



# Norges veterinærhøgskole

*Tracing the origin of farmed Atlantic salmon escapees by DNA  
parentage assignment: optimizing methods and real-life validation  
studies*

## **Sluttrapport for FHF prosjektnummer 900706**

Oslo, desember 2013

Prosjektpartnere: Norges veterinærhøgskole (NVH), Centre for Integrative Genetics (CIGENE)  
ved Universitetet for Miljø- og Biovitenskap (UMB), Biobank AS, AquaGen AS og Marelife  
Services AS.

Rapportens forfattere: Randi I Krontveit (NVH), Matthew Peter Kent (CIGENE), Harald Grove  
(CIGENE), Paul Johan Midtlyng (NVH)

## Innhold

Sammendrag .....	3
Summary .....	4
Innledning .....	5
Prosjektets bakgrunn og omfang.....	5
Organisering.....	5
Problemstilling og formål .....	5
Milepæler .....	7
Prosjektgjennomføring.....	8
Arbeidspakker og ansvarlige.....	8
Avvik.....	8
Oppnådde resultater, konklusjon .....	9
WP 1: Metodeoptimalisering .....	12
WP 1.1 (Nofima) Kostnadsoptimalisering og standardisering av DNA-ekstraksjon og konservering.....	12
WP 1.2. (CIGENE) Utvikling av SNP markør panel, genotypeprotokoll og programvare..	13
WP 1.3 (CIGENE) Databaseformat for lagring av SNP-profiler .....	15
WP 2: (Biobank AS) Innsamling, ekstraksjon og distribusjon av DNA til valideringsstudie 1 og 2.....	15
WP 3: (CIGENE) Valideringsstudie 1 (V1): SNP basert tilordning blant foreldre som er full eller halvøsken.....	16
WP 4: (CIGENE) Valideringsstudie 2 (V2): Tilordning ved et stort antall potensielle foreldrefisk.....	17
WP 5: (Nofima) Valideringsstudie 3 (V3): Eksklusjonsstyrke med hensyn på villfisk .....	18
WP 6: (NVH) Prosjektadministrasjon .....	18
Konklusjon.....	18
Leveranser.....	19
Kvalitetssikring av prosjektgjennomføring og resultater.....	19

## Sammendrag

Dette prosjektet var forankret i arbeidet med å utvikle metoder for å skille rømt fisk fra villfisk og spore oppdrettslaks tilbake til anlegget de rømte fra. Formålet med dette prosjektet har dermed vært standardisering av protokoller for DNA konservering og ekstraksjon samt å utvikle et SNP-basert markørpanel for tilordning av mistenkt rømt oppdrettsfisk til sine oppdrettsforeldre. SNP-panelet ble evaluert på tre måter: tilordning når foreldre er hel- eller halvsøsken, tilordning ved mange potensielle foreldre og eksklusjonsstyre med hensyn på villfisk. Standardisering av metoder for DNA ekstraksjon mellom laboratorier og metode for DNA ekstraksjon er viktig for genotypingsresultatet. Et SNP panel med 60 SNPer ble utviklet og viste 97 % vellykket tilordning når foreldre var hel- eller halvsøsken og med 72 % vellykket tilordning til gyldig foreldrepår når det var mange potensielle foreldre. Ingen villfiskprøver ble tilordnet.

Resultatene fra disse test-eksperimentene viser at SNP genetiske markører kan tilordne fisk til sine rettmessige foreldre med stor grad av nøyaktighet. Vi har identifisert en prosedyre for DNA-ekstraksjon og genotyping som er både kostnads- og tidseffektivt.

I praksis betyr dette at en mistenkt rømt oppdrettsfisk kan spores tilbake til opprinnelsepunktet (lokalitet) med svært høy sikkerhet, men implementering av et slikt system vil kreve at to vilkår er oppfylt; (i) den genetiske profilen til alle potensielle foreldre må være registrert i en database, og (ii) produsenter som leverer genetisk materiale (befruktede egg) til lokaliteter må sikre at innen et forhåndsbestemt geografisk område må ikke to lokaliteter få startmateriale fra nøyaktig samme foreldre. Siden produsentene allerede rutinemessig genotyper stamfisken og har full kontroll over sine krysningsskjemaer, mener vi at disse vilkårene kan oppfylles.

## Summary

This project focused on developing methods for distinguishing escaped farmed fish from wild fish and tracing escapees back to their farm of origin. Specific elements within this project have been to standardize protocols for DNA preservation and extraction as well as developing a SNP marker panel for assigning suspected escaped farmed fish to their breeding parents. Starting with a set of 5650 Atlantic salmon SNPs, a subset of 59 was chosen based on their informativeness and physical distribution for use in performing genetic assignment of offspring to parents. Although we chose to continue with this set of 59, there are hundreds of other potential candidate SNPs which could be used for tracing. The SNP panel was tested in three ways; **validation study 1** attempted assignment when the number of offspring exceeded parents and included offspring samples not belonging to any of the provided parents, **validation study 2** included few offspring but many possible parents, and finally **validation study 3** explored the power of the SNP set to excluded wild fish. The project found that standardization of DNA extraction methods between laboratories and the exact method of DNA extraction has an impact on power of assignment. Furthermore, this particular SNP panel demonstrated a good power to assign offspring to parents with high precision and was able to avoid erroneously assigning wild fish.

