



Rømming 1-2023

Sporing av rømt oppdrettstorsk fanget i Nordland januar 2023

Quintela, M., Dahle, G., Sørvik, A.G.E., Wennevik, V., Glover*, K. A.

* = kontaktperson

Havforskningsinstituttet, Postboks 1870, Nordnes, 5817 Bergen
17.02.2023

Innledning og konklusjon

Denne rapporten beskriver genetiske og statistiske analyser av prøver fra antatt rømt oppdrettstorsk innsamlet fra Meløy i Nordland i perioden 11-13 januar 2023, samt analyse av prøver fra en oppdrettslokalitet for torsk i regionen. Rapporten estimerer sannsynlighet for at den rømte torsken stammer fra anlegget, basert på DNA-analyser. Undersøkelsen ble initiert av Fiskeridirektoratet. Prøvene ble analysert av Havforskningsinstituttet i Bergen.

De ulike analysene viste en svært stor grad av genetisk likhet mellom de rømte torskene og fisk fra alle merdene i anlegget. Spesielt vil vi framheve slektskapsanalysen som med høy sannsynlighet identifiserte helsøskenpar mellom fisk fra alle tre prøver fra anlegget og de rømte torskene. Basert på disse analysene, konkluderer vi derfor at de rømte torskene med svært høy sannsynlighet stammer fra dette anlegget.

Materiale og metode

Materiale

Totalt ble det fanget inn prøver fra 100 fisk som var antatt å være rømt oppdrettstorsk, men bare 78 ble levert videre for DNA-analyser. Det ble også samlet inn otolitter, tatt bilder og målt lengde av noen av fiskene. Fiskene ble fanget av to lokale fiskere. Fiskeridirektoratet tok prøver av fangsten, og dette prøvesettet fikk samlavn RF og individnavn RF_1 til RF_78. Otolittene

ble undersøkt og basert på gitte kriterier ble det bekreftet at disse fiskene stammer fra oppdrett.

Referanseprøver (baselineprøver) av oppdrettstorsk for de genetiske analysene ble samlet fra to merder ved anlegget, i tillegg til en prøve fra utsortert død fisk, på det eneste anlegget i området som har torsk av tilsvarende størrelse til de rømte torskene. Baseline prøvene ble kalt for M1_2_3_5_7 (død fisk fra anlegget), M2 og M4 (to merder ved anlegget) (**Tabell 1**). Dødfisk prøven ble kalt for M1, og fikk individ navn M1_1 til M1_48 osv. Det er oppgitt fra oppdretteren, at all fisk ved anlegget har utgangspunkt egg fra ett enkelt gytekar.

For å kunne gi et perspektiv på genetisk differensiering mellom vill og domestisert torsk, ble en prøve av 50 villtorsk inkludert på noe av analysene. Disse prøvene stammer fra torsk samlet inn på Sørlandet, og er derfor ikke helt representative for lokal vill torsk i Meløy. Det presiseres at dette har ingen betydning for å kunne fastslå graden av likhet mellom rømt torsk og prøvene fra anlegget. Vi tar med denne prøven i noen av analysene for å vise i hvilken grad oppdrettstorsken skiller seg fra en generell villfiskprøve.

Genetiske analyser

Prøvene ble registrert i en database ved Havforskningsinstituttet og ble tildelt nummer (**Tabell 1**). DNA analyser ble utført i laboratoriet i Bergen i tidsrommet 23.01.2023-06.02.2023. DNA ble isolert med Qiagen DNeasy isoleringskit, og 21 DNA mikrosatelitter-markører (*i.e.* Gmo19, Gmo35, Gmo37, Gmo8, Tch11, Gmo132, Gmo2, Gmo3, Gmo34, Tch13, GmoC303, GmoC305, GmoC78, GmoC80, GmoC83, GmoC127, GmoC272, GmoC274, GmoC343, GmoG12, GmoG25a) ble analysert for alle individ i sammensatte multiplex (genotyping-betingelser tilgjengelig på forespørsel). I tillegg ble Pantophysin (PAN) analysert med QPCR. PCR produkter ble analysert på en ABI 3730XL Genetic Analyser og fragmentstørrelsen i basepar (bp) ble beregnet med en 500LIZTM størrelsesstandard. Automatisk scorete alleler (varianter av mikrosatellittene) ble kontrollert manuelt av to forskere og, etter kvalitetskontroll, ble totalt 261 individer beholdt for videre statistiske analyser etter å ha akseptert maksimalt <25 % manglende genotyper.

