</sub> and BC<sub>F</sub>). The mean  $q$ -value between adjacent classes was used as a threshold to classify simulated individuals; i.e. in the simplest scenario, the threshold was  $q = 0.5$ , and those individuals below or above the threshold were assigned to farm and wild categories, respectively; in the second scenario, the thresholds were 0.25 and 0.75, so those individuals below 0.25 or above 0.75 were assigned to farm or wild categories, respectively, while those between 0.75 and 0.25 were assigned to F1. In the last scenario, the thresholds were set at 0.125, 0.375, 0.675 and 0.875 defining the 5 categories: farmed, BC<sub>F</sub>, F1, BC<sub>W</sub> and wild following the same criteria. The percentage of well-classified individuals to each category was obtained for each set of SNPs and graphically represented to visualize the gaining in accuracy as the number of markers increased. The process was iterated 20 times, and the mean values of assignment to each category were obtained across replicates.

Finally, considering the amount of markers necessary to correctly classify simulated farmed and wild fish in the simplest scenario (see 'Results: Designing a practical tool for evaluation of farming impact in the wild'), a subset of SNPs was selected to test their capacity for discriminating individuals of farmed ancestry in the wild (real scenario). STRUCTURE

was used with the USEPOPINFO option activated and  $W_{ATL}$ ,  $F_{ORI1}$  or  $F_{ORI2}$  as reference samples. Additionally, the Bayesian method of Rannala & Mountain (1997) implemented in GENECLASS 2.0 (Piry et al. 2004) was also used to compare its assignment success with STRUCTURE using the same group of individuals and the same subset of markers.

## RESULTS

### Identification and genetic characterization of farmed and wild fish

Using STRUCTURE, 2 differentiated clusters were detected in farmed samples (Fig. 1), suggesting 2 main domestication origins, which agrees with previous information provided by the companies. Hence, farmed samples were divided into  $F_{ORI1}$  and  $F_{ORI2}$  for subsequent analyses including only pure individuals according to STRUCTURE data ( $q_{ORI1} > 0.9$  and  $q_{ORI2} < 0.1$ ). STRUCTURE analysis without *a priori* information and including all wild and farmed samples suggested that 106 individuals collected in the wild presented some degree of farmed ancestry (see next section) (Fig. 2). These individuals were removed from the wild collection for further evaluation of genetic diversity and differentiation between wild and farmed populations. Samples from the Atlantic Ocean ( $N = 502$ ) and Baltic Sea ( $N = 78$ ) were grouped separately to constitute pure Atlantic and Baltic reference groups (Table 1) in accordance with their distinct genetic constitution (Prado et al. 2018). Among wild samples, the Atlantic Ocean showed higher genetic diversity than the Baltic Sea with all estimators, while among farming groups,  $F_{ORI2}$  showed higher diversity than  $F_{ORI1}$  (Table 1). Globally, wild samples showed higher diversity than farmed ones, although the difference was more pronounced for the mean number

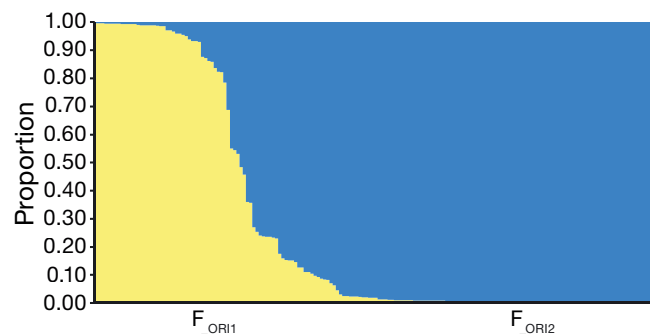


Fig. 1. STRUCTURE results for farmed samples of turbot showing 2 population units ( $K$ ) indicative of 2 genetic origins for domesticated individuals (yellow:  $F_{ORI1}$ , blue:  $F_{ORI2}$ )

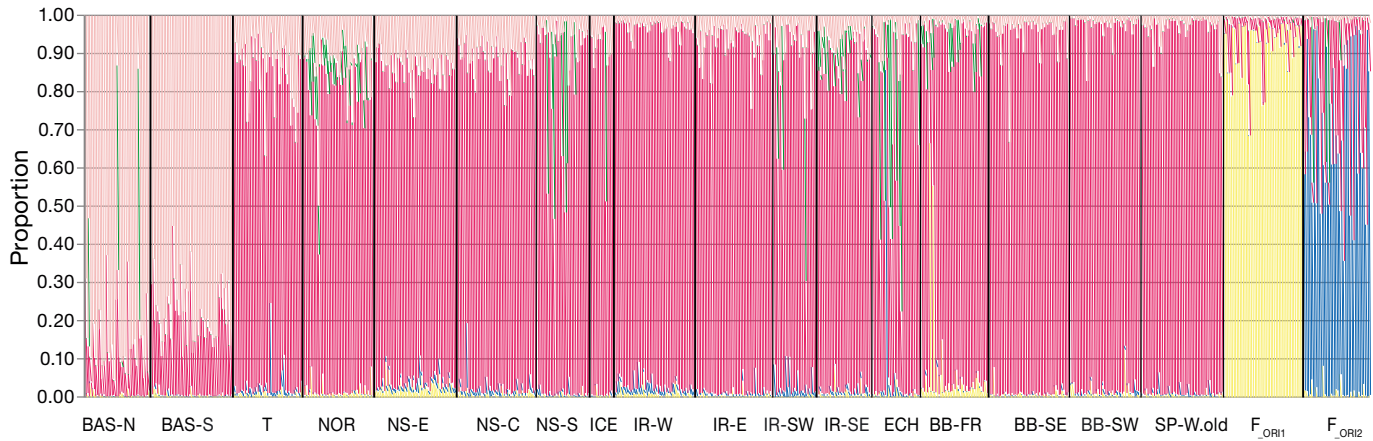


Fig. 2. STRUCTURE results for all wild and farmed samples of turbot genotyped for 755 SNPs, showing  $K = 5$  different genetic clusters, each represented with a different colour. Population codes in the x-axis, where each vertical bar corresponds to a single individual: BAS (Baltic Sea)-N (North), -S (South); T (transition area between the Atlantic and the Baltic Sea; Skagerrak location); NOR (Norway Sea); NS (North Sea)-E (East), -C (Central), -S (South); ICE (Iceland); IR (Ireland)-W (West), -E (East), -SW (South West), -SE (South East); ECH (English Channel); BB (Biscay Bay)-FR (France), -SE (South East), -SW (South West); SP-W.old (Spain West old sample); F\_ORI1, F\_ORI2: farmed turbot from 2 distinct origins

of alleles ( $N_a$ ). It is well accepted that genetic drift affects this parameter before it is obvious from heterozygosity (Luikart et al. 1998), and accordingly, farm-based founder effects and small effective population sizes would explain this observation.

Highly significant differentiation ( $F_{ST}$ ;  $p < 0.001$ ) was observed in all pairwise comparisons between the 4 groups considered (Table 2) and between all wild samples and the 2 farmed origins (Table S2 in Supplement 1). As expected, and according to the

Atlantic origin of the turbot farms, the Atlantic group differentiated less from the farmed samples than the Baltic group. Among the farmed samples, F\_ORI1 showed a higher differentiation from the Atlantic group ( $F_{ST} = 0.060$ ) than F\_ORI2 ( $F_{ST} = 0.039$ ). DAPC confirmed these results (Fig. 3).