This project provides a solid proof of principle that SNPs can be used for assignment, however we believe that optimization of the SNP panel, especially the inclusion of additional SNPs, and of the algorithm used for automated assignment is necessary to be able to attain 100% assignment while at the same time allowing for some genotyping error. The practical benefit of such optimization would be that a suspected escaped farmed-fish could be traced to its point of origin (farm) with very high confidence, however implementation of this system would require that two conditions are first met, (i) the genetic profile of all potential parents must be registered within a database, and (ii) producers delivering genetic material (fertilized eggs) to farms must ensure that within a to-be-determined geographic range no two farms receive start material from the exact same parents. Since producers already routinely sample their brood-stock and have complete control over their crossing schemes, we believe that these conditions can be met.

## **Innledning**

### **Prosjektets bakgrunn og omfang**

Et langsiktig mål for norsk akvakulturnæring er å holde antallet rømt oppdrettsfisk på et minimum. I tillegg ønskes det å redusere, og om mulig unngå, negative effekter av rømlinger på villfisk og annen fauna. Som en del av dette arbeidet er det et mål å utvikle kostnadseffektive metoder for å skille rømt fisk fra villfisk og spore oppdrettslaks tilbake til anlegget de rømte fra. Dette prosjektet har evaluert genetisk sporing i form av farskaps- og morskapstesting basert på «single nucleotide polymorfisms» (SNPer).

### **Organisering**

Prosjektet ble gjennomført av et konsortium bestående av Centre for Integrative Genetics (CIGENE) ved Universitetet for Miljø- og Biovitenskap (UMB), Norges veterinærhøgskole (NVH), Biobank AS, AquaGen AS og Marelife Services AS.

Prosjektet var koordinert med et parallelt prosjekt ved Nofima (FHF 900708) som har brukt en alternativ metode (mikrosatellitter) til genetisk sporing. Det har vært felles gjennomføring av flere arbeidspakker.

Prosjektet har hatt en felles styringsgruppe med det parallelle prosjekt ved Nofima (FHF 900708) bestående av:

- Petter Arnesen, Marine Harvest
- Sissel Kjølglum, AquaGen AS
- Vidar Lund, Raumagruppen
- Håvard Bakke, SalmoBreed

Kontaktperson hos FHF har vært Kjell Maroni.

### **Problemstilling og formål**

Hovedmålet med dette prosjektet har vært å forbedre og validere ytelsen til SNP-basert gentesting når det gjelder å spore mistenkt rømt oppdrettsfisk som fanges i fjord eller elv tilbake

til sine biologiske oppdrettsforeldre. Hovedmålet var delt inn i delmål for å dokumentere metodens ytelse og egnethet:

- Beskrivelse av standardisering og kostnadseffektivisering av laboratorieprotokoller for DNA ekstraksjon og lagring
- Seleksjon av informative SNP markører og valg av analysestrategier for kostnadseffektiv tilordning
- Skaffe krevende sett med prøvesett til valideringsstudier
- Validering av SNP-basert tilordning i krevende situasjoner som når foreldre er hel- eller halvsøsken og blant mange potensielle foreldre
- Vurdere eksklusjonsstyrke med hensyn på villfisk

Prosjektet var forventet å bli sentralt for å kunne etablere et DNA-basert sporingssystem for all norsk oppdrettslaks i løpet av prosjektperioden. Den SNP-baserte sporingssystemet har høy sikkerhet og er kostnadseffektiv, tilpasningsdyktig og teknologisk robust. SNP-basert sporing vil være klar til å implementeres på et tidlig tidspunkt i forhold til alternative metoder. Prosjektet antas å kunne bidra til en analyse-infrastruktur og forbedrede teknologiske løsninger som kan

- understøtte bruken av laksegenomet på andre områder innen havbruk
- bidra til økt vitenskapelig presisjon i estimering av både utbredelse og biologisk konsekvens av krysninger mellom vill laks og rømt oppdrettslaks
- tilrettelegge for integrering av prøver fra stamfisk av vill laks og stimulere til samarbeid mellom forskning innen vill laks og akvakultur
- forbedre mulighetene til å overvåke forvaltningen av ville laksestammer

## Milepæler

Milepæler med tilhørende leveranser er angitt i Tabell 1..

