Vol. 8: 131–145, 2016
doi: 10.3354/aei00164

AQUACULTURE ENVIRONMENT INTERACTIONS
Aquacult Environ Interact

Published February 25

Evaluating genetic traceability methods for captive-bred marine fish and their applications in fisheries management and wildlife forensics

Jonas Bylemans[1,2,*], Gregory E. Maes[1,3], Eveline Diopere[1], Alessia Cariani[4], Helen Senn[5], Martin I. Taylor[6], Sarah Helyar[7], Luca Bargelloni[8], Alessio Bonaldo[9], Gary Carvalho[10], Ilaria Guarniero[9], Hans Komen[11], Jann Th. Martinsohn[12], Einar E. Nielsen[13], Fausto Tinti[4], Filip A. M. Volckaert[1], Rob Ogden[14]

[1]Laboratory of Biodiversity and Evolutionary Genomics, University of Leuven, Ch. Deberiotstraat 32, 3000 Leuven, Belgium

[2]Institute for Applied Ecology, University of Canberra, Canberra, ACT 2612, Australia

[3]Centre for Sustainable Tropical Fisheries and Aquaculture, Comparative Genomics Centre, College of Marine and Environmental Sciences, Faculty of Science and Engineering, James Cook University, Townsville, 4811 QLD, Australia

[4]Department of Biological, Geological and Environmental Sciences, University of Bologna, Ravenna 48123, Italy

[5]WildGenes Laboratory, Royal Zoological Society of Scotland, Edinburgh EH12 6TS, UK

[6]School of Biological Sciences, University of East Anglia, Norwich NR4 7TJ, UK

[7]Institute for Global Food Security, Queen's University Belfast, Belfast BT9 5BN, UK

[8]Department of Public Health, Comparative Pathology, and Veterinary Hygiene, University of Padova, Viale dell'Università 16, 35020 Legnaro, Italy

[9]Department of Veterinary Medical Sciences DIMEVET, University of Bologna, 40064 Bologna, Italy

[10]Molecular Ecology and Fisheries Genetics Laboratory, School of Biological Sciences, Environment Centre Wales, Bangor University, Bangor, Gwynedd LL57 2UW, UK

[11]Animal Breeding and Genomics Centre, Wageningen University, PO Box 338, 6700AH Wageningen, The Netherlands

[12]JRC.G.4 – Maritime Affairs, Institute for the Protection and Security of the Citizen (IPSC), Joint Research Centre (JRC), European Commission, Via Enrico Fermi 2749, 21027 Ispra VA, Italy

[13]Section for Marine Living Resources, National Institute of Aquatic Resources, Technical University of Denmark, Vejlsøvej 39, 8600 Silkeborg, Denmark

[14]TRACE Wildlife Forensics Network, Royal Zoological Society of Scotland, Edinburgh EH12 6TS, UK

ABSTRACT: Growing demands for marine fish products is leading to increased pressure on already depleted wild populations and a rise in aquaculture production. Consequently, more captive-bred fish are released into the wild through accidental escape or deliberate releases. The increased mixing of captive-bred and wild fish may affect the ecological and/or genetic integrity of wild fish populations. Unambiguous identification tools for captive-bred fish will be highly valuable to manage risks (fisheries management) and tracing of escapees and seafood products (wildlife forensics). Using single nucleotide polymorphism (SNP) data from captive-bred and wild populations of Atlantic cod *Gadus morhua* L. and sole *Solea solea* L., we explored the efficiency of population and parentage assignment techniques for the identification and tracing of captive-bred fish. Simulated and empirical data were used to correct for stochastic genetic effects. Overall, parentage assignment performed well when a large effective population size characterized the broodstock and escapees originated from early generations of captive breeding. Consequently, parentage assignments are particularly useful from a fisheries management perspective to monitor the effects of deliberate releases of captive-bred fish on wild populations. Population assignment proved to be more efficient after several generations of captive breeding, which makes it a useful method in forensic applications for well-established aquaculture species. We suggest the implementation of a case-by-case strategy when choosing the best method.

KEY WORDS: Aquaculture · Conservation genetics · Escapees · Fisheries management · Wildlife forensics

*Corresponding author: Jonas.Bylemans@canberra.edu.au

Publisher: Inter-Research · www.int-res.com

INTRODUCTION

Aquaculture is one of the fastest growing food-producing sectors and will remain so in the foreseeable future due to a growing human demand for animal protein and lipids (Braithwaite & Salvanes 2010) and the limits that have been reached for wild-capture fisheries production (FAO Fisheries and Aquaculture Department 2014). This has led to various challenges related to the aquaculture industry, including organic, chemical and pharmaceutical pollution (Seymour & Bergheim 1991), infectious diseases (Murray & Peeler 2005), feed supply (Naylor et al. 2000, 2009, Natale et al. 2013) and escapees (Kitada et al. 2009, Glover 2010, Glover et al. 2011, Noble et al. 2014).

Accidental escapees (Bekkevold et al. 2006, Glover et al. 2013, Noble et al. 2014) or deliberate releases (Bell et al. 2008, Kitada et al. 2009) of captive-bred marine fish may impact the environment, and the ecological and genetic integrity of wild fish populations (Braithwaite & Salvanes 2010, Laikre et al. 2010). First, a decrease in genetic diversity, and consequently a lower evolutionary potential, has been observed in wild marine fish populations which have been invaded by captive-bred conspecifics (Hindar et al. 1991, Weir & Grant 2005, Glover et al. 2013). Given that recent studies have indicated surprisingly fine-scale local genetic adaptation in marine fish (André et al. 2011, Nielsen et al. 2012, Vandamme et al. 2014), the introgression of captive-bred fish can be detrimental to the long-term survival of wild fish populations. Second, introgression might disrupt adaptive gene complexes, which reduces the fitness of hybrids and in turn may compromise the persistence of locally adapted populations (McGinnity et al. 2003, Danancher & Garcia-Vazquez 2011, Lamaze et al. 2013). Managing and mitigating risks and assessing the impacts of released/escaped captive-bred fish on local wild populations are thus of utmost importance to ensure the long-term sustainability of aquaculture and fisheries industries. Third, aquaculture companies might have legal obligations to report escapees and failure to comply with these regulations might result in fines (Glover 2010). As such, the ability to trace back escapees to the farm of origin constitutes a highly valuable asset in delivering evidence for legal action (Glover et al. 2008, Glover 2010). Finally, an increase in international trade and consumer awareness in recent decades has highlighted the need for accurate labelling of seafood products. Mislabelling to increase profits has been extensively documented in the seafood industry (Jacquet & Pauly 2008, Hanner et al. 2011, Mariani et al. 2014). Given that market prices of wild-caught marine fish species are generally higher than aquaculture sourced fish, fraudulent labelling captive-bred fish as 'wild-caught' may increase income for the perpetrator (Cline 2012, Warner et al. 2013). Hence, genetic identification methods for farmed and wild marine fish species would be extremely valuable in aquaculture and fisheries management and wildlife forensics.

For a large variety of commercially reared species, escapees and deliberate releases have been reported (Liao et al. 2003, Bell et al. 2008, Jensen et al. 2010, Danancher & Garcia-Vazquez 2011). However, due to their long breeding history and the availability of genetic tools, research on tracing and quantifying escapees has focused mainly on salmonids (Glover 2010, Glover et al. 2013) and only recently on sea bass and sea bream (Arechavala-Lopez et al. 2013, Somarakis et al. 2013, Brown et al. 2015). Extending standardized traceability methods to other commercially exploited marine fish species will thus advance research into the effects of escapees and restocking programmes.

The lack of a long breeding history in most cultured marine fish species complicates the genetic discrimination between wild and captive-bred marine fish, especially when the identification of the hatchery of origin is required. The recent domestication history of many marine fish results in similar allele frequencies in captive-bred and wild populations, which lowers the discrimination power of genetic markers (Duarte et al. 2007). Likewise, the absence of long-term selective breeding programmes reduces the likelihood of finding species-specific 'domestication' markers (Karlsson et al. 2011, Gjedrem et al. 2012). Stochastic and selective breeding processes in aquaculture and recent developments in genetic traceability tools can however facilitate discrimination between captive-bred and wild fish. The common use of a relatively small broodstock and the unwanted high variance in reproductive success within the hatchery will result in increased genetic differentiation between captive-bred and wild populations and a lower genetic diversity within the captive-bred population (Porta et al. 2006a,b, 2007). Within the marine environment, provided that a solid genetic baseline is available, wild fish can be individually assigned to their region and/or population of origin with high precision using gene-associated single nucleotide polymorphism (SNP) markers (Nielsen et al. 2012). Genetic background information is increasingly available for commercially important

fish species (Nielsen et al. 2009, Abadía-Cardoso et al. 2013, Clemento et al. 2014), which makes the use of simulation studies possible to assess the discrimination power of existing genetic markers for wild and captive-bred fish. Finally, while the rate of genetic drift at neutral markers depends on the effective population size (N_e) of the broodstock, SNP markers associated with important aquaculture traits (such as growth and disease resistance) are subjected to directional selection which will increase the degree of differentiation between wild and captive-bred populations (Glover et al. 2010). Such markers may introgress at different rates compared to selectively neutral markers (Lamaze et al. 2012, Hohenlohe et al. 2013), thus providing crucial insights into both the fitness and molecular consequences of escapees and restocking programmes.

Multiple approaches are available for identifying and discriminating between captive-bred and wild marine fish (Manel et al. 2005). The 2 main methods used to date are individual assignment (IA) and parentage-based tagging (PBT) (Manel et al. 2005, Jones et al. 2010). Most commonly used, IA methods rely on allele frequency differences between populations to assign an individual to its most likely source (Ogden 2008, Glover 2010, Nielsen et al. 2012). However, in order to achieve highly robust assignments, IA requires some level of genetic differentiation between populations and extensive genetic reference data (Manel et al. 2005, Nielsen et al. 2012). In contrast, PBT utilizes the genetic variation within the complete data to determine the most likely parental pair for a particular genotype and can achieve high assignment success even when genetic differentiation among populations is insufficient for IA (Jones & Ardren 2003, Steele et al. 2013).

Our study focuses on Atlantic cod *Gadus morhua* L., 1758 and sole *Solea solea* L., 1958, 2 commercially important fish of the Northeast Atlantic Ocean for which extensive genetic resources are available (Nielsen et al. 2012). Both species have a widespread distribution across the Northeast Atlantic Ocean, and their high commercial value has resulted in an increased interest in captive-breeding programmes, restocking, stock enhancement and sea ranching (Howell 1997, Kjesbu et al. 2006, Björnsson 2011). More specifically, declines in wild-caught Atlantic cod and advances in captive breeding and feed formulation have led to an increase in global aquaculture production, reaching 22 000 tons in 2010 (Rosenlund & Skretting 2006, Thurstan et al. 2010, FAO Fisheries and Aquaculture Department 2015a). Although cod aquaculture has recently decreased due to large catches on the northern fishing grounds (FAO Fisheries and Aquaculture Department 2015a), the use of traditional cage farming in cod aquaculture and the substantial interest in stock enhancement and sea ranching programmes continues to represent a significant risk for interactions between wild and hatchery-reared cod (Bekkevold et al. 2006, Jørstad et al. 2008, Björnsson 2011). Similarly, recent advances and changing economic perspective have increased the interest in sole aquaculture, with production peaking at 125 tons in 2010 but decreasing in recent years (Howell 1997, Imsland et al. 2003, FAO Fisheries and Aquaculture Department 2015b). Although intensive land-based recirculation systems are currently preferred in flatfish aquaculture, there is considerable interest to reduce production costs through less intensive systems (e.g. cage farming, stock enhancement and sea ranching) (Brown 2002, Kitada & Kishino 2006, Sparrevohn & Støttrup 2007). Hence, for both focal species, there is a considerable risk of introgression between captive-bred individuals and local wild populations.