Statistiske analyser

Analysene ble utført i henhold til protokollen for sporing av rømt oppdrettsfisk ved Havforskningsinstituttet. Totalt antall alleler og rikdom (A_r - også referert til som allelisk

diversitet), dvs. en beregning av gjennomsnittlig antall alleler per locus når antall individer er korrigeret for, ble beregnet med programmet MSA mens observert (H_o) genetisk heterozygositet ble beregnet med GenAEx.

Analysen av slektskap mellom individene ble utført ved hjelp av programmet COLONY. I dette programmet benyttes sannsynlighetsmetoder for å vurdere mulig slektskap mellom individer, både mulige foreldre/avkom og eventuell hel- eller halvsøsken, basert på informasjon fra de genetiske markørene som er analysert. I analysen ble det tatt høyde for polygami (dvs. en fisk kan gyte med flere andre), meget høy presisjon for de beregnede sannsynligheter, og lange runs.

Populasjonsstrukturen ble undersøkt ved hjelp av ulike metoder. Først, ble forholdet mellom prøvene undersøkt ved diskriminantanalyse av hovedkomponenter (DAPC) implementert i programmet ADEGENET hvor grupper ble definert ved hjelp av geografisk plassering av prøvene. Kryssvalideringsfunksjon ble brukt både for å beregne det optimale antallet hovedkomponenter og diskriminantfunksjoner som skal beholdes. I tillegg, ble genetisk differensiering mellom prøver testet ved estimering av parvis F_{ST} beregnet med programmet ARLEQUIN. F_{ST} er et mål på populasjonsdifferensiering; dvs. genetisk avstand mellom populasjoner (forskjeller i allelfrekvens mellom to populasjoner) på grunn av genetisk struktur. Korreksjonen for flere sammenligninger for F_{ST} P -verdier ble utført gjennom Falsk Discovery Rate (FDR). Programmet STRUCTURE ble også brukt til å identifisere genetiske grupper i prøvematerialet. Dette programmet kan identifisere genetiske clusterer/grupper i genetiske datasett ved å påvise allelfrekvensforskjeller i dataene. Programmet forsøker å finne den oppdelingen av individene i det antall genetiske clusterer som gir høyest sannsynlighet. I analysen tester man et antall mulige clusterer for å finne antallet grupperinger som er mest sannsynlig. STRUCTURE ble brukt med en modell som antar blanding og korrelerte allelfrekvenser uten å bruke populasjonsinformasjon i analysene. Det ble gjort med en innbrenningsperiode bestående av 100 000 replikasjoner og 1 000 000 MCMC-iterasjoner. Disse ble utført for $K=2$ til $K=4$ clusterer.

Sammenligning og tilordning av de rømte individene til baselineprøvene ble gjennomført ved to metoder: 1) Direkte tilordning til baselineprøvene ble utført med programmet GeneClass2 ved hjelp av Rannala & Mountains beregningsmetode, og 2) Ekskludering fra baselineprøver på signifikansnivå 0.001 og 0.05 som også ble utført med det samme programmet med Rannala & Mountains simuleringsalgoritme med 10000 simulerte individ og $\alpha=0.001$. Direkte tilordning plasserer et individ i den baselineprøven som er genetisk mest lik (uansett absolutt grad av likhet). Direkte tilordning tar ikke med i betraktningen at i

noen situasjoner, slik som i alle rømmingsepisoder, vil ikke alle kilder som teoretisk sett kan ha gitt opphav til rømlingene, være representert. Følgelig er det viktig å få et mål for genetisk likhet mellom disse rømlingene og baselineprøvene. Dette oppnår man ved "Eksklusjons basert simulering" kalkulert for ulike grader av sannsynlighet. Denne metoden ekskluderer hvert individ i tur og orden fra hver av baselineprøvene ut fra sannsynligheten for at et gitt individ tilhører baselineprøven.

Resultater og diskusjon

Genotyping kvalitetskontroller

Hver mikrosatellit locus ble vurdert i forhold til manglende data. Locus Gmo127 hadde 43.2 % manglende data og ble derfor fjernet fra videre analyser. PAN locus ble også fjernet på grunn av at det var monomorf, *i.e.* alle individer hadde genotype AA og dette locuset ga derfor ingen informasjon til videre statistiske analyser.