A total of 25 outlier loci candidates for divergent selection in captivity were identified between wild turbot from the Atlantic Ocean and each of the 2

Table 1. Genetic diversity estimated for pure wild (Baltic Sea, Atlantic) and farmed (F\_ORI1, F\_ORI2) turbot based on 755 SNPs.  $H_e$ : expected heterozygosity;  $P_{99}$ : percentage of polymorphic loci;  $N_a$ : mean number of alleles per locus; SD: standard deviation

Sample	N	$H_e$ (SD)	$P_{99}$	$N_a$ (SD)
Baltic Sea	78	0.086 (0.140)	0.425	1.527 (0.500)
Atlantic	502	0.095 (0.138)	0.559	1.833 (0.373)
F_ORI1	48	0.087 (0.148)	0.393	1.393 (0.489)
F_ORI2	41	0.096 (0.149)	0.501	1.501 (0.500)

Table 2. Pairwise  $F_{ST}$  matrix between pure wild and farmed groups of turbot based on 755 SNPs. All  $F_{ST}$  values were significant after sequential Bonferroni correction ( $p < 0.005$ )

	Baltic Sea	Atlantic	F_ORI1	F_ORI2
Baltic Sea	–			
Atlantic	0.004	–		
F_ORI1	0.084	0.060	–	
F_ORI2	0.051	0.039	0.049	–

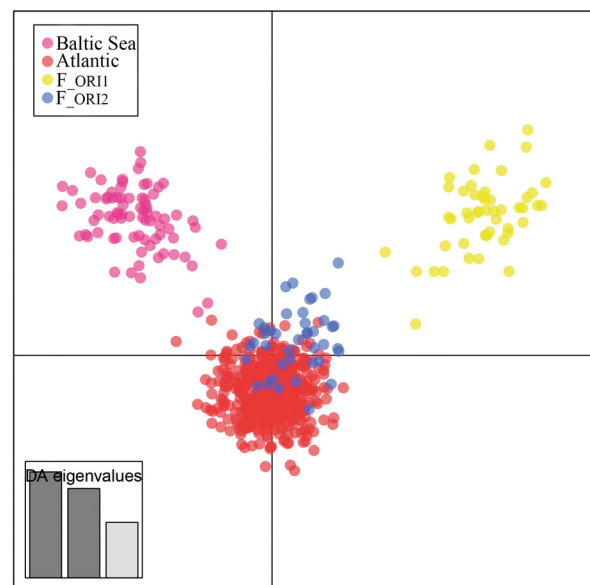


Fig. 3. Clustering of pure wild and farmed (see Table 1) samples of turbot genotyped for 755 SNPs using discriminant analyses of principal components (DAPC). Insert: Discriminant analysis (DA) eigenvalues of the retained discriminant functions

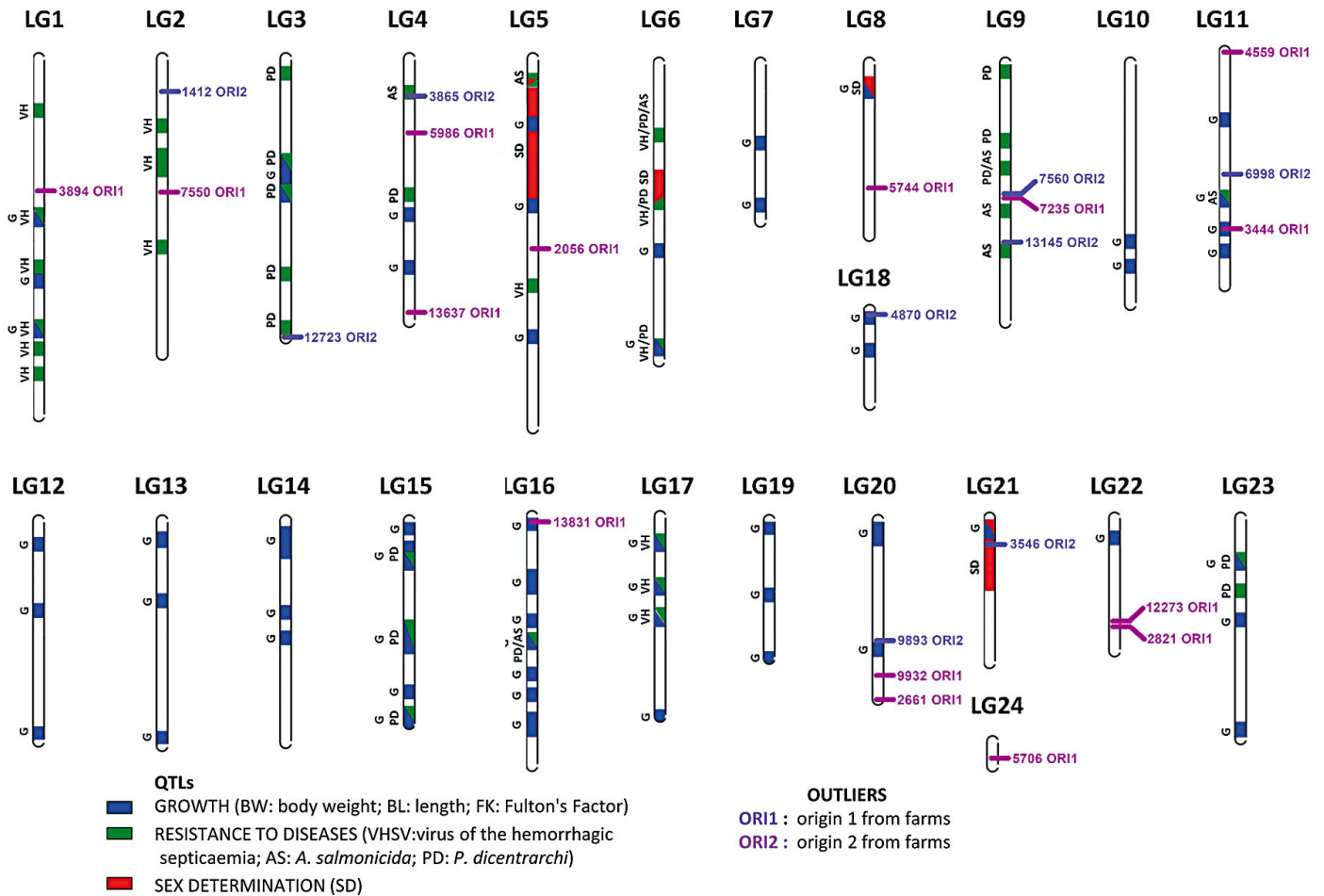


Fig. 4. Position of outlier loci related to domestication in farmed broodstock ( $F_{OR1}$  and  $F_{OR2}$ ) in the linkage groups (LG) of the turbot genetic map (from LG1 to LG23; Bouza et al. 2012)

farmed samples (Fig. S1 in Supplement 1). Among these, 2 were identified by BAYESCAN and 23 suggested by LOSITAN, and a total of 24 were anchored to the turbot genetic map (Fig. 4). Comparison with  $F_{OR1}$  returned 16 outliers, while 9 were detected with  $F_{OR2}$ . Although no markers were shared between farms, 7  $F_{OR1}$  and 6  $F_{OR2}$  outliers co-localized in the same linkage group (LG), and 2 pairs were very closely linked at LG9 and LG22, respectively, suggesting convergent selection/domestication at the farms. Two outlier loci (SNP codes: 7550 and 3865; Fig. 4, Table S5 in Supplement 2 at [www.int-res.com/articles/suppl/q010p447\\_supp2.xls](http://www.int-res.com/articles/suppl/q010p447_supp2.xls)) were previously reported to be associated with divergent selection (Prado et al. 2018). Three additional loci (4870, 13145 and 13831) lay very close (<1 Mb) to other outliers related to divergent selection in wild populations as reported by Vilas et al. (2010, 2015) and Prado et al. (2018). Furthermore, 5 and 3 outliers were located within the confidence intervals of QTLs associated

with growth and disease resistance, respectively (Martínez et al. 2016) (Fig. 4).

In this study, we specifically mined the turbot genome close to 2 regions at LG9 and LG22, where 2 outliers related to divergent selection were closely linked and, thus, consistently related to farm-based selection (Table 3). The pair of outliers at LG9 (7235ORI1 and 7560ORI2) was separated by 0.4 Mb, while the pair at LG22 was separated by 1.1Mb (2821ORI1 and 12273ORI1) (Fig. 4). Locus 7235ORI1 was located in an intronic region of the gene paralemmin 3 (*PALM3*), a gene related to toll signalling pathway, and 7560ORI2 was located in an intronic region of *SWI/SNF* related, matrix associated, actin-dependent regulator of chromatin, subfamily A, member 4 (*SMARCA4*), related to regulation of viral processes. Mining around these outliers ( $\pm 0.250$  Mb) showed GO terms related to immune system and metabolic process. Two genes within the mined region encoding for the alpha-1-acid glycoprotein 1

Table 3. Outlier loci for divergent selection between  $W_{ATL}$  and farmed samples ( $F_{OR1}$ ,  $F_{OR2}$ ) at LG9 and LG22 in turbot. Statistical significance per locus is presented in the following order: BAYESCAN/LOSITAN programs (– means not significant); LG: linkage group; SNP: single nucleotide polymorphism; asterisks indicates posterior probabilities (P) of 0.95 (\*) and 0.99 (\*\*) for BAYESCAN and a confidence threshold of 0.99 (\*\*) for LOSITAN; gene/function: official gene symbol and function; data mining: relevant genes identified  $\pm$  250 kb around outliers. QTL: quantitative trait loci; AS: resistance to *Aeromonas salmonicida*; PD: resistance to *Philasterides dicentrarchi*

