Report 12/2014 • Published February 2014

SporLaks – Industry-wide tracing of Norwegian farmed Atlantic salmon. Final report

Matthew Baranski, Céleste Jacq, Sten Karlsson, Ben Hayes and Robbert Blonk

Nofima is a business oriented research institute working in research and development for aquaculture, fisheries and food industry in Norway.

Nofima has about 400 employees.

The main office is located in Tromsø, and the research divisions are located in Averøy, Bergen, Stavanger, Sunndalsøra, Tromsø and Ås.

Company contact information:
Tel: +47 77 62 90 00
E-mail: post@nofima.no
Internet: www.nofima.no

Business reg.no.:
NO 989 278 835 VAT

Main office in Tromsø:
Muninbakken 9–13
P.O.box 6122 Langnes
NO-9291 Tromsø

Ås:
Osloveien 1
P.O.box 210
NO-1431 ÅS

Stavanger:
Måltidets hus, Richard Johnsensgate 4
P.O.box 8034
NO-4068 Stavanger

Bergen:
Kjerreidviken 16
P.O.box 1425 Oasen
NO-5828 Bergen

Sunndalsøra:
Sjølseng
NO-6600 Sunndalsøra

Averøy:
Ekkilsøy
NO-6530 Averøy

Report

ISBN: 978-82-8296-171-4 (printed)
ISBN: 978-82-8296-172-1 (pdf)
ISSN 1890-579X

Title: **SporLaks – Industry-wide tracing of Norwegian farmed Atlantic salmon. Final report**	*Report No.:* 12/2014
	Accessibility: **Open**
Author(s)/Project manager: Matthew Baranski, Céleste Jacq, Sten Karlsson, Ben Hayes and Robbert Blonk	*Date:* 05 February 2014
Department:	*Number of pages and appendixes:* 24+17
Client: Fiskeri- og Havbruksnæringens Forskningsfond (FHF)	*Client's ref.:* FHF # 900708
Keywords: Atlantic salmon, genetics, tracing	*Project No.:* 10067

Summary/recommendation:

The foundation of this project was to use parentage assignment methods in order to trace farmed escapees back to the farm or company at the source of the escape with 100 % precision. For this purpose we aimed to use highly informative microsatellite DNA markers. Parentage assignment using microsatellites is well established for many species and is widely used in forensic medicine, but for Atlantic salmon there is a lack of high-quality and highly-informative microsatellite markers to use for this purpose. To rectify this, we mined the Atlantic salmon genome to discover highly informative markers of similar quality to those used in establish forensic panels. From the 30,000 microsatellite markers we found, we chose the best markers based on their characteristics and compiled two highly informative microsatellite panels, MP10 (12 markers) and MP11 (10 markers). Upon testing these panels across a diverse sample set, the number of alleles in each was found to be 236 (MP10) and 244 (MP11); and the combined exclusion probabilities are 1.00000 (100%). Results from empirical and simulated assignments using these markers show they are extremely powerful and the combination of both panels will achieve a level of accuracy approaching 100 % in tracing escaped farmed salmon from Norwegian Atlantic salmon aquaculture industry.

One limitation of this approach is the practical and logistic challenges involved in sampling the approximately 50 000 broodstock in production in Norway. For this to occur, standardised methods for sampling, identification and DNA extraction need to be in place. We evaluated different methods for sampling using biopsy tools and barcode-labelled vials that are promising for an industry-wide application. We further evaluated common DNA extraction methods using different tissue and preservation methods across different laboratories and found that there was a statistically significant effect on the DNA quality based upon both the type of DNA extraction and the laboratory performing extraction. An important decision for industry-wide implementation will be whether to extract higher quality DNA for archival purposes (i.e. that can be used for more demanding applications later) or lower quality DNA that is used for the sole purpose of microsatellite genotyping.

From the results of this study, we can conclude that DNA tracing of escaped farmed salmon using microsatellite DNA is feasible, however there are logistical challenges that must be met in terms of the dissemination of genetic material throughout the industry. In practise, an optimal solution would be to limit the genetic material in order that different companies operating within a limited geographic range do not obtain genetic material from the same source. Further, a database needs to be established in which the identification and dissemination of each of the 50 000 broodstock is recorded along with their microsatellite DNA profile.

Norwegian summary

Laks som rømmer er ansett som en miljømessig trusselfaktor overfor ville bestander av laksefisk. Derfor er det viktig at man kan både skille mellom vill- og oppdrettslaks i naturen og spore rømt oppdrettslaks tilbake til eieren. Registrering av distribusjon av rogn fra bestemte foreldrepar til matfiskprodusentene, og deretter sporing av rømt fisk tilbake til sine foreldre og derfor eieren med bruk av DNA markører er en måte å oppnå dette. Optimalisert sett av svært robust mikrosatellitt DNA-markører er mye brukt for sporing i rettsmedisinske tester, og i dette prosjektet vi benyttet laksens genomsekvens for å finne mikrosatellitt markører som kunne lages til en test som er like effektiv og robust i laks. Over 30 000 mikrosatellitt sekvenser ble oppdaget i genomet og basert på strenge kvalitetskriterier, 81 primerpar ble testet. To multiplekser ble utviklet: MP10 med 12 markører og hittil 236 alleler observert, og MP11 med 10 markører og hittil 244 alleler observert. Både simuleringer og empiriske tester ble gjennomført i prøver av kjent stamtavle for å teste tilordningspotensiale. Simuleringer viser til entydig tilordning som nærmer seg 100% og de empiriske resultatene viser en 99.8 % tilordning av avkom til minst en riktig foreldrefisk med bruk av en av disse multipleksene (MP10). Dessuten, ingen villfisk ble feiltilordnet til kjent oppdrettsforeldre. Optimalisering av vevsprøver innsamling protokoller og DNA ekstrahering ble også gjennomført i prosjektet. Finneprøver lagret i sprit i prøveglass ferdig merket med strekkoder og kompatibel med automatisering i labben virker som en bra løsning. DNA ekstrahering med 'Chelex' metoden ført til gode resultater med mikrosatellitt genotyping, men hvis det er behov for langtidslagring av høykvalitets DNA, andre metoder er anbefalt. Med bruk av optimalisert lab protokoller inkl. Q5 PCR enzym (New England Biolabs) i kombinasjon med en hurtig 40 minutters PCR, kan vi enkelt generer genotyper fra vevsprøver i så lite som fire timer. Selv om logistikken rundt implementering av DNA-basert sporing over hele næringen er en utfordring, sporing av rømt oppdrettslaks med mikrosatellitt DNA markører virker lovende.

Table of Contents

1 Introduction

1.1 Project background

Initiatives from both the Fishery and Aquaculture Industry Research Fund (FHF) and the Norwegian Seafood Federation (FHL) have stated that from 2012 every production Atlantic salmon shall be marked in such a way that each escapee can be traced back to a responsible owner/company. This demands a national traceability scheme where the assignment of escaped fish must be both highly reliable and efficient.

Several strategies for industry-wide traceability in Atlantic salmon have been suggested, and yet each has limitations. A traceability scheme using DNA markers has shown particular promise (Hayes *et al.* 2005, Håstein *et al.* 2001), and has been implemented in the farmed salmon industry in Norway to trace escaped fish back to their cage of origin in the case of local escape events (Glover *et al*. 2008; Glover 2010). Although several of these strategies have been shown to be effective, such methods are limited if the escapees have travelled a considerable distance; also the vast number of production cages spread over the country adds further complexity, logistical demands and therefore costs. Furthermore, these methods are better at exclusion, rather than confirmation, of a particular cage as the source of the escape. This is largely because the current dissemination of fish material may result in several cages with similar genetic composition. The PAR strategy by Hayes *et al.* (2005) modelled the assignment of individual fish to parents at the multiplier level; however, this strategy is limited in its ability to assign offspring back to a specific owner/company due to no legally imposed constraint on the current dissemination structure of material from breeding nuclei to smolt producers and grow-out farms (i.e. offspring from the same parental pair may be distributed to multiple farms).

Breeding companies including Aqua Gen and SalmoBreed have biobanks of tissue samples from the nucleus parents and offspring (e.g. disease challenge test offspring) from several generations in their breeding program. While this is not precisely the structure (numbers of families and number of individuals per family) of the multiplier-production-escapee scheme, the multi-generational samples available will provide an equivalent means to test the effectiveness of the proposed traceability scheme prior to implementation on an industry-wide scale.

1.2 Project organisation

The project team consisted of Nofima (project leader), the Norwegian Institute for Nature Research (NINA), IMARES (Wageningen University, the Netherlands) and the Department of Environment and Primary Industries (Australia). The project was led by Dr. Matthew Baranski (Nofima), and was carried out in coordination with a parallel FHF project investigating DNA tracing using a different type of genetic marker (FHF 900706). Several work packages were carried out as collaborations between these two projects.

A common steering group was appointed to both projects, consisting of: Petter Arnesen, Marine Harvest; Sissel Kjøglum, AquaGen AS; Vidar Lund, Raumagruppen; Håvard Bakke, Salmobreed AS. The contact person in FHF was Kjell Maroni.

2 Research questions and project aims

The main goal of this research project is to develop, validate and scientifically document the performance of microsatellite DNA markers to trace suspected escaped farmed salmon caught in the wild back to their parents. Secondary objectives were as follows:

- To develop and evaluate efficient laboratory protocols for sample handling and storage
- To identify a set of microsatellite markers of high quality and polymorphism
- To combine these microsatellites into 'multiplexes' of at least 12 markers for efficient genotyping
- To empirically evaluate the power of parentage assignment using these markers using a sample set comprising parents and offspring of known pedigree
- To exclude assignment of non-related wild salmon to any parents in this pedigree
- To assess the feasibility of the method at the 'industry wide scale' using empirical data generated by the project

3 Project design and methods

Our overall strategy for implementing traceability of escaped farmed fish is shown in Figure 1. In short, this scheme involves physically marking all production fish (methods assessed in other FHF projects) to enable rapid and simple identification of farmed escapees, sampling and genotyping of all multiplier parents with DNA markers, and to limit the distribution of eggs from a particular female to only one grow-out company. Genotyping of an escaped farmed fish will then permit the tracing of that fish back to its multiplier parents and hence to the grow-out company that received offspring of that female parent.

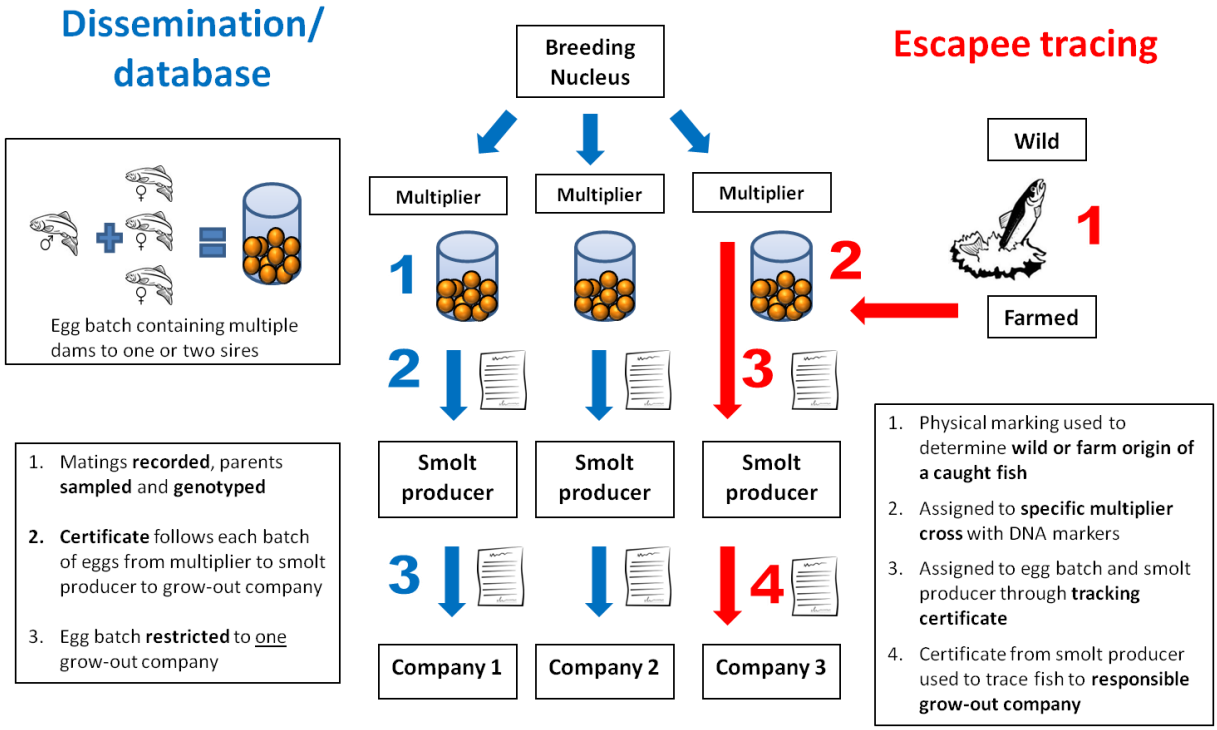

Figure 1 A suggested strategy for implementing traceability of escaped fish

Initially, it was intended that this project would cover all elements of the described scheme. However, the scope of the project was subsequently reduced to focus on the development and testing of the microsatellite markers, together with assessment of sampling, preservation and DNA extraction protocols that form the 'sample to genotype' part of the overall strategy.

Five work packages were carried out within the project.

3.1 Work Package 1: Optimisation of sampling, transport, DNA extraction and storage methods with large-scale industry implementation in mind

To ensure the feasibility of an industry-wide DNA tracing project, optimised sampling, tracking, transport, DNA isolation and storage methods are needed. In this work package we evaluated various methodologies and scenarios to optimise steps from sampling to DNA storage.

DNA quality and quantity can vary widely depending on the type and amount of tissue used, the method of preservation of the tissue, and the method used to isolate the DNA. We evaluated the DNA quality and quantity obtained from three tissue types: adipose fin, pectoral fin, scale. In

addition, we evaluated different preservation methods for these tissues: Freezing at or below -20 °C, 70 % lab-grade ethanol, household methylated spirits (Norwegian: rødsprit).