**Tabell 1.** Milepæler med tilhørende leveranser

<b>Milepæl med leveranse</b>	<b>Frist</b>	<b>Ansvarlig</b>
Etablering av konsortieavtale	31.01.2012	Marelife
Etablering av prosjektets websider	31.01.2012	Marelife
Rapport om optimalisering av prøvetaking og DNA-ekstraksjon	31.08.2012	CIGENE (Nofima)
Distribusjon av like DNA-alikvoter for valideringsstudie 1 til CIGENE og Nofima	31.03.2012	CIGENE, Biobank AS
Distribusjon av like DNA-alikvoter for valideringsstudie 2 til CIGENE og Nofima	30.09.2012	CIGENE, Biobank AS
Rapport SNP-basert tilordning av avkom fra halv- og helsøsken	30.09.2012	NVH
Sammenlikning av SNP- og mikrosatellittbasert tilordning fra halv- og helsøsken	30.11.2012	NVH
Faglig halvveisrapport	31.12.2012	NVH
Rapport SNP-basert tilordning blant mange foreldrefisk	29.02.2013	NVH
Sammenlikning av SNP- og mikrosatellittbasert tilordning blant mange foreldre	31.05.2013	NVH
Rapport eksklusjonsstyrke ved SNP analyse i forhold til villfisk	31.05.2013	CIGENE
Sammenlikning av SNP- og mikrosatellittbasert eksklusjonsstyrke overfor villfisk	31.08.2013	NVH
Faglig sluttrapport og faktaark	31.12.2013	NVH
Administrativ/økonomisk sluttrapport	31.12.2013	NVH

## Prosjektgjennomføring

For å evaluere genetisk sporing av individer til foreldre ved hjelp av SNPer, ble det anvendt tre prøvesett som simulerer krevende analysesituasjoner som kunne tenkes oppstå i praktisk bruk:

- a) Foreldrefisk som er hel- og halvsøsken
- b) Et stort antall potensielle foreldrefisk
- c) Validering slik at villfisk ikke feilaktig tilordnes oppdrettspopulasjonen

## Arbeidspakker og ansvarlige

Prosjektet var organisert i seks arbeidspakker (WP):

- **WP 1** metodeoptimalisering. Hovedansvarlig Nofima (FHF-prosjekt 900708) med CIGENE og Biobank AS som deltakere
- **WP 2** Innsamling, ekstraksjon og distribusjon av DNA til valideringsstudie 1 og 2. Hovedansvarlig Biobank AS
- **WP 3** Valideringsstudie 1: tilordning blant foreldrefisk som er hel- eller halvsøsken. Hovedansvarlig CIGENE med AquaGen AS, Biobank AS og NVH som deltagere
- **WP 4** Valideringsstudie 2: tilordning ved et stort antall potensielle foreldrefisk. Hovedansvarlig CIGENE med AquaGen AS, Biobank AS og NVH som deltagere
- **WP 5** Valideringsstudie 3: eksklusjonsstyrke med hensyn på villfisk. Hovedansvarlige Nofima og CIGENE med NVH som deltager
- **WP 6** Prosjektadministrasjon. Hovedansvarlig har vært NVH etter overtagelse fra Marelife Services AS.

## Avvik

Uforutsette maskinvareproblemer ved laboratoriene forsinket arbeidet både høsten 2012 og våren 2013 og ga en viss utsettelse av leveransene i forhold til plan og milepæler. Det var også forsinkelser i arbeidet med et parallelt prosjekt om mikrosatellitt-basert sporing som hadde innvirkning på sammenlikningen av disse to prosjektenes resultater. Slutføringen av prosjektet ble imidlertid ikke forsinket. Administrativt ansvar ble overført fra Marelife Services AS til NVH i mars 2012. Det var i tillegg skifte i administrativt ansvarlig personell på NVH. Dette har hatt en viss betydning for kontinuiteten i arbeidet, men vi mener det ikke har påvirket sluttresultatet.



## Oppnådde resultater, konklusjon

Dette prosjektet har søkt å demonstrere at ved hjelp av moderne DNA analyser, er det mulig å skille rømt oppdrettslaks i elv fra villfisk og å spore rømt laks tilbake til produsent. Prosjektet ønsket videre å vise at dette var gjennomførbart både teknisk og økonomisk og i tillegg hadde høy sikkerhet (spesielt hvis resultatene skal anvendes til å plassere rettslig ansvar for en rømming).

Prosjektet startet med identifisering av SNPer (60) som var jevnt representert på alle kromosomene til laks og som samtidig viste et høyt nivå av polymorfisme hos de tre viktigste norske avlsstammene (Mowi, AquaGen, SalmoBreed). Disse markørene utgjør en svært liten del av de hundretusener av spesifikke SNPer i laksegenomet, og valget av disse spesifikke markørene er således ikke det eneste mulige valget av SNP markørsett. Fleksibilitet og overflod av SNPer betyr at alternative markørpaneler relativt lett kan tas i bruk ikke bare for tilordning av fisk, men også for å få viktig informasjon om andre egenskaper hos fisken hvis det var behov for det. Vårt opprinnelige mål var å kombinere det maksimalt mulige antall SNP markører i en enkelt analyse satt til maksimal kostnadseffektivitet og reduserte prosesseringskostnader på laboratoriet. Dette målet ble bare delvis oppnådd, og av tekniske grunner ble SNPene prosessert i to sett som inneholdt 31 og 29 SNPer. I en storskala anvendelse vil en slik dobbelprosessering være uhensiktsmessig av logistiske og økonomiske grunner. Imidlertid tror vi det er en løsning på dette og at prosessering i ett trinn er mulig.