LG	SNP	Farmed	Statistical significance	Region	Gene / function	Data mining	Gene ontology (GO) terms	QTLs
9	7560_71	$F_{OR2}$	–/**	Intron	<i>PALM3</i>	ORM1 <sup>a</sup> ; LILRB3 <sup>b</sup>	Metabolism & immunity	AS <sup>c</sup> ; PD <sup>d</sup>
	7235_80	$F_{OR1}$	–/**	Intron	Toll signaling pathway <i>SMARCA4</i>			
22	12273_75	$F_{OR1}$	–/**	Noncoding	Regulation viral process		Metabolism & immunity	
	2821_6	$F_{OR1}$	–/**	Noncoding				

<sup>a</sup>Cairns et al. (2008), <sup>b</sup>Castro & Tefalla (2015), <sup>c</sup>Rodríguez-Ramilo et al. (2011), <sup>d</sup>Rodríguez-Ramilo et al. (2013)

(*ORM1*) and the leukocyte immunoglobulin-like receptor subfamily B member 3 (*LILRB3*) have been linked to fish immunity (Cairns et al. 2008, Castro & Tafalla 2015). *ORM1* encodes for a key acute phase plasma protein, classified as an acute-phase reactant, while *LILRB3* for a protein which binds to MHC class I molecules on antigen-presenting cells and transduces a negative signal that inhibits stimulation of the immune response. Both loci at LG22 (2821OR1 and 12273OR1) were at intergenic regions but lay within a region where immune system and metabolic process GO terms were identified (Table 3). No significant enrichment was detected at either of the regions at FDR < 5%.

### Identification and impact of farmed fish on wild populations

$K = 5$  was suggested as the best number of clusters when including all wild,  $F_{OR1}$  and  $F_{OR2}$  samples

without reference populations (Fig. 2). However, the assignment distributions on the plot clearly showed the presence of 4 main clusters related to their geographical or farming origin (Baltic, Atlantic,  $F_{OR1}$  and  $F_{OR2}$ ), while the 5th green cluster (Fig. 2) was mixed in some populations. This cluster was mainly observed in the Atlantic area, especially at ECH, NS-S (North Sea-South), and less frequently at NOR and BB-FR (Biscay Bay France), but also at  $F_{OR2}$ , which could suggest a farmed origin. Accordingly, we defined a wild Atlantic reference sample ( $W_{ATL}$ ) using only those samples without any trace of putative farmed ancestry as suspected in NS-E, IR-W, IR-E, BB-SE and BB-SW. Self-assignment tests of  $W_{ATL}$  samples showed a minimum admixture coefficient of  $q_W = 0.919$  (see Table S3 in Supplement 2). Then, we applied the USEPOPINFO option implemented in STRUCTURE assuming  $K = 3$  and using  $W_{ATL}$ ,  $F_{OR1}$  and  $F_{OR2}$  as reference samples. Following a conservative criterion, we decided on a cut-off of  $q < 0.9$  to identify individuals of farmed

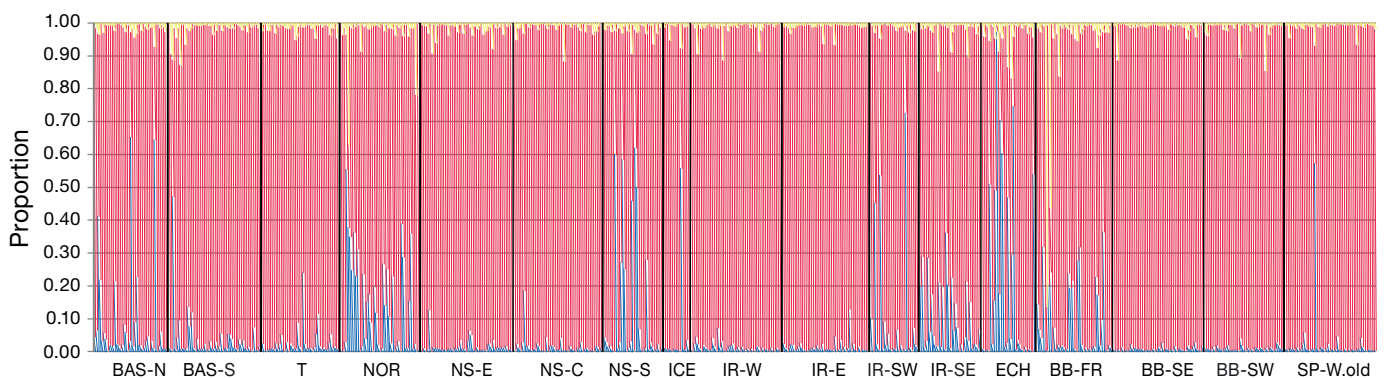


Fig. 5. STRUCTURE results for all wild samples of turbot genotyped for the whole panel of 755 SNPs. Data shows 3 population units ( $K$ ) corresponding to the 3 reference samples (red:  $W_{ATL}$ , yellow:  $F_{OR1}$ , blue:  $F_{OR2}$ ). See Fig. 2 legend for description of abbreviations



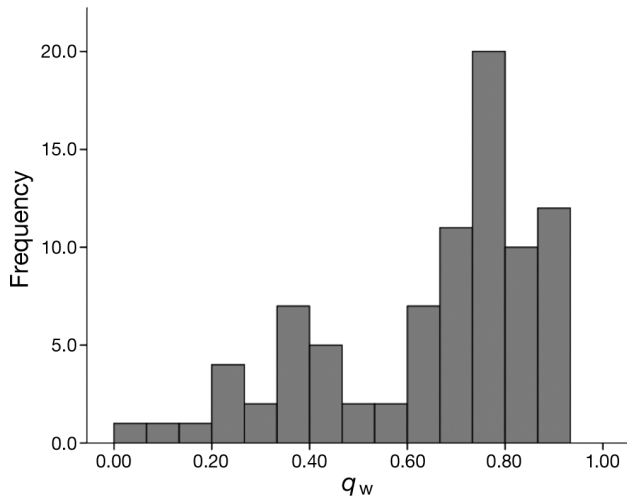


Fig. 6. Distribution of  $q_w$  values (membership coefficients to the wild cluster) of all turbot individuals with some degree of farm ancestry detected with the 755 SNPs in the Atlantic Ocean and Baltic Sea

ancestry considering the minimum self-assignment detected ( $q_w = 0.919$ ). Results supported connection of the previously observed green cluster (Fig. 2) to  $F_{\text{OR}12}$ , and a total of 106 individuals of farmed ancestry (15.5%) were identified (Fig. 5, Table S3). One fish from ECH ( $q_{\text{OR}12} = 0.973$ ) was presumably a pure farmed fish, but most fishes of farmed ancestry showed intermediate  $q$ -values, putatively representing different hybrid categories (F1, BCs or beyond), and were mainly detected in IR-SE (Ireland South East) (13; 43.3%), NOR (18; 41.9%), BB-FR (15; 36.6%), ECH (10; 34.5%) and NS-S (8; 25.0%) (Fig. 5, Table S3). The distribution of  $q_w$  values of fish with farmed ancestry (Fig. 6) was sharply asymmetric, and the main mode was located at 0.74, corresponding with  $BC_{\text{W}}$ . These features suggest that farmed individuals survive and, more importantly, reproduce with local wild fish.

### Farmed escapees vs. restocking

The old and new samples from the Spanish coast (SP.old and SP.new, respectively) were compared as a case study to evaluate the relative influence of aquaculture escapees and restocking. SP.old, only affected by escapees, showed a single individual of farmed ancestry ( $q_w = 0.661$  assigned to  $F_{\text{OR}12}$ ; 2%) and other fish close to the limit ( $q_w = 0.884$ ,  $q_{\text{OR}12} = 0.094$ ). SP.new, presumably impacted by escapees and restocking, included 5 individuals of farmed ancestry, 3 related to  $F_{\text{OR}11}$  and 2 to  $F_{\text{OR}12}$  (Table S4 in Supplement 2; Fig. S2 in Supplement 1; 10.4%).

Thus, although limited by the sample size and the level of introgression, our data support an increase in the number of individuals of farmed ancestry after restocking. Data also suggest that escapees from farms which are land-based have low impact in the wild.