We evaluated various biopsy tools as a means of obtaining relatively equal amounts of tissue, and for their ease of use and handling. A meeting with OS ID (www.osid.no/139.OS-ID-Ear-Tags.html) facilitated a trial of a system that couples biopsy sampling with tissue preservation and sample coding. A further meeting with FluidX (www.fluidx.eu) facilitated a trial of barcoded sample tubes and sample racks in addition to barcode reading systems that allow the simultaneous scanning of up to 96 samples in a rack. Various biopsy punch tools from generic laboratory suppliers were evaluated for their ease of use.

We evaluated the quality and quantity of DNA obtained using two common DNA extraction methods: Chelex-100 (Yue and Orban, 2005), and a salting-out extraction method (Miller *et al.*, 1988). These methods were modified to accommodate extraction in 96-well plates. A standard 96-well PCR plate was used for the Chelex extraction, and a standard 96-well deep well plate was used for the salting-out method. DNA was extracted at three participating institutions (Nofima, CIGENE and Biobank AS) and the DNA was evaluated in terms of quality and quantity by a single laboratory to ensure a thorough comparison.

3.2 Work Package 2: Development of two highly efficient microsatellite marker multiplexes, and optimisation of genotype and data analysis

Existing (published) and new microsatellite markers mined from the Atlantic salmon genome sequence were used to develop optimised multiplexes of markers. The software package QDD was used to scan the Atlantic salmon genome sequence for microsatellite markers meeting criteria of optimal repeat length, repeat motif and having non-repetitive flanking sequence. These stringent criteria were used to narrow down a set of highly promising candidate markers. Primer sets for these candidate loci were obtained and genotyping was performed on a small number of individuals in order to identify markers that were polymorphic and produced high quality genotypes. Markers that produced poor quality genotypes or were monomorphic were subsequently rejected. The remaining pool of high quality markers were then amplified in different combinations in order to identify multiplexes that maximised the total number of markers, level of polymorphism and ease of analysis. Laboratory optimisation of the multiplexes was performed evaluation of different PCR reagents and cycling protocols. GeneMarker software parameters we optimised so that analysis of genotype data could be automated to the maximum possible extent.

3.3 Work Package 3: Refinement of simulation schemes for industry traceability using marker data from WP2

Simulations of traceability schemes (Hayes et al. 2005) have indicated the feasibility of using DNA markers as a traceability tool for the Norwegian salmon industry. However, for such simulations to reflect the current day situation in the Norwegian industry, they need to be refined using updated parameters, and need to be extended to model a full-size industry situation. In this work package, empirical marker data (allele frequencies and levels of relatedness) generated in WP2 and WP4, together with additional genotyping analysis performed for breeding company partners, was used in stochastic simulations assessing the assignment power of the markers at the industry wide scale. Specific details on the aims, methods and results are provided in Appendices 3 and 4.

3.4 Work Package 4: Testing of the discriminatory power of the microsatellite marker multiplex for correctly assigning offspring to parents and excluding wild fish with a 'blinded' sample set

The microsatellite multiplexes were tested for their assignment/exclusion power using the 'blinded' sample sets common to both the microsatellite and SNP assessment efforts (provided by FHF 900706). The 'blinded' sample sets consisted of parents and offspring of known pedigree or known egg batches. Genotyping of the optimised microsatellite multiplex was carried out, and using the provided list of offspring, sires and dams, and known crossing information, parentage assignment was carried out using Cervus software (Kalinowski *et al.*, 2007). The known pedigree (validation study 1) or known egg batches (validation study 2) were subsequently used to confirm the accuracy of these assignments. Discrepancies were compared to the SNP marker based results from FHF 900706.

3.5 Work Package *5:* Final report and development of an implementation guide for industry

This final report will form the foundation of an implementation guide for industry, through summarising the findings of several of the most important elements of the DNA marker based tracing pipeline. A comprehensive implementation guide for industry will need to build on these conclusions, adding missing elements of the logistical chain. These logistical elements primarily consist of a method and protocol for physical tracking (and perhaps restricted dissemination) of egg batches and establishment of a common database that contains information on broodstock genotypes and the dissemination of their offspring.

3.6 Co-operation plan between 900708 (Nofima) and 900706 (NVH)

This document describes the coordination between the two projects, in order to minimise duplication of research and maximise the ability to address the common goal. Co-operative tasks represent work packages where direct collaboration was undertaken between the two projects. The relevant work package numbers are given to indicate these tasks that are described in more detail in the respective project descriptions. A description of the co-operative tasks is provided in Figure 2.

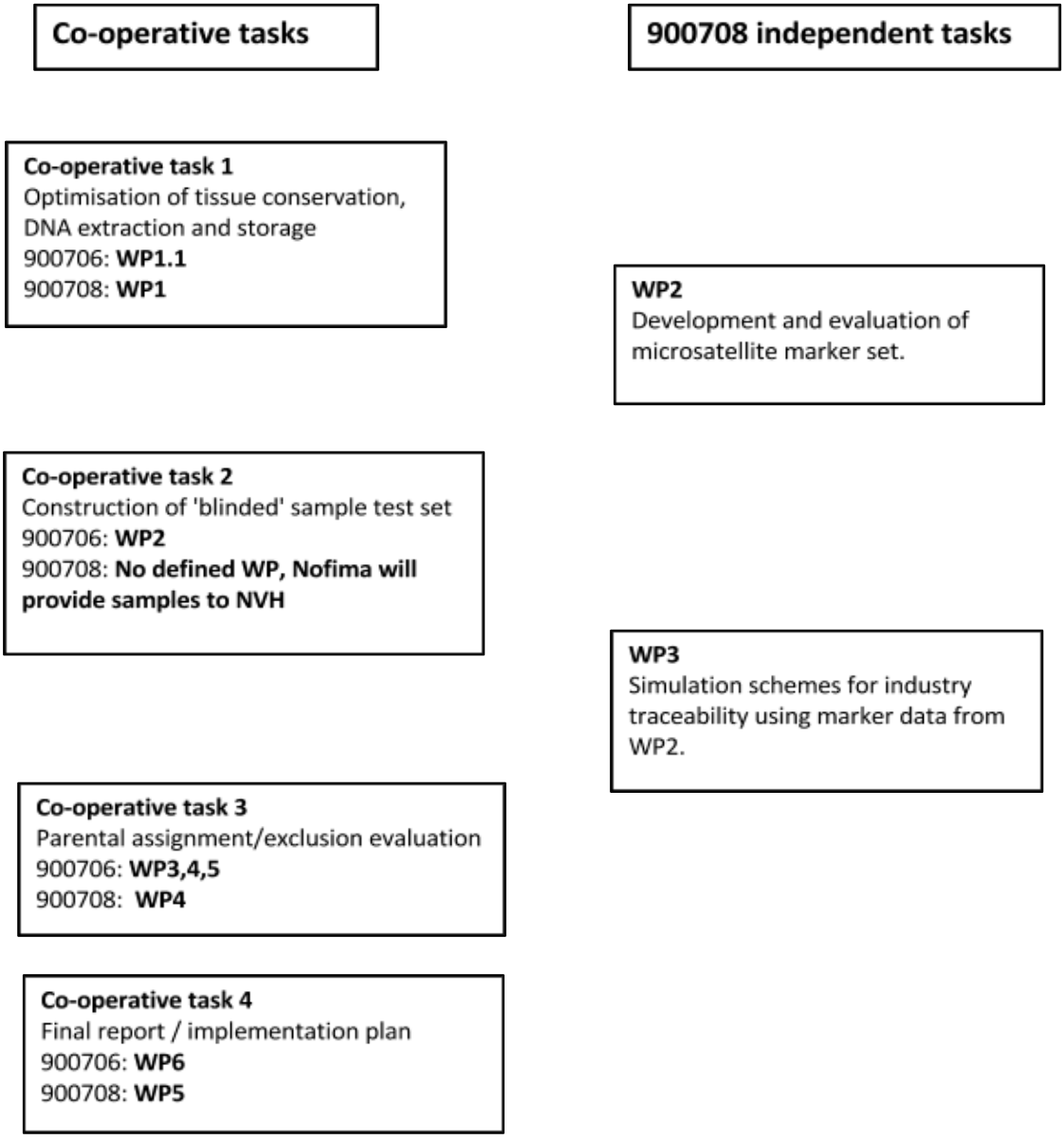

Figure 2 Descriptions of co-operative tasks in projects 900706 and 900708

3.6.1 Co-operative task 1: Optimisation of sampling (tissue conservation), transport, DNA extraction, and storage

Description: Efficient sampling methods will be evaluated, optimal storage and transport protocols developed, and optimised DNA extraction methods developed and integrated into high-throughput workflows. Standard operating procedures and protocols documented for broad use (multipliers, fisherman etc.).

Responsible: Nofima

Participants: CIGENE and Biobank AS

Timeframe: 1/1/2012 - 1/1/2013

3.6.2 Co-operative task 2: Construction of 'blinded' sample test set for parentage assignment evaluation

Description: A single sample set will be developed for assessment of the two different DNA marker types, and will be primarily constructed by project 900706, with additional samples provided by

900708. The sample set will be structured so that it represents a 'challenging' assignment situation, including closely related candidate parents and offspring. Wild samples will be included to test the ability of marker panels to 'exclude' false assignment to a parental set. Partners will contribute available samples from breeding companies in order to make the most representative sample set possible.

Responsible: Biobank AS

Participants: Nofima, NVH, CIGENE.

Timeframe: 1/1/2012 - 1/1/2013

3.6.3 Co-operative task 3: Parental assignment/exclusion evaluation

Description: Development and evaluation of software and analysis procedures for assessing the effectiveness of markers for parentage assignment and exclusion. Various existing programs to be evaluated and workflows developed for efficient data handling and processing. Parentage assignment and exclusion analysis using SNP and microsatellite data from joint sample set.

Responsible: CIGENE

Participants: Nofima and NVH

Timeframe: 1/7/2012 - 1/7/2013

3.6.4 Co-operative task 4: Final reports - implementation plan for industry

Description: All results and developed protocols will be documented in a final report of each project, which together can be extended into an implementation plan for industry, including detailed cost evaluations and logistical requirements. This will enable industry members to make a clear assessment of the feasibility of DNA marker based tracing as a means of addressing the defined traceability issue.

Responsible: Marelife Services, Nofima

Participants: NVH

Timeframe: 1/7/2013 - 31/12/2013

4 Results

4.1 Work Package 1*:* Optimisation of sampling, transport, DNA extraction and storage methods with large-scale industry implementation in mind

4.1.1 Sampling and storage

In collaboration with project 900706, we have evaluated a range of different solutions for sampling, preservation, and DNA extraction. Given that a comprehensive examination of the logistical chain from sampling to genotyping was not included in this project, it was not possible to establish a standard protocol for sampling. Nevertheless, we have identified several promising solutions for different elements in this chain.

OS ID® (www.osid.no/139.OS-ID-Ear-Tags.html) supplies tags and tag-reading equipment for the livestock and aquaculture industries. In addition they have developed a highly-automated system for tissue sampling that combines biopsy sampling, tissue preservation and sample identification. Further, the OS ID® tissue sampling unit (TSU) contains a preservation buffer that can be incorporated into a DNA isolation procedure, minimising the risk of contamination by transferring the sample prior to DNA extraction. OS ID® supply an automated de-capping instrument which allows de-capping of 96 sampling tubes in a standard 96-well plate. Every tube is labelled with a 2-D barcode which negates human error in writing down the identification code of the sample. One advantage of the OS ID® TSU is that there is virtually no risk of cross-contamination of the tissue sample as the biopsy punch releases the tissue into the tube with preservation buffer and seals it in a single process. In addition this system can be combined with PIT tagging of broodstock, minimising extra stress and handling on the animals. One down-side of this system is the size of the biopsy punch, which can be cumbersome in a real-life sampling environment for fish where water and fish slime may cause difficulty in operation. However, this system was initially developed for terrestrial livestock species (cows, pigs, sheep, etc) and may be able to be modified for aquatic species. An image and description of the OS ID® TSU can be found here: www.osid.no/3479.3472.DNA-Tissue-sampling.html.

FluidX (www.fluidx.eu) is a company that provides sample storage consumables and equipment. They provide barcoded sample tubes in either or both 2-D and 3-D barcode formats, in addition to barcoded sample racks and barcode scanners. Sample tubes can be decapped in an automated fashion, and barcode scanners are available that read either each tube individually, or read every tube in an entire plate by scanning the 3D barcode at the bottom of the tube. Further, the tubes are certified for freeze storage and barcode scanning functions despite frost on the tubes. An additional product from FluidX is a sample tracking software that is suitable for storing and tracking barcoded samples and the database is customisable for laboratory-specific fields allowing one to define freezers, organise samples into groups, etc. For further details of the FluidX EasyTrack software refer to: (www.fluidx.eu/easytrack-sample-tracking-software.html).

Both the OS ID® TSU and the tubes provided by FluidX can easily be incorporated with PIT tag readings, creating a secure link between fish ID and sample ID. In addition, a sample tracking database such as that supplied by FluidX can enable secure tracking of samples from the sampling point through the DNA extraction process and storage, minimising human errors associated with recording sample identification codes and storage locations.

Various biopsy tools were evaluated for tissue sampling, including Miltex 3 mm punches with plungers, and Harris Unicore (3 mm) punches either with, or without, plungers. In addition, a generic craft single-hole punch was evaluated. Biopsy punches can be used multiple times with cleaning between each sample by cutting a hole through filter paper and rinsing thoroughly with alcohol and deionised water; yet it was found that punches which lacked a plunger caused difficulties dislodging the tissue, even after cleaning. Discussions with industry breeding companies confirmed that 3 mm diameter punches were suitable in terms of easy of tissue cutting and amount of DNA obtained.