Manglende data per individ ble deretter vurdert for datasettet med 18 mikrosatelitt-loci. For de aller fleste individene var det <5% manglende data. Men for elleve individer var det 25-35% manglende genotyper og disse individene ble fjernet fra videre analyser (*i.e.* individ M1_17, M1_25, M2_20, M2_44, RF_35, RF_38, RF_41, RF_46, RF_48, RF_55, RF_63). Totalt antall individer per prøve med akseptabel datakvalitet er vist i **Tabell 2**.

Genetisk diversitet

Totalt antall alleler og allelisk rikdom var på samme nivå i prøvene av de rømte torskene og i prøvene fra anlegget (**Tabell 2**). I prøven av villtorsk var nivået langt høyere for disse parameterne og vi observerte dobbelt så høye verdier, noe som demonstrerer tap av genetisk diversitet under domestiseringsprosessen. Tilsvarende, var observert heterozygositet lavere i prøvene fra anlegget og rømte torskene sammenlignet med prøvene av villtorsk. Disse resultatene bekrefter at de antatt rømte individene med høy sannsynlighet ikke er ville individer, noe som bekrefter vurderingen ut fra otolittene.

Slektskap mellom rømte individer og individer i baselineprøver

En metode for å undersøke om det er sannsynlig at rømte individer kommer fra en konkret kilde er å analysere slektskap mellom individer. I en oppdrettsmerd vil det ofte være et begrenset antall familier representert, og ved hjelp av analyser i programmet COLONY kan man undersøke om det er sannsynlige helsøsken- og halvsøskenpar mellom rømt fisk og fiskene i en merd. Programmet foretar en sannsynlighetsanalyse for slektskap mellom individer basert på variasjonen i DNA-markører. Vi gjennomførte slike analyser for å undersøke mulig slektskap mellom de rømte individene og fiskene fra anlegget, og fant mange hel- og halvsøskenpar både blant de rømte fiskene men også mellom de rømte fiskene og alle merdene (**Fig. 1**). Når sannsynlighet >0.8 ble akseptert, fantes 34 halvsøskenfamilier med variabelt antall medlemmer, fra 2 til 14 medlemmer per familie (**Tabell 3**). Sytten av disse familiene hadde medlemmer i ulike merder i anlegget, og fire familier ble funnet blant de rømte fiskene. Mer interessant var at det fantes 13 familier med medlemmer både i merdene og i prøven av antatt rømt fisk. En av familiene hadde medlemmer fra hver av merdene i anlegget og i prøven av antatt rømt fisk. Dette funnet er et svært tydelig resultat som demonstrerer et nært genetisk slektskap mellom RF og prøvene fra anlegget.

Genetisk differensiering mellom prøvene (både baselineprøver, RF og vill torsk)

Genetisk forskjell mellom individer og samleprøver av individer ble undersøkt ved hjelp av flere ulike statistiske teknikker. Blant disse er parameteren F_{ST} , som er et mål for genetisk distanse. Jo høyere verdi for F_{ST} , desto større genetisk distanse mellom prøvene. I dette prøvematerialet fant vi de største parvise genetiske forskjellene målt som F_{ST} mellom vill torsk og resten av prøvene (**Tabell 4**), med verdier mellom 0.062-0.068, og alle var signifikant forskjellige fra hverandre ved $P < 0.0001$. I motsetning til dette, var den genetiske forskjellen mellom merdene og RF 16 ganger mindre (F_{ST} fra null til 0.005). Nivået på differensiering mellom merder fra samme anlegg var den samme som mellom antatt rømt torsk og merdene. De tre høyeste parvise estimatene av F_{ST} (0.005) var statistisk signifikant forskjellige ved $\alpha = 0.05$ etter FDR, men ingen av ved $\alpha = 0.01$. Dette betyr at den samlede prøven av rømt fisk var genetisk lik prøvene fra anlegget.