### Designing a practical tool for evaluation of farming impact in the wild

Allele frequencies of the reference samples  $W_{\text{ATL}}$ ,  $F_{\text{OR}11}$  and  $F_{\text{OR}12}$  (Table S5 in Supplement 2) were used to estimate pairwise  $F_{\text{ST}}$  values per locus between the 3 samples for the 100 most divergent SNPs (Table S6 in Supplement 2).  $F_{\text{ST}}$  values ranged from 0.748 to 0.045 for  $W_{\text{ATL}}$  vs.  $F_{\text{OR}11}$  (mean = 0.177), from 0.461 to 0.068 for  $W_{\text{ATL}}$  vs.  $F_{\text{OR}12}$  (mean = 0.161) and from 0.625 to 0.071 (mean = 0.166) for  $F_{\text{OR}11}$  vs.  $F_{\text{OR}12}$ . Only 9 loci were shared between the 3 comparisons (Fig. S3 in Supplement 1), and 20 (9 + 11 shared between  $F_{\text{OR}11}$  and  $F_{\text{OR}12}$ ) were useful to discriminate between farmed and wild fish. These values show the difficulty in developing a single marker tool for identifying farmed turbot in the wild, irrespective of their origin, and thus, we simulated separately the power of markers to detect individuals of farmed ancestry for  $F_{\text{OR}11}$  and for  $F_{\text{OR}12}$ .

In the simplest scenario, where only pure wild and farmed individuals were involved (Fig. 7a,b), a 100% assignment was achieved with 40 and 80 SNPs in the  $F_{\text{OR}11}$  (Fig. 7a) and  $F_{\text{OR}12}$  (Fig. 7b) scenarios, respectively. Furthermore, with only 7 and 9 markers, respectively, a 90% success was achieved. In the second scenario, further including F1 hybrids, pure farmed and wild individuals were successfully assigned with a probability > 95% using the 100 SNPs, while a rate of ~90% was attained for F1 hybrids in both  $F_{\text{OR}11}$  and  $F_{\text{OR}12}$  scenarios (Fig. 7c,d). In this case, an assignment level > 80% was achieved for the 3 categories with 24 and 50 SNPs. In the more complex scenario, adding BCs (Fig. 7e,f), assignment remained high with the 100 SNPs for pure wild individuals (>90%), was moderate for the pure farmed (>80% for  $F_{\text{OR}11}$  and >70% for  $F_{\text{OR}12}$ ), but performed poorly for BCs and F1 hybrids, especially when involving  $F_{\text{OR}12}$  (Fig. 7f). In this scenario, 22 and 42 SNPs would be necessary for correctly assigning >80% of wild individuals. It should be noted that whatever the scenario, these last 2 sets of SNPs may suffice for a reasonable discrimination between wild vs. farmed ancestry individuals, irrespective of their category.

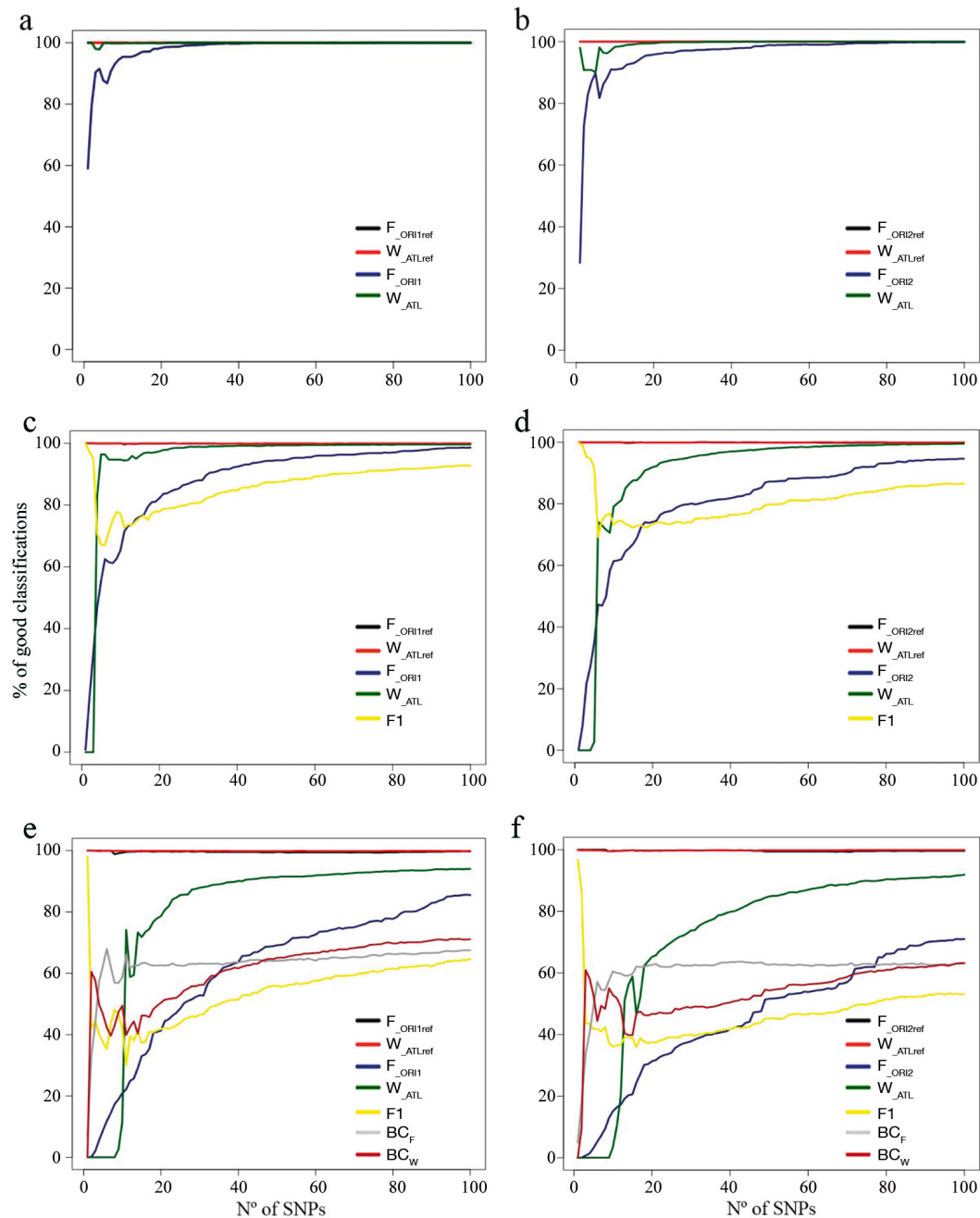


Fig. 7. Classification success of turbot pertaining to simulated populations and hybrid categories using up to 100 SNPs.  $F_{\text{ORI1}}$ : farmed origin 1;  $F_{\text{ORI2}}$ : farmed origin 2;  $W_{\text{ATL}}$ : pure wild population from the Atlantic region; F1: first generation of hybrids between wild and farmed; BC<sub>F</sub>: backcross between F1 and the respective farmed origin; BC<sub>W</sub>: backcross between F1 and the pure wild population; ref: wild and farmed reference genotypes; F, W, F1, BCs: genotypes generated by simulation for testing. Different scenarios were tested: (a,b) wild and farmed; (c,d) wild, farmed and F1 hybrid; (e,f) wild, farmed, F1 and BCs; (a,c,e) tests with  $F_{\text{ORI1}}$ ; (b,d,f) tests with  $F_{\text{ORI2}}$

### Testing the SNP tool in a real scenario

Finally, we contrasted assignment success of the small sets of SNPs resulting from the simulation (40 SNPs for  $F_{\text{ORI1}}$  and 80 SNPs for  $F_{\text{ORI2}}$ ) with regard to the full dataset (755 SNPs) in the real scenario

(Atlantic Ocean and Baltic Sea; Table S3). Here, we also used GENECLASS 2.0 (Piry et al. 2004) to compare the performance of different statistical approaches. Using STRUCTURE, 22 of 106 individuals of farmed ancestry detected with the full SNP dataset were not detected with either the 40 SNPs

( $F_{\text{OR1}}$ : 9 of 18) or the 80 SNPs ( $F_{\text{OR2}}$ : 13 of 75) subsets. Furthermore, 8 individuals of  $F_{\text{OR1}}$  and 14 of  $F_{\text{OR2}}$  ancestry were false positives in comparison with the full SNP dataset. Most discordances were in the limit of the established threshold ( $q_w = 0.9$ ), and only 5 of 43 misclassified individuals (11.6%) showed a  $q_w < 0.8$  (Fig. S4a in Supplement 1). Globally, a 6.4% discordance was detected between the full 755 SNPs vs. the SNP subsets (44 of 686), mainly related to the 40 SNP set ( $F_{\text{OR1}}$  ancestry). This suggests that, despite the simulation outcome, the number of SNPs could be enlarged to increase the power of the SNP tool to identify individuals of  $F_{\text{OR1}}$  ancestry. When STRUCTURE vs. GENECLASS 2.0 was applied with the aforementioned SNP subsets, a global discordance of 6.9% was detected (47 of 686), which was lower for individuals of  $F_{\text{OR1}}$  ancestry (5 out of 20 detected; 33.3%) than for  $F_{\text{OR2}}$  ancestry (42 of 91; 46.2%). Although discordances were again at the limit of the  $q_w$  distribution, the proportion of  $q_w < 0.8$  was higher than that detected with STRUCTURE (11.6%; Fig. S4b).