Det er verdt å merke seg at etter dette prosjektet ble igangsatt, har CIGENE og AquaGen AS utviklet nye SNP-ressurser og har i dag tilgang til > 600 000 validerte SNP-analyser. Disse kan lett undersøkes og utnyttes for å identifisere ytterligere markører og enten supplere det eksisterende panelet eller brukes til å designe et helt nytt panel med andre egenskaper i tillegg (f.eks. SNPer i gener, SNPer i QTL etc.).

I tillegg til å optimalisere valget av genetiske markører, er det undersøkt og funnet den mest optimale metoden for konservering av prøver og ekstrahering av DNA egnet for genetisk testing. Kommersielle DNA ekstraksjonssett er kjent for å være robuste og generere utmerket DNA, men kostnadene per prøve kan være så høy som NOK 50,-. Vi har testet tre konserveringsmetoder og to ekstraksjonsmetoder og funnet at konservering av vevsprøver i rødsprit etterfulgt av presipitasjon er en svært god kombinasjon og med en omtrentlig kostnad på NOK 15,- per prøve.

Andre konserveringsmetoder og ekstraksjonsmetoder kan føre til svakere genotypingsresultater og slik redusere styrken til analysen.

Det ble utført tre eksperimenter (V1, V2, V3) for å teste SNP basert tilordning av avkom til foreldre. **V1** skulle tilordne et sett med 520 avkom til 229 foreldre uten kjennskap til sant slektskap, og samtidig ikke tilordne 40 ubeslektede yngel (dvs. negative kontroller). I **V2** skulle relativt få avkom (279) tilordnes et stort sett av potensielle foreldre (496). I **V3** undersøkte vi om 96 prøver fra villaks (som representerer 15 forskjellige elver i Norge) ville bli feilaktig tilordnet noen av de genotypedede foreldrene i V1. I både V1 og V2 ble prøvene først tilordnet foreldre basert utelukkende på genotypingsdata. Deretter ble informasjon i form av kjente krysninger inkludert for å løse situasjoner der det var mer enn en «gyldig» foreldrepar (dvs. ugyldige krysninger ble identifisert). Hoveddelen (79 %) av avkommene i V1 ble tilordnet entydig til to foreldre uten den ekstra krysningsinformasjonen. Tilordningsprosenten økte til 86 % ved inkludering av krysningsinformasjon. Avkom som ikke ble tilordnet omfattet gruppen negative kontroller (8 %), prøver med et utilstrekkelig antall genotyper (2,8 %) og prøver som ikke samsvarte med noen av de oppgitte krysningene (3 %). Validering av de unike trioer (mor, far, avkom) viste en nøyaktighet på 98,4 % når tilordningen ble sjekket mot stamtavleinformasjon («fasit»). For V2 finnes ingen «fasit» i så måte at et individ kan tilordnes et bestemt foreldrepar, men det foreligger informasjon om gyldige krysninger (dvs. gyldige foreldrepar). Når krysningsinformasjon ble lagt til, ble 86,4 % av avkommene tilordnet et gyldig foreldrepar.

Et søsterprosjekt som undersøker nytten av mikrosatellitter (FHF prosjekt 900708) for genetisk tilordning av fisk har fått nesten identisk tilordning ved bruk av nøyaktig samme prøvesett.

Konklusjonen er dermed at presisjonen for V2 vil være høy. V3 søkte å vurdere om det aktuelle SNP-panelet som syntes korrekt å tilordne oppdrettsfisk til sine foreldre, ville ukorrekt tilordne villfisk til kommersiell stamfisk. En tidligere artikkel (Karlsson et al, 2011, Mol Ecol Res, 11 (1), 247-253) hvor deltakere i dette prosjektet er medforfattere, viste at et lite sett med SNP markører hadde potensiale for å differensiere villfisk fra oppdrettsfisk med høy sikkerhet. Markørene i dette settet ble imidlertid valgt ut spesielt for dette formålet, og en slik differensiering var ikke et av nøkkelutvalgskriteriene i vårt SNP-panel design. Likevel fant vi at ved å bruke de samme eksklusjonsprinsippene for villfisk som for oppdrettsfisk for å eliminere falske positive tilordninger, ble ingen villfisk tilordnet foreldre i våre datasett. Analysene ble

gjort på et relativt lite sett med prøver ( $n = 96$ ), men resultatene antyder likevel at uten noen ytterligere informasjon (f.eks. merking) er det mulig å skille villfisk fra oppdrettsfisk og samtidig tilordne oppdrettsfisken til riktige foreldre.

Fra et industrisynspunkt tror vi at genetiske markører i form av SNPer representerer et godt alternativ for å spore antatt rømt laks tilbake til opprinnelsessted. De viktigste årsakene til dette er:

- (i) Det er fiskens naturlige DNA som brukes som et «fingeravtrykk» og dermed er det ikke nødvendig å introdusere kunstige tilsetninger eller sporstoffer i dietten og en unngår problemer i forhold til mattrygghet.
- (ii) En DNA-prøve kan tas enkelt ved å ta en liten finnebiopsi som ikke har noen langvarig negativ innvirkning på fisken og som lett kan sendes til et laboratorium for analyse.
- (iii) Det er ikke nødvendig å ta prøver av eller behandle hver enkelt fisk i produksjon, men av de fiskene som brukes som foreldre og av mistenkte rømlinger.
- (iv) Sammensetningen av det SNP-panelet som brukes for tilordning kan tilpasses slik at det også omfatter SNPer assosiert med spesielle produksjonsegenskaper og/eller genetisk introgresjon. Dette betyr at det er mulig å gjøre tilordning og samtidig få interessante vitenskapelige data som understøtter og letter industriens bruk av ressursene innen laksegenomet.
- (v) De fleste av de større kommersielle produsentene av genetisk materiale foretar allerede prøvetaking av stamfiskmaterialet i avlsarbeidet og genotyping av foreldre og opparbeidelse av den nødvendige databasen vil kunne gjennomføres raskt.