En annen måte å illustrere genetisk forskjell mellom prøver på er ved hjelp av diskriminantanalyse av hovedkomponenter (DAPC). Analysen ble utført ved å inkludere de rømte fiskene (RF), baselineprøvene (M1, M2, M3) og ville fiskene. Etter kryssvalidering ble

120 hovedkomponenter (PCs) beholdt og første aksene (PC1), som forklarte den 94.5% av variasjonen, skilte tydelig villfiskene fra oppdrettsfisk og de antatt rømte fiskene (**Fig. 2a**). Den andre aksene (PC2) forklarte bare 2.2% av variasjonen og kunne ikke skille mellom de oppdrettstorsk og antatt rømt torsk. I neste analyse ble vill fisk ikke inkludert for å øke oppløsningen og bedre illustrere eventuelle forskjeller mellom antatt rømt fisk og fisken i anlegget (**Fig. 2b**). Denne siste analysen viser bedre den store over overlappende graden av likhet mellom baselineprøvene og den antatt rømte torsken.

Den tredje metoden som var brukt for å beregne genetisk forskjell mellom prøvene var clusteranalyse i STRUCTURE. Et strukturdiagram viser genetisk likhet mellom individer og sett med prøver. Hver fisk blir representert med en vertikal linje som kan bestå av et eller flere farger. Hver genetisk gruppe blir angitt med en farge, og tar høyde for at en samlet prøve evt. kan bestå av fisk av forskjellige opphav (både genetisk blanding dvs. krysninger mellom grupper og fysisk blanding). I denne saken ble den største genetiske forskjellen oppdaget mellom vill torsk og resten av prøvene som tilsvarer to veldig distinkte clusterer (**Fig. 3a**). Antatt rømte torsk kunne ikke skilles fra oppdrettsfisken ved $K=2$ clusterer. Når antall clusterer/grupper settes til 3, finner vi stor genetisk homogenitet mellom RF og baselineprøvene fra anlegget, men en rekke individer stikker seg ut. De vises som mørkeblå linjer i **Fig. 3b**. Dette nivået av understrukturering skyldes slektskap mellom individene da alle disse 14 fiskene er søsken. De ble identifisert som medlemmer av samme familie (se **Tabell 3**) og de er fordelt på tvers av de tre merdene i anlegget, og de rømte fiskene. STRUKTUR analysen støtter alle de tidligere analysene og viser svært stor grad av genetisk likhet mellom de rømte torskene og torsk fra anlegget. Ikke minst, identifisering av søsken mellom RF og prøvene fra anlegget støtte slektskaps analysen allerede presenterte.

Simulering av genetisk tilordning (assignment) mellom prøvene

Muligheten for å kunne bruke genetiske metoder for å identifisere opphavet til rømt fisk øker med den genetiske forskjellen mellom baselineprøvene. Genetisk tilordning ble simulert mellom de ulike baselineprøvene fra oppdrettsanlegget, for å teste om det var tilstrekkelig genetisk forskjell mellom merdene til å kunne identifisere de rømte fiskene tilbake til riktig merd. Graden av genetisk differensiering mellom baselineprøvene (målt som F_{ST} verdi) og potensialet for genetisk tilordning er tett koblet.

Graden av selvtilordning mellom baselineprøvene var veldig lav; *i.e.* bare 52 av 140 individer (37.1%) ble riktig identifisert tilbake sin opprinnelige merd ved bruk av genetisk

informasjon. Dette var forventet fordi det er oppgitt at alt genetisk materiale fra anlegget stammer fra samme gytekar, og ingen genetisk forskjell ble observert mellom merdene fra anlegget.

Direkte tilordning av alle 71 rømte fiskene til de enkelte merdene er presenterte (**Fig. 4**). Individene ble tilordnet nesten jevnt til merdene, dette er fordi det er stor grad av likhet mellom merdene og da ikke mulig å si hvilke merd fiskene har evt rømt fra. Men, ~95 % av rømlingene kunne ikke utelukkes fra å ha sin opprinnelse i noen av de tre testede merdene ved $P < 0.001$. Igjen, dette tilordningsanalyse bekrefte alle tidligere resultater presenterte, at det er en svært stor grad av genetisk likhet mellom de rømte fiskene og fisk på anlegget.

Tabell 1. Oversikt over baselineprøvene fra anleggene, rømt fisk og vill fisk. Det er til sammen 3 oppdrett prøver fra 1 anlegg.