## DISCUSSION

Evaluating the presence of farmed individuals and their impact on wild populations is a primary issue for developing a sustainable aquaculture industry and for providing guidelines for fisheries management. Currently, turbot are listed as 'Near Threatened' in Europe (Golani et al. 2011) due to declines in wild catches and historical population reductions (Bouza et al. 2014, FAO 2016). In such a context, the genetic differentiation of farmed turbot might represent a serious risk for native populations if farm escapees occur or restocking with farmed individuals is carried out. To evaluate the impact of farmed fish in nature, a genetic characterization of wild populations, both regarding sampling representativeness of the distribution area as well as genotyping a panel of markers covering the whole genome, is essential. This information was previously reported in a study by Prado et al. (2018), where the main units for turbot management were defined and further refined using information of outlier loci related to adaptive variation. Here, the other necessary part of the work has been tackled through characterization of the farmed broodstock used for production and restocking. Comparison of wild and farmed pools allowed reasonable accuracy in establishing the impact of these activities in the wild as well as the development and validation of a useful tool for future monitoring.

## Genetic characterization of farmed turbot: differentiation from wild populations

Loss of genetic variability due to genetic drift and selection is a common concern in aquaculture (Danancher & Garcia-Vazquez 2011) because breeding populations are vulnerable to inbreeding depression in the long term (Chavanne et al. 2016). Although genetic diversity was globally lower in farm broodstock than in wild populations of turbot, this pattern was more related to the loss of rare allele variants than to global heterozygosity, which indicates an appropriate management of breeding programs in this species (Chavanne et al. 2016, Martínez et al. 2016). Loss of rare alleles has been commonly reported for farmed fish in comparison to the wild populations (Danancher & Garcia-Vazquez 2011, Glover et al. 2017) and, specifically, in turbot (Vilas et al. 2015).

Few generations of domestication and selection can generate important genetic differentiation between native and farmed stocks, as observed in salmonids (Blanchet et al. 2008, Glover et al. 2013) and flatfish (Danancher & Garcia-Vazquez 2011). Studies comparing genetic differentiation of farmed turbot broodstock with wild populations revealed pairwise  $F_{\text{ST}}$  values ranging from 0.047 to 0.075 for microsatellites and from 0.037 to 0.085 for SNPs (Bouza et al. 2002, Coughlan et al. 1998, Estoup et al. 1998, Vilas et al. 2015). In our study, 5 generations of selection of turbot (Janssen et al. 2017) produced similar differentiation with the original Atlantic populations (average  $F_{\text{ST}} = 0.059$ ). Slightly higher values were observed when comparing farmed and wild Atlantic salmon *Salmo salar* L. with  $F_{\text{ST}}$  values ranging from 0.020 to 0.090 (Glover et al. 2013), reaching an average of 0.094 when pooling wild and farmed salmon (Karlsson et al. 2011). This is in agreement with a higher number of generations of selection that Atlantic salmon has been exposed to.

While natural selection promotes adaptation of individuals to their particular environment, farmers routinely choose those individuals performing better for traits desirable for commercial production (Chavanne et al. 2016). Since the 1990s, turbot breeding programs have been mainly focused on increasing growth rate, but also weak fish or fish showing morphological abnormalities are routinely discarded (Martínez et al. 2016). This has led to farmed fish growing 65% faster than wild ones at the same age (Janssen et al. 2017). Although disentangling the genetic basis of traits subjected to selection is a difficult task (Martínez et al. 2016, Glover et al. 2017), our results suggested consistent outlier loci related to

artificial selection, as previously reported (Vilas et al. 2015). Some of these markers are related to growth and disease resistance, which agrees with the selection imposed at the farms. Furthermore, 2 outlier loci detected between wild and farmed turbot in our study had been previously related to divergent selection between Atlantic vs. Baltic/Black Sea populations (Prado et al. 2018). A similar result was reported by Vilas et al. (2015), who detected common outlier loci for divergent selection when comparing Atlantic vs. Baltic and Atlantic vs. farm populations. These authors suggested convergent evolution for growth due to artificial selection on the farms and to differences in temperature and salinity in the wild. In our study, we identified 2 consistent new genomic regions related to artificial selection pinpointed by 2 pairs of tightly linked outlier loci at LG9 and LG22. The region at LG9 involved both  $F_{\text{OR1}}$  and  $F_{\text{OR2}}$  outlier loci located within immune-related genes, which increases their reliability as true fingerprints for selection in captivity. Our data support the hypothesis that farmed turbot is adapting to the farm-specific rearing conditions. Consequently, farmed fish, originating from escapees or restocking programs, survive and reproduce in the wild at a reduced fitness, although further research will be necessary to confirm this issue (Bylemans et al. 2016).

### Identification of farmed turbot in the wild

Identification of individuals of farmed ancestry in the wild is challenging because farmed broodstock were founded with wild individuals not many generations ago. These challenges have been observed for other marine species subject to aquaculture and escapes, such as Atlantic cod *Gadus morhua* (Glover et al. 2011). In the case of turbot, 5 generations of selection have passed since breeding programs started with a broodstock of Atlantic origin, which has led to differentiation across the whole genome ( $F_{\text{ST}}$  close to 6%), but also at specific genomic regions underlying target traits for selection. Moreover, genetic differentiation of wild Atlantic turbot ( $W_{\text{ATL}}$ ) was higher for  $F_{\text{OR1}}$  than for  $F_{\text{OR2}}$ , the 2 farmed gene pools identified in our study. Information indicates that important restocking with farmed turbot has taken place since the 1990s and that released turbot can survive and reproduce in the wild (Bouza et al. 2014). This means that for ~25 yr (~5 to 6 generations; 4 yr generation interval), farmed turbot has been backcrossing with wild fish, the main component of natural populations, and accordingly, a nearly

admixed composition, biased towards wild fish  $q_{\text{W}}$ , is expected. In this scenario, a cut-off to identify fishes of farmed ancestry should be established, which unavoidably brings about statistical errors. More specifically, wild individuals might be wrongly assigning to farmed ancestry (Type I) and farmed individuals to wild ancestry (Type II). We followed a conservative criterion to identify fishes of farmed ancestry and established a threshold of  $q_{\text{W}} = 0.9$  considering that the minimum  $q_{\text{W}}$  detected after self-assignment of wild fish to their Atlantic origin was  $q_{\text{W}} = 0.921$  (238 individuals pertaining to NS-E, IR-W, IR-E, BB-SE and BB-SW). Following this approach, we are likely underestimating the proportion of fish of farmed ancestry in our study, but at the same time diminishing the likelihood of misclassifying wild fish as farmed.