In this study, we aimed to evaluate the utility of IA and PBT approaches to discriminate between captive-bred marine fish and natural fish populations. To achieve this, we used a combination of simulated and empirical SNP datasets to perform a series of assignment experiments in each species, across a range of potential scenarios. The level of genetic differentiation between captive-bred and wild marine fish will vary due to: (1) the number of captive-bred generations (F_n) prior to escape or release, (2) the number of broodstock and the strength of reproductive variance between broodstock individuals, which both influence N_e, and (3) genetic (and geographical) differences between the hatchery population and locally occurring wild populations with which the escapees will intermingle. We investigated each of these potential variables to evaluate their relative impact on assignment power under IA and PBT approaches. In addition, the effect of (in)complete reference samples was also assessed given that the availability and representative nature of reference samples will also affect traceability outcomes.

From the outset, we anticipated that increasing F_n, decreasing N_e and a distinct genetic origin of the broodstock will all favour IA, given that IA relies on the realized level of genetic differentiation between populations to make robust assignments. On the other hand, the performance of PBT will be negatively impacted by those parameters that reduce the genetic diversity within the captive-bred population (i.e. high F_n and low N_e) due to the difficulty of

excluding candidate parents from real parents. Therefore, in addition to evaluating the relative performance of the 2 approaches, we were interested in examining possible thresholds of F_n and N_e across which the optimal approach for determining fish origin actually changes.

MATERIALS AND METHODS

Sampling

Wild samples of 10 Atlantic cod and 14 sole populations have been previously collected from European waters and genotyped (Nielsen et al. 2012). An Atlantic cod broodstock ($A_{cod\text{-}BS}$) (n = 92) sourced from the ICES region 27.V.b2 – Faroe Bank was sampled from the Fiskaaling aquaculture research station (Faroe Islands). Atlantic-sourced (ICES 27.IV.c – Southern North Sea) sole were sampled from a Dutch experimental breeding farm, SOLEA in IJmuiden, and consisted of 2 full-sib families with 4 broodstock individuals ($A_{sole\text{-}BS}$) (n = 4) and their first-generation offspring (F_n = 1) ($A_{sole\text{-}F1}$) (n = 92) (Blonk et al. 2009). Captive-bred sole samples originating from the Mediterranean Sea (FAO 37.2.1 – North Adriatic) were obtained from a pilot farm of the Laboratory of Aquaculture, Department of Veterinary Medical Sciences of the University of Bologna, Italy, and included samples from a broodstock ($M_{sole\text{-}BS}$) (n = 26) and first-generation offspring (F_n = 1) ($M_{sole\text{-}F1}$) (n = 96), obtained from 4 batch spawnings ($M_{sole\text{-}F1\text{-}B1}$, $M_{sole\text{-}F1\text{-}B2}$, $M_{sole\text{-}F1\text{-}B3}$ and $M_{sole\text{-}F1\text{-}B4}$). More details on all populations used in this study are found in Supplement 1 (www.int-res.com/articles/suppl/q008p131_supp.pdf).

Genotypic data

Gene-associated SNP markers were available for: the wild populations of Atlantic cod (1258 SNPs), the wild populations of sole (427 SNPs), $A_{cod\text{-}BS}$ (427 SNPs), $A_{sole\text{-}BS}$ and $A_{sole\text{-}F1}$ (423 SNPs) (Table 1) (Nielsen et al. 2012, Diopere et al. 2014). Additional genotyping of the $M_{sole\text{-}BS}$ and $M_{sole\text{-}F1}$ samples was conducted using VeraCode™ technology on the BeadExpress platform (Illumina), following the manufacturer's instructions. Of the 427 available SNPs, the 192 most informative SNPs, showing high genetic discrimination values (F_{ST} values) between the Mediterranean populations, were genotyped (Nielsen et al. 2012). Quality assessment and genotype calling was performed using GenomeStudio v.2009.2 software (Illumina). Three individuals from the $A_{sole\text{-}BS}$, initially genotyped with the wild populations using the SAM assay (GoldenGate, GG) on the iScan platform (Illumina) and with the highest GG call rate for the selected panel (Diopere et al. 2014), were included as cross-platform genotyping controls to ensure comparability between the archived and newly generated data.

Marker selection

In order to obtain marker panels with sufficient assignment power and to ensure that they are easily transferrable between laboratories, a subset of 96 highly informative SNPs were selected based on the practical limitations of common genotyping platforms (Supplement 2 at www.int-res.com/articles/suppl/q008p131_supp.pdf). Given that cod data were only used in IA analyses (see 'Tracing escapees' below), and the ability of markers to distinguish between populations provides a good indication of their power in IA analyses, the available SNPs for cod were ranked based on the pairwise F_{ST} values calculated among the wild cod populations using FSTAT v.2.9.3 (Goudet 1995). The Atlantic and Mediterranean sole data (wild and aquaculture) were used in both IA and PBT. To maximize the traceability power of selected sole SNPs for IA, markers were first ranked based on the pairwise F_{ST} values obtained from comparisons between the wild Atlantic and Mediterranean populations respectively. PBT analyses, on the other hand, require markers with a high genetic variability within a population to make robust assignments. Consequently, a second ranking of markers was based on their polymorphic information content (PIC) calculated with Cervus v.3.0 (Marshall et al. 1998) using the com-

Table 1. Available and newly generated single nucleotide polymorphism (SNP) genotypic datasets for wild and captive-bred populations of *Solea solea* and *Gadus morhua*. Atl-Aqua = aquaculture population sourced from the Atlantic Ocean, Med-Aqua = aquaculture population sourced from the Mediterranean Sea

Species	Origin	SNP genotypic data	
		Available	Source
Sole	Wild	427	Nielsen et al. (2012)
	Atl-Aqua	423	Diopere et al. (2014)
	Med-Aqua	181	Current study
Cod	Wild	1258	Nielsen et al. (2012)
	Atl-Aqua	427	Nielsen et al. (2012)

bined data from the respective broodstocks ($A_{sole\text{-}BS}$ and $M_{sole\text{-}BS}$) and their genetically similar wild populations (GER and ADR1 respectively). The top 96 ranking SNPs were used in all further analyses and further reduced genotypic datasets were used in the assignment analyses to determine the assignment power of the selected loci (Table 2).

This selection procedure for highly informative markers is unlikely to suffer from high-grading bias (Anderson 2010a, Waples 2010) for 3 reasons. First, assignment power was estimated from a different (holdout) set of samples to those used for SNP selection. Second, outlier SNPs were defined using initially high sampling sizes (n ≈ 40) from various geographical locations, which reduces the effects of random sampling errors (Nielsen et al. 2012). Third, the use of 2 rigorous outlier detection methods and annotation information provides confidence that the high F_{ST} values of the selected markers are more likely to result from diversifying selection (i.e. real differentiation) rather than being at the extremes of a neutral marker F_{ST} distribution (Waples 2010, Nielsen et al. 2012).

Simulations of hatchery data

To formally evaluate the individual and combined impacts of F_n, N_e and the availability of reference data on the traceability efficiency of IA and PBT analyses, various breeding scenarios were simulated for both species. Data were simulated using the previously selected 96 SNPs of the $A_{cod\text{-}BS}$ and $M_{sole\text{-}BS}$ (including 7 individuals that died before reproduction) genotypes for cod and sole respectively (Fig. 1). An initial parental broodstock ($P_{1\text{-}SIM}$) and 4 offspring generations ($F_{1\text{-}SIM}$, $F_{2\text{-}SIM}$, $F_{3\text{-}SIM}$ and $F_{4\text{-}SIM}$) were simulated under the assumption of perfect Hardy-Weinberg (HW) equilibrium. Different simulation series were performed using various N_e values (N_e = 5, 10, 20 or 50) to simulate drift due to reproductive variance. HYBRIDLAB v.1.0 (Nielsen et al. 2006) was used to simulate offspring genotypes used in the IA analyses. For the PBT analyses, Nookie v.1.0 (Anderson 2014) was used to simulate offspring genotypes because it generates individual genotypes 'bred' from specific parental pairs which are required for parentage assignment in simulated generations, rather than simply simulating individuals from a pool of population allele frequencies. Comparisons of genetic diversity showed that datasets generated through both programs were comparable (Supplement 3 at www.int-res.com/articles/suppl/q008p131_supp.pdf).

Comparative data analyses

A detailed comparison of the traceability results based on both the simulated and empirical datasets is important to determine the optimal traceability approach for a specific scenario. To be able to compare simulated and empirical results, population genetic parameters (F_{ST} values, observed heterozygosity H_{obs} and expected heterozygosity H_{exp}) associated with each dataset have to be understood as they will strongly influence the traceability power of the datasets. With the most comprehensive empirical data being available for the Mediterranean captive-bred sole, population genetic parameters were calculated for the broodstock ($M_{sole\text{-}BS}$ and $P_{1\text{-}SIM}$ [N_e = 5, 10, 20, 50]) and first-generation offspring ($M_{sole\text{-}F1}$ and $F_{1\text{-}SIM}$ [N_e = 5, 10, 20, 50]). The Northern Adriatic population (ADR1), as the original source of $M_{sole\text{-}BS}$, was included in the analysis as a reference. Genetic diversity (H_{obs} and H_{exp}) was calculated for each independent dataset ($M_{sole\text{-}BS}$, $M_{sole\text{-}F1\text{-}B1}$, $M_{sole\text{-}F1\text{-}B2}$, $M_{sole\text{-}F1\text{-}B3}$, $M_{sole\text{-}F1\text{-}B4}$, $P_{1\text{-}SIM}$ [N_e = 5, 10, 20, 50], $F_{1\text{-}SIM}$ [N_e = 5, 10, 20, 50] and ADR1) using

Table 2. *Solea solea* and *Gadus morhua*. Datasets used in population (IA = individual assignment) and parentage (PBT = parentage-based tagging) analysis to test for effects of sampling regimes and traceability scenarios. Sampling Regime 1: reference data of the aquaculture population is available for the parental generation. Sampling Regime 2: reference data of the aquaculture population is limited to the founding broodstock. Scenario A: aquaculture broodstock originated from a genetically distinct population than the local wild populations. Scenario B: aquaculture broodstock originated from a local, genetically similar wild population. The number of single nucleotide polymorphisms (SNPs) used in each analysis is indicated between parentheses. SD = simulated data, EAD = empirical Atlantic data, EMD = empirical Mediterranean data, na = not applicable

Traceability method	Species	Sampling Regime		Scenario	
		1	2	A	B
IA	Sole	SD (96, 52)	SD (96)	na	EAD (96, 30, 1)
	Cod	SD (96)	SD (96)	na	na
PBT	Sole	SD (96, 48)	na	EAD (50, 30, 21) EMD (40, 35, 30)	EAD (50, 35, 30) EMD (40, 35, 30)
	Cod	na	na	na	na