Den største fordel ved genetisk tilordning er at hendelser med rømming kan spores tilbake til lokaliteten hvor rømmingen skjedde. Det er imidlertid også andre nyttevirkninger av genetisk sporing. For det første har tilordningen en høy grad av rettsmedisinsk presisjon som betyr at det er svært liten sjanse for at en uskyldig blir holdt ansvarlig for en rømming. For det andre gir denne typen sporing fleksibilitet slik at de genetiske markørene vi foreslår kan velges slik at den gir ytterligere vitenskapelig informasjon. Dette vil ha størst relevans for stamfiskprøver hvor jo alle av nødvendighet blir genotypet for å bygge opp en foreldredatabase.

Prosjektet var organisert i seks arbeidspakker (WP) og resultater fra hver enkelt WP følger nedenfor.

## **WP 1: Metodeoptimalisering**

Hovedansvarlig i denne arbeidspakken har vært Nofima (FHF-prosjekt 900708) med CIGENE og Biobank AS som deltakere.

### **WP 1.1 (Nofima) Kostnadsoptimalisering og standardisering av DNA-ekstraksjon og konservering**

Diskusjoner med avlselskapene har ført til at punchbiopsi fra huden på rygg eller finne i størrelse 3 mm i diameter anses hensiktsmessig i praksis, og slike prøver er teknisk enkle å håndtere på laboratoriet. Fra villfisk har skjell (med påsittende hudceller) vært brukt som prøvemateriale. Det har vært gjort arbeid med å vurdere ulike DNA ekstraheringsmetoder i forhold til prosesseringstid, pris og kvalitet. Det har vært møter med "OS ID" (<http://www.osid.no/16.OS-ID-Startsiden.html>) og "Fluidx" (<http://www.fluidx.eu/>) angående utstyr til prøvetaking og lagring av prøver, og flere gode kandidater er identifisert for vurdering på laboratoriet og hos stamfiskprodusentene.

Et multifaktorielt design for lagring av prøver og DNA-ekstraksjon ble utformet. Dette inkluderte vevsprøver fra et lite antall fisk og inkluderte prøver fra bukfinne, fettfinne og skjell. Biologiske replikater fra hver prøve ble frosset eller konservert ved romtemperatur med etanol eller rødsprit. Prøvene ble distribuert til tre ulike laboratorier hvor DNA ble ekstrahert ved hjelp av enten Chelex ([http://en.wikipedia.org/wiki/Chelex\\_100](http://en.wikipedia.org/wiki/Chelex_100)) eller etanolpresipitasjon ([http://en.wikipedia.org/wiki/Ethanol\\_precipitation](http://en.wikipedia.org/wiki/Ethanol_precipitation)). Etter ekstraksjon ble DNA returnert til CIGENE og genotypet ved hjelp av det 60-plex SNP panelet som ble utviklet i dette prosjektet. Hver prøve og behandling fikk en score 0-60 basert på antall avlesbare genotyper. Prøver med lav score, som gjenspeiler få lesbare genotyper, har per definisjon dårlig kvalitet i forhold til de med høyere score. Dataene ble analysert med en generalisert lineær regresjonsmodell med gammafordeling. Vevstype (skjell, bukfinne, fettfinne) og konserveringsmetode (frost, etanol, rødsprit) gjorde ingen forskjell for genotype score, mens både ekstraksjonsmetode og laboratorium hadde en statistisk signifikant effekt (Tabell 2).

**Tabell 2.** Effekt av vevstype, konserveringsmetode, DNA-ekstraksjonsmetode og laboratorium på antall genotyper (genotype score)

Variabel	Effekt	p-verdi
Vevsprøve (skjell, bukfinne, fettfinne)	Ingen signifikant endring i antall genotyper ved de ulike vevstypene	0,7
Konservering (frysing, etanol, rødsprit)	Ingen signifikant endring i antall genotyper ved de ulike konserveringsmetodene	0,8
DNA ekstraksjonsmetode (Chelex, etanol presipitasjon)	Presipitasjon gav i gjennomsnitt 3,7 flere lesbare genotyper sammenliknet med Chelex	0,018
Laboratorium (A, B, C)	Laboratorium A og C produserte 7 - 8 flere lesbare genotyper enn laboratorium B	0,001

Dette resultatet indikerer at presipitasjon bør velges som metode for DNA-ekstraksjon, foran Chelex. Chelex er en hurtig og kostnadseffektiv metode, men gir grove DNA preparater med høyt forurensningsnivå. Resultatet indikerer videre at standardisering og kvalitetssikring av protokollen for DNA ekstraksjon er nødvendig for å redusere variasjon i kvalitet, særlig dersom flere laboratorier skal utføre slike analyser parallelt.