Our results strongly support the presence of fish of farmed ancestry in the wild, mostly in the Atlantic Ocean, where aquaculture activity takes place, but, specifically, in the North Sea, where important restocking has been carried out in the last decades (Delbare & De Clerck 2000, Stottrup et al. 2002, Bouza et al. 2014). Admixture coefficients indicate that a few fish might be pure  $F_{\text{OR1}}$  or  $F_{\text{OR2}}$ , but most of them would represent hybrid classes of diverse ancestry biased towards a wild constitution (mean  $q_{\text{W}} = 0.74$ ). Such findings suggest that farmed fish not only survive in the wild, but also reproduce. As suggested in previous studies (Delbare & De Clerck 2000, Stottrup et al. 2002), reproduction of farmed fishes in the wild leads to introgression in native populations. Our results are very consistent for the most differentiated farm pool ( $F_{\text{OR1}}$ ) but suggest some caution with the interpretation of  $F_{\text{OR2}}$  because of its lower differentiation from wild populations and the limited information on the fish used for restocking activities in the North Sea. It is well known that large-scale releases have taken place in the North Sea, for example close to 400 000 turbot in Denmark between 1991 and 1998 (Stottrup et al. 2002), and the available information indicates the farmed origin of these fish (Bouza et al. 2014). Turbot released off the Belgian Coast, supplied by the France Turbot–Adrien Group company on Noirmoutier Island, survived well and dispersed widely, mainly to the Central North Sea and a small number into the English Channel (Delbare & De Clerck 2000). These are the areas where higher introgression was detected in our study. All in all, the data suggest that a proportion of ~15% of turbot caught in the wild were of farmed ancestry.

Introgression of mal-adapted domesticated conspecifics may lead to outbreeding depression (Allen-

dorf et al. 2010), changes in life-history characteristics (Bolstad et al. 2017), losses of genetic variability (Danancher & Garcia-Vazquez 2011) and losses in inter-population genetic differentiation (Glover et al. 2012). Genetic introgression and its impact represent a general concern in conservation genetics (Allendorf et al. 2012) and has been widely studied, especially with salmonids of commercial importance in Europe (Heino et al. 2015, Glover et al. 2017). As these species have been domesticated for a longer period than turbot, their impact on native populations is expected to be higher. Karlsson et al. (2011) found high introgression levels (up to 42.25%) in Norwegian wild salmon due to farm escapees. Only 3 to 5 generations of introgression between farmed and wild salmon have significantly changed the genetic integrity of the native populations, including possible reductions in fitness and adaptations to wild environmental conditions (Bourret et al. 2011, Bolstad et al. 2017). The consequences of introgression with farmed strains will mainly depend on the divergence between wild and farmed strains (Chavanne et al. 2016) and the natural structuring of wild populations (Danancher & Garcia-Vazquez 2011). These observations suggest that farmed turbot might disturb wild populations through the loss of local adaptations and outbreeding depression, and therefore, a mitigation plan should be established.

#### **Farmed turbot in the wild: aquaculture escapees vs. restocking**

Escapees from farms are thought to be limited in the case of turbot. Most fish are produced onshore, but some leakage of larvae might occur from production units in open circuits (Bouza et al. 2014). Furthermore, fry might escape from experimental cages that have been set up in the open sea (D. Chavarrías pers. comm.). Hence, it is not easy to disentangle the impact of farming and restocking since both activities have co-occurred across the Atlantic region. Moreover, no monitoring of either escapees or intentional releases in the areas close to the farming facilities has been carried out.

Here, we analyzed a specific scenario, where 2 samples collected before and after the restocking performed after the *Prestige* oil spill in Galicia (2006–2011) were available. Galicia is the main turbot production region in Europe (80%) and, consequently, an appropriate location to evaluate the impact of putative aquaculture escapees. The older sample (2002) could only be affected by aquaculture activi-

ties (since the 1980s in Galicia), while the more recent sample (2010) could be affected both by aquaculture and restocking. Our preliminary results suggest that farming escapees are scarce and that most individuals of farming ancestry originate from restocking. The observation is congruent with observations in the North Sea, where restocking seems to be the main source of turbot of farmed ancestry.

#### **Validating a genetic tool for monitoring**

Our simulation supports that a small set of SNPs suffices for the identification of individuals of farmed ancestry in the Atlantic scenario analyzed here, and it can be routinely used for monitoring wild populations in case of the suspicion of escapees or restocking. Presumably, 40 to 80 SNPs should suffice to assign pure wild or farmed fish, but an increasing number of SNPs would be necessary in case of a more complex scenario with hybrids. The pitfall is that the sets of SNPs to check for  $F_{OR1}$  or  $F_{OR2}$  hardly overlap because both farm pools differ so much. Consequently, a single panel could be used if an *a priori* hypothesis exists on the origin of escaped fish. Alternatively, one should genotype with both SNP panels and independent testing for each of the 2 farm pools. Although the identification of the type of hybrid is challenging, especially in the real scenario, discrimination of wild vs. farmed ancestry fishes could be reasonably addressed with small SNP panels for  $F_{OR1}$  and  $F_{OR2}$ , respectively. Analysis of the frequency distribution of the admixture coefficient ( $q$ ) in the real scenario might provide information on the history and the nature of the introgression process ( $q$  modes, mean, SD), very useful for management decisions. It should be borne in mind that our results are based on a representative sample in a specific generation of the broodstock from the 4 main turbot companies developing breeding programs and that additional assessments should be carried out routinely in the future, both considering genetic changes due to the selection and genetic drift, as well as the foundation of new companies. The comparison of the 40 and 80 SNP tools ( $F_{OR1}$  and  $F_{OR2}$ , respectively) with the full dataset in the real scenario rendered a high classification success (~95%), although this figure could be improved by increasing the number of SNPs, especially when  $F_{OR1}$  individuals are involved. Moreover, the use of the Bayesian and the multilocus likelihood approaches did not improve the interpretation. Hence, the first approach is recommended for routine evaluations.

## Management implications

Hatchery turbot have been widely used to enhance recruitment in wild populations, especially in the North Sea. Additionally, turbot of farmed origin have been used for habitat restoration programs, such as those following the *Prestige* oil spill off the NW Spanish coast. Our results strongly support that farmed turbot survive and reproduce in the wild, and consequently, caution should be taken considering that fitness of wild populations may be lowered through hybridization (Castellani et al. 2018). The estimated contribution of turbot of farmed ancestry is not very high in the Atlantic region (15.5%) but raises concerns in specific populations in the North Sea and adjacent areas, where it reached up to 40%. In principle, restocking and supplementation programs are not an option, especially with farmed fish, because the highly selected strains differ too much from the wild populations. However, restocking and supplementation with wild and outbred broodstock might be considered, provided that it is the only option left, and that progeny from captive fish mimic natural conditions (Milot et al. 2013) and that scientific follow-up is included. Restocking is common practice in freshwater, especially with salmonids, and to a much lesser extent in coastal habitats for stock enhancement (Wada et al. 2012) and community restoration (e.g. the *Prestige* oil spill). Because turbot lives in a rather heterogeneous environment and is locally adapted (Vilas et al. 2010, 2015, Vandamme et al. 2014, Prado et al. 2018), broodstock should be sourced locally. Genetic monitoring of natural populations with the SNP panels could also be used to monitor natural populations and to check for intentional or unintentional releases (Glover 2010). Our data suggest that the impact of farm escapees on wild populations are likely negligible, but areas of production should be regularly monitored. In the future, it is important to obtain a more refined picture of introgression in the wild to verify its short- and long-term population impacts, such as competition, demographical changes or genetic degradation affecting the evolutionary potential of the population. In this sense, it would be of great importance to increase our knowledge on the fitness of introgressed turbot populations, and for this purpose, performing common garden experiments could help reveal the level of (mal)adaptive variation of farmed individuals and their hybrids in comparison to the wild ones.

**Acknowledgements.** The authors thank all AquaTrace partners, Dr. Sara Vandamme and the Institute for Agricultural and Fisheries Research (ILVO) for sampling support. Authors are also indebted to Dr. K. A. Glover and 2 anonymous reviewers for their constructive comments on an earlier version of this manuscript. The project was funded by the 7th Framework Programme for research (FP7) under 'Knowledge-Based Bio-Economy — KBBE', Theme 2: 'Food, Agriculture and fisheries, and Biotechnologies' Project identifier: FP7-KBBE-2012-6-singlestage Grant agreement no.: 311920 'The development of tools for tracing and evaluating the genetic impact of fish from aquaculture: AquaTrace' and the Spanish Regional Government Xunta de Galicia GRC2014/010. Ciência sem Fronteiras/CAPES – Brazil supported the fellowship for the stay of F.D.P. at USC. The bioinformatic resources of the Centro de Supercomputación de Galicia (CESGA) were used for SNP genotyping.