4.1.2 Tissue type and preservation and DNA extraction

A multi-factorial design for storage and extraction of DNA was made with the three participating laboratories (Nofima, CIGENE, Biobank AS). Adipose fin, pectoral fin and scale samples were taken from a number of fish at Nofima's research station at Averøy in April 2012. Biological replicates of each sample were preserved by either freezing at or below -20 degrees, or at room temperature in 70% laboratory grade ethanol or household methylated spirits (Norwegian: rødsprit). After 12 months of storage, DNA was sent to each of the three laboratories where DNA was extracted following identical protocols (Chelex-100® or salting-out methods). DNA quality was assessed by CIGENE on an agarose gel in addition to concentration measurements using picogreen and nanodrop. The gel image indicated that DNA of high molecular weight (HMY) was not obtained in samples extracted by the Chelex-100® method. Nevertheless DNA extracted using this method was of sufficient quantity and quality for amplification by PCR and subsequent genotyping by microsatellites. DNA extracted using the salting-out procedure was of higher molecular weight, however this method is more labour-intensive compared to the Chelex-100® method. It is important to note that although we did not compare commercial kit-based methods for this DNA extraction evaluation, a pilot-test of a commercial kit indicated intermediate results when compared with Chelex-100® and the salting-out procedure (Figure 3). The performance of commercial DNA extraction kits for various downstream applications has been evaluated extensively in the scientific literature (e.g. Claassen *et al.*, 2013; Fahle and Fischer, 2000; Smith *et al.*, 2003). The different extraction methods were not evaluated in detail for microsatellite genotyping quality in this project as both methods are routinely used in our laboratory with good results. Likewise, all tissue types are routinely used in this laboratory and there has been no noticed difference in microsatellite genotyping quality associated with different tissue. For practical purposes, it is recommended that either an adipose or other fin sample is used rather than a scale sample as the scales are more difficult to place into a standard 96-well PCR plate for DNA extraction. In terms of preservation method, there was no significant difference in DNA quality for the same tissue type preserved using freezing, laboratory grade ethanol (70%) or household methylated spirits. CIGENE conducted further tests on the effect of tissue type, preservation method and extraction method on the quality of SNP genotyping. They found no significant effect of preservation method or tissue type, but did find a significant effect of the DNA extraction method and the laboratory performing the extraction on the SNP quality. For details of these results refer to the final report by project FHF 900706.

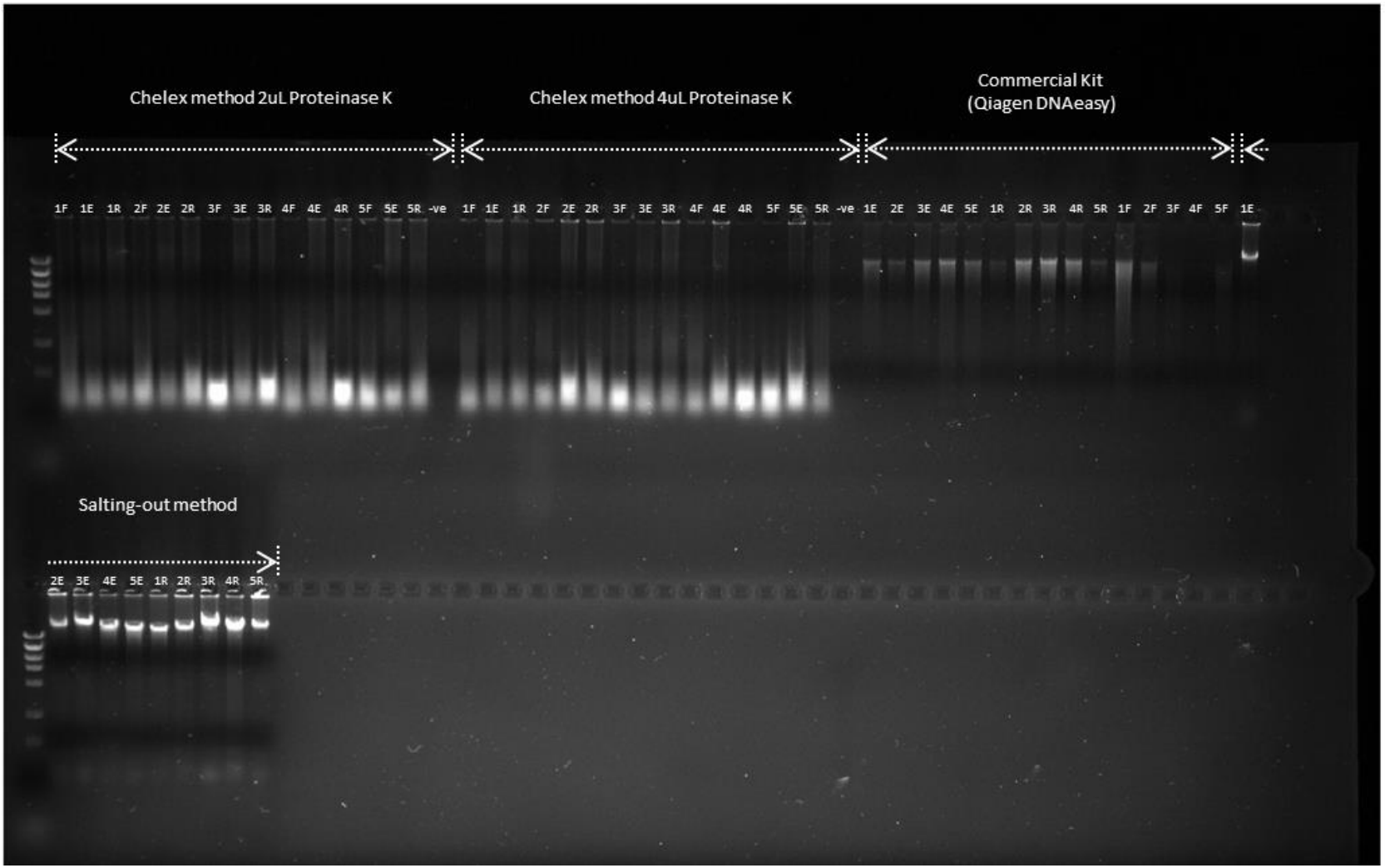

Figure 3 *A comparison of different DNA extractions using the Chelex-100® method with either 2 µL proteinase K, 4 µL proteinase K; a commercial DNA extraction kit (Qiagen DNAEasy) and a salting-out method.*

The main conclusion from the DNA extraction assessment is that the methods typically offer a compromise between sample quality and processing speed. Fast and cheap methods like Chelex-100® perform well for microsatellite genotyping, but downstream DNA quality is not as high, limiting use for other applications (such as whole-genome sequencing). However, it has not been established whether there is a need to maintain high quality DNA from 50,000+ broodstock samples each year. This should be discussed with industry representatives to identify whether this is desired or not. Late in the project timeframe, we also evaluated a DNA extraction method (Nexttec™ http://www.nexttec.biz/) that appears to offer rapid and high quality DNA extraction, potentially combining the benefits of two other methods evaluated. A gel-image of DNA extracted from 3 mm biopsies of fin clips using the Nexttec™ column kit is shown in Figure 4 and indicates DNA of high molecular weight for all 10 samples (bright band); the pale bands in the well are likely RNA as no RNAase digestion was performed. The significant difference in DNA quality that was attributed to the different laboratories indicates that a standard protocol and quality controls are necessary if several laboratories will be performing the DNA extractions on an industrial scale.

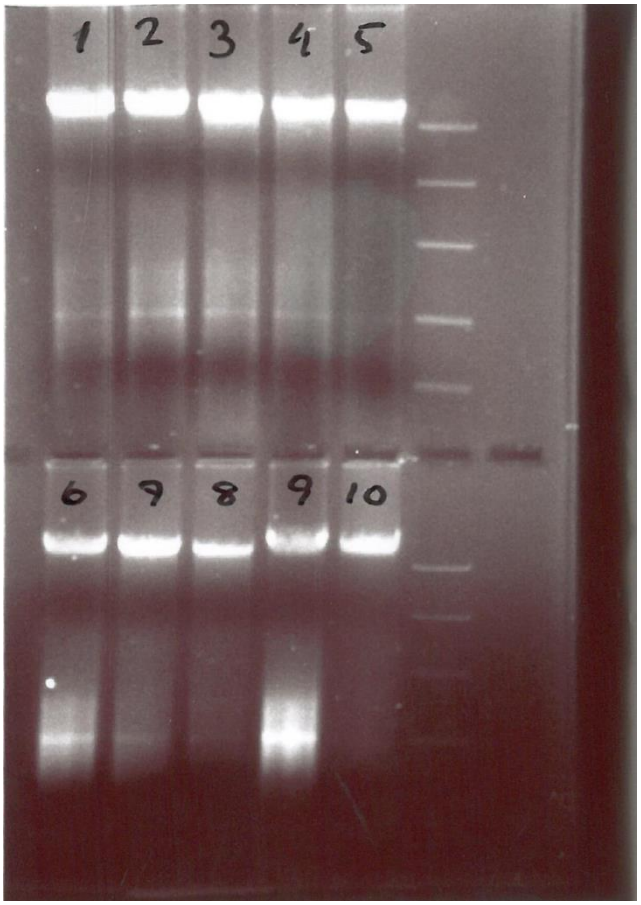

Figure 4 A gel image of DNA from 10 fish extracted using a Nexttec™ column kit

4.2 Work Package 2*:* Development of two highly efficient microsatellite marker multiplexes, and optimisation of genotype and data analysis

4.2.1 Development of microsatellite panels

Over 30,000 microsatellite sequences were discovered in the Atlantic salmon genome sequence assemblies. Based on strict criteria of repeat unit type (perfect not compound) repeat unit length (around 14 repeats) and absence of mononucleotide stretches in primer regions, 81 primer sets were ordered. These 81 primer sets were tested individually on 8 individuals to determine whether the marker was polymorphic and to assess amplification quality. We excluded markers that appeared to amplify multiple regions or those which had a signature that would be difficult for automated scoring of genotypes (e.g. many stutter peaks). Of the 81 markers, 69 amplified and were polymorphic. From the resulting 69 microsatellites and two previously described microsatellite markers (Sssp2216, Paterson et al, 2004; SsaA124, King et al 2005), we selected optimal microsatellites and attempted to fit these together in a multiplex panel using the four common dye labels from Applied Biosystems (6-Fam, Vic, NED, PET). A multiplex panel needs to be optimised in order that microsatellites of the same colour do not have overlapping allele ranges, and as wild salmon typically possess a large amount of genetic variation, we used wild salmon samples to determine the range of alleles at each microsatellite. DNA from 89 salmon from 19 populations throughout Norway was used for this purpose; Figure 5 shows the location and number of wild salmon used. The multiplexes have been trialled in different laboratories and on different genetic analysers to ensure cross-amplification works with different instrumentation and laboratory protocols. The multiplex panels were continually optimised according to the allele ranges, until we obtained a panel that appears to be stable in terms of number and ranges of alleles. 12 markers were ultimately combined into the multiplex “MP10” and primer concentration ratios were optimized to produce amplicons showing relatively equal RFU (relative fluorescence units) intensities on ABI 3130xl and 3730xl Genetic Analyzers (Figure 6). In parentage assignment studies, the total number of alleles available in the population is more important than the number of markers for accuracy of assignment. We assessed the number of alleles present in wild Norwegian Atlantic salmon, in addition to approximately 830 broodstock salmon from 3 of the 4 Atlantic salmon breeding programs in Norway, and a total of 236 different alleles have been observed for the 12 markers in MP10. An additional 10 marker microsatellite panel “MP11” was designed following the same procedures above and has been tested on approximately 500 wild Norwegian salmon from six populations in Norway, giving a total of 244 alleles; further testing of MP11 with more diverse salmon is recommended to ensure no overlap of allele ranges.

Figure 5 *Locations of wild Atlantic salmon provided by NINA for testing in this project. Numbers in parentheses represent the number of samples provided from each population; * some samples were excluded as they were found to be salmon and trout hybrids*

When developing a marker set for use in parentage or relatedness assignment it is important to be able to predict the power of the marker set to accurately predict the assignment. Polymorphic information content (PIC) is the probability that an individual will be heterozygous (i.e. have two different alleles) at a particular marker, and is often used as an indicator of the usefulness of genetic markers in parentage assignment. The exclusion probability (EP) is the probability that a marker will exclude a false (unrelated) parent as being related to the individual being tested. For both multiplexes, we calculated the PIC and the EP for both a single parent assignment and for a parent pair assignment for each of the markers using the program Cervus (www.fieldgenetics.com; (Kalinowski *et al.*, 2007). Table 1 lists each of the markers in both multiplexes and gives information on the number of alleles, allele ranges, the dyes used in the multiplex, PIC, EP, and the number of individuals assessed to derive these values. As can be seen from Table 1 both MP10 and MP11 are highly polymorphic and have an extremely high power of exclusion for both single parent (EP-1P) and parent pair (EP-2P) scenarios, effectively 100% for each multiplex. It is important to note however that exclusion power for a marker set assumes that the families to be tested are independent from each other and that no errors in genotyping occur; thus exclusion power will diminish with closely related families and with the presence of scoring errors. For this reason, simulations of the power of MP10 to assign correct parents under a scenario representing the Norwegian salmon industry were conducted; the results of this are discussed in section 4.3. For more details on the microsatellite markers developed in this project, refer to Appendix 1: Microsatellite multiplex information.