### **WP 1.2. (CIGENE) Utvikling av SNP markør panel, genotypeprotokoll og programvare**

For å identifisere mest mulig informative SNP-markører til sporing ble 756 laks fra de mest vanlige oppdrettspopulasjonene i Norge (Aqua Gen, SalmoBreed og Mowi (Marine Harvest)) genotypet med en SNP-chip utviklet hos CIGENE. Denne chipen inneholder mer enn 5000 SNPer. Utvalgskriteriene for “sporing-SNPer” var høy frekvens av begge alleler ( $MAF > 0,45$ ) i alle tre populasjonene samt en god fordeling over hele genomet (3 - 5 SNPer per kromosom). I tillegg ble det valgt ut 3 SNP markører fra det mitokondrielle genomet. Til sammen ble 114 SNPer identifisert som potensielle kandidater for inkludering i et SNP basert sporingspanel.

Genotyping av sporingspanelet ble utviklet på teknologiplattformen utviklet av Sequenom ([www.sequenom.com](http://www.sequenom.com)). Dette er en teknologi basert på PCR-amplifikasjon pluss minisekvensering og deteksjon ved hjelp av massespektroskopi (MALDI-TOF). Nyutvikling i teknologien har økt plattformens deteksjonsvindu slik at man nå har større mulighet til å kombinere mange (>40) SNPer i et og samme design. For å minimere genotypingskostnader ble

det prioritert å evaluere én multipleks-protokoll med 60 SNPer. Småskala uttesting på kjent familiemateriale viste at 59 av disse markørene inkludert tre mitokondrielle SNPer, produserte robuste data. Det ble imidlertid funnet at ved storskala semiautomatisert genotyping medførte det høye antall av SNPer en generelt lavere signalstyrke. Initialt var en passende løsning på dette å kjøre halvparten av SNP-analysene separat for de innledende stegene i analysen (PCR og minisekvensering) og for deretter å slå sammen med massespektroskopianalysene.

Dataanalysen søker i hovedsak å sammenligne genotyper fra en SNP hentet fra testing av en enkelt avkomsprøve (eller en antatt rømling) mot registrerte foreldregenotyper for den samme SNP'en, og å identifisere en umulig match (Mendelske feil). Et eksempel på en umulig match er at et avkom med AA-genotype ikke kan komme fra en forelder med BB. Prosessen gjentas så for alle de 59 markørene og identifiserer alle mulige (feilfrie) foreldre for et enkelt avkom. Ytterligere informasjon i form av krysningsinformasjon over registrerte parringer kan deretter brukes til å validere matchingen og finne det riktige foreldrepåret i de tilfeller der flere potensielle foreldrepår er identifisert. Modellen er basert på «eksklusjons-prinsipper» ved at den prøver å eliminere umulige koblinger for å finne den riktige. Det er nødvendig med et minimum av 40 genotype sammenligninger for å utføre elimineringen, og i en tilordning vil ikke avkomsprøver med færre enn 40 rapporterte genotyper bli inkludert i analysen. Årsaker til et lavt antall genotyper kan være dårlig DNA kvalitet eller kvantitet, eller feil under bearbeiding i laboratoriet. Et viktig poeng er imidlertid at vi har funnet at det er nødvendig med en balanse mellom hvor mye vekt som legges på Mendelske feil i forhold til krysningsinformasjon. Hvis det for eksempel er bevis for to Mendelske feil i en tilordning, vil det lede til eksklusjon av en bestemt avkom-forelderkonstellasjon i favør av en annen som bare viser en feil. Krysningsinformasjon kan støtte den førstnevnte tilordningen og avvise den sistnevnte. I et slikt scenario oppstår spørsmålet om vi skal tolerere to Mendelske feil i favør av en gyldig kryssning, eller kreve at det ikke er mer enn en feil, men samtidig forslå en ugyldig kryssning? I virkeligheten vil både laboratorie-genererte genotyper og produsent-levert krysningsinformasjon inneholde feil, og det er svært utfordrende å utvikle et sett med «regler» som pålitelig styrer valgene i en automatisk algoritme. Dette er ytterligere vist ved det faktum at antallet genotyper som oppnås for en enkelt prøve varierer. Mens minste tillatte er 40 kan de andre prøvene ha 41 eller 59, og det å tillate to Mendelske uoverensstemmelser i en prøve med 41 genotyper gir en relativt sett større usikkerhet enn i en prøve med 59 genotyper. Resultatene fra valideringsstudie

1 og 2 har tilordningskategorier som viser disse vanskelighetene. I begge studiene var det et visst antall prøver som det var vanskelig å tilordne til et foreldrepar med den informasjonen som var tilgjengelig. Slike er klassifisert som «ikke tilordnet».