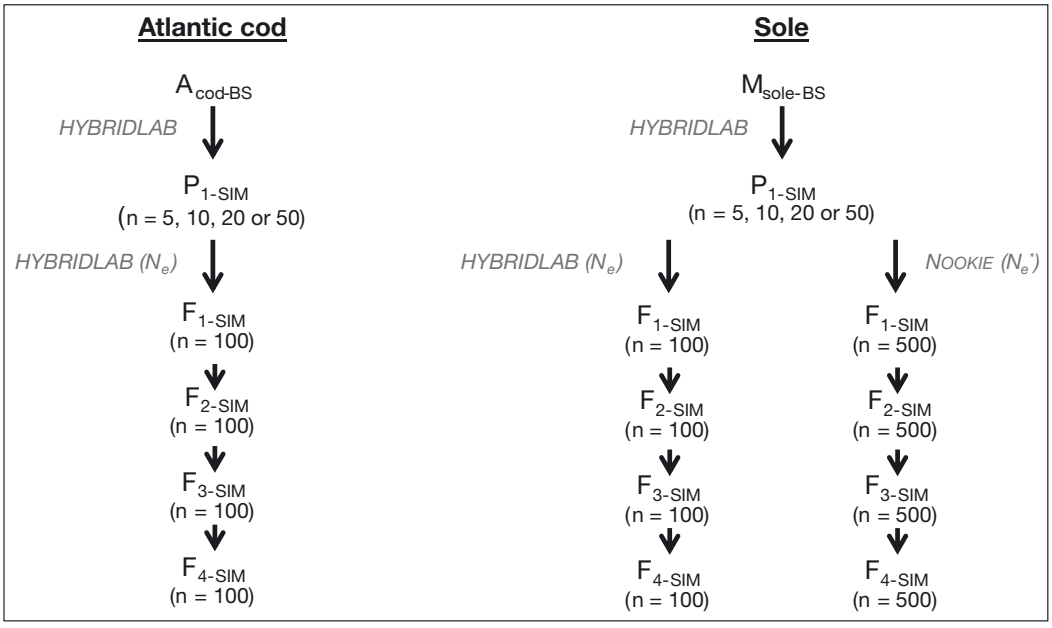

Fig. 1. *Gadus morhua* and *Solea solea*. Simulations used to generate captive-bred offspring genotypes. HYBRIDLAB v.1.0 (Nielsen et al. 2006) was used to generate the initial broodstock ($P_{1\text{-SIM}}$). Offspring genotypes used in individual assignment (IA) and parentage-based tagging (PBT) analyses were simulated with HYBRIDLAB and Nookie v.1.0 (Anderson 2014), respectively. $A_{cod\text{-}BS}$ = Atlantic cod broodstock, $M_{sole\text{-}BS}$ = Mediterranean sole broodstock, $F_{1\text{-}SIM}$, $F_{2\text{-}SIM}$, $F_{3\text{-}SIM}$, $F_{4\text{-}SIM}$ = 4 offspring generations, n = number of individuals, N_e (effective population size) = 5, 10, 20 or 50 and N_e^* = 4 or 50

Genetix v.4.05 (Belkhir et al. 2004). The realized levels of genetic differentiation between the simulated and empirical datasets was evaluated by calculating pairwise F_{ST} values and performing a discriminant analysis of principal components (DAPC) with the adegenet package in R v.3.0.2 (Jombart 2008, R Development Core Team 2010).

In addition to the comparative data analyses, the results also allow us to determine the N_e of the $M_{sole\text{-}BS}$. Using the H_{obs} and H_{exp} values obtained for $M_{sole\text{-}F1\text{-}B1}$, $M_{sole\text{-}F1\text{-}B2}$, $M_{sole\text{-}F1\text{-}B3}$ and $M_{sole\text{-}F1\text{-}B4}$, the N_e can be calculated for $M_{sole\text{-}BS}$ during each batch spawning event using the equation from Luikart & Cornuet (1999):

$$N_e = H_{exp}/[2(H_{obs} - H_{exp})] \quad (1)$$

Tracing escapees

Assignment efficiency is strongly influenced by the realized levels of genetic differentiation between the captive-bred and wild populations (IA) and the amount of genetic variability within the captive-bred population (PBT). By using the simulated datasets which are characterized by differences in the N_e and F_n, 2 parameters that significantly affect genetic differentiation and genetic variability, the effects of these changes could be evaluated. In addition, the origin of the captive-bred population will also influence the traceability outcomes. To assess the effects of genetic dissimilarities between captive-bred and wild populations, 2 traceability scenarios were used: A, the broodstock originated from a genetically distinct population than the local wild populations; and B, the broodstock originated from a local, genetically similar wild population (Table 2).

From a forensic perspective, the ability to assign captive-bred fish back to their origin will be influenced by the nature and availability of reference samples, which may be challenging in well-established aquaculture species (Glover et al. 2009). For the purpose of our study, 2 simplified sampling regimes were used to evaluate the effect of missing data from previous captive-bred generations: Sampling Regime 1, in which data from the parental generation, which produced the escapees, is available, and escapees can thus be assigned to their parental generation or to the wild populations; Sampling Regime 2, in which data is restricted to the founding broodstock (often the case in operational hatcheries) and escapees can only be assigned to the founding broodstock or the wild populations. The lack of multiple captive-bred generations in the empirical data restricted the analyses of the empirical data to Sam-

pling Regime 1. Furthermore, PBT relies on the identification of parent–offspring relationships and will thus only be valuable under the assumptions of Sampling Regime 1.

IA and PBT analyses were performed using both simulated and empirical datasets (see Table 2). For the analyses of the simulated data, escapees were assumed to be flagged (i.e. genotypes of escapees are known) to obtain a baseline traceability efficiency, while for the analysis of the empirical data, escapees were mixed within a single wild population to create a more realistic scenario. IA analyses used the simulated datasets of both species and the empirical Atlantic sole data. PBT analyses were performed using sole data only, as Atlantic cod family data was unavailable.

IA analysis

IA analyses were performed with GeneClass2 v.2.0 using the 'assign/exclude population as origin of individuals' option (Piry et al. 2004). The threshold value was set to p = 0.05 and only individuals assigned to a population with rank 1 were considered. The probability of an individual being assigned to all possible reference populations was calculated using the Monte Carlo re-sampling method (Paetkau et al. 2004).

Using the simulated data of both species, assignment efficiency was evaluated under both sampling regimes. Input data for assignments consisted of 100 $F_{n\text{-}SIM}$ genotypes (escapees) which could be assigned to either wild populations or their captive-bred population (i.e. their parental generation $F_{(n\text{-}1)\text{-}SIM}$ or their founding broodstock $P_{1\text{-}SIM}$ for Sampling Regime 1 or 2 respectively).

For the analyses of the empirical data, 20 individuals from $A_{sole\text{-}F1}$ (10 from each full-sib family) representing the escapees were randomly selected and mixed with a genetically similar wild population (Scenario B) originating from the Belgian coast (BEL). Genotypes contained within this mixed population and the neighbouring wild populations (STO, GER, NOR, ENG, IS and GAS; see Supplement 1) could then be assigned to the remaining $A_{sole\text{-}F1}$ individuals.

PBT analysis

The parent–offspring relationships within the empirical aquaculture samples were obtained from previous studies (Blonk et al. 2009) and additional parentage testing (Supplement 4 at www.int-res.com/articles/suppl/q008p131_supp.pdf). Only the SNP genotypes of individuals for which reliable parent-offspring relationships could be obtained were used in further analyses to ensure that the effectiveness of PBT could be formally evaluated. PBT analyses were performed with the software SNPPIT v.1.0 (Anderson 2010b), using only genotypic information (i.e. sex, age, year of sampling, etc. were considered unknown) and a genotyping error rate of 0.5% per allele.

Using the simulated data and the wild populations of sole, the effect of N_e and F_n on the assignment success was evaluated under the assumption of Sampling Regime 1. Input files consisted of a list of putative parents (all wild populations and $F_{(n\text{-}1)\text{-}SIM}$) and offspring to be assigned ($F_{n\text{-}SIM}$) (i.e. the escapees).

Empirical analyses were performed with the Atlantic and Mediterranean sole data to determine the influence of the origin of the broodstock (Scenario A or B) on the traceability efficiency. Under Scenario A, the input file of putative parents contained the genotypes of $A_{sole\text{-}BS}$ or $M_{sole\text{-}BS}$ mixed with their respective source population (i.e. GER and ADR1 respectively). The offspring to be assigned contained a mixed population of 20 randomly selected $A_{sole\text{-}F1}$ or $M_{sole\text{-}F1}$ individuals added to genetically different wild populations (i.e. IS and THY respectively) and the remaining wild populations. In the case of Scenario B, assignment input was similar with the exception that the 20 randomly selected $A_{sole\text{-}F1}$ or $M_{sole\text{-}F1}$ individuals were mixed with a genetically similar wild population (i.e. BEL and ADR2 respectively).

RESULTS

Sampling and genotyping

Following complementary genotyping of the sole samples ($M_{sole\text{-}F1}$) with 192 SNPs, 181 SNPs passed the initial quality assessment. Of these, a panel of 96 highly informative SNP markers was selected and used in the analyses. An overview of all 96 selected SNPs used in the traceability analyses can be found in Supplement 2. Based on the re-genotyping of the 3 $A_{sole\text{-}BS}$ individuals at 181 loci, a genotyping discordance rate of 1.2% was obtained. Hence, in all further analyses, a genotyping error of 1% was used as an approximation.

Comparative data analyses

The comparative analyses of overall genetic diversity (H_{obs} and H_{exp}) showed no strong deviation between H_{obs} and H_{exp} in the $P_{1\text{-}SIM}$ (N_e = 5, 10, 20, 50) and $F_{1\text{-}SIM}$ (N_e = 5, 10, 20, 50) data (Fig. 2). However, in the $M_{sole\text{-}F1}$ data, a heterozygote excess was observed ($H_{obs} > H_{exp}$), suggesting that within $M_{sole\text{-}BS}$, a low number of individuals contributed to the next generation. Based on this heterozygote excess, the N_e is estimated to be 2.16, 2.10, 1.58 and 1.67 for $M_{sole\text{-}F1\text{-}B1}$, $M_{sole\text{-}F1\text{-}B2}$, $M_{sole\text{-}F1\text{-}B3}$ and $M_{sole\text{-}F1\text{-}B4}$ respectively.

Pairwise F_{ST} values and the DAPC show that both $M_{sole\text{-}BS}$ and ADR1 have a similar genetic composition (Fig. 3; Supplement 5 at www.int-res.com/articles/suppl/q008p131_supp.pdf). However, strong genetic differentiation is observed between $M_{sole\text{-}F1}$ and their population of origin ($M_{sole\text{-}BS}$ and ADR1). A comparison of the simulated data ($P_{1\text{-}SIM}$) and the wild populations ($M_{sole\text{-}BS}$ and ADR1) shows an increase in genetic differentiation when a strong bottleneck was applied (from N_e = 50 to N_e = 5), and the same pattern can be observed in the derived $F_{1\text{-}SIM}$ samples. Furthermore, the F_{ST} values are generally higher between the $M_{sole\text{-}F1}$ batches ($M_{sole\text{-}F1\text{-}B1}$, $M_{sole\text{-}F1\text{-}B2}$, $M_{sole\text{-}F1\text{-}B3}$, $M_{sole\text{-}F1\text{-}B4}$) than between the $F_{1\text{-}SIM}$ data (Fig. 3), and the same pattern can be observed with the DAPC (i.e. $F_{1\text{-}SIM}$ clusters are positioned closer together than $M_{sole\text{-}F1}$ clusters). One exception is the low genetic differentiation between $M_{sole\text{-}F1\text{-}B3}$ and $M_{sole\text{-}F1\text{-}B4}$, which is due to the same parents having produced these batches (Supplement 4). The results suggest that the simulated data provides a good baseline (broodstock under HW equilibrium) for the validation of the traceability methods under real-life scenarios.

Tracing escapees

IA analysis

The success rate of correctly assigning escapees to the previous aquaculture generation (Sampling Regime 1) ranged from 73 to 100% across all simulated datasets (Fig. 4). Results clearly indicate that the assignment success increased with increasing genetic drift (smaller N_e) and increasing generational distance from the original broodstock generation (higher F_n). Under the assumptions of Sampling Regime 2, the assignment success increased with increasing genetic drift, but no change in assignment success was observed with increasing generational distance from the broodstock (Fig. 4). However, in sole, an increasing F_n resulted in a decrease of assignment performance when a large effective population size (N_e = 50) was employed.

The population assignment analyses based on the empirical data of the Atlantic farmed sole and their neighbouring wild populations revealed that 81% of escapees were correctly assigned using 1 SNP (average assignment score: 40), while a 100% assignment was achieved with only 30 SNPs (average assignment score: 100).