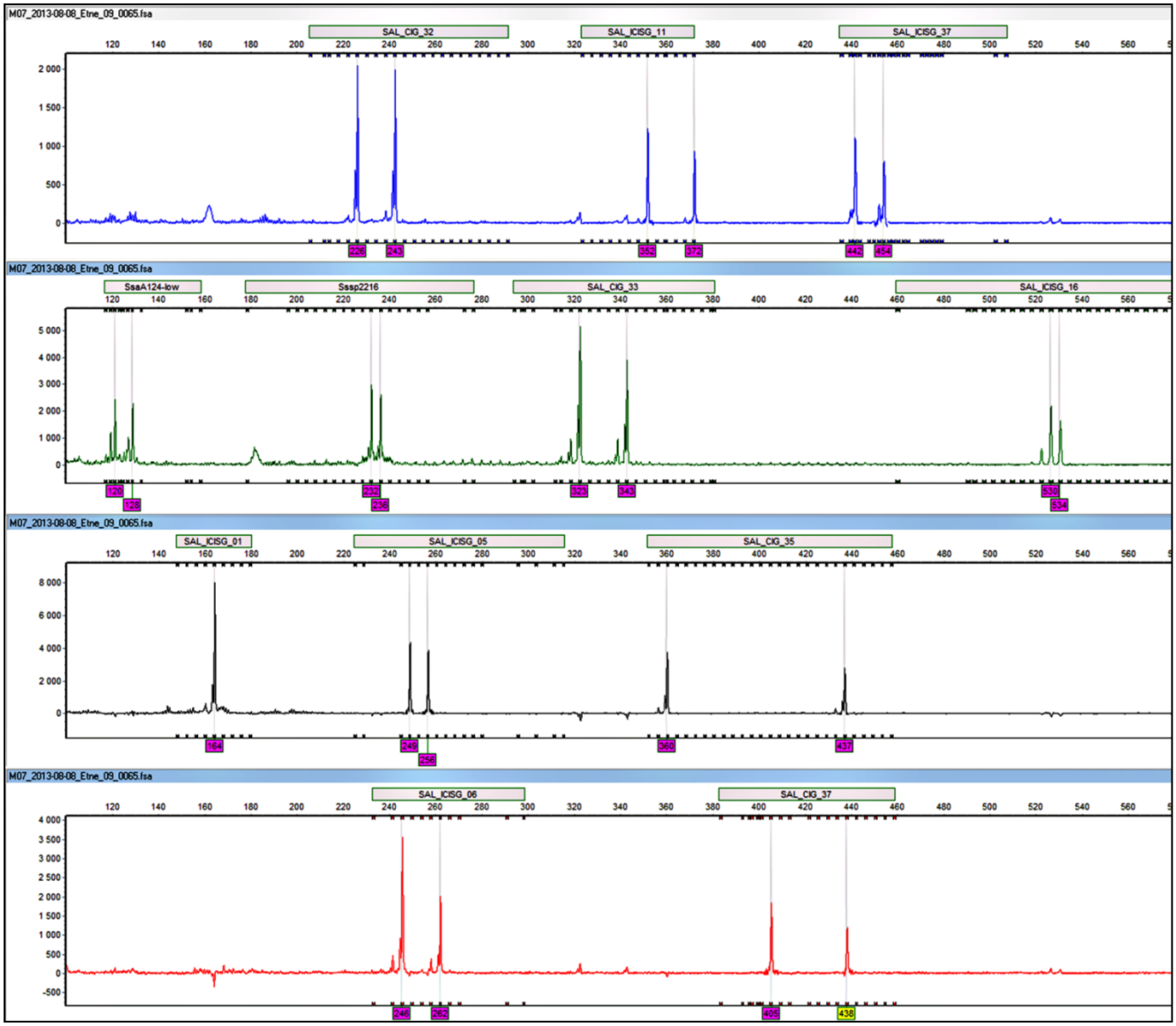

Figure 6 *Microsatellite genotypes for a single individual using MP10. PCR was genotyped on an ABI 3730xl and scored using the software GeneMarker.*

Table 1 *Information for microsatellite markers in multiplexes 10 and 11 (MP10 and MP11). The dye refers to standard dye from ABI; k refers to the number of alleles identified; N refers to the number of individuals assayed; PIC is the polymorphic information content; EP-1P is the exclusion probability for one parent; EP-2P is the exclusion probability for a parent pair.*

Locus	Dye	k	*N*	allele range	PIC	EP-1P	EP-2P
MP10							
SAL-CIG-32	6-FAM	23	1778	206-295	0.879	0.6380	0.9250
SAL-ICISG-11	6-FAM	17	1775	320-372	0.832	0.5360	0.8700
SAL-ICISG-37	6-FAM	19	1764	436-480	0.624	0.2640	0.6620
SsaA124-low*	VIC	15	1786	114-162	0.549	0.2100	0.5150
Sssp2216**	VIC	17	1742	196-265	0.894	0.6700	0.9380
SAL-CIG-33	VIC	24	1787	299-387	0.884	0.6480	0.9300
SAL-ICISG-16	VIC	29	1770	456-599	0.912	0.7150	0.9550
SAL-ICISG-01	NED	9	1721	148-180	0.766	0.4240	0.7880
SAL-ICISG-05	NED	21	1389	225-315	0.801	0.4810	0.8310
SAL-CIG-35	NED	33	1406	344-461	0.936	0.7820	0.9740
SAL-ICISG-06	PET	12	1788	225-295	0.739	0.3960	0.7840
SAL-CIG-37	PET	17	1774	393-459	0.869	0.6120	0.9110
Total MP10		236				1.0000	1.0000
MP11							
SAL-ICISG-19	6-FAM	9	545	110-133	0.717	0.3580	0.7180
SAL-CIG-28	6-FAM	18	524	147-222	0.866	0.6040	0.9070
SAL-CIG-29	6-FAM	19	540	233-296	0.799	0.4850	0.8390
SAL-CIG-38	6-FAM	32	523	414-555	0.927	0.7570	0.9680
SAL-CIG-31	VIC	48	536	203-391	0.931	0.7710	0.9720
SAL-CIG-27	NED	30	527	120-263	0.909	0.7080	0.9520
SAL-ICISG-10	NED	47	478	293-484	0.955	0.8410	0.9870
SAL-ICISG-07	PET	26	506	183-322	0.938	0.7860	0.9750
SAL-CIG-21	PET	5	497	374-383	0.42	0.1270	0.3640
SAL-ICISG-38	PET	10	385	455-476	0.576	0.2210	0.5760
Total MP11		244				1.0000	1.0000
Total MP11 + MP10		480				1.0000	1.0000

4.2.2 Optimising of PCR conditions

Genotyping efficiency is critical if tens of thousands of samples are to be genotyped rapidly, therefore we tested the newly developed markers with a new generation of PCR reagents. Q5® Hot-start high-fidelity DNA polymerase is a product developed by New England Biolabs® Inc. This is a high-fidelity polymerase with a genotyping error rate more than 100-fold lower than that of the traditional form of polymerase used in PCR (Taq DNA polymerase). Further, this new generation polymerase is extremely fast, (10 s/kb), allowing the potential for very rapid PCR cycles. We designed our primers with high annealing temperatures, in order to enable a 2-step PCR cycle thus further reducing the time required for PCR. The PCR program used for both the MP10 and MP11 multiplexes is given in Table 2; for relative concentrations of PCR primers and reagents for both multiplexes refer to Appendix 2: PCR reagent volumes and ratio of primers in MP10 and MP11.

Table 2 PCR cycle settings used for MP10 and MP11 with Q5 Hot-start high-fidelity DNA polymerase

Step	Temperature	Time
Initial denaturation	95 °C	2 minutes
33 cycles	95 °C	8 seconds
	72 °C	45 seconds
Final extension	72 °C	2 minutes

4.3 Work Package 3: Refinement of simulation schemes for industry traceability using marker data from WP2

Two simulation studies were performed by collaborators at DEPI (Australia) and Imares (The Netherlands). In general, it can be concluded from these simulation studies that based on the number and level of informativeness of alleles in MP10, this marker set is powerful enough to trace individual salmon back to their farm of origin, assuming that for each farm, the crosses provided are known, and that one cross is only provided to one farm. As expected, the effects of close relationships among the broodstock and genotype scoring errors affect the power of assignment. However, these simulations were performed on a single marker set (MP10) and it is believed that with inclusion of MP11 for ambiguously allocated individuals, or a combination of genetic markers and physical markers, the power of assignment would be effectively 100 %. The detailed reports from DEPI and Imares are provided in Appendix 3 and Appendix 4, respectively.

4.4 Work Package 4: Testing of the discriminatory power of the microsatellite marker multiplex for correctly assigning offspring to parents and excluding wild fish with a 'blinded' sample set

Two validation sets were compiled by NVH; the first set represented a scenario with many offspring (N = 520) from few potential parents (N = 230), in addition to 40 unrelated offspring. The second set represented few offspring (N = 279) from many possible parents (N = 496). These were *blind* sample sets, (i.e. no information regarding the relationship was provided). A third validation study used wild Atlantic salmon samples supplied from NINA (Figure 5) and the parental samples from validations sets 1 and 2, the aim being to ensure that the wild fish could be not be assigned to any of the farmed parents. Both Cigene and Nofima received identical plates of DNA from these validation sets. Genetic assignment was performed in Cervus using sex information for parents. Assignment was carried out according to suggestions in Kalinowski et al (2007), allowing 1% genotyping errors. A null allele is an allele that fails to be amplified during PCR due to errors in PCR or due to a mutation in the priming region of the DNA sequence of the individual. Null alleles can be a major cause of errors in parentage assignments as a heterozygous individual may appear homozygous if one allele fails to amplify. Due to the possibility of null alleles in the dataset a single mismatch error was typically allowed in the assignments. Crossing information was provided for the parents and only parent pairs that represented known crosses were accepted in the genetic assignment. For offspring that could not be assigned to a known cross, the assignment was repeated to assign to single parents; in this case the parent assigned was that with zero mismatches or the highest likelihood calculated in Cervus.

4.4.1 Validation study 1: Assignment of many offspring to few families

All offspring were successfully genotyped, although genotype quality was poor for one individual and only 10 of the 12 markers could be scored. All sires and all but one dam were successfully genotyped. The results from Validation set 1 are presented in Table 3. From this table it can be seen that the

majority of offspring (99.8 %) assigned to one or more parent, although 12 of these (2 %) assigned to incorrect parents according to the pedigree provided. 480 offspring assigned to two parents with zero mismatches, 23 assigned with a single mismatch, suggesting null allele(s) may be present in one of the MP10 markers. The offspring that had poor quality genotypes assigned to a parent cross but had 2 mismatches across the trio, an exception was made for this individual as there was a higher likelihood that human error in genotype scoring attributed to the mismatches. For the 15 offspring that assigned to a single parent, these assigned at all 12 loci with zero mismatches. We re-evaluated the results for the 12 offspring that assigned to parent pairs that were incorrect according to the pedigree and could still exclude the supposed true parents. Further investigation of the well positions in the 96-well plate suggest human error in recording of sample ID may be the cause of this discrepancy in pedigree assignment and genetic assignment, as the error for each of the 12 offspring followed a pattern consistent with the mix-up of well positions. Further, the results from the genetic assignments of these 12 offspring were concordant with those obtained using SNPs in project 900706. Despite this discrepancy, the MP10 microsatellite panel proves nearly 100 % effective in assignment, and also effective in non-assignment of unrelated individuals. It should be noted that DNA was of insufficient quality for genotyping of a single dam, and this was a reason why some individuals assigned to a single parent. It may also be worth noting that a genetic anomaly in one family resulted in triploid offspring (containing three chromosomes instead of the usual two). Assignment nonetheless was successful for these offspring and it appeared that they inherited both copies of their mother's chromosomes and a single copy of their father's. This demonstrates the robustness of genetic assignment with the microsatellite panel even in light of atypical inheritance.

Table 3 Results of parentage assignment in validation study 1

	Unassigned	Assigned to a valid parental cross	Assigned to a single parent	Correct according to pedigree
520 offspring	1 (0.19%)	504 (96.9%)	15 (2.88%)	507 (97.5%)
40 unrelated offspring	40 (100%)			100%

4.4.2 Validation study 2: Assignment of few offspring to a large number of potential parents

All offspring, all sires and all but one dam were successfully genotyped for the majority of the markers in MP10; genotype quality was poorer than in Validation set 1, and after repeated genotyping some offspring and some parents could still not be scored with certainty for all 12 markers. Due to the possibility of null alleles in the MP10 dataset, in addition to the poor genotype quality of some of the individuals, we allowed two parent-offspring mismatches for this dataset. Results for Validation study 2 are presented in Table 4. In this validation study, an overview of parent crosses was supplied for the analysis but no pedigree information other than the egg batch was supplied for checking of results. Table 4 shows that the assignment to valid parental crosses was lower than that for validation study 1; also the number of assignments to single parents was greater. Four offspring assigned to a valid parental cross with greater than two mismatches across the trio (sire-offspring-dam) and one offspring assigned to a non-valid cross with greater than two mismatches. In total, 10 offspring assigned to non-valid crosses. Nevertheless, when assessing the assignment to egg batches, 269 (96%) of the offspring assigned to the correct egg batch, and just 7 (2.5%) assigned to an incorrect egg batch. For the two offspring that assigned to multiple valid crosses, each of the possible crosses was the correct egg batch. The lower assignment power in this validation study is likely due to the close relatedness among potential parents, and it is possible that some of the mismatches and the multiple assignments are due to assignment to a closely related non-parent. This validation study shows that high assignment power can be achieved by assigning to egg batches, even when the close relatedness of the parents results in lower parental assignment

Table 4 Results of parentage assignment in validation study 2

	Assigned to valid parental cross with greater than 2 mismatches	Assigned to a single valid parental cross	Assigned to a single parent	Assigned to multiple valid crosses	Assigned to non-valid parental crosses
279 offspring	4 (1.43%)	209 (74.9%)	59 (21.1%)	2 (0.72%)	10 (3.58%)

4.4.3 Validation study 3: Exclusion strength of MP10 with 89 wild salmon and farmed parents

All 89 wild salmon were genotyped and the parental genotypes were obtained from validation sets 1 and 2. Using the program Cervus, we forced assignment of each of the wild fish to the most likely parents; however, no wild salmon could be assigned to any of the farmed parents with confidence. The average number of markers with mismatches in comparisons of wild salmon and parents was 6.3, i.e. there was discordance at approximately 53 % of the microsatellites for each trio.