## LITERATURE CITED

- ✦ Allendorf FW, Hohenlohe PA, Luikart G (2010) Genomics and the future of conservation genetics. *Nat Rev Genet* 11:697–709
- Allendorf FW, Luikart GH, Aitken SN (2012) *Conservation and the genetics of populations*, 2nd edn. Wiley-Blackwell, Chichester
- ✦ Antao T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G (2008) LOSITAN: a workbench to detect molecular adaptation based on a  $F_{st}$ -outlier method. *BMC Bioinformatics* 9:323
- ✦ Araguas RM, Vera M, Aparicio E, Sanz N, Fernández-Cebrián R, Marchante C, García-Marín JL (2017) Current status of the brown trout (*Salmo trutta*) populations inside eastern Pyrenees genetic refuges. *Ecol Freshwat Fish* 26:120–132
- Arechavala-Lopez P, Toledo-Guedes K, Izquierdo-Gomez D, Šegvi-Bubi T, Sanchez-Jerez P (2018) Implications of sea bream and sea bass escapes for sustainable aquaculture management: a review of interactions, risks and consequences. *Rev Fish Sci Aqua* 26:214–234
- ✦ Arias J, Sánchez L, Martínez P (1995) Low stocking incidence in brown trout populations from Northwestern Spain monitored by *LDH-5\** diagnostic marker. *J Fish Biol* 47:170–176
- ✦ Blanchet S, Páez DJ, Bernatchez L, Dodson JJ (2008) An integrated comparison of captive-bred and wild Atlantic salmon (*Salmo salar*): implications for supportive breeding programs. *Biol Conserv* 141:1989–1999
- ✦ Bolstad GH, Hindar K, Robertsen G, Jonsson B and others (2017) Gene flow from domesticated escapes alters the life history of wild Atlantic salmon. *Nat Ecol Evol* 1:124
- ✦ Bourret V, O'reilly PT, Carr JW, Berg PR, Bernatchez L (2011) Temporal change in genetic integrity suggests loss of local adaptation in a wild Atlantic salmon (*Salmo salar*) population following introgression by farmed escapees. *Heredity* 106:500–510
- ✦ Bouza C, Presa P, Castro J, Sánchez L, Martínez P (2002) Allozyme and microsatellite diversity in natural and domestic populations of turbot (*Scophthalmus maximus*) in comparison with other Pleuronectiformes. *Can J Fish Aquat Sci* 59:1460–1473
- Bouza C, Hermida M, Pardo BG, Vera M and others (2012) An Expressed Sequence Tag (EST)-enriched genetic map of turbot (*Scophthalmus maximus*): a useful frame-

- work for comparative genomics across model and farmed teleosts. *BMC Genet* 13:54
- Bouza C, Vandamme S, Hermida M, Cabaleiro S, Volckaert F, Martínez P (2014) AquaTrace species leaflet turbot (*Scophthalmus maximus*). <https://fishreg.jrc.ec.europa.eu/web/aquatrace/leaflets/turbot> (accessed 5 June 2017)
- ✦ Bylemans J, Maes GE, Diopere E, Cariani A and others (2016) Evaluating genetic traceability methods for captive-bred marine fish and their applications in fisheries management and wildlife forensics. *Aquacult Environ Interact* 8:131–145
- ✦ Cairns MT, Johnson MC, Talbot AT, Pemmasani JK and others (2008) A cDNA microarray assessment of gene expression in the liver of rainbow trout (*Oncorhynchus mykiss*) in response to a handling and confinement stressor. *Comp Biochem Physiol Part D Genomics Proteomics* 3:51–66
- ✦ Castellani M, Heino M, Gilbey J, Araki H, Svåsand T, Glover KA (2018) Modeling fitness changes in wild Atlantic salmon populations faced by spawning intrusion of domesticated escapees. *Evol Appl* 11:1010–1025
- Castro R, Tafalla C (2015) Overview of fish immunity. In: Peatman BHB (ed) *Mucosal health in aquaculture*. Academic Press, San Diego, p 3–54
- ✦ Chavanne H, Janssen K, Hofherr J, Contini F and others (2016) A comprehensive survey on selective breeding programs and seed market in the European aquaculture fish industry. *Aquacult Int* 24:1287–1307
- ✦ Coughlan JP, Inslan AK, Galvin PT, Fitzgerald RD, Naevdal G, Cross TF (1998) Microsatellite DNA variation in wild populations and farmed strains of turbot from Ireland and Norway: a preliminary study. *J Fish Biol* 52:916–922
- ✦ Danancher D, Garcia-Vazquez E (2011) Genetic population structure in flatfishes and potential impact of aquaculture and stock enhancement on wild populations in Europe. *Rev Fish Biol Fish* 21:441–462
- Delbare D, De Clerck R (2000) Release of reared turbot in Belgian coastal waters as a tool for stock enhancement. *ICES CM* 2000/O:2
- ✦ Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour* 4:359–361
- ✦ Ellis T, Hughes RN, Howell BR (2002) Artificial dietary regime may impair subsequent foraging behaviour of hatchery-reared turbot released into the natural environment. *J Fish Biol* 61:252–264
- ✦ Estoup A, Gharbi K, SanCristobal M, Chevalet C, Haffray P, Guyomard R (1998) Parentage assignment using microsatellites in turbot (*Scophthalmus maximus*) and rainbow trout (*Oncorhynchus mykiss*) hatchery populations. *Can J Fish Aquat Sci* 55:715–725
- ✦ Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* 14:2611–2620
- ✦ Exadactylos A, Rigby MJ, Geffen AJ, Thorpe JP (2007) Conservation aspects of natural populations and captive-bred stocks of turbot (*Scophthalmus maximus*) and Dover sole (*Solea solea*) using estimates of genetic diversity. *ICES J Mar Sci* 64:1173–1181
- ✦ Excoffier L, Lischer HE (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* 10: 564–567
- FAO (2016) The state of world fisheries and aquaculture 2016. Contributing to food security and nutrition for all. FAO, Rome
- ✦ Figueras A, Robledo D, Corvelo A, Hermida M and others (2016) Whole genome sequencing of turbot (*Scophthalmus maximus*): a fish adapted to demersal life. *DNA Res* 23:181–192
- ✦ Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and co-dominant markers: a Bayesian perspective. *Genetics* 180: 977–993
- ✦ Froehlich HE, Runge CA, Gentry RR, Gaines SD, Halpern B (2018) Comparative terrestrial feed and land use of an aquaculture-dominant world. *Proc Natl Acad Sci USA* 115:5295–5300
- ✦ Glover KA (2010) Forensic identification of fish farm escapees: the Norwegian experience. *Aquacult Environ Interact* 1:1–10
- ✦ Glover KA, Dahle G, Jorstad KE (2011) Genetic identification of farmed and wild Atlantic cod, *Gadus morhua*, in coastal Norway. *ICES J Mar Sci* 68:901–910
- ✦ Glover KA, Quintela M, Wennevik V, Besnier F, Sørvik AGE, 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:e43129
- ✦ Glover KA, Pertoldi C, Besnier F, Wennevik V, Kent M, Skaala Ø (2013) Atlantic salmon populations invaded by farmed escapees: quantifying genetic introgression with a Bayesian approach and SNPs. *BMC Genet* 14:74
- ✦ Glover KA, Solberg MF, McGinnity P, Hindar K and others (2017) Half a century of genetic interaction between farmed and wild Atlantic salmon: status of knowledge and unanswered questions. *Fish Fish* 18:890–927
- ✦ Golani D, Kada O, Nouar A, Quignard JP, Cuttelod A (2011) *Scophthalmus maximus*. The IUCN Red List of Threatened Species 2011: e.T198731A9089507 (accessed 5 June 2017)
- ✦ Habtemariam BT, Arias A, Garcia-Vazquez E, Borrell YJ (2015) Impacts of supplementation aquaculture on the genetic diversity of wild *Ruditapes decussatus* from northern Spain. *Aquacult Environ Interact* 6:241–254
- ✦ Heino M, Svasand T, Wennevik V, Glover KA (2015) Genetic introgression of farmed salmon in native populations: quantifying the relative influence of population size and frequency of escapees. *Aquacult Environ Interact* 6: 185–190
- ✦ Iglesias J, Ojea G, Otero JJ, Fuentes L, Ellis T (2003) Comparison of mortality of wild and released reared 0-group turbot, *Scophthalmus maximus*, on an exposed beach (Ría de Vigo, NW Spain) and a study of the population dynamics and ecology of the natural population. *Fish Manag Ecol* 10:51–59
- ✦ Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23:1801–1806
- ✦ Janssen K, Chavanne H, Berentsen P, Komen H (2017) Impact of selective breeding on European aquaculture. *Aquaculture* 472:8–16
- ✦ Jombart T, Ahmed I (2011) Adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* 27: 3070–3071
- ✦ Karlsson S, Moen T, Lien S, Glover KA, Hindar K (2011)