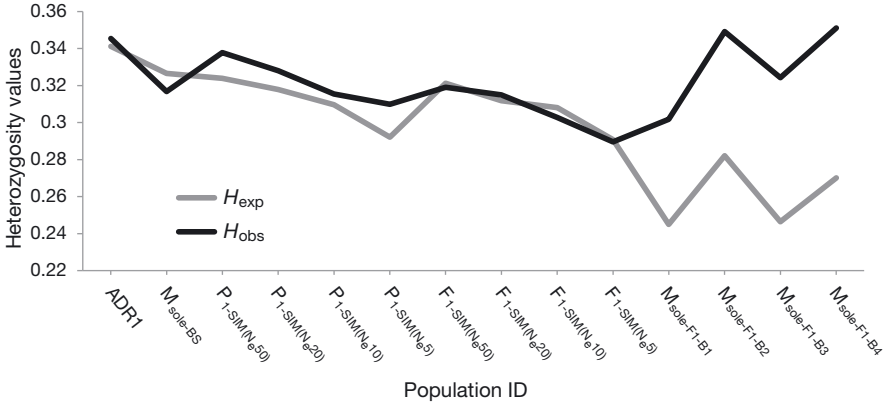

Fig. 2. Observed (H_{obs}) and expected (H_{exp}) heterozygosity values in population samples of Mediterranean sole *Solea solea*. ADR1 = natural population from Adriatic Sea; $M_{sole\text{-}BS}$ = broodstock composed of wild fish from Adriatic Sea; $P_{1\text{-}SIM(N_eX)\ (X = 5\text{–}50)}$ = simulated parental populations at varying degrees of effective population size N_e; $F_{1\text{-}SIM(N_eX)\ (X = 5\text{–}50)}$ = simulated first offspring populations at varying degrees of effective population size N_e; $M_{sole\text{-}F1\text{-}BX\ (X = 1\text{–}4)}$ = 4 actual offspring batches from aquaculture

PBT analysis

PBT analyses (SNPPIT) using the simulated data of sole showed that a panel of 48 SNP loci was sufficient to obtain an assignment success of ≥99%. Assignment success decreased (i.e. increasing number of non-excluded parent-offspring trios) with an increasing number of breeding generations (F_n), especially when N_e was small (Table 3).

The PBT results based on the empirical sole data show that under Scenario A, a dataset of 30 and 40 highly polymorphic SNPs was sufficient to trace back the Atlantic and Mediterranean aquaculture escapees, respectively (Table 4). Under the assumptions of Scenario B, a total of 35 highly poly-

	M_{soleBS}	$M_{soleF1\text{-}B1}$	$M_{soleF1\text{-}B2}$	$M_{soleF1\text{-}B3}$	$M_{soleF1\text{-}B4}$	$P_{1\text{-}SIM}(N_e50)$	$P_{1\text{-}SIM}(N_e20)$	$P_{1\text{-}SIM}(N_e10)$	$P_{1\text{-}SIM}(N_e5)$	$F_{1\text{-}SIM}(N_e50)$	$F_{1\text{-}SIM}(N_e20)$	$F_{1\text{-}SIM}(N_e10)$	$F_{1\text{-}SIM}(N_e5)$	ADR1
M_{soleBS}	0.000													
$M_{soleF1\text{-}B1}$	0.078	0.000												
$M_{soleF1\text{-}B2}$	0.062	0.145	0.000											
$M_{soleF1\text{-}B3}$	0.057	0.161	0.080	0.000										
$M_{soleF1\text{-}B4}$	0.043	0.135	0.053	0.010	0.000									
$P_{1\text{-}SIM}(N_e50)$	0.002	0.068	0.061	0.050	0.040	0.000								
$P_{1\text{-}SIM}(N_e20)$	0.006	0.092	0.076	0.058	0.047	0.003	0.000							
$P_{1\text{-}SIM}(N_e10)$	0.009	0.085	0.076	0.061	0.048	0.005	0.007	0.000						
$P_{1\text{-}SIM}(N_e5)$	0.012	0.072	0.070	0.053	0.044	0.006	0.011	0.010	0.000					
$F_{1\text{-}SIM}(N_e50)$	0.002	0.048	0.046	0.035	0.029	0.002	0.003	0.003	0.003	0.000				
$F_{1\text{-}SIM}(N_e20)$	0.005	0.059	0.049	0.035	0.029	0.005	0.001	0.003	0.003	0.007	0.000			
$F_{1\text{-}SIM}(N_e10)$	0.007	0.058	0.056	0.043	0.036	0.008	0.004	0.000	0.002	0.008	0.008	0.000		
$F_{1\text{-}SIM}(N_e5)$	0.017	0.070	0.074	0.053	0.047	0.020	0.012	0.005	0.000	0.022	0.022	0.016	0.000	
ADR1	0.010	0.070	0.058	0.062	0.050	0.011	0.015	0.015	0.013	0.011	0.016	0.021	0.030	0.000

Fig. 3. Pairwise F_{ST} values matrix among the simulated and empirical datasets of Mediterranean aquaculture sole *Solea solea*. □: $F_{ST} = 0$; □: $0 < F_{ST} < 0.01$; □: $0.01 \leq F_{ST} < 0.05$; ■: $0.05 \leq F_{ST} < 0.1$; ■: $0.1 \leq F_{ST}$. See Fig. 2 for abbreviations

morphic SNPs were sufficient for the identification of all aquaculture escapees for both broodstocks (Table 4).

DISCUSSION

Our study shows that a panel of highly informative, gene-associated SNP markers can discriminate between wild and captive-bred marine fish, even without extensive domestication of the species of interest. Furthermore, the results show that IA and PBT analyses can both be valuable tools for wildlife forensics and fisheries management, depending on the genetic history of the relevant captive populations.

Potential of SNP markers for traceability

Biallelic SNP markers are generally considered less informative than microsatellite markers. However, SNPs are highly abundant and evenly distributed throughout the genome (Morin et al. 2004).

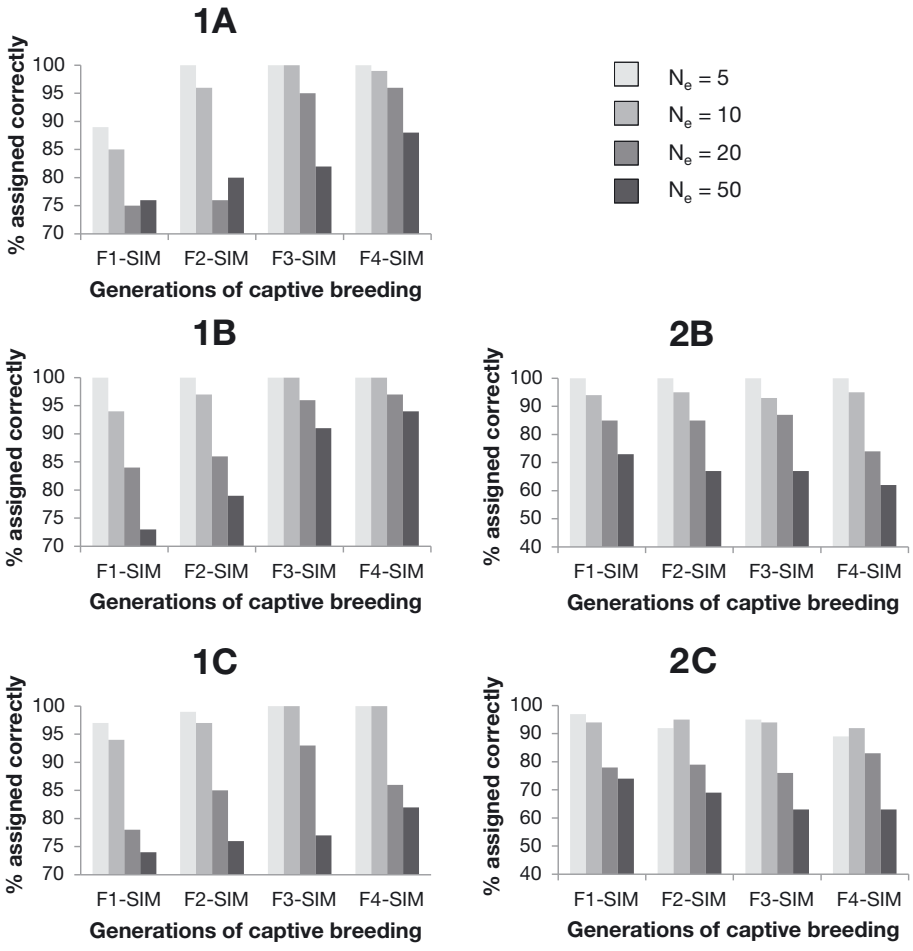

Fig. 4. *Gadus morhua* and *Solea solea*. Percentage of correctly assigned individuals in individual assignment (IA) analysis using simulated datasets. 1A, 1B, 1C: Sampling Regime 1 (reference data of aquaculture population is available for parental generation); 2B, 2C: Sampling Regime 2 (reference data of aquaculture population is limited to the founding broodstock); with results based on (A) sole data using 52 single nucleotide polymorphisms (SNPs), (B) sole data using 96 SNPs, (C) cod data using 96 SNPs. $F_{1\text{-SIM}}$, $F_{2\text{-SIM}}$, $F_{3\text{-SIM}}$, $F_{4\text{-SIM}}$ = 4 offspring generations, N_e = effective population size

Hence, low polymorphism levels can be compensated through the development of a large number of gene-associated SNPs which can detect even small population genetic differences (Nielsen et al. 2012). Additionally, SNP genotyping can be highly automated and does not require extensive calibrations for marker exchange (Hauser & Seeb 2008). These characteristics make SNP markers ideal for the development of universally applicable genetic traceability tools, which inherently rely on the availability of robust reference data (Helyar et al. 2011, Nielsen et al. 2012). In the case of tracing captive-bred marine fish, SNPs can be used to detect subtle genetic differences between wild and captive-bred populations, even after just a few generations of captive breeding. Consequently, there is ample opportunity to use SNP-based tracing in fisheries management and wildlife forensics. From a management perspective, SNPs can be employed to monitor the effects of accidental/deliberate releases of captive-bred fish on wild populations. SNP-based tracing will also have forensic applications, as it will be a useful tool in the fight against mismanagement practices in aquaculture and the mislabelling of seafood products, since universal markers for the identification of captive-bred individuals can be developed (Karlsson et al. 2011).

Applications of IA and PBT analyses

Our results demonstrate that IA and PBT perform optimally under different scenarios. The performance of IA analyses improves with increased genetic differentiation between the aquaculture and wild populations as a result of increased generational breeding (high F_n) and/or a low N_e in the broodstock. PBT analyses, on the other hand, perform better when a high N_e characterizes the broodstock and/or generational breeding is low. This is as expected, given that candidate parents are less likely to be excluded from being the real parents due to loss of genetic diversity (low N_e and/or high F_n). As a result of the performance differences, the suitability of IA

Table 3. *Solea solea*. Parentage-based tagging (PBT) analysis using software package SNPPIT to identify escapees based on simulated sole data. $F_{n\text{-}SIM}$ = number of captive-bred generations that were simulated; $F_{1\text{-}SIM}$, $F_{2\text{-}SIM}$, $F_{3\text{-}SIM}$, $F_{4\text{-}SIM}$ = 4 offspring generations that were simulated; N_e = effective population size; na = not applicable

Loci	N_e	$F_{n\text{-}SIM}$	% assigned to correct population	Proportion assignments with $p > 0.05$	Number of non-excluded parentage trios ($\times 10^3$)	Number of non-excluded trios from the wrong population ($\times 10^3$)
96	50	$F_{1\text{-}SIM}$	100	0.00	0.69	0.05
		$F_{2\text{-}SIM}$	100	0.04	4.98	0.07
		$F_{3\text{-}SIM}$	100	0.04	5.95	0.09
		$F_{4\text{-}SIM}$	100	0.04	7.70	0.13
	4	$F_{1\text{-}SIM}$	100	0.00	0.54	0.04
		$F_{2\text{-}SIM}$	100	0.34	471.06	0.18
		$F_{3\text{-}SIM}$	100	0.28	400.90	1.84
		$F_{4\text{-}SIM}$	na	na	na	na
48	50	$F_{1\text{-}SIM}$	99	0.01	6.65	1.94
		$F_{2\text{-}SIM}$	100	0.11	81.82	5.18
		$F_{3\text{-}SIM}$	99	0.10	122.12	8.41
		$F_{4\text{-}SIM}$	100	0.25	240.94	10.69
	4	$F_{1\text{-}SIM}$	100	0.00	1.72	1.14
		$F_{2\text{-}SIM}$	99	0.11	81.83	5.18
		$F_{3\text{-}SIM}$	na	na	na	na
		$F_{4\text{-}SIM}$	na	na	na	na

and PBT analyses is strongly dependent on the ultimate goal of genetic tracing studies. Hence, our results are important for wildlife forensics and fisheries management to determine the optimal assignment strategy.