4.5 Work Package *5:* Final report and development of an implementation guide for industry

Within the scope of this project, it has not been possible to develop an integrated implementation guide for the industry. Such a guide would need a comprehensive description of the logistics of the physical tracking of egg batches from producers to grow-out farmers, something that was not covered in this project due to a separate effort being undertaken concurrently. Nevertheless, we have demonstrated the feasibility of a streamlined protocol for sampling and genotyping consisting of the following steps:

1. Sampling of fin tissue samples into laboratory compatible (96 format) barcoded storage tubes
2. Preservation in fixative such as ethanol

High throughput DNA extraction (Chelex-100 ® suitable for microsatellite genotyping, but alternative methods feasible if higher quality archival DNA is desired)

High throughput genotyping with the developed microsatellite multiplex

- Genotyping of a second multiplex in a subset of samples if offspring cannot be unambiguously assigned

Figure 7 displays the evaluated the time required for each of the above steps for 384 samples for a single multiplex; for genotyping of an additional multiplex panel, only steps 2 and three in the diagram need to be repeated.

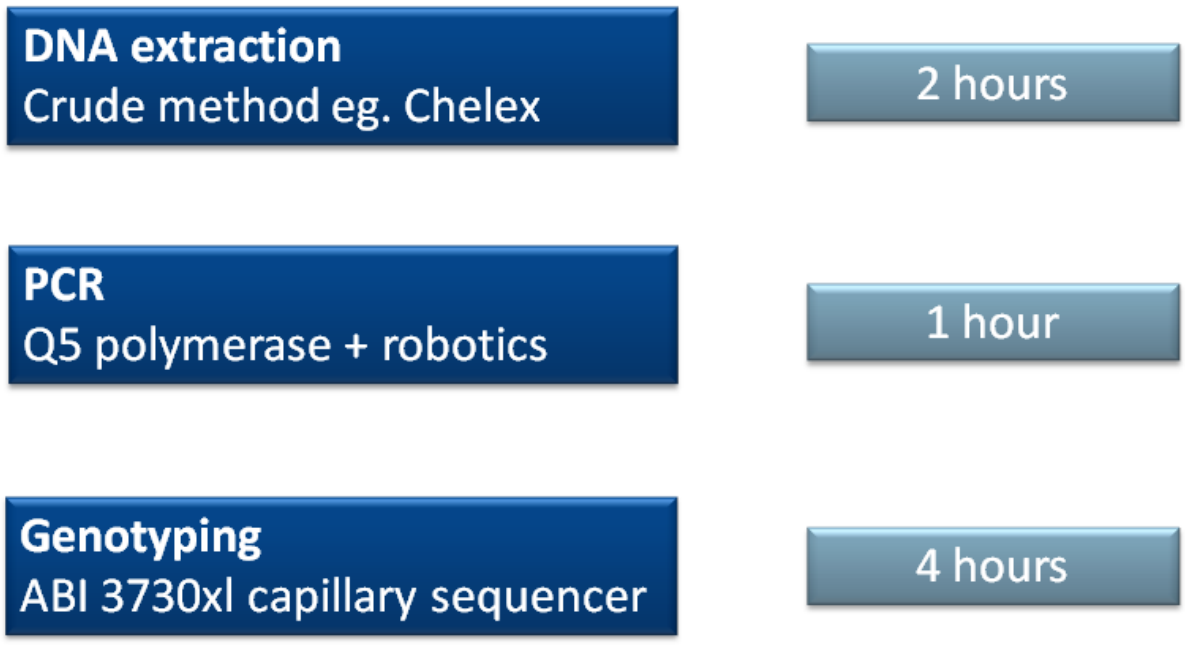

Figure 7 Approximate timeline for obtaining microsatellite genotypes from tissue from 384 fish

At the industry-wide scale of 50 000+ broodstock each year, a dedicated, high throughput laboratory with considerable experience and capacity will be required. It is suggested that a competitive tender process is the optimal way to identify such a laboratory. Nofima could consult and interact with the laboratory during the establishment stage in order to facilitate knowledge transfer of the results of this project to large scale implementation.

Furthermore, it was not possible to perform a full economic assessment for the implementation of this scheme, as such an evaluation is dependent on including all elements of the logistical chain, and a number of these were not within the scope of the project. In addition, the costing is highly dependent on the pricing of the laboratory chosen to perform the genotyping and analysis, and the ultimate scale of this effort. Nevertheless, consumable and running costs evaluated in this project are in the range of 25 – 50 NOK per sample, depending on the extraction method used.

4.6 Long term use of results and utilitarian value to the industry

The collection and storage of all broodstock and escapees DNA and their genotypes will be a tremendous resource for future studies of interactions between wild and farmed Atlantic salmon. It will potentially enable the identification of multiple generations of farmed fish (tracing offspring back to grandparents, indicating potential increased persistence of various genotypes) as well as different types of hybrids and backcrosses and which presently cannot be identified using the method of Karlsson et al. (2011). Therefore, this approach also secures a long term biobank that will be of unprecedented value as a resource for future studies. This resource could also prove to be a valuable tool for monitoring levels of genetic variation at a genomic level over subsequent generations of artificial selection in the breeding programs.

This proposal presents a solution for tracing escaped farmed fish from Norwegian producers, however there is the possibility of escaped fish from the Faroe Islands and Scotland. If successfully implemented, this traceability scheme could be a strong candidate for implementation in these locations to serve both national and international interests.

The DNA based traceability scheme proposed herein would also provide a means of protecting the genetic material from the breeding companies (Gjerde et al., 2011), and facilitate traceability further down the market chain. This will enable traceability and verification of ownership in the event of the occurrence of genetic defects, disease agents or contaminated raw materials.

5 Deliverables

5.1 Detailed overview of deliverables in this project according to the project plan

A list of the deliverables as listed in the project plan are given below.

5.1.1 Deliverables 1 and 2

1. Standard protocol for tissue sampling
2. Report on comparison of different DNA extraction methods

We assessed various methods for tissue sampling, tissue preservation, DNA extraction in Work Package 1. Transportation and storage of DNA was evaluated by Biobank AS in the corresponding project 900706. We found that tissue sampling can be achieved best by taking 3 mm biopsy sections from a fin and storage in alcohol (either 70% ethanol or household methylated spirits). The method of DNA extraction and the laboratory performing the DNA extraction had significant effects on the SNP genotyping quality. 'Rough' methods of DNA extraction are satisfactory for genotyping for microsatellites, however the DNA may not be of sufficient quality for long-term storage. The method of DNA extraction utilised under an industry-wide DNA tracing scheme will require consideration of the method of genotyping to be utilised (e.g. microsatellite or SNP) and the level and time of storage of the DNA, however for the direct purpose of genotyping microsatellites, a rapid and cheap method like the 'Chelex' method is more than sufficient.

5.1.2 Deliverables 3-5

3. Standard protocol for high-throughput DNA extraction for genotyping
4. Optimisation of microsatellite multiplexes
5. Optimised program and analysis parameters and workflow

In Work Package 2 we developed a high-throughput pipeline for genotyping of microsatellite panels for parentage testing. Under this pipeline, DNA can be extracted from 384 fish via a quick DNA extraction method such as Chelex-100® in 2 hours; PCR amplification using new-generation reagents (Q5® Hot start high-fidelity DNA polymerase, New England BioLabs® Inc. and using a, robotics platform for liquid handling can be achieved in 1 hour; genotyping using a 96 capillary genetic analyser such as the ABI 3730 xl and optimised scoring settings in genotyping software GeneMarker® can be achieved in 4 hours. Thus this pipeline can transform fin clips from 384 fish to DNA genotypes in approximately 7 hours.

The microsatellite multiplexes developed in this project provide exceptional power for parentage assignment, with 480 alleles identified to-date. These multiplexes provide further potential in the aquaculture and fisheries industries and for the management of wild populations.

5.1.3 Deliverable 6

6. Simulated assessment of DNA based tracing with use of marker data from WP2

The assessment of the MP10 microsatellite multiplex developed in Work Package 2 for use in an industry-scale tracing scenario was done by researchers at Imares, The Netherlands, and DEPI,

Australia. Results from these simulations confirmed the MP10 multiplex was powerful for parentage assignment, however close relationships among potential parents and genotyping errors will affect the power of assignment. Nevertheless, the predicted proportions of correct assignments were high, and the combination of the second multiplex panel MP11 for ambiguous assignments will likely recover the remaining correct assignments. The results from these simulations are presented in two separate reports in sections 0 and 0.

5.1.4 Deliverable 7

7. Report on parentage testing with microsatellite multiplexes

The efficiency of the microsatellite panel MP10 on parentage testing was assessed in Work Package 4 using blind tests of samples with a known pedigree (validation study 1) or known egg batch (validation study 2). The microsatellite MP10 panel proved extremely effective in parentage assignment, assigning close to 100 % of the offspring to their correct parents in validation study 1. In validation study 2, the close relationships among the potential parents provided a challenge for the MP10 panel, yet the panel was still effective in assigning greater than 95 % of the offspring to the correct egg batches. The inclusion of MP11 would effectively double the power of MP10 and allow assignment of the offspring that were incorrectly assigned using the single microsatellite panel.

5.1.5 Deliverable 8

8. Report on comparison of microsatellite and SNP markers.

This project ran in close collaboration with project 900706, and tests of the microsatellites developed in this project and tests of the SNP panel developed in project 900706 were performed on the same DNA in an effort to compare the results and efficiency of the markers for parentage testing. The results of the two marker sets broadly agree, although the microsatellite MP10 panel appeared to have slightly higher levels of assignment than the SNP panel. Further details of the results of the SNP panel are provided in the final report from 900706.

5.1.6 Deliverable 9

9. Implementation plan including protocols and cost analysis for tracing on an industry-wide level.

The logistical challenges to a genetic tracing scheme like the one presented in this report will likely require changes to the dissemination of genetic material from multiplier to hatchery and grow-out facilities (e.g. Figure 1). Such challenges are to be evaluated by the industry in relation to the other five tracing projects that were funded by this call, and it may be that a combination of tracing methods is adapted by the industry. As such, it is not possible to develop an implementation plan in detail, other than the suggestion in Figure 1. We developed a highly efficient pipeline for protocols and procedures in order to allow the microsatellite genotyping to be performed as quickly and accurately as possible. This included optimising the microsatellite choice and software analysis parameters to enable automated scoring of microsatellites that require little manual editing. An estimate of the cost of microsatellite genotyping for parentage analysis is provided in section 4.5 of this report.

5.2 Publications and planned publications

Ongoing results of this project have been presented at a number of meetings.

10. Baranski M., Jacq C. and Hånes Kirste K. (2013) Development and optimization of a microsatellite 14-plex for parentage assignment in Atlantic salmon (*Salmo salar*). Aquaculture Europe. Trondheim, Norway. August 9-12.

11. Baranski, M., Jacq, C., Karlsson, S. (2013) SporLaks - Industry-wide tracing of Norwegian farmed Atlantic salmon. FHF seminar sporing og merking av laks. Gardermoen, Norway November 19-20.

12. Jacq, C. (2014) Ny test for DNA-sporing av rømt oppdrettslaks. Fiskeseminar 2014 Førde, Norway January 25.

An abstract for a further presentation has been submitted to World Aquaculture Society 2014, Adelaide Australia.

At least two peer-reviewed scientific publications are planned as a result of this work (putative titles provided):

1. 'Two optimized multiplexes for the parentage assignment and relatedness analysis in Atlantic salmon'
2. 'Assessment of power of a microsatellite multiplex for parentage assignment in Atlantic salmon at an industry wide level using simulations'

Popular-science articles have also been presented on this project:

1. *Vil spore rømt oppdrettslaks*. www.kyst.no 13.03.12

2. *Sporer rømt laks.* Dagens Næringsliv, 03.03.12

3. *Ny test for DNA-sporing av rømt oppdrettslaks*, written by Reidun Lilleholt Kraugerud www.Intrafish.no 19.12.2013

4. *Scientists reveal new DNA test to track escaped salmon*. (English translation of the above article) www.Intrafish.com 02.01.2014

5.3 Further dissemination of results and areas of further study

Application of a DNA based tracing method, be it microsatellite or SNP based, will require a close cooperation with industry and governing bodies. In all likelihood a large commercial laboratory will need to be contracted to perform DNA extraction and genotyping. If desired, the project team will cooperate with a commercial provider in order to implement the findings of this project.

6 Quality control of project and results

The project group has had regular meetings both internally and with the collaborating project group to monitor progress and coordinate common activities. Alterations to the original project plan were made to ensure the overall goals were reached, and when appropriate this was performed in cooperation with project 900706.

7 References

Claassen, S., du Toit, E., Kaba, M., Moodley, C., Zar, H.J., Nicol, M.P., 2013. A comparison of the efficiency of five different commercial DNA extraction kits for extraction of DNA from faecal samples. Journal of Microbiological Methods. 94, 103-110.

Fahle, G.A., Fischer, S.H., 2000. Comparison of Six Commercial DNA Extraction Kits for Recovery of Cytomegalovirus DNA from Spiked Human Specimens Journal of Clinical Microbiology. 38, 3860-3863.

Glover, K., Skilbrei, O., Skaala, O., 2008. Genetic assignment identifies farm of origin for Atlantic salmon (*Salmo salar* L.) escapees in a Norwegian fjord. ICES Journal of Marine Science. 65, 1 - 9.

Glover, K.A., 2010. Forensic identification of fish farm escapees: The Norwegian experience. Aquaculture Environment Interactions. 1, 1-10.

Hayes B.J., Sonesson A.K., Gjerde B., 2005. Evaluation of three strategies using DNA markers for traceability in aquaculture species. Aquaculture, 250: 70.

Håstein, T., Hill, B.J., Berthe, F., Lightner, D.V., 2001. Traceability of aquatic animals. Revue scientifique et technique (International Office of Epizootics). 20, 564-583.

Kalinowski, S.T., Taper, M.L., Marshall, T.C., 2007. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Molecular Ecology. 16, 1099-1106.

King, T.L., Eackles, M.S., Letcher, B.H., 2005. Microsatellite DNA markers for the study of Atlantic salmon (*Salmo salar*) kinship, population structure, and mixed-fishery analyses. Molecular Ecology Notes. 5, 130-132.

Miller, S.A., Dykes, D.D., Polesky, H.F., 1988. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Research. 16, 1215.