### **WP 1.3 (CIGENE) Databaseformat for lagring av SNP-profiler**

For å implementere SNP genotyping (eller også mikrosatellitt-basert genotyping) av fisk for tilordning til en spesifikk populasjon kreves det informasjon fra flere kilder inkludert (i) genotyper fra alle potensielle foreldre, (ii) kryssningsoversikter fra avlsselskaper og (iii) informasjon om distribusjonen av rogn/ungel til oppdrettsselskap og lokalitet. Denne informasjonen må være helhetlig på nasjonalt nivå, oppdateres kontinuerlig og vil kreve en betydelig satsing fra norsk oppdrettsnæring. Problemstillingen er imidlertid ikke unik for fiskeoppdrettsbransjen, men ligner mye på den som gjelder for produksjons- og distribusjonskjeden for legemidler, mat og mange andre industrivarer. Vi går derfor ut fra at relativt velprøvde prinsipper og løsninger for elektronisk sporing av varepartier kan tilpasses dette behovet.

Det er ikke innenfor rammen til dette prosjektet å etablere en slik database, men det er mulig å spekulere over hvilke data og strukturer som vil være nødvendig. For eksempel vil de tilordningsaktivitetene som er beskrevet i dette prosjektet kreve spesialiserte algoritmer samt stor regnekraft og dessuten tilpassede databasestrukturer som inneholder identiteter, metadata og genotyper. Egenskapene til de programmene og filene vi har brukt i løpet av dette prosjektet kan brukes i utviklingen av en slik database.

### **WP 2: (Biobank AS) Innsamling, ekstraksjon og distribusjon av DNA til valideringsstudie 1 og 2**

Hovedansvarlig for denne arbeidspakken har vært Biobank AS. To prøvesett med DNA fra kjente foreldrepopulasjoner og vevsprøver fra simulerte «rømlinger» (tilhørende eller ubeslektet avkom) ble sammenstilt og kodet (blindet) i 96 brønners sett. Hvert prøvesett forelå i identiske dubletter som ble sendt til CIGENE for bruk i SNP valideringsstudie 1 og 2 (V1 og V2 i dette prosjektet) og til Nofima for bruk i mikrosatellitt-analyse (FHF-prosjekt 900708). CIGENE har etablert protokoll for kostnadseffektiv storskala DNA-ekstraksjon som ble brukt i dette prosjektet. Vevsprøver (finneklipp, 3 mm i diameter, 1 mm tykk) leveres i 96-brønners plater

inneholdende 94 % etanol som konserveringsmiddel. Den automatiserte DNA-ekstraksjonsprotokollen CIGENE bruker, leverer høykvalitets DNA på 24 timer til en konkurransedyktig pris (omtrent NOK 15,- per stk. inkludert reagenser og lønn).

### **WP 3: (CIGENE)Valideringsstudie 1 (V1): SNP basert tilordning blant foreldre som er full eller halvsøsken**

Hovedansvarlig for denne arbeidspakken var CIGENE med AquaGen AS, Biobank AS og NVH som deltagere. Det første prøvesettet ble designet av NVH og bestod av 229 kjente foreldre og 520 avkom fra AquaGens populasjon, samt 40 urelaterte lakeyngel (kontroller). Prøvene ble blindet (anonymisert) og sendt til CIGENE for genotyping og analyse (tilordning til foreldre). Informasjon om foreldrenes kjønn og kryssinger var også tilgjengelig, mens informasjon om riktig stamtavle ble holdt hemmelig inntil analysene var foretatt.

Genotypingen var effektiv for alle foreldre, men 16 av 560 avkom ga ikke et tilstrekkelig antall SNP-genotyper, trolig på grunn av dårlig DNA kvalitet. Tabellen nedenfor viser antall avkom tilordnet et foreldrepar (assigned), og antall avkom som ikke kunne tilordnes foreldre (unassigned). Flesteparten av ikke tilordnede fiskene utgjøres av de 40 negative (ubeslektede) kontrollene i prøvesettet, og de resterende 8 utgjøres av fisker som ikke tilordnes entydig til et foreldrepar.

**Tabell 3.** Resultater fra valideringsstudie 1 for 560 avkom

<b>Kategori</b>	<b>Beskrivelse</b>	<b>Antall</b>
Tilordnet «Assigned»	Avkom entydig tilordnet til 2 foreldre	496
Ikke tilordnet «Unassigned»	Avkom ikke tilordnet unikt til et gyldig foreldrepar ELLER ingen gyldige foreldre identifisert	48
Ikke konklusiv	For få genotyper for tilordning	16

Der alle metadata (kjønn, foreldre, kryssinger) brukes ble 496 av prøvene entydig tilordnet. Dette tilsvarer 98,4 % av avkommene siden 40 individer ikke hadde foreldre i datasettet og siden det ikke var mulig å få gode genotyper fra 16 prøver (på grunn av dårlig DNA kvalitet).