A common goal of fisheries management is the preservation or restoration of commercially important fish populations to levels which will produce a long-term maximum sustainable yield (MSY) (FAO Fisheries and Aquaculture Department 2008). Since the number of overexploited marine fish populations continues to increase, stock enhancement and sea ranching programmes have become popular management actions (Bell et al. 2008, FAO Fisheries and Aquaculture Department 2012). Consequently, the release of first-generation captive-bred juvenile fish which are genetically similar to the local wild populations has increased (Bell et al. 2008). Given that PBT analyses have a high identification efficiency for first-generation escapees and can detect hybridization between wild and captive-reared conspecifics, they can be used to jointly evaluate the levels of introgression (enforcement action) and the efficiency of restocking, stock enhancement and sea ranching programmes (management action).

Robust, forensically validated and universally applicable traceability tools can also be used in wildlife forensics to support legal actions against mismanagement of aquaculture facilities, which increases the chance of escapees, or the mislabelling of seafood products for financial profits (Ogden 2008, Glover 2010, Hanner et al. 2011). Our results indicate that both IA and PBT are potentially valuable provided that the aquaculture history of the species of interest is taken into account. IA analyses are a powerful tool for species with a long aquaculture history since cap-

Table 4. *Solea solea*. Parentage-based tagging (PBT) approach using software package SNPPIT for identifying escapees based on the empirical sole aquaculture data. Scenario A: broodstock originated from a genetically different population than the local wild populations, Scenario B: broodstock originated from a local wild population. SNP = single nucleotide polymorphism

Broodstock origin	Scenario	Number of SNPs	Escapees		Natural individuals	
			% assigned to both parents	% significantly assigned	% assigned to at least 1 parent	% significantly assigned
Atlantic Ocean	A	50	100	100	6	0
	A	30	100	100	21	0
	A	21	85	35	26	2
	B	50	100	100	6	0
	B	35	100	100	15	0
	B	30	100	95	19	0
Mediterranean Sea	A	40	100	100	52	0
	A	35	100	95	51	0
	A	30	95	80	74	0.7
	B	40	100	100	84	0
	B	35	100	100	81	0
	B	30	95	80	90	0

tive breeding has resulted in a strong genetic differentiation between captive-bred and wild fish populations (Bekkevold et al. 2006, Karlsson et al. 2011). However, most marine fish species have only recently been bred in captivity and thus forensic tools need to be able to differentiate between genetically similar captive-bred species and wild conspecifics. Our findings suggest that PBT can be used for these recently domesticated fish species, since assignment success was high after only a single generation of captive breeding. This is in line with expectations, since PBT was originally developed to identify the source of salmon released into rivers and is thus capable of differentiating between genetically similar hatchery populations (Anderson & Garza 2006). Genetic assignment methods have already been successfully applied in a forensic context (Wong & Hanner 2008, Glover 2010). However, real-life situations often complicate genetic tracing studies (Glover et al. 2009). As such, the presence of multiple (genetically similar) putative source farms and the lack of extensive genetic reference data will reduce the assignment efficiency of both IA and PBT. Although the latter is less problematic for IA analysis, PBT unequivocally requires genotypic information from all parental individuals that have contributed to the subsequent generation. The increased use of genetic broodstock management and selective breeding programmes might partially resolve this but the feasibility of using PBT in a forensic context remains controversial (Blonk et al. 2010, Vandeputte et al. 2011).

Validation of traceability approaches

Validating traceability methods requires a detailed comparison between expected (simulations) and observed (empirical) results. The assignment success rates of the analyses based on the $F_{1\text{-SIM}}$ and the empirical data reveal that overall, a higher success rate is obtained in the empirical analyses. The fact that relatively more SNP makers are needed for unambiguous assignments in the simulated data can be explained by a high reproductive skew in real aquaculture production ($A_{\text{sole-BS}}$ and $M_{\text{sole-BS}}$), which is difficult to simulate with currently available software packages. From the N_e values estimated based on the observed heterozygote excess in the $M_{\text{sole-F1}}$, we conclude that on average, 2 parental individuals contributed to each offspring batch, and these finding are supported by the results from the additional parentage testing (Supplement 4). Furthermore, comparing the genetic differentiation between the empirical and simulated data (DAPC) suggests that within the $M_{\text{sole-BS}}$, an N_e of between 5 and 10 is the most likely, which is supported by the N_e estimates found in the $A_{\text{sole-BS}}$ by Blonk et al. (2009).

Other evidence supporting the methodology employed here arises from the comparison of current results with earlier studies. Vandeputte et al. (2011) recorded a decrease in the assignment power when comparing theoretical, simulated and empirical parentage assignments using microsatellite data. However, our study has clearly indicated that large-scale SNP genotyping (i.e. genome scan) combined with a selection procedure for highly informative gene-associated markers (high F_{ST} values and PIC) can increase the assignment power in empirical studies. This is consistent with the findings of previous studies which recorded similarly high assignment efficiencies with only a small number of markers (Nielsen et al. 2012). Hence, the methodology presented here will be valuable for future traceability studies where sufficient genetic background information is available for the species of interest. With low-cost high-throughput genotyping-by-sequencing methods now available to be implemented in breeding programmes (Davey et al. 2011), the cost of developing a large battery of markers should not impede applications to fisheries management and wildlife forensics.

CONCLUSIONS

This study has evaluated the relative power of parentage-based tagging (PBT) and individual assignment (IA) for identifying the population of origin of marine aquaculture fish under a range of scenarios, highlighting the benefits and disadvantages of each. PBT potentially offers the strongest line of traceability evidence, as the identification of a specific parental pair with high confidence is likely to be more powerful than a combined population assignment and exclusion approach under IA, particularly where aquaculture and wild populations have not diverged significantly. The results presented here have shown that PBT analyses will be particularly valuable in fisheries management to evaluate the genetic effects and the impact of accidental and/or deliberately released captive-bred fish. However, current aquaculture practices restrict the practical application of PBT due to the requirement for complete broodstock sampling; consequently, in most marine fish aquaculture scenarios, IA analyses are considered to be of more practical use for future

traceability applications. Ultimately, the availability of genetic background information and the aim of the study will determine whether IA or PBT will be the method of choice.

Acknowledgements. We thank Fiskaaling aquaculture research station (Faroe Islands) for providing samples of farmed cod. Additionally, we thank R. J. W. Blonk for providing samples and parentage information of the hatchery-reared Atlantic sole. Research was funded by the European Commission Joint Research Centre (JRC) through the AQUAGEN project (https://aquagen.jrc.ec.europa.eu; tender No. IPSC_2010_04_11_NC) and benefited from the EU FP7 project FishPopTrace (KBBE-2007-212399) and AquaTrace (KBBE-311920). E.D. and G.E.M. were funded by the Research Foundation-Flanders (FWO-Vlaanderen) during the course of this research.

LITERATURE CITED

Abadía-Cardoso A, Anderson EC, Pearse DE, Garza JC (2013) Large-scale parentage analysis reveals reproductive patterns and heritability of spawn timing in a hatchery population of steelhead (*Oncorhynchus mykiss*). Mol Ecol 22:4733–4746

Anderson EC (2010a) Assessing the power of informative subsets of loci for population assignment: standard methods are upwardly biased. Mol Ecol Resour 10:701–710

Anderson EC (2010b) Computational algorithms and user-friendly software for parentage-based tagging of Pacific salmonids. Final report to Pacific Salmon Commission's Chinook Technical Committee. NOAA Southwest Fisheries Science Center, Santa Cruz, CA

Anderson EC (2014) Simple simulation of Mendelian inheritance of unlinked markers from specified matings and populations. Source code for Nookie software. http://github.com/eriqande/nookie

Anderson EC, Garza JC (2006) The power of single-nucleotide polymorphisms for large-scale parentage inference. Genetics 172:2567–2582

André C, Larsson LC, Laikre L, Bekkevold D and others (2011) Detecting population structure in a high gene-flow species, Atlantic herring (*Clupea harengus*): direct, simultaneous evaluation of neutral vs putatively selected loci. Heredity 106:270–280

Arechavala-Lopez P, Fernandez-Jover D, Black KD, Ladoukakis E, Bayle-Sempere JT, Sanchez-Jerez P, Dempster T (2013) Differentiating the wild or farmed origin of Mediterranean fish: a review of tools for sea bream and sea bass. Rev Aquacult 5:137–157

Bekkevold D, Hansen MM, Nielsen EE (2006) Genetic impact of gadoid culture on wild fish populations: predictions, lessons from salmonids, and possibilities for minimizing adverse effects. ICES J Mar Sci 63:198–208

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (2004) GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Lab génome, Popul Interact CNRS Umr 5000, Université de Montpellier II, Montpellier

Bell JD, Leber KM, Blankenship HL, Loneragan NR, Masuda R (2008) A new era for restocking, stock enhancement and sea ranching of coastal fisheries resources. Rev Fish Sci 16:1–9

Björnsson B (2011) Ranching of wild cod in 'herds' formed with anthropogenic feeding. Aquaculture 312:43–51

Blonk RJW, Komen J, Kamstra A, Crooijmans RPMA, van Arendonk JAM (2009) Levels of inbreeding in group mating captive broodstock populations of Common sole, (*Solea solea*), inferred from parental relatedness and contribution. Aquaculture 289:26–31

Blonk RJW, Komen H, Kamstra A, van Arendonk JAM (2010) Estimating breeding values with molecular relatedness and reconstructed pedigrees in natural mating populations of common sole, *Solea solea*. Genetics 184: 213–219

Braithwaite VA, Salvanes AGV (2010) Aquaculture and restocking: implications for conservation and welfare. Anim Welf 19:139–149

Brown N (2002) Flatfish farming systems in the Atlantic region. Rev Fish Sci 10:403–419

Brown C, Miltiadou D, Tsigenopoulos C (2015) Prevalence and survival of escaped European seabass *Dicentrarchus labrax* in Cyprus identified using genetic markers. Aquacult Environ Interact 7:49–59

Clemento AJ, Crandall ED, Garza JC, Anderson EC (2014) Evaluation of a single nucleotide polymorphism baseline for genetic stock identification of Chinook salmon (*Oncorhynchus tshawytscha*) in the California Current large marine ecosystem. Fish Bull 112:112–130

Cline E (2012) Marketplace substitution of Atlantic salmon for Pacific salmon in Washington State detected by DNA barcoding. Food Res Int 45:388–393

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

Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat Rev Genet 12:499–510

Diopere E, Maes GE, Komen H, Volckaert FAM, Groenen MAM (2014) A genetic linkage map of sole (*Solea solea*): a tool for evolutionary and comparative analyses of exploited (flat)fishes. PLoS ONE 9:e115040

Duarte CM, Marbá N, Holmer M (2007) Rapid domestication of marine species. Science 316:382–383

FAO Fisheries and Aquaculture Department (2008) FAO technical guidelines for responsible fisheries. FAO, Rome

FAO Fisheries and Aquaculture Department (2012) The state of world fisheries and aquaculture. FAO, Rome

FAO Fisheries and Aquaculture Department (2014) The state of world fisheries and aquaculture—opportunities and challenges. FAO, Rome

FAO Fisheries and Aquaculture Department (2015a) Species fact sheets: *Gadus morhua* (Linnaeus, 1758). FAO, Rome

FAO Fisheries and Aquaculture Department (2015b) Species fact sheets: *Solea solea* (Quensel, 1806). FAO, Rome