Paterson, S., Piertney, S., Knox, D., Gilbey, J., Verspoor, E., 2004. Characterization and PCR multiplexing of novel highly variable tetranucleotide Atlantic salmon (*Salmo salar* L.) microsatellites. Molecular Ecology Notes. 4, 160-162.

Smith, K., Diggle, M.A., Clarke, S.C., 2003. Comparison of Commercial DNA Extraction Kits for Extraction of Bacterial Genomic DNA from Whole-Blood Samples. Journal of Clinical Microbiology. 41, 2440-2443.

Yue, G.H., Orban, L., 2001. A simple and affordable method for high-throughput DNA extraction from animal tissues for polymerase chain reaction. Electrophoresis. 26, 3081–3083

Appendix 1: Microsatellite multiplex information

*Table 1 Primer sequences and dyes used in both MP10 and MP11; * redesigned primers from SsaA124 (King et al., 2005); ** Sssp2216 primers described in Paterson et al (2004).*

	Forward primer	Reverse primer	Dye
MP10			
SAL-CIG-32	ACCAGCCTGCTTAGCTCCTCTCAGC	TGGCATCAACCTGATTCCTGGTCTACGG	6-FAM
SAL-ICISG-11	ACCCACAGCTTTGGTGTTGTCAGC	GCGTTGGGTTGTGTAACAGGCTGC	6-FAM
SAL-ICISG-37	TCTCACAACACCCTGCCTGTGCCA	AGGACACCTGGCTGTCGTCTGTGT	6-FAM
SAL-CIG-33	TGCCTAGGGACAAGCCTTTGCCAC	TGGCACCAGGAGTGGTTGAGTTTGC	VIC
SsaA124-low*	GCACCTGACTTCTATTCCAGTAGCGCA	GCAGCAGGGCCAGTGGGTTTAACT	VIC
Sssp2216**	GGCCCAGACAGATAAACAAACACGC	GCCAACAGCAGCATCTACACCCAG	VIC
SAL-ICISG-16	GCAGTGTGCCGCACTTCAACAAGC	GGTGCTGGAACGAAAGACAGACAGAGG	VIC
SAL-ICISG-05	ACGCCAGACATAAACCGTCCCAAGT	GGGTGTGGGTTGACTTGAGTTGCTG	NED
SAL-ICISG-01	TGTGTCATGTGTTGTGTCACCTGTGC	TGCCAGAAGTGATCCGGGCCTTCA	NED
SAL-CIG-35	ACACCGGCATAGCATGTTGCTCCA	TGTGGCCTCTGGGTCAATCACAGGT	NED
SAL-CIG-37	TTCCTTCCCAGGCGGTAGCTGAAGG	ACTCCTTCTCTTTGCTGGCACCCG	PET
SAL-ICISG-06	CCACCAACCCACCAGGTGTTGAACC	TGCCCACATGGACCAGACCAGCTT	PET
MP11			
SAL-CIG-29	CCGAGCCAGCTGACTGAACCTGGAT	AGGGCCATTACAGAGTCCCACTGGC	6-FAM
SAL-CIG-28	AGTCCGCATACAGGCCCTGGTGAA	TGCTGTTAGTTAGCCACCTTTGTTGGCA	6-FAM
SAL-ICISG-19	AGGCAAAGGGTTACAGAGGTGAGCG	ACAGAAAGGACAACGCCCGGGTCA	6-FAM
SAL-CIG-38	GGTCATTAGCCAAGCAGCTCCACTGT	TTTAGGTTGGGCTGTGCTGTGGCT	6-FAM
SAL-CIG-31	AGCACATGCACTCTTCCTTCGGCT	ATCAGAGCCCTACTGAGCCAGCCA	VIC
SAL-CIG-27	CGCAGACACGCATGGACTCAGACA	AGTCGTGTTGGGAGATGGGAGTGCC	NED
SAL-ICISG-10	CCTGTCCAACAGGGTAAGAGGTCAGGA	ACTGTGCCATGCATCTGACCAGCC	NED
SAL-ICISG-38	TGTGTCCTGATGCCCGATGCCTGA	TCTCATTCTCCCGGTGTTGCTCTACG	PET
SAL-ICISG-07	GGCCGGTTGCTATGTTGCCAAGCTAT	TCTGCTCCACAAGGGACTGCTGCT	PET
SAL-CIG-21	CGTGGTGTTCTGTGATCCATCGTGGG	AGCCCTGTTTCCCGTGACATGGTG	PET

Appendix 2: PCR reagent volumes and ratio of primers in MP10 and MP11.

Table 2 *Volumes of reagents used in a high-fidelity hot-start Q5 PCR reaction are given below, volumes are given per sample and the primer mix volume is relative to the concentration factor in Table 3.*

Reagent	Volume (µL)
Q5 2x High-Fidelity Master Mix	2.5
Primer mix	0.025
dH20	1.475
DNA	1
TOTAL volume	5

Table 3 *Concentration factors for primers in MP10 and MP11, the working primer concentration for the following calculations is 25 µM*

Primer	**concentration factor**
MP10	
SAL-CIG-32	0.5
SAL-ICISG-11	1.1
SAL-ICISG-37	2.0
SsaA124-low	0.5
SSsp2216	1.2
SAL-CIG-33	0.6
SAL-ICISG-16	1.8
SAL-ICISG-01	0.8
SAL-ICISG-05	0.7
SAL-CIG-35	0.7
SAL-ICISG-06	0.4
SAL-CIG-37	1.5
MP11	
SAL-ICISG-19	0.9
SAL-CIG-28	2.5
SAL-CIG-29	0.25
SAL-ICISG-10	3.0
SAL-CIG-31	1.5
SAL-CIG-27	0.2
SAL-CIG-38	1.2
SAL-ICISG-07	1.5
SAL-CIG-21	1.0
SAL-ICISG-38	2.0

Appendix 3 – WP3 simulation study 1

Report from Ben Hayes, DEPI Australia.

Simulation studies to assess accuracy of parentage assignment and accuracy of distinguishing farmed from wild fish with a panel of 12 microsatellites

Aim

The aim of this study was to determine the value of a panel of 12 microsatellites, with known allele frequencies in a variety of populations, for parentage assignment and distinguishing wild from farmed fish. The panel was evaluated under a range of scenarios, including different genotyping error rates, and different levels of relatedness among individuals in the farmed population.

Methods

The number of alleles and frequencies of these alleles at each of the 12 microsatellites were the averages for these microsatellites across several farmed populations. An *in-silico* base population was created by sampling alleles according to these allele frequencies for 3000 fish. In subsequent generations, offspring were created by sampling a sire and a dam, then randomly sampling either the paternal allele or the maternal allele for the sire as the paternal allele of the progeny, and likewise for the maternal allele sampled from the dam.

For the farmed fish population (1500 individuals), a half sib structure was simulated, with one male mating with two dams. Starting with the same base population, a wild population was also generation by selecting parents at random to generate a population of 1500 fish. After the base population, there was no gene flow between the farmed and wild populations. Twelve generations of breeding were simulated. In the default simulation, for the farmed population 150 females were mated to 75 males, with 10 offspring per female. These parameters were chosen for the default because they gave an average relatedness among individuals that was similar to real farmed populations (Jacq, pers comm). The simulation program is described in more detail in Hayes (2005).

In the final generation, offspring were to the farmed or wild population, and if assigned to the farmed population, to parents. The probabilities that the fish came from any of the possible pairs of parents was calculated, following Letcher and King (2001). For each marker, the probability that an offspring with the genotype A_iA_j is derived from parents with genotype A_aA_b and AcAd, was: $Pr\{[A_iA_j]|(A_aA_b).(A_cA_d)\}=T(i|ab)T(j|cd)+T(j|ab)T(i|cd)$ where $T(i|ab)=Pr([A_i]|(A_aA_b),(A_cA_d))=½(a=i)+½(b=i)$, and (a=i) and (b=i) are Boolean operator that give the value of one if the allele value of a equals the allele value of i, or zero otherwise. If the offspring was a homozygote $Pr([A_iA_j]|(A_aA_b),(A_cA_d))$ is divided by two. The global likelihood for the offspring conditional on the parental pair is the product of all single locus likelihoods. The most likely parents are the pair with the highest highest global likelihood. If all the global likelihoods are zero, the fish was considered to be of wild origin.

The effect of genotyping error on the accuracy of parentage assignment was investigated. The most common genotyping error is to call a heterozygote as one of the homozygotes. This was simulated by setting a randomly chosen heterozygote genotype in the offspring to a homozygote of one of the alleles the offspring carried, if $x<Pr$, where x is a number sampled from a uniform distribution (0-1)

and Pr is the rate of genotype error. Genotyping errors from 1% up to 5% were simulated, though this upper level is perhaps unlikely in practise.

The effect of inbreeding on accuracy of assignment, and proportion correctly assigned to farmed or wild populations, was investigated by decreasing the number of parents in the farmed population and increasing the number of progeny. The ratio of males to females was always 1:2. The degree of relatedness in the population was assessed with the COANCESTRY program using Dyadic maximum likelihood (Wang 2011).

All results are from five replicate simulations.

Results

The accuracy of assignment to parents was 99 % when all 12 markers on the panel were used, Figure 1, with a rapid decay in accuracy as the number of markers was reduced. The accuracy of assigning individuals to populations (eg farmed or wild), was 100 % provided all 12 markers were used.

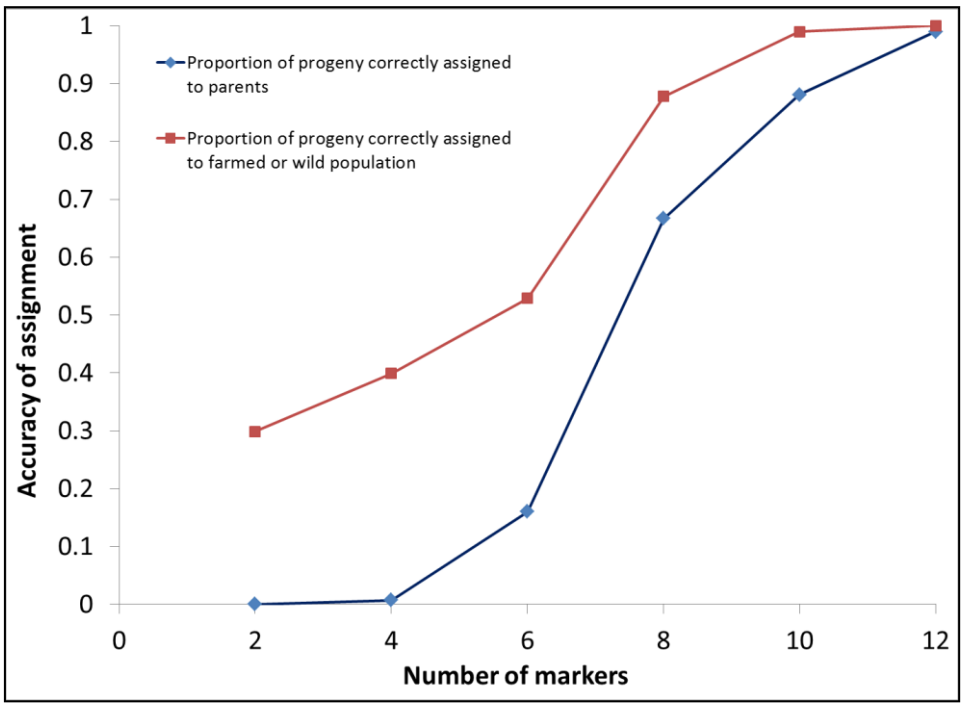

Figure 1 *Accuracy of assigning individuals to parents and accuracy of assigning individuals to populations (farmed or wild) with an increasing number of markers. In both cases accuracy of assignment is the proportion of individuals correctly assigned.*

All subsequent results are based on the full 12 marker panel.

Genotyping errors reduced both the accuracy of parentage assignment in the farmed population, and the accuracy of assigning individuals to either the farmed or wild populations, Figure 2. The effect of genotyping error was relatively small when the error rate was 2 % or less, however accuracy of parentage assignment decreased by 6.5 % to 93.5 % when genotyping error rates were 5 %. The effect on accuracy of assignment of individuals to wild or farmed populations was negligible.

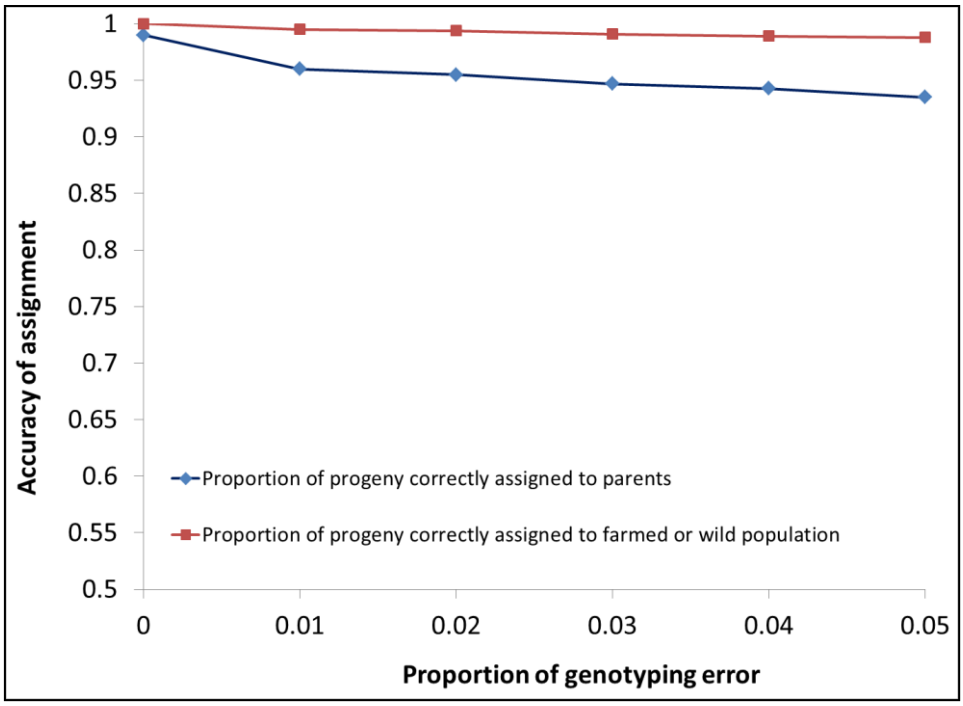

Figure 2 *The effect of genotyping error on accuracy of parentage assignment in the farmed population and accuracy of assignment of individuals to either the wild or farmed populations.*

The effect of the level of relatedness in the farmed population on the accuracy of parentage assignment, and assignment of individuals to the farmed or wild populations was also assessed. High levels of relatedness will decrease the accuracy of parentage assignment, as genotypes of potential parents will be more likely to be the same or similar. In fact the average relatedness of the population (reflecting the level of population inbreeding) will be a key parameter determining the accuracy of parentage assignment (or alternatively how many markers are required to achieve accurate assignment). We have estimates of average relatedness in real farmed populations from Jacq (pers comm) using the 12 marker panel. These were 0.071 for Ra and MH populations, and 0.058 for the AG population, estimated using the COANCESTRY program (Wang 2011). Our default simulation gave levels of relatedness slightly above this, at 0.09. So 99 % (Figure 1) is likely a conservative estimate of the level of inbreeding. At a 0.05 level of average relatedness in the simulated population, the accuracy of parentage assignment was very close to 100 %, Figure 3. At higher levels of relatedness, accuracy of assignment declined rapidly however.