Når resultatene fra V1 ble sjekket mot fasiten (pedigree fra avlsselskapet) fant vi avvik mellom SNP-tilordning og den skriftlige informasjonen for 37 av de tilordnede fiskene. Tjuesju (27) av disse var med sikkerhet tilordnet et foreldrepar som ikke var korrekt i følge krysningsinformasjonen fra avlsselskapet. De samme 27 fiskene ble også identifisert i et parallelt, men uavhengig prosjekt som anvendte mikrosatellitter (FHF 900708). Dette kan tolkes som at det foreligger feil i pedigree og ikke at det nødvendigvis er feil på grunn av SNP genotypingen. I de resterende 10 prøvene som ble tilordnet, men som ikke ble støttet av krysningsinformasjonen, var det varierende grad av usikkerhet i dataene (f.eks. manglende genotyper). Dette gjorde det vanskelig å konkludere med at uoverensstemmelsene skyldtes feil enten i pedigree eller i genotypingen.

#### **WP 4: (CIGENE) Valideringsstudie 2 (V2): Tilordning ved et stort antall potensielle foreldrefisk**

Hovedansvarlig for denne arbeidspakken var CIGENE med AquaGen AS, Biobank AS og NVH som deltagere. To eksperimenter har blitt gjennomført på AquaGens avlsmateriale. Det ene datasettet omfatter tilordning av et relativt lite antall avkom (n=279) til en stor mengde potensielle foreldre (n=496; WP 4, V2). I dette prøvesettet forelå oversikt over gyldige kryssninger av mødre og fedre, men ingen kobling til individuelt avkom utover rognbatch. Resultater for dette prøvesettet er vist under i Tabell 4.

**Tabell 4.** Resultater fra valideringsstudie 2 for 279 avkom

<b>Kategori</b>	<b>Beskrivelse</b>	<b>Antall</b>
Tilordnet «Assigned»	Avkom entydig tilordnet til 2 foreldre	241
Ikke tilordnet «Unassigned»	Avkom ikke tilordnet unikt til et gyldig foreldrepar ELLER ingen gyldige foreldre identifisert	30
Ikke konklusiv	For få genotyper for tilordning	8

Ved hjelp av metadata (kjønn og krysningsinformasjon) var det mulig å tilordne 241 (86,4 %) fisker entydig til unike foreldrepar. For 40 av disse det var en uenighet med krysningsinformasjonen, men fravær av betydelige Mendelske feil argumenterte for tilordning.

En relativt stor andel av prøvene (30) kunne ikke bli tilordnet noen individuelle foreldre selv om de hadde et tilstrekkelig antall genotyper. En mulig årsak til dette kan være at de aktuelle foreldrene ikke var inkludert i prøvesettet. Åtte prøver (2,9 %) ble ikke tilordnet på grunn av dårlig genotyping.

### **WP 5: (Nofima) Valideringsstudie 3 (V3): Eksklusjonsstyrke med hensyn på villfisk**

Hovedansvarlig for denne arbeidspakken var Nofima med NVH og CIGENE som deltagere. Nofima har fått villfiskprøver fra Norsk institutt for naturforskning (NINA) bestående av 5 fisker fra 19 forskjellige elver. Disse prøvene ble genotypet og fiskene forsøkt tilordnet til foreldrene i prøvesettet til V1 (n=230). Ingen fisker ble tilordnet, og dette indikerer at SNP panelet kunne ekskludere villfisk med høy styrke.

### **WP 6: (NVH) Prosjektadministrasjon**

Etter avtale med MareLife Services AS har NVH tatt over prosjektadministrasjonen som omfatter inngåelse av konsortieavtale, etablering og drift av prosjektets internettsider ([www.laksesor.no](http://www.laksesor.no)), organisering av kommunikasjonen om framdrift og faglige resultater, organisering av møter i prosjektgruppen og styringsgruppen samt utarbeidelse av periodiske rapporter og sluttrapport til FHF.

### **Konklusjon**

SNP basert teknologi viser en god evne for å tilordne oppdrettsfisk til sine foreldre, og det nåværende SNP panelet har styrke til å skille villfisk og oppdrettsfisk. Det er viktig å standardisere og optimalisere metodikkene for DNA ekstraksjon. Videre implementering av SNP basert sporing vil dra nytte av at et SNP markør panel re-designes med inklusjon av flere markører og/eller for å oppnå bedre ytelse for enkeltanalyser. Rutinemessig DNA-testing av laks ved hjelp av SNP-teknologi utføres i stor skala i dag, inkludert prøvetaking, prøveoppbevaring, DNA ekstraksjon, genotyping og rapportering av data. En nøye gjennomtenkt design av SNP

paneler kan skape merverdi ved at det kan inkluderes markører som ikke bare gir mulighet for sporing av fisk, men som også kan gi informasjon om stamfisken til produsentene.

## **Leveranser**

Oversikt over leveranser

- Faglig halvveisrapport desember 2012
- Rapport resultater SNP basert tilordning av avkom fra halv- og helsøsken og SNP basert tilordning blant mange foreldre, mai 2013
- Sluttrapport desember 2013
- Faktaark desember 2013
- Administrativ/økonomisk sluttrapport desember 2013

## **Kvalitetssikring av prosjektgjennomføring og resultater**

Administrativt ansvarlig har hatt jevnlig kontakt med prosjektgruppa for oppfølging av prosjektgjennomføringen. Det har vært holdt møter jevnlig for oppfølging av analyser, resultater og avvik fra prosjektplanen.