Gjedrem T, Robinson N, Rye M (2012) The importance of selective breeding in aquaculture to meet future demands for animal protein: a review. Aquaculture 350–353: 117–129

Glover KA (2010) Forensic identification of fish farm escapees: the Norwegian experience. Aquacult Environ Interact 1:1–10

Glover KA, Skilbrei OT, Skaala Ø (2008) Genetic assignment identifies farm of origin for Atlantic salmon *Salmo salar* escapees in a Norwegian fjord. ICES J Mar Sci 65:

912–920
Glover KA, Hansen MM, Skaala Ø (2009) Identifying the source of farmed escaped Atlantic salmon (*Salmo salar*): Bayesian clustering analysis increases accuracy of assignment. Aquaculture 290:37–46
Glover KA, Hansen MM, Lien S, Als TD, Høyheim B, Skaala Ø (2010) A comparison of SNP and STR loci for delineating population structure and performing individual genetic assignment. BMC Genet 11:2 doi:10.1186/1471-2156-11-2
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, 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 doi: 10.1186/1471-2156-14-74
Goudet J (1995) FSTAT (version 1.2): a computer program to calculate F-statistics. J Hered 86:485–486
Hanner R, Becker S, Ivanova NV, Steinke D (2011) FISH-BOL and seafood identification: geographically dispersed case studies reveal systemic market substitution across Canada. Mitochondrial DNA 22:106–122
Hauser L, Seeb JE (2008) Advances in molecular technology and their impact on fisheries genetics. Fish Fish 9: 473–486
Helyar SJ, Hemmer-Hansen J, Bekkevold D, Taylor MI and others (2011) Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges. Mol Ecol Resour 11:123–136
Hindar K, Ryman N, Utter F (1991) Genetic effects of cultured fish on natural fish populations. Can J Fish Aquat Sci 48:945–957
Hohenlohe PA, Day MD, Amish SJ, Miller MR and others (2013) Genomic patterns of introgression in rainbow and westslope cutthroat trout illuminated by overlapping paired-end RAD sequencing. Mol Ecol 22:3002–3013
Howell BR (1997) A re-appraisal of the potential of the sole, *Solea solea* (L.), for commercial cultivation. Aquaculture 155:355–365
Imsland AK, Foss A, Conceicão LEC, Dinis MT and others (2003) A review of the culture potential of *Solea solea* and *S. senegalensis.* Rev Fish Biol Fish 13:379–407
Jacquet JL, Pauly D (2008) Trade secrets: renaming and mislabeling of seafood. Mar Policy 32:309–318
Jensen Ø, Dempster T, Thorstad EB, Uglem I, Fredheim A (2010) Escapes of fishes from Norwegian sea-cage aquaculture: causes, consequences and prevention. Aquacult Environ Interact 1:71–83
Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403–1405
Jones AG, Ardren WR (2003) Methods of parentage analysis in natural populations. Mol Ecol 12:2511–2523
Jones AG, Small CM, Paczolt KA, Ratterman NL (2010) A practical guide to methods of parentage analysis. Mol Ecol Resour 10:6–30
Jørstad KE, van der Meeren T, Paulsen OI, Thomsen T, Thorsen A, Svåsand T (2008) 'Escapes' of eggs from farmed cod spawning in net pens: recruitment to wild stocks. Rev Fish Sci 16:285–295
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
Kitada S, Kishino H (2006) Lessons learned from Japanese marine finfish stock enhancement programmes. Fish Res 80:101–112
Kitada S, Shishidou H, Sugaya T, Kitakado T, Hamasaki K, Kishino H (2009) Genetic effects of long-term stock enhancement programs. Aquaculture 290:69–79
Kjesbu OS, Taranger GL, Trippel EA (2006) Gadoid mariculture: development and future challenges. ICES J Mar Sci 63:187–191
Laikre L, Schwartz MK, Waples RS, Ryman N (2010) Compromising genetic diversity in the wild: unmonitored large-scale release of plants and animals. Trends Ecol Evol 25:520–529
Lamaze FC, Sauvage C, Marie A, Garant D, Bernatchez L (2012) Dynamics of introgressive hybridization assessed by SNP population genomics of coding genes in stocked brook charr (*Salvelinus fontinalis*). Mol Ecol 21:2877–2895
Lamaze FC, Garant D, Bernatchez L (2013) Stocking impacts the expression of candidate genes and physiological condition in introgressed brook charr (*Salvelinus fontinalis*) populations. Evol Appl 6:393–407
Liao IC, Su MS, Leaño EM (2003) Status of research in stock enhancement and sea ranching. Rev Fish Biol Fish 13: 151–163
Luikart G, Cornuet JM (1999) Estimating the effective number of breeders from heterozygote excess in progeny. Genetics 151:1211–1216
Manel S, Gaggiotti OE, Waples RS (2005) Assignment methods: matching biological questions with appropriate techniques. Trends Ecol Evol 20:136–142
Mariani S, Ellis J, O'Reilly A, Bréchon AL, Sacchi C, Miller DD (2014) Mass media influence and the regulation of illegal practices in the seafood market. Conserv Lett 7: 478–483
Marshall TC, Slate J, Kruuk LE, Pemberton JM (1998) Statistical confidence for likelihood-based paternity. Mol Ecol 7:639–655
McGinnity P, Prodohl P, Ferguson K, Hynes R and others (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 B Biol Sci 270:2443–2450
Morin PA, Luikart G, Wayne RK (2004) SNPs in ecology, evolution and conservation. Trends Ecol Evol 19:208–216
Murray AG, Peeler EJ (2005) A framework for understanding the potential for emerging diseases in aquaculture. Prev Vet Med 67:223–235
Natale F, Hofherr J, Fiore G, Virtanen J (2013) Interactions between aquaculture and fisheries. Mar Policy 38: 205–213
Naylor RL, Goldburg RJ, Primavera JH, Kautsky N and others (2000) Effect of aquaculture on world fish supplies. Nature 405:1017–1024
Naylor RL, Hardy RW, Bureau DP, Chiu A and others (2009) Feeding aquaculture in an era of finite resources. Proc Natl Acad Sci USA 106:15103–15110
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
Nielsen EE, Cariani A, Aoidh E, Maes GE and others (2012) Gene-associated markers provide tools for tackling ille-

gal fishing and false eco-certification. Nat Commun 3: 851–856

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

Ogden R (2008) Fisheries forensics: the use of DNA tools for improving compliance, traceability and enforcement in the fishing industry. Fish Fish 9:462–472

Paetkau D, Slade R, Burden M, Estoup A (2004) Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. Mol Ecol 13:55–65

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

Porta J, Porta JM, Martínez-Rodríguez G, Alvarez MC (2006a) Genetic structure and genetic relatedness of a hatchery stock of Senegal sole (*Solea senegalensis*) inferred by microsatellites. Aquaculture 251:46–55

Porta J, Porta JM, Martínez-Rodríguez G, Alvarez MC (2006b) Development of a microsatellite multiplex PCR for Senegalese sole (*Solea senegalensis*) and its application to broodstock management. Aquaculture 256:159–166

Porta J, Porta JM, Martínez-Rodríguez G, Alvarez MC (2007) Substantial loss of genetic variation in a single generation of Senegalese sole (*Solea senegalensis*) culture: implications in the domestication process. J Fish Biol 71:223–234

R Development Core Team (2010) R : a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna

Rosenlund G, Skretting M (2006) Worldwide status and perspective on gadoid culture. ICES J Mar Sci 63:194–197

Seymour EA, Bergheim A (1991) Towards a reduction of pollution from intensive aquaculture with reference to the farming of salmonids in Norway. Aquacult Eng 10:73–88

Somarakis S, Pavlidis M, Saapoglou C, Tsigenopoulos C, Dempster T (2013) Evidence for 'escape through spawning' in large gilthead sea bream *Sparus aurata* reared in commercial sea-cages. Aquacult Environ Interact 3: 135–152

Sparrevohn CR, Støttrup JG (2007) Post-release survival and feeding in reared turbot. J Sea Res 57:151–161

Steele CA, Anderson EC, Ackerman MW, Hess MA, Campbell NR, Narum SR, Campbell MR (2013) A validation of parentage-based tagging using hatchery steelhead in the Snake River basin. Can J Fish Aquat Sci 70: 1046–1054

Thurstan RH, Brockington S, Roberts CM (2010) The effects of 118 years of industrial fishing on UK bottom trawl fisheries. Nat Commun 1:15, doi:10.1038/ncomms1013

Vandamme SG, Maes GE, Raeymaekers JAM, Cottenie K and others (2014) Regional environmental pressure influences population differentiation in turbot (*Scophthalmus maximus*). Mol Ecol 23:618–636

Vandeputte M, Rossignol MN, Pincent C (2011) From theory to practice: empirical evaluation of the assignment power of marker sets for pedigree analysis in fish breeding. Aquaculture 314:80–86

Waples RS (2010) High-grading bias: subtle problems with assessing power of selected subsets of loci for population assignment. Mol Ecol 19:2599–2601

Warner K, Timme W, Lowell B, Hirshfield M (2013) Oceana study reveals seafood fraud nationwide. Oceana. http://oceana.org/en/news-media/publications/reports/oceana-study-reveals-seafood-fraud-nationwide

Weir LK, Grant JWA (2005) Effects of aquaculture on wild fish populations : a synthesis of data. Environ Rev 13: 145–168

Wong EHK, Hanner RH (2008) DNA barcoding detects market substitution in North American seafood. Food Res Int 41:828–837

Editorial responsibility: Marianne Holmer, Odense, Denmark

Submitted: June 18, 2015; Accepted: December 10, 2015
Proofs received from author(s): February 8, 2016

The following supplements accompany the article

Evaluating genetic traceability methods for captive bred marine fish and their applications in fisheries management and wildlife forensics

Jonas Bylemans*, Gregory E. Maes, Eveline Diopere, Alessia Cariani, Helen Senn, Martin I. Taylor, Sarah Helyar, Luca Bargelloni, Alessio Bonaldo, Gary Carvalho, Ilaria Guarniero, Hans Komen, Jann Th. Martinsohn, Einar E. Nielsen, Fausto Tinti, Filip A.M. Volckaert, Rob Ogden

*Corresponding author: Jonas.Bylemans@canberra.edu.au

Aquaculture Environment Interactions 8: 131–145 (2016)

SUPPLEMENTS

The supplementary material includes a complete list of all samples (Supplement 1) and all genetic markers (Supplement 2) used in the analysis. Details about the comparative analyses between the simulated datasets and an overview of the parentage analyses performed to reconstruct parent-offspring relationships within the farmed sole samples are given (Supplements 3 and 4, respectively). Additionally, the results of the Discriminant Analysis of Principal Components (Supplement 5) are provided.

Supplement 1. Sampling details.