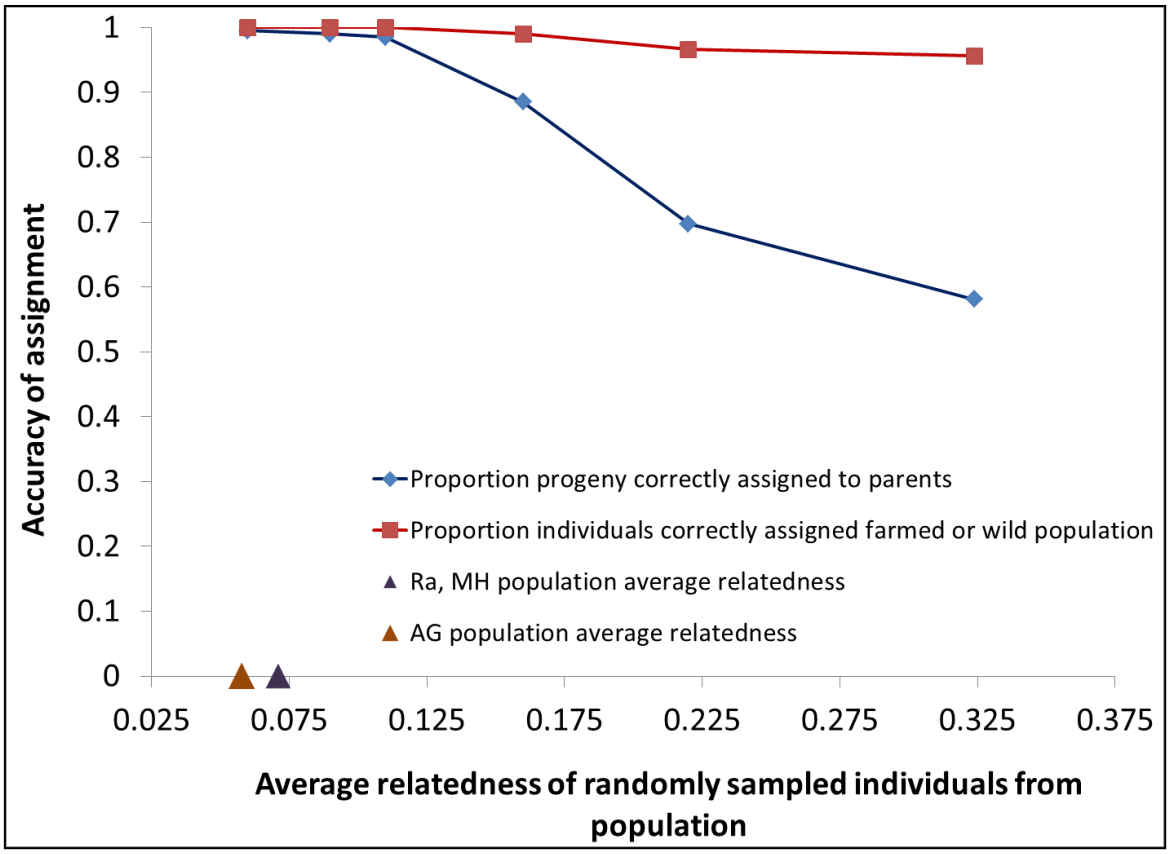

Figure 3 *Effect of level of relatedness on accuracy of parentage assignment and accuracy of assigning individuals to the farmed or wild populations.*

Conclusions

Our simulation results, using the number of alleles and allele frequencies for the 12 marker panel from real populations as a starting point, suggest close to 100 % accuracy for both parentage assignment and assignment individuals to the wild or to farmed populations can be achieved. This was at a level of relatedness, a key factor affecting assignment accuracy, for the simulated population that was close to several real farmed populations.

Genotyping errors will decrease assignment accuracy, as will increased relatedness of the farmed population (eg fewer parents used). However, within bound for these parameters that are likely to be observed in practice (eg <5 % genotyping error, 0.09 level of relatedness), high accuracies of assignment can still be achieved.

The 12 microsatellite marker will be a useful tool for parentage assignment and traceability applications.

References

Wang J, 2011. coancestry: A program for simulating, estimating and analysing relatedness and inbreeding coefficients. Molecular Ecology Resources, 11: 141–145.

Hayes BJ, Sonesson AK, Gjerde B, 2005. Evaluation of three strategies using DNA markers for traceability in aquaculture species. Aquaculture, 250: 70.

Appendix 4 – WP3 simulation study 2

Salmon tracing: Genotyping to trace back escapees from salmon aquaculture

R.J.W. Blonk

Report number ~nummer~*If it is a confidential report then mention here* [confidential]

~Foto (aan te leveren door projectleider)~

IMARES Wageningen UR

(IMARES - Institute for Marine Resources & Ecosystem Studies)

Client: Nofima AS
P.O. Box 6122
NO-9291 Tromsø, Norway

Publication date: 31-01-2014

IMARES is:

- an independent, objective and authoritative institute that provides knowledge necessary for an integrated sustainable protection, exploitation and spatial use of the sea and coastal zones;
- an institute that provides knowledge necessary for an integrated sustainable protection, exploitation and spatial use of the sea and coastal zones;
- a key, proactive player in national and international marine networks (including ICES and EFARO).

P.O. Box 68
1970 AB IJmuiden
Phone: +31 (0)317 48 09 00
Fax: +31 (0)317 48 73 26
E-Mail: imares@wur.nl
www.imares.wur.nl

P.O. Box 77
4400 AB Yerseke
Phone: +31 (0)317 48 09 00
Fax: +31 (0)317 48 73 59
E-Mail: imares@wur.nl
www.imares.wur.nl

P.O. Box 57
1780 AB Den Helder
Phone: +31 (0)317 48 09 00
Fax: +31 (0)223 63 06 87
E-Mail: imares@wur.nl
www.imares.wur.nl

P.O. Box 167
1790 AD Den Burg Texel
Phone: +31 (0)317 48 09 00
Fax: +31 (0)317 48 73 62
E-Mail: imares@wur.nl
www.imares.wur.nl

A_4_3_2-V13.3

Contents

Summary

The overall objective of the project is to assign an escaped salmon back to the farm responsible for the escape with near 100% accuracy. In this report, the potential of a set of genetic markers to assign an escaped salmon was determined for a set of 12 polymorphic microsatellite markers, provided by Nofima, and by using stochastic simulation. Also, the effect of different numbers of sires, and the effect of pooling of multiple sires in crosses was determined.

The effect of the number of sires included was as expected with less sires resulting in lower allocation power. However, for the currently believed common numbers of sires and dams used for production stocks this still resulted in very high allocation power. The effect of pooling of 3 sires in one cross was small, thus leaving room for the salmon breeding companies to put higher selection intensities in their breeding program.

In general, it can be concluded that, based on the genetic data provided, the current set of polymorphic microsatellite markers is enough to trace back most individual salmon back to their farm of origin, assuming that for each farm, the crosses provided are known, and that one cross is only provided to one farm. However, to be 100% accurate, the set of markers needs to be enlarged for ambiguously allocated individuals or a combination of genetic markers and a phenotypic marker such as a tag could be considered.

The Sub delivery provided described in this report is linked to WP3 in the execution of the project "Industry-wide tracing of Norwegian farmed Atlantic salmon".

Introduction

Parental allocation methods using genetic markers and software programs to perform parental allocation have been extensively described in literature (Duchesne *et al.*, 2002; Marshall *et al.*, 1998; Wang, 2004). These programs typically use genotypes of individual offspring and putative parents to reconstruct pedigrees. The success of these methods strongly depends on the quality of the genetic markers as determined by potential presence of null alleles, heterozygosity of the markers in the population, number of alleles and allele frequencies. To test the allocation power of a set of markers, most software programs provide options for simulation of populations, based on provided information on the marker quality. However, the currently available programs are not able to simulate large populations, such as true commercial population sizes, and specific breeding structures.

The overall objective of the project is to assign an escaped salmon back to the farm or company responsible for the escape with near 100% accuracy. In this report, the potential of genetic markers to assign an escaped salmon back to its original farm was assessed using basic population genetic data from a part of the commercial salmon breeding population in Norway. To assess the potential of genetic markers in a realistic situation, the power to allocate an individual back to both its parents (and thus to its farm of origin) was determined for a set of 12 polymorphic microsatellite markers in a simulated population nearing the realistic commercial size in Norway and its underlying breeding structures. Also, the effect of different numbers of sires, and the effect of pooling of multiple sires in crosses was determined.

Materials and Methods

Information of the markers used

Information on the markers was provided by NOFIMA and based on a part of the current commercial breeding population of Atlantic salmon (*Salmo salar*) in Norway (table 1).

Table 1 Locus name, number of alleles and observed heterozygosity in a part of the current commercial breeding population of Atlantic salmon (Salmo salar) in Norway.

Locus name	Na	Het obs
SAL_CIG_32	22	0.8432
SAL_ICISG_11	16	0.7947
SAL_ICISG_37	15	0.6571
SsaA124-low	11	0.5307
Sssp2216	16	0.8476
SAL_CIG_33	24	0.8287
SAL_ICISG_16	24	0.8771
SAL_ICISG_01	9	0.7591
SAL_ICISG_05	17	0.7731
SAL_CIG_35	29	0.8363
SAL_ICISG_06	10	0.8008
SAL_CIG_37	16	0.8631

Stochastic simulation of populations

To determine allocation success of the selected set of markers (table 1), a stochastic simulation program was written in R (R Development Core Team, 2008). In this program, given the marker information provided, random genotypes were generated for a defined number of sires and dams. To define population structure, different mating schemes were defined, including pooling of 1 or 3 males with one female.

Allocation power

To determine the allocation power of a set of markers within a given population size and structure, *matching of parental alleles in pairwise compared crosses of parents* was scored for each locus. For

example, consider a 1 locus model where 2 crosses are compared. Here, alleles *a* in cross 1 are represented by a_i for sire *i* and by a_k, for dam *k,* where alleles *a* in cross 2 are represented by a_j for sire *j* and by a_l for dam *l*. Matching of one or more alleles was scored as $I(a_i,a_j) + I(a_k,a_l)$ and as $I(a_i,a_l) + I(a_k,a_j)$. Here, *I* is the identity between alleles, and *I* was set 1 when one or more of the compared alleles are identical, and zero otherwise. Cases with at least one comparison equal to 2 were considered as one match for this particular locus.

Moreover, to obtain the total matching of loci in a pairwise compared cross, matching of parental alleles was summed for all loci and from this the allocation power of the set of markers follows. For example: in a situation with 12 tested loci and where two pairwise compared crosses of parents have 0 matching loci (i.e. 0 matches), these particular crosses will always produce offspring with different genotypes. In contrast, when two pairwise compared crosses of parents have 12 matches (a "full match"), these particular crosses can both produce genotypic identical offspring, at least for these markers.

Obviously, when comparing genotypes of offspring from two crosses in which the crossed parents have a full match on all loci, it will not always be possible to determine the true parents (and the farm of origin when individual crosses are only kept at one farm) with 100% certainty for all individual offspring. In contrast, for all pairwise compared crosses with at least one non-matching locus, in principle there will be enough information (unique alleles) to unambiguously allocate individuals to the parents and thus the farm of origin. Therefore this method was considered a good measure for the allocation power of the set of markers in this particular application.

Within each simulated population, the number of matches was calculated for all possible pairwise compared crosses. From this, the frequency (i.e. the number of pairwise compared crosses) for all possible numbers of matches (0 until N_{loci}) in the population was determined. To determine the allocation power of a certain set of markers in a given situation, the frequency of full matches was particularly considered.

Scenario's analysed

So assess effects of the number of loci used, number of sires and dams in the population, the population structure and pooling of males in crosses, several scenarios were simulated (see table *2*). The scenarios with 33.000 dams and 660 sires, i.e. mating one sire on 50 dams, were assumed to represent the current situation of commercial salmon production stocks in Norway, whereas the scenarios with 33.000 dams and 330 sires, i.e. mating one sire on 100 dams, and pooling of 3 males were assumed to represent the potential future situation. Each scenario was replicated 10 times, and means and standard deviations of the frequencies for all potential number of matches were calculated.

In scenario A and C, 17 alleles per marker were used and random heterozygosity was programmed. In scenario B and D, the number of alleles and the heterozygosity was based on a genotyped part of the current commercial breeding population of Atlantic salmon in Norway (data provided by Nofima). In these analyses effects of null alleles and erroneous genotyping are neglected.

Table 2. Simulated scenarios to assess allocation power of genetic markers to assign an escaped salmon back to its original farm.