- Generic genetic differences between farmed and wild Atlantic salmon identified from a 7K SNP-chip. *Mol Ecol Resour* 11:247–253
- Karlsson S, Diserud OH, Moen T, Hindar K (2014) A standardized method for quantifying unidirectional genetic introgression. *Ecol Evol* 4:3256–3263
- Lischer HEL, Excoffier L (2012) PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics* 28:298–299
- Luikart G, Allendorf FW, Cornuet JM, Sherwin WB (1998) Distortion of allele frequency distributions provides a test for recent population bottlenecks. *J Hered* 89:238–247
- Mariño JC, Pérez G, Rodríguez JL, Cid E, Fernández B, Graña MI (2009) Repoblación de rodaballo (*Scophthalmus maximus*, L.) en la Costa de Galicia. Libro de Resúmenes. XII Congreso Nacional de Acuicultura, Madrid
- Martínez P, Robledo D, Rodríguez-Ramilo ST, Hermida M and others (2016) Turbot (*Scophthalmus maximus*) genomic resources: application for boosting aquaculture production. In: MacKenzie S, Jentoft S (eds) *Genomics in aquaculture*. Elsevier, London, p 131–163
- Milot E, Perrier C, Papillon L, Dodson JJ, Bernatchez L (2013) Reduced fitness of Atlantic salmon released in the wild after one generation of captive breeding. *Evol Appl* 6: 472–485
- Narum SR, Hess JE (2011) Comparison of FST outlier tests for SNP loci under selection. *Mol Ecol Resour* 11:184–194
- Nielsen EE, Bach LA, Kotlicki P (2006) HYBRIDLAB (Version 1.0): a program for generating simulated hybrids from population samples. *Mol Ecol Notes* 6:971–973
- Nielsen EE, Hemmer-Hansen J, Larsen PF, Bekkevold D (2009) Population genomics of marine fishes: identifying adaptive variation in space and time. *Mol Ecol* 18: 3128–3150
- Noble TH, Smith-Keune C, Jerry DR (2014) Genetic investigation of the large-scale escape of a tropical fish, barramundi *Lates calcarifer*, from a sea-cage facility in northern Australia. *Aquacult Environ Interact* 5:173–183
- Paulsen H, Stottrup JG (2004) Growth rate and nutritional status of wild and released reared juvenile turbot in southern Kattegat, Denmark. *J Fish Biol* 65:210–230
- Pauly E, Zeller D (2017) Comments on FAOs *State of the World Fisheries and Aquaculture* (SOFIA 2016). *Mar Policy* 77:176–181
- Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLOS ONE* 7:e37135
- Piry S, Alapetite A, Cornuet JM, Paetkau D, Baudouin L, Estoup A (2004) GENECLASS2: a software for genetic assignment and first-generation migrant detection. *J Hered* 95:536–539
- Prado FD, Fernández-Cebrián R, Hashimoto DT, Senhorini JA, Foresti F, Martínez P, Porto-Foresti F (2017) Hybridization and genetic introgression patterns between two South American catfish along their sympatric distribution range. *Hydrobiologia* 788:319–343
- Prado FD, Vera M, Hermida M, Bouza C and others (2018) Parallel evolution and adaptation to environmental factors in a marine flatfish: implications for fisheries and aquaculture management of the turbot (*Scophthalmus maximus*). *Evol Appl* 11:1322–1341
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959
- Pritchard VL, Erkinaro J, Kent MP, Niemelä E, Orell P, Lien S, Primmer CR (2016) Single nucleotide polymorphisms to discriminate different classes of hybrid between wild Atlantic salmon and aquaculture escapees. *Evol Appl* 9: 1017–1031
- Rodríguez-Ramilo S, Toro MA, Bouza C, Hermida M, Pardo BG, Cabaleiro S, Martínez P (2011) QTL detection for *Aeromonas salmonicida* resistance related traits in turbot (*Scophthalmus maximus*). *BMC Genomics* 12:541
- Rodríguez-Ramilo ST, Fernández J, Toro MA, Bouza C and others (2013) Uncovering QTL for resistance and survival to *Philasterides dicentrarchi* in turbot (*Scophthalmus maximus*). *Animal Genetics* 44:149–157
- R Development Core Team (2014) The R project for statistical computing, Version 3.2.3. R Foundation for Statistical Computing, Vienna
- Rannala B, Mountain JL (1997) Detecting immigration by using multilocus genotypes. *Proc Natl Acad Sci USA* 94: 9197–9201
- Rice WR (1989) Analyzing tables of statistical tests. *Evolution* 43:223–225
- Robledo D, Hermida M, Rubiolo JA, Fernández C, Blanco A, Bouza C, Martínez P (2017) Integrating genomic resources of flatfish (Pleuronectiformes) to boost aquaculture production. *Comp Biochem Physiol Part D Genomics Proteomics* 21:41–55
- Rousset F (2008) GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Mol Ecol Resour* 8:103–106
- Saura M, Carabaño MJ, Fernández A, Doeschl-Wilson A and others (2017) Genetic basis of susceptibility and tolerance for scuticilociliatosis in turbot. Oral Communication presented at the Aquaculture Europe 17 Congress, Dubrovnik
- Seeb JE, Carvalho G, Hauser L, Naish K, Roberts S, Seeb LW (2011) Single-nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in nonmodel organisms. *Mol Ecol Resour* 11:1–8
- Shimada Y, Shikano T, Merilä J (2011) A high incidence of selection on physiologically important genes in the three-spined stickleback, *Gasterosteus aculeatus*. *Mol Biol Evol* 28:181–193
- Souche EL, Hellemans B, Babbucci M, Macaoidh E and others (2015) Range-wide population structure of European sea bass *Dicentrarchus labrax*. *Biol J Linn Soc* 116:86–105
- Sparrevoorn CR, Nielsen A (2003) Estimating size dependent diet, advection, diffusion, catchability and mortality applying a normal distribution model to spreading of released turbot. *J Fish Biol* 63(S1):257
- Stottrup JG, Sparrevoorn CR, Modin J, Lehmann K (2002) The use of releases of reared fish to enhance natural populations—a case study on turbot *Psetta maxima* (Linne, 1758). *Fish Res* 59:161–180
- Taranger GL, Karlsen Ø, Bannister RJ, Glover KA and others (2015) Risk assessment of the environmental impact of Norwegian Atlantic salmon farming. *ICES J Mar Sci* 72: 997–1021
- Toledo-Guedes K, Sanchez-Jerez P, Benjumea ME, Brito A (2014) Farming-up coastal fish assemblages through a massive aquaculture escape event. *Mar Environ Res* 98: 86–95
- Vandamme SG, Maes G, Raeymaekers J, Cottenie K and



- others (2014) Environmental selective pressure reinforces historical differentiation in turbot (*Scophthalmus maximus*). *Mol Ecol* 23:618–636
- ✦ Vilas R, Bouza C, Vera M, Millán A, Martínez P (2010) Variation in anonymous and EST-microsatellites suggests adaptive population divergence in turbot. *Mar Ecol Prog Ser* 420:231–239
- ✦ Vilas R, Vandamme SG, Vera M, Maes GE, Bouza C, Volkcaert FAM, Martínez P (2015) A genome scan for candidate genes involved in the adaptation of turbot (*Scophthalmus maximus*). *Mar Genomics* 23:77–86
- ✦ Viñas L, Franco MA, Soriano JA, González JJ, Ortiz L, Bayona JM, Albaigés J (2009) Accumulation trends of petroleum hydrocarbons in commercial shellfish from the Galician coast (NW Spain) affected by the Prestige oil spill. *Chemosphere* 75:534–541
- ✦ Wada T, Kamiyama K, Shimamura S, Mizuno T, Nemoto Y (2012) Effectiveness of stock enhancement of a rare species, spotted halibut *Verasper variegatus*, in Fukushima, Japan. *Aquaculture* 364–365:230–239
- ✦ Waples RS (1999) Dispelling some myths about hatcheries. *Fisheries* (Bethesda, Md) 24:12–21
- ✦ Worm B, Hilborn R, Baum JK, Branch TA and others (2009) Rebuilding global fisheries. *Science* 325:578–585

*Editorial responsibility: Kevin Glover,  
Bergen, Norway*

*Submitted: March 26, 2018; Accepted: August 20, 2018  
Proofs received from author(s): September 21, 2018*