Table S1: Summary information on location, position, number of individuals and sampling year for the empirical samples used in the traceability analysis for Atlantic cod, Atlantic and Mediterranean populations of sole. Sample code refers to the abbreviations used for the population samples in the EU FP7 project FISHPOPTRACE. Sample type indicates the method employed to obtain samples; A) scientific cruise or scientific collection in case of aquaculture populations, B) contracted collection by commercial fishermen. (na = information not available)

Species	ICES/FAO region and sampling location	Sample code	Sample type	Latitude	Longitude	Number of individuals	Sampling year
Solea solea	Wild populations						
	27.III.a - Skagerak and Kattegat						
	Belt Sea	STO	A	55.65	10.76	40	2007
	27.IV.b - Central North Sea						
	German Bight	GER	A	54.52	7.89	40	2007
	27.IV.c - Southern North Sea						
	Norfolk	NOR	A	52.92	2.24	28	2008
	Belgian Coast	BEL2009	A	51.22	2.83	24	2009
	Thames Estuary	THA	A	51.47	1.33	40	2007
	27.VII.d - Eastern English Channel						
	Eastern English Channel	ENG	A	50.78	1.48	40	2008
	27.VII.a - Irish Sea						
	Bristol Channel	IS	A	52.21	-5.33	40	2008
	27.VIII.a - Bay of Biscay - North						
	Pertuis Breton	GAS	A	45.92	-1.69	40	2009
	37.1.3 - Sardinia						
	Viareggio, Northern Tyrrhenian Sea	THY	A	43.30	9.54	40	2009
	37.2.1 - North Adriatic						
	Chioggia Lagoon, North Adriatic	ADR1	A	44.73	13.27	40	2009
	37.2.2 - Ionian						

	South Adriatic Albanian Coast	ADR3	B	41.28	19.13	14	2000
	South Adriatic Italian Coast	ADR2	A	42.02	15.40	19	2000
	37.3.1 - Aegean						
	Gulf of Kavala, Northern Greece	GRE	A	40.85	24.49	40	2009
	37.3.2 - Levant						
	Turkish Coast	TUR2009	A	36.75	33.87	27	2009
	Aquaculture populations						
	Solea BV, the Netherlands						
	Broodstock	$A_{BS\text{-}SOLE}$	na	na	na	4	2003-2005
	Offspring	$A_{F1\text{-}SOLE}$	na	na	na	92	2006
	UNIBO DVPHAP, Italy						
	Broodstock	$M_{BS\text{-}SOLE}$	na	na	na	26	2006
	Offspring	$M_{F1\text{-}SOLE}$					
	Batch 1	$M_{F1\text{-}SOLE}$-B1	na	na	na	24	2008
	Batch 2	$M_{F1\text{-}SOLE}$-B2	na	na	na	24	2008
	Batch 3	$M_{F1\text{-}SOLE}$-B3	na	na	na	24	2008
	Batch 4	$M_{F1\text{-}SOLE}$-B4	na	na	na	24	2009
Gadus morhua	Wild populations						
	27.IV.a - Northern North Sea						
	Northern North Sea	MF03	A	58.00	-3.00	39	2003
	27.IV.b - Central North Sea						
	Northeastern North Sea	NO07	B	57.75	5.50	40	2007
	Southern North Sea	SC06	B	54.29	0.02	40	2006
	27.V.a - Icelandic Grounds						
	Iceland south, offshore	IS	A	63.20	-19.30	39	2002
	27.V.b1 - Faroe Plateau						
	Faroe Plateau	FP02	A	62.53	-6.16	40	2002
	27.V.b2 - Faroe Bank						
	Faroe Bank	FB02	A	60.95	-8.49	40	2002
	27.VII.a - Irish Sea						
	Irish Sea	IR06	A	54.62	-5.46	39	2006
	27.VII.d - English Channel						
	English Channel	EK05	A	50.79	0.48	40	2005
	27.VII.f - Bristol Channel						
	Celtic Sea	CS98	B	50.50	-5.16	39	1998
	27.XII.a - Norwegian Sea						
	Lofoten (NEAC)	SK03	A	68.35	12.14	39	2003
	Aquaculture populations						
	Fiskeaaling A/S, Faeroe Islands						
	Broodstock	$A_{BS\text{-}COD}$	B	na	na	92	2009

Supplement 2. Identification codes and NCBI accession numbers of the SNP loci used

Table S2: Overview of the 96 SNP markers used in the analysis of the Atlantic cod and sole samples.

Atlantic cod samples					
SNP ID	**Accession number**	**SNP ID**	**Accession number**	**SNP ID**	**Accession number**
Gm349_1196	rs119054515	cgpGmo-S973	rs119056013	cgpGmo-S1926	rs119055995
cgpGmo-S248a	rs119056000	cgpGmo-S693	rs119055530	cgpGmo-S626b	rs119055882
HbBeta1_1	NA	cgpGmo-S510	rs119055475	cgpGmo-S1098	rs119056384
cgpGmo-S2122	rs119055203	cgpGmo-S1708	rs119056382	cgpGmo-S2187	rs119056259
cgpGmo-S1406	rs119054982	cgpGmo-S459	rs119056202	cgpGmo-S209	rs119056237
cgpGmo-S1644	rs119055520	cgpGmo-S426	rs119055254	Gm240_0209	rs119054548
Gm1154_0166	rs119054629	Gm375_0144	rs119054503	cgpGmo-S13b	rs119056051
cgpGmo-S316	rs119055374	cgpGmo-S1001	rs119055621	cgpGmo-S2058	rs119055972
Gm0738_0160	rs119054733	cgpGmo-S917	rs119055827	cgpGmo-S740	rs119056364
cgpGmo-S1046	rs119055980	cgpGmo-S1391	rs119055071	cgpGmo-S1740	rs119055375
cgpGmo-S1205	rs119055597	cgpGmo-S224	rs119055825	cgpGmo-S87	rs119055065
cgpGmo-S252	rs119055474	cgpGmo-S703	rs119056289	cgpGmo-S831	rs119055880

cgpGmo-S936	rs119056133	Gm1156_0573	rs119054627	cgpGmo-S1024	rs119055966
cgpGmo-S251	rs119056385	Gm394_0364	rs119054493	cgpGmo-S1085a	rs119056019
Gm1339_0238	rs119054574	cgpGmo-S474	rs119055080	cgpGmo-S78	rs119055665
cgpGmo-S1112	NA	cgpGmo-S1076a	rs119056360	cgpGmo-S18	rs119056212
cgpGmo-S689	rs119055763	Hsp90	rs267733128	Gh_2_1	NA
cgpGmo-S742b	rs119055042	cgpGmo-S1094	rs119056042	cgpGmo-S1497	rs119055607
cgpGmo-S430a	rs119055124	cgpGmo-S1338	rs119056054	cgpGmo-S471	rs119055806
cgpGmo-S261b	rs119055116	Gm374_0856	rs119054504	cgpGmo-S2182	rs119055658
cgpGmo-S1751	rs119055210	cgpGmo-S594	rs119055234	cgpGmo-S1219b	rs119056340
cgpGmo-S535b	rs119055277	cgpGmo-S944	rs119055207	cgpGmo-S1051	rs119054998
cgpGmo-S814b	rs119055258	cgpGmo-S968	rs119055670	cgpGmo-S965	rs119056203
cgpGmo-S466	rs119055814	cgpGmo-S1978	rs119056150	cgpGmo-S241	rs119055894
cgpGmo-S875b	rs119055511	cgpGmo-S1418	rs119055063	cgpGmo-S1664	rs119055669
cgpGmo-S879	rs119055741	cgpGmo-S312	rs119055737	cgpGmo-S544	rs119055570
cgpGmo-S624	rs119055557	cgpGmo-S1104	rs119055218	cgpGmo-S127	rs119055331
cgpGmo-S408	rs119055633	cgpGmo-S760	rs119056399	cgpGmo-S2093	rs119056129
cgpGmo-S1743	rs119055928	Gm1002_0428	rs119054666	cgpGmo-S1423a	rs119055003
cgpGmo-S350	rs119056090	cgpGmo-S967b	rs119055979	cgpGmo-S1085b	rs119055638
cgpGmo-S905	rs119055666	cgpGmo-S1362	rs119056422	cgpGmo-S603	rs119055501
Rhod_1_1	NA	cgpGmo-S515	rs119055332	cgpGmo-S2229	rs119055872
Atlantic sole samples					
SNP ID	**Accession number**	**SNP ID**	**Accession number**	**SNP ID**	**Accession number**
SNP1012	ss1026565503	SNP1355	ss1026565675	SNP520	ss1026565844
SNP1018	ss1026565506	SNP1388	ss1026565687	SNP570	ss1026565857
SNP1030	ss1026565516	SNP1400	ss1026565690	SNP600	ss1026565867
SNP1033	ss1026565518	SNP1413	ss1026565697	SNP642	ss1026565879
SNP1038	ss1026565521	SNP147	ss1026565713	SNP652	ss1026565883
SNP1068	ss1026565537	SNP1472	ss1026565715	SNP725	ss1026565899
SNP1070	ss1026565539	SNP1478	ss503772168	SNP726	ss1026565900
SNP1091	ss1026565547	SNP1489	ss1026565719	SNP73	ss1026565901
SNP1106	ss1026565552	SNP1496	ss1026565723	SNP776	ss1026565918
SNP1114	ss503772271	SNP1512	ss503772216	SNP779	ss1026565919
SNP1125	ss1026565561	SNP1519	ss1026565731	SNP788	ss1026565925
SNP1127	ss1026565562	SNP1531	ss1026565736	SNP809	ss1026565936
SNP1129	ss503772195	SNP1536	ss1026565737	SNP821	ss1026565945
SNP1137	ss1026565567	SNP1546	ss1026565739	SNP831	ss1026565948
SNP1159	ss1026565577	SNP184	ss1026565753	SNP844	ss1026565952
SNP1160	ss1026565578	SNP199	ss1026565758	SNP845	ss1026565953
SNP1169	ss1026565584	SNP220	ss1026565764	SNP850	ss1026565955
SNP1184	ss1026565590	SNP228	ss503772147	SNP855	ss1026565959
SNP1190	ss1026565593	SNP235	ss503772240	SNP864	ss1026565966
SNP1191	ss1026565594	SNP276	ss1026565777	SNP877	ss1026565969
SNP1200	ss1026565600	SNP284	ss503772263	SNP88	ss1026565972
SNP1213	ss503772234	SNP35	ss1026565787	SNP898	ss1026565979
SNP1262	ss1026565626	SNP376	ss1026565794	SNP899	ss1026565980
SNP1269	ss1026565629	SNP383	ss1026565796	SNP915	ss503772245
SNP1293	ss1026565639	SNP386	ss1026565797	SNP920	ss1026565986
SNP1294	ss1026565640	SNP398	ss1026565805	SNP923	ss503772160
SNP1320	ss1026565658	SNP399	ss1026565806	SNP932	ss503772200
SNP1331	ss1026565661	SNP418	ss1026565809	SNP935	ss1026565993
SNP1337	ss1026565665	SNP455	ss1026565823	SNP948	ss1026565998
SNP134	ss1026565666	SNP464	ss1026565827	SNP963	ss1026566005
SNP1343	ss1026565668	SNP488	ss1026565835	SNP977	ss1026566014
SNP1346	ss1026565669	SNP499	ss503772166	SNP992	ss1026566020
Mediterranean sole samples					
SNP ID	**Accession number**	**SNP ID**	**Accession number**	**SNP ID**	**Accession number**
SNP1003	ss503772179	SNP1319	ss1026565656	SNP609	ss1026565870
SNP1010	ss1026565501	SNP1359	ss503772231	SNP633	ss1026565875
SNP1022	ss1026565509	SNP1376	ss1026565680	SNP638	ss1026565877
SNP1024	ss1026565511	SNP1383	ss1026565686	SNP640	ss1026565878
SNP1029	ss1026565515	SNP1388	ss1026565687	SNP645	ss1026565880

SNP1031	ss1026565517	SNP1404	ss1026565691	SNP652	ss1026565883
SNP1033	ss1026565518	SNP1415	ss503772192	SNP7	ss503772184
SNP1046	ss1026565526	SNP1432	ss1026565702	SNP726	ss1026565900
SNP1052	ss503772171	SNP1436	ss1026565703	SNP747	ss1026565905
SNP106	ss1026565532	SNP1439	ss1026565705	SNP749	ss1026565907
SNP1060	ss1026565533	SNP1491	ss1026565720	SNP750	ss1026565908
SNP1070	ss1026565539	SNP1492	ss1026565721	SNP767	ss1026565912
SNP1074	ss1026565540	SNP1496	ss1026565723	SNP776	ss1026565918
SNP1091	ss1026565547	SNP1512	ss503772216	SNP780	ss503772187
SNP1114	ss503772271	SNP1519	ss1026565731	SNP788	ss1026565925
SNP1129	ss503772195	SNP158	ss1026565742	SNP800	ss503772266
SNP1137	ss1026565567	SNP201	ss1026565760	SNP806	ss1026565934
SNP1160	ss1026565578	SNP228	ss503772147	SNP844	ss1026565952
SNP1182	ss1026565588	SNP232	ss1026565767	SNP848	ss1026565954
SNP1184	ss1026565590	SNP235	ss503772240	SNP850	ss1026565955
SNP1190	ss1026565593	SNP246	ss1026565768	SNP879	ss1026565971
SNP1203	ss503772190	SNP275	ss1026565776	SNP88	ss1026565972
SNP1214	ss503772211	SNP350	ss503772258	SNP891	ss1026565978
SNP1236	ss503772209	SNP357	ss1026565789	SNP90	ss1026565981
SNP1240	ss1026565615	SNP394	ss1026565802	SNP914	ss503772228
SNP1250	ss503772203	SNP418	ss1026565809	SNP920	ss1026565986
SNP1251	ss1026565620	SNP422	ss1026565811	SNP925	ss1026565988
SNP1260	ss1026565624	SNP464	ss1026565827	SNP928	ss1026565989
SNP1261	ss1026565625	SNP466	ss1026565829	SNP935	ss1026565993
SNP1284	ss1026565635	SNP486	ss1026565834	SNP953	ss1026566001
SNP1302	ss1026565644	SNP503	ss1026565839	SNP962	ss503772213
SNP1310	ss1026565650	SNP520	ss1026565844	SNP992	ss1026566020

Supplement 3. Comparative analysis between the simulated datasets.