Scenario	Dams	Sires	Pooling sires	Loci	Na	Het obs
A	1000	50,100	1	3	17	Random
	1000	50,100	1	6	17	Random
	1000	50,100	1	9	17	Random
	1000	50,100	1	12	17	Random
B	1000	50,100,200,500	1	12	Sample	Sample
C	33000	660	1	3,6,9,12	17	Random
D	33000	660	1,3	12	Sample	Sample
	33000	330	1,3	12	Sample	Sample

Pooling sires = the number of sires mated with one female. Na = number of alleles; Het obs = heterozygosities used; Sample = information based on genotyped part of the current commercial breeding population of Atlantic salmon in Norway; For scenarios A and C, the number of alleles was set to 17 (the average number of alleles in the real sampled population;

Results

Distribution of frequencies for all possible numbers of matches in populations of 1000 dams, 50 sires and 12 markers (scenario A) are shown in figure *1*. In this situation, 17 alleles per marker were used and random heterozygosity was programmed. In this situation, most of the compared pairs in the simulated population show 4 to 5 matches, whereas almost no pairs had either 0, 10, 11 or 12 matches.

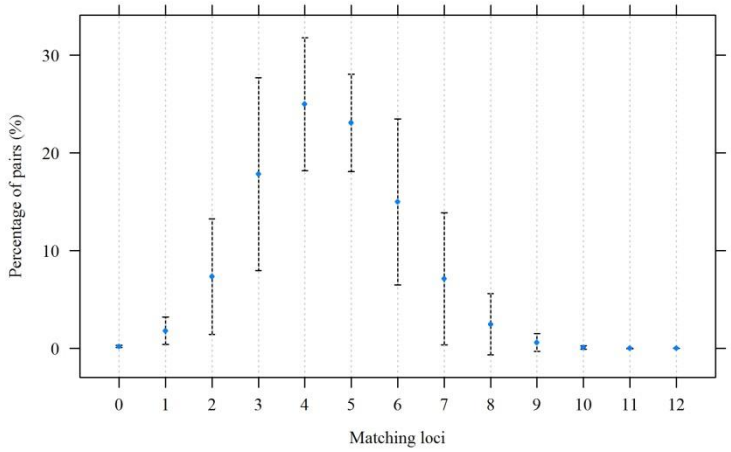

Figure 1. Distribution of frequencies (percentage of all possible pairwise compared crosses) for all possible numbers of matches (matching loci), in a population of 1000 dams, 50 sires and 12 loci, all with 17 alleles and random heterozygosity.(Scenario A)

The effect of the number of used loci on the percentage of full matches in a population of 1000 dams with 50 or 100 sires is shown in figure *2*. In this scenario, 17 alleles and random heterozygosity were programmed (scenario A). From figure *2* it is clear that, although standard deviations are high, most full matches occur with 3 to 6 loci, indicating that both for cases with 50 and 100 sires, such sets of markers are less suitable to accurately allocate fish to farms, at least in the current population size and structure.

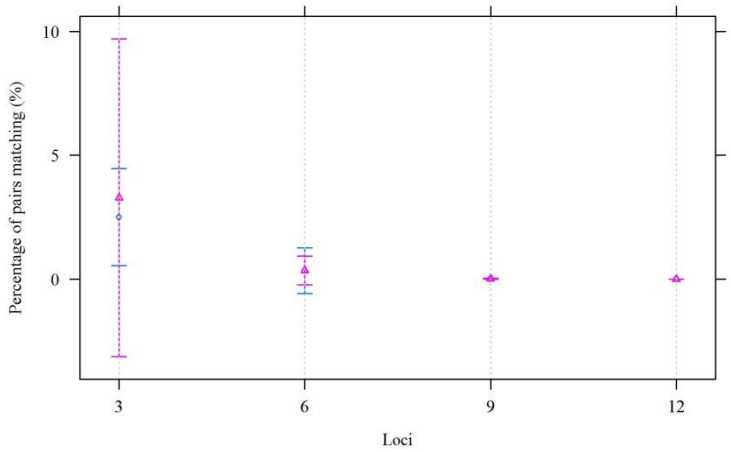

Figure 2. The relation between percentage of full matches and the number of used loci in a population of 1000 dams with 50 or 100 sires, 17 alleles and random heterozygosity (scenario A).

The effect of the number of sires on the percentage of full matches in a population of 1000 dams genotyped for 12 markers with sampled numbers of alleles and heterozygosity (scenario B) is shown in figure 3. From this figure it can be seen that at this population size, relatively small numbers of sires imply higher frequency of full matches, thus corresponding to lower allocation power. This is probably caused by the fact that the same number of dams, using more sires implies production of more families and thus more divergent genotypes, resulting is a smaller chance of matching loci.

When comparing results from scenario A and B, at 50 and 100 sires, the effect of using a realistic marker set can be determined as the marker set as used in scenario B is based on information of a genotyped part of the current commercial breeding population of Atlantic salmon, whereas A is not. When using 50 sires, the percentages of full matching were 0.001502 (sd=0.002095) for a situation with equal numbers of alleles and random heterozygosity (scenario A) whereas for a more realistic situation (scenario B), a much lower value 0.000701 (sd = 0.000380) was found. This implies that the realistic markers set is more informative. However, when using 100 sires, results were much less different with percentages of 0.000350 (sd = 0.000192) for scenario A and percentages of 0.000561 (sd = 0.000363) for B. Very likely the effect of the number of sires and families plays a role in the latter.

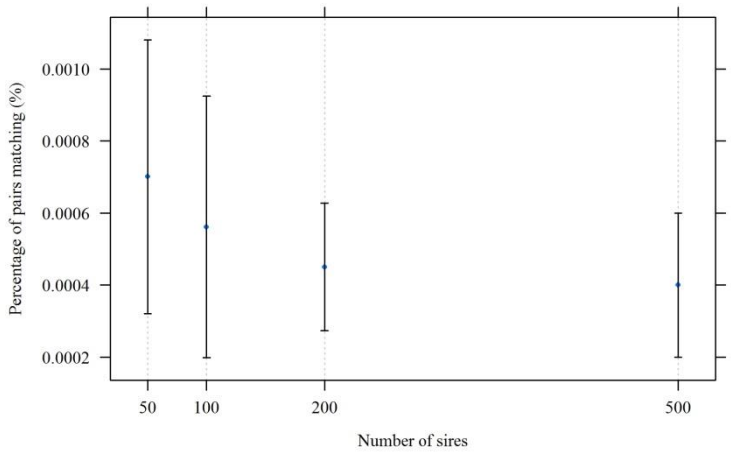

Figure 3. Relation between the percentage of full matches and the number of sires, in a population of 1000 dams, with 12 markers and sampled numbers of alleles and heterozygosity (scenario B).

The effect of the number of used loci on the percentage of full matches in a population of 33,000 dams with 660 sires, i.e. mating one sire on 50 dams, is shown in figure 4. In this scenario, 17 alleles and

random heterozygosity were programmed (scenario C). From figure 4 it can be seen that most full matches occur with 3 loci. This shows that with large realistic population sizes, really more than 6 markers are required to accurately allocate escaped fish to parents and farms. The percentages of full matches where 0.008641 (sd = 0.018309) for 9 loci and 0.000143 (sd = 0.000256) for 12 loci.

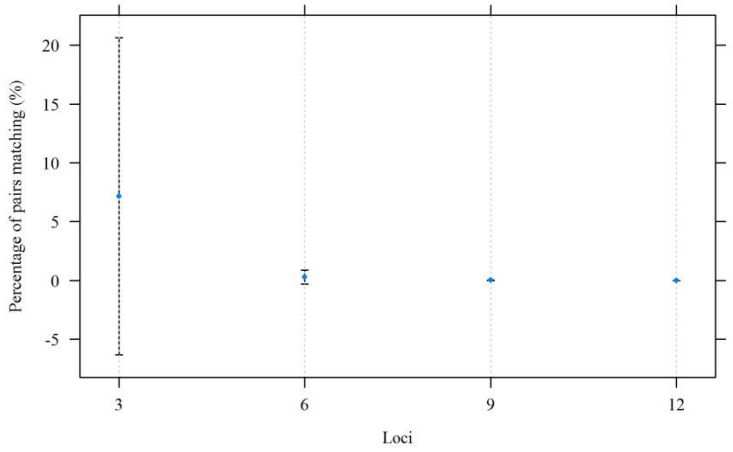

Figure 4. The relation between percentage of full matches and the number of used loci in a population of 33,000 dams with 660 sires, 17 alleles and random heterozygosity (scenario C).

The effect of number of sires in a population of 33,000 dams, and pooling of males per cross (scenario D), can be seen in figure *5*. This is the most realistic scenario. It can be seen that in general, using more sires in the population, increases the allocation power of the set of genetic markers. However, in both cases, percentage of full matches is very low, but dropping from 0.00028 (sd = 0.000026) with the use of 330 sires to 0.000211 (sd = 0.000022) for 660 sires. The effect of pooling of males, i.e. mixing sperm of 3 males and then crossing with 50 or 100 females is very limited, mainly showing its effect in populations with 660 males.

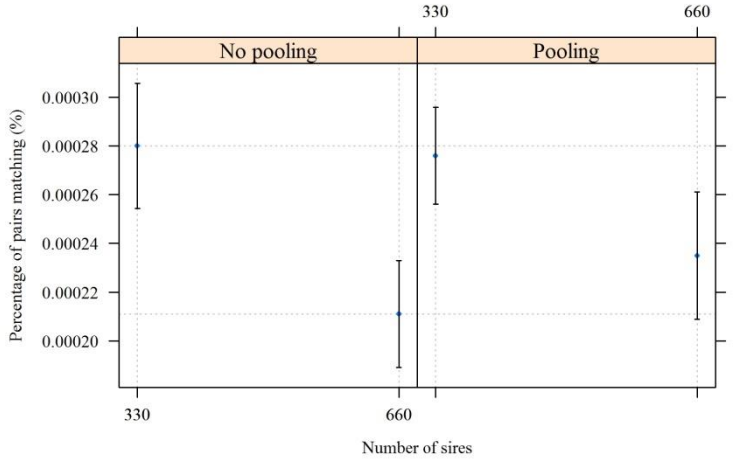

Figure 5. The relation between percentage of full matches and the number of used sires (330 vs 660) in a population of 33,000 dams, with 12 loci and sampled numbers of alleles and heterozygosity (scenario D).

Conclusions

The effect of number of loci included in the set or the number of sires in the population, as well as total population size was as expected with less loci, less sires or smaller populations (i.e. less families) resulting in lower allocation power. However, for common numbers of sires and dams used for production stocks in Norway the proposed set of 12 highly polymorphic markers this still resulted in very high allocation power. The effect of pooling of 3 sires in one cross was small, thus suggesting that there is still room for the salmon breeding companies to put higher selection intensities on their selected stock.

In general, it can be concluded that, based on the genetic data provided, the current set of polymorphic microsatellite markers is able to accurately trace back most individual salmon back to their farm of origin, assuming that for each farm the crosses provided are known, and that one cross is only provided to one farm. However, with a likely scenario where 33,000 dams and 660 sires without pools are used to produce the production stocks in Norway, a percentage of full matches of crosses of 0.000211% can be expected. In practice this means that with 33,000 dams in the population, ((33,000^2-33,000)/2 ≈) 544 $*10^6$ pairwise comparisons can be made, implying approximately 1150 pairs with full matches that produce some offspring with identical genotypes. Thus, in this case one can expect between 1,150 and 2,230 families in a population (i.e. 3-7% of the families produced, see table 3) that include at least some offspring with similar genotypes.

Table 3. Numbers of families (crosses) producing at least some offspring with full matching genotypes with other families, in different scenarios.

Dams	Sires	Pooling	Full matches		Pairwise comparisons	Families with full match			
			mean%	sd		min	max	min%	max%
33000	660	1	0.000211	0.000022	544,483,500	1,151	2,302	3.5%	7.0%
33000	330	1	0.000280	0.000026	544,483,500	1,523	3,046	4.6%	9.2%
33000	660	3	0.000235	0.000026	1,633,450,500	3,835	7,669	11.6%	23.2%
33000	330	3	0.000276	0.000020	1,633,450,500	4,506	9,011	13.7%	27.3%

Pooling = number of sires mated with 1 female.

Should these offspring escape from a farm, it would be difficult to unambiguously allocate them back to the farm of origin. It is therefore suggested that to be 100% accurate, either the set of markers should be enlarged, at least for the animals that ambiguous allocation to parents and farms, and/or a combination of genetic markers and a phenotypic markers such as a tag should be considered.

Recommendations:

1) Full genotypes of all used parents, and mating schemes are required to determine the true power of the set of markers.
2) A larger set of genetic markers is needed to more accurately trace back individuals to their parents and farm of origin.

Quality Assurance

IMARES utilises an ISO 9001:2008 certified quality management system (certificate number: 124296-2012-AQ-NLD-RvA). This certificate is valid until 15 December 2015. The organisation has been certified since 27 February 2001. The certification was issued by DNV Certification B.V. Furthermore, the chemical laboratory of the Fish Division has NEN-EN-ISO/IEC 17025:2005 accreditation for test laboratories with number L097. This accreditation is valid until 1th of April 2017 and was first issued on 27 March 1997. Accreditation was granted by the Council for Accreditation.

References

Duchesne, P., Godbout, M.H., Bernatchez, L., 2002. PAPA (Package for the Analysis of Parental Allocation) : A computer program for simulated and real parental allocation. Molecular Ecology Notes. 2, 191-194.

Marshall, T.C., Slate, J., Kruuk, L.E.B., Pemberton, J.M., 1998. Statistical confidence for likelihood-based paternity inference in natural populations. Molecular Ecology. 7, 639-655.

R Development Core Team, 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, URL http://www.R-project.org.

Wang, J., 2004. Sibship Reconstruction From Genetic Data With Typing Errors. Genetics. 166, 1963-1979.

ISBN 978-82-8296-171-4 (printed)
ISBN 978-82-8296-172-1 (pdf)
ISSN 1890-579X