The effect of the simulation software on the overall genetic diversity of the simulated datasets was assessed to evaluate the comparability between the datasets. Since no strong deviation between H_{obs} and H_{exp} were observed in the simulated data, H_{obs} could be used as a proxy for the overall genetic diversity within the datasets. Values of H_{obs} were calculated with Genetix v4.05 for the simulated data generated with HYBRIDLAB and NOOKIE. The data of the two most extreme simulation series (N_e = 5 (4 for the PBT) and N_e = 50) were used and the H_{obs} was calculated for each series and for all simulated generations ($P_{1\text{-SIM}}$, $F_{1\text{-SIM}}$, $F_{2\text{-SIM}}$, $F_{3\text{-SIM}}$ and $F_{4\text{-SIM}}$).
Results (Figure S1) show that both programs performed similarly and yielded similar estimates of H_{obs} for all simulated data sets. Furthermore, as expected, a decline in H_{obs} with an increasing number of captive bred generations ($F_{n\text{-SIM}}$) was observed and this decline was more pronounced at low N_e.

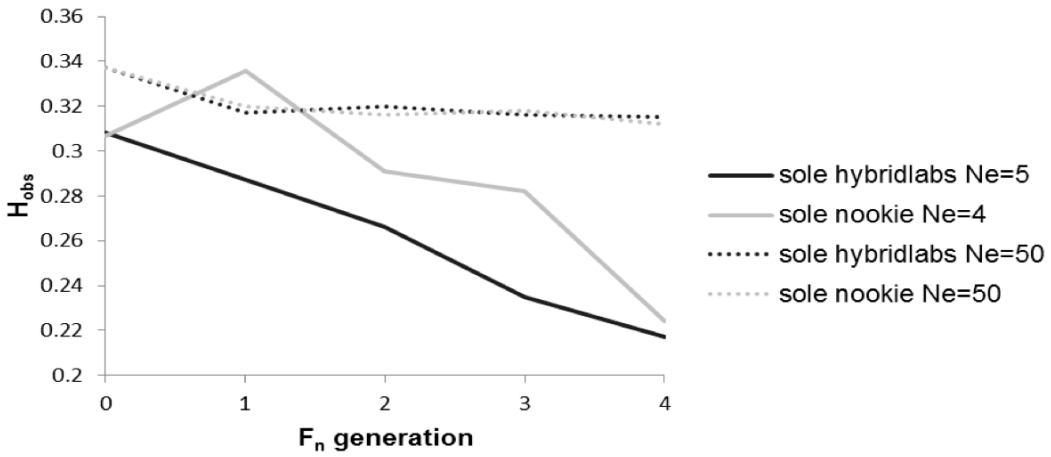

Figure S1: Plot of the H_{obs} in the simulated datasets of sole for the two most extreme effective population sizes (N_e = 5 (4) and N_e = 50) and simulated with the software NOOKIE v. 1.0 and HYBRIDLAB v.1.0. $F_0 = P_{1\text{-SIM}}$ and F_n ($n \neq 0$) = $F_{n\text{-SIM}}$ (see text).

Supplement 4. Determining parent-offspring relations in the empirical aquaculture samples.

Parent-offspring relations in the captive bred Atlantic sole

In order to validate the earlier obtained parent-offspring relations, parentage analysis were performed using the SNP genotype data of the Atlantic farmed sole and the software program CERVUS v3.0. Parentage analyses were performed using default parameters. Results of the analysis show that a minimum of 21 highly polymorphic SNPs was sufficient to obtain the same full-sib family structure as in Blonk *et al.* (2009) under strict confidence levels (95%). Furthermore, increasing the number of SNP used in the analysis did not result in another outcome (Table S3). Hence, parent-offspring relations as defined by Blonk *et al.* (2009) do reflect the real mating pattern within the captive bred Atlantic sole samples and thus can be used to evaluate the efficiency of the traceability methods employed in the further analysis.

Table S3: Percentage correctly assigned offspring using Cervus for parentage analysis using the SNP data from the Atlantic farmed sole samples.

Number of SNPs	% correctly assigned offspring under a 95% confidence level
50	100
30	100
21	100
20	98
15	89

Parent-offspring relations in the captive bred Mediterranean sole

Reconstructing the parent-offspring relations within the Mediterranean farmed sole samples was complicated by the absence of SNP genotypes for five broodstock individuals that contributed to the F_1. However, all 24 candidate parents and a subset of F_1's were genotyped at seven selected microsatellite markers (Table S4). By comparing parentage analysis based on both SNP and microsatellite datasets the most successful parental individuals can be determined and the missing SNP genotypes of highly reproductive parents can be reconstructed.

Table S4: Overview of the seven microsatellite markers for which genotyping data was available for the Mediterranean farmed sole samples.

Marker ID	Reference
F8-ICA9	Iyengar *et al.* (2000)
F8F8-IGAA7	Iyengar *et al.* (2000)
F8-ITG11	Iyengar *et al.* (2000)
F8-IIGT15	Iyengar *et al.* (2000)
F13-II8/4/7	Iyengar *et al.* (2000)
Sos(AC)6	Garoia *et al.* (2006)
Sos(AC)45	Garoia *et al.* (2006)

Analysis based on the microsatellite data

Complete parent-offspring information was obtained with an initial parentage analysis using the microsatellite data and the software package CERVUS. From these results two parental individuals could be identified that were not SNP-genotyped but did have a relatively high contribution to the F_1 generation. Firstly, a female individual (mother7) did have a high reproductive success within $M_{F1\text{-}sole}$-B2, $M_{F1\text{-}sole}$-B3 and $M_{F1\text{-}sole}$-B4. Additionally, a male individual (father3) could be identified as a successful spawner in $M_{F1\text{-}sole}$-B2.

Analysis based on the SNP data

The software package COLONY (Jones and Wang, 2010) was used to perform an initial parentage assignment analysis using all available loci (181 good quality SNPs) of the Mediterranean farmed sole samples. COLONY relies on sibship reconstruction to determine parent-offspring relations and potential parental genotypes can be

incorporated into the analysis. COLONY will thus determine parent-offspring relations taking into account the genotypes of potential parents but also taking into account that some parental genotypes might not be incorporated. Hence, it is possible to determine whether or not some missing parental genotypes did contribute to the F_1 generation. In addition, COLONY can also be used to reconstruct the genotypes of these missing parental genotypes. The initial analysis based on the SNP data indicated that one missing female genotype (mother #1) was highly successful within $M_{F1\text{-}sole}$-B2, $M_{F1\text{-}sole}$-B3 and $M_{F1\text{-}sole}$-B4 and one missing male genotype (father *3) had a relatively high reproductive success in $M_{F1\text{-}sole}$-B2. Since these results are highly comparable with the results obtained from the parentage analysis based on the microsatellite data we concluded that mother7 = mother #1 and father3 = father *3. The SNP genotypes of these individuals were subsequently reconstructed with COLONY and added to the $M_{BS\text{-}sole}$ genotypes.
Using the complete SNP dataset of the Mediterranean farmed sole ($M_{BS\text{-}sole}$ + 2 reconstructed parental genotypes and $M_{F1\text{-}sole}$), new parentage assignment analysis were performed. Before parentage assignment was performed, all SNPs deviating from HW-equilibrium were excluded and a total of 62 SNPs remained. Both COLONY and CERVUS were employed to determine parent-offspring relations based on the 62 SNPs. All relations determined by both software packages under strict (95%) confidence were considered to be the effective parent-offspring trios; these were used in PBT analysis. Relatedness could be reconstructed for 38 individuals of which 34 F1's and 4 broodstock individuals (Table S5).

Table S5: Overview of the number of offspring per batch for which both parental genotypes could be identified with sufficient confidence.

	Father 3				Father 34			
Batch ID	B1	B2	B3	B4	B1	B2	B3	B4
Mother 7	0	7	0	0	0	0	13	6
Mother 10	0	8	0	0	0	0	0	0

References

Blonk RJW, Komen J, Kamstra A, Crooijmans RPMA, van Arendonk JAM (2009). Levels of inbreeding in group mating captive broodstock populations of Common sole, (*Solea solea*), inferred from parental relatedness and contribution. Aquaculture 289: 26–31.

Supplement 5. Discriminant Analysis of Principal Components

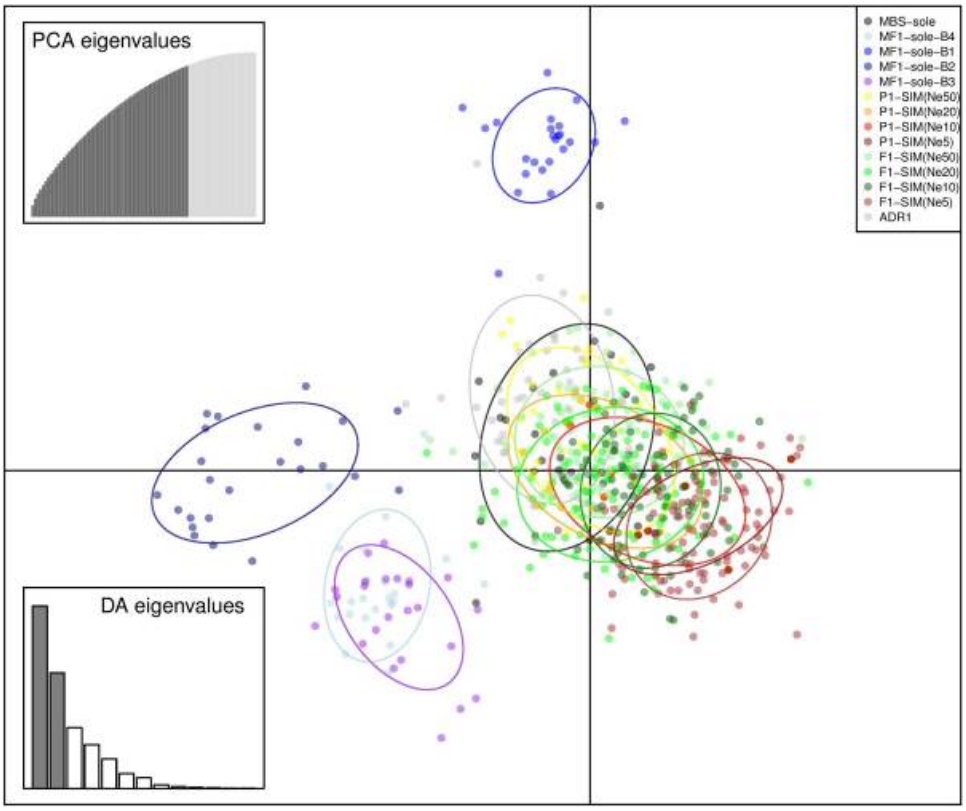

Figure S2: Discriminant Analysis of Principal Components (DAPC) plot for all empirical and simulated Mediterranean aquaculture samples and the wild population ADR1.