	Prøve	Lokalitet	Merd	Innsamlet	Vekt (g)	Rognleverandør	Gruppe
M1_1 til M1_48*	M1**	Meløy i Norland	1,2,3,5 og 7	18.01.2023		Havlandet Marin Yngel	Miks
M2_1 til M2_48*	M2	Meløy i Norland	2	18.01.2023		Havlandet Marin Yngel	a
M4_1 til M4_48*	M4	Meløy i Norland	4	18.01.2023	3.3	Havlandet Marin Yngel	a
RF1-78	RF	Norland		11-13.01.2023			
Vill***	Vill	Sørlandet					

Merknader:

*Individuelle navn ble forenklet på denne måten for å forhindre at statistiske programmer avkuttet dem.

**Prøve M1-2-3-5-7 kalles M1 for enkelhet.

***En prøve av vill torsk ble inkludert for perspektiv.

Tabell 2. Oppsummeringsstatistikk: Antall individer i hver prøve etter kvalitetskontroll av genotyping, antall alleler, allelisk rikdom (A_r) basert på et minimum antall av 46 individer og observert heterozygositet (H_o , gjennomsnitt \pm SE).

Prøve	Type	Antall individer	Antall alleler	Allelisk rikdom	H_o
M1	Oppdrett	46	155	7.7	0.672 \pm 0.043
M2	Oppdrett	46	149	7.4	0.676 \pm 0.049
M4	Oppdrett	48	150	7.5	0.654 \pm 0.047
RF	Rømt	71	157	7.5	0.693 \pm 0.044
Vill	Vill	50	298	14.3	0.758 \pm 0.044

Tabell 3. COLONY resultater: Helsøsken familier med sannsynlighet for inkludering >0.8. Fet skrifttype viser familien med medlemmer på tvers av alle anlegg og rømt fisk. Lilla viser oppdrettsindivider og gult viser rømte torsk.

Sannsynlighet	Medlem1	Medlem2	Medlem3	Medlem4	Medlem5	Medlem6	Medlem7	Medlem8	Medlem9	Medlem10	Medlem11	Medlem12	Medlem13	Medlem14	Medlem15
0.999	M1_46	M2_30													
0.911	M1_26	M2_40													
0.777	M1_3	M4_25													
0.996	M1_37	M4_21													
0.995	M1_14	M1_21													
0.997	M1_40	M2_22													
0.998	M1_32	M2_9													
0.997	M1_24	M1_43	M4_4												
0.980	M1_6	M2_12	M2_32												
0.967	M1_1	M1_18	M2_47												
1	M4_7	M4_8	M4_40												
1	M1_38	M2_29	M4_33	M4_43											
0.998	M4_13	M4_22	M4_26	M4_48											
0.997	M4_5	M4_9	M4_41	M4_46											
0.999	M1_13	M1_20	M2_21	M4_27											
1	M2_5	M2_6	M2_8	M2_11	M2_42										
1	M1_11	M2_13	M2_45	M4_12	M4_17										
0.989	RF_15	RF_71													
0.999	RF_65	RF_73													
0.994	RF_39	RF_56													
0.999	RF_18	RF_23	RF_24	RF_31	RF_78										
0.999	M4_38	RF_47													
0.998	M4_45	RF_76													
0.992	M4_36	RF_60													
0.966	M2_17	RF_66													
0.952	M2_7	RF_20													
0.934	M1_23	RF_37													
0.862	M4_30	RF_54													
0.803	M2_28	RF_3													
0.753	M2_23	RF_6													
0.953	M1_34	RF_21	RF_22												
0.999	M1_39	RF_53	RF_72												
1	M2_16	RF_4	RF_10	RF_27	RF_77										
1	M1_29	M1_41	M1_45	M2_19	M4_11	M4_29	M4_34	RF_7	RF_25	RF_32	RF_45	RF_58	RF_67	RF_75	

Tabell 4. Parvise F_{ST} verdier (genetisk distanse – nede til venstre), og tilhørende P -verdier (oppe til høyre) mellom prøvene. P -verdier med fet skrift er signifikante etter FDR-korreksjon på $\alpha=0.05$ og den tilsvarende F_{ST} verdien er også uthevet med fet skrift for å lette identifiseringen. På $\alpha=0.01$, er den eneste betydelige differensieringen mellom vill torsk og resten av prøvene

	M1	M2	M4	RF	Vill
M1	*	0.588	0.026	0.064	0.000
M2	0.000	*	0.036	0.012	0.000
M4	0.005	0.005	*	0.029	0.000
RF	0.003	0.005	0.004	*	0.000
Vill	0.068	0.065	0.062	0.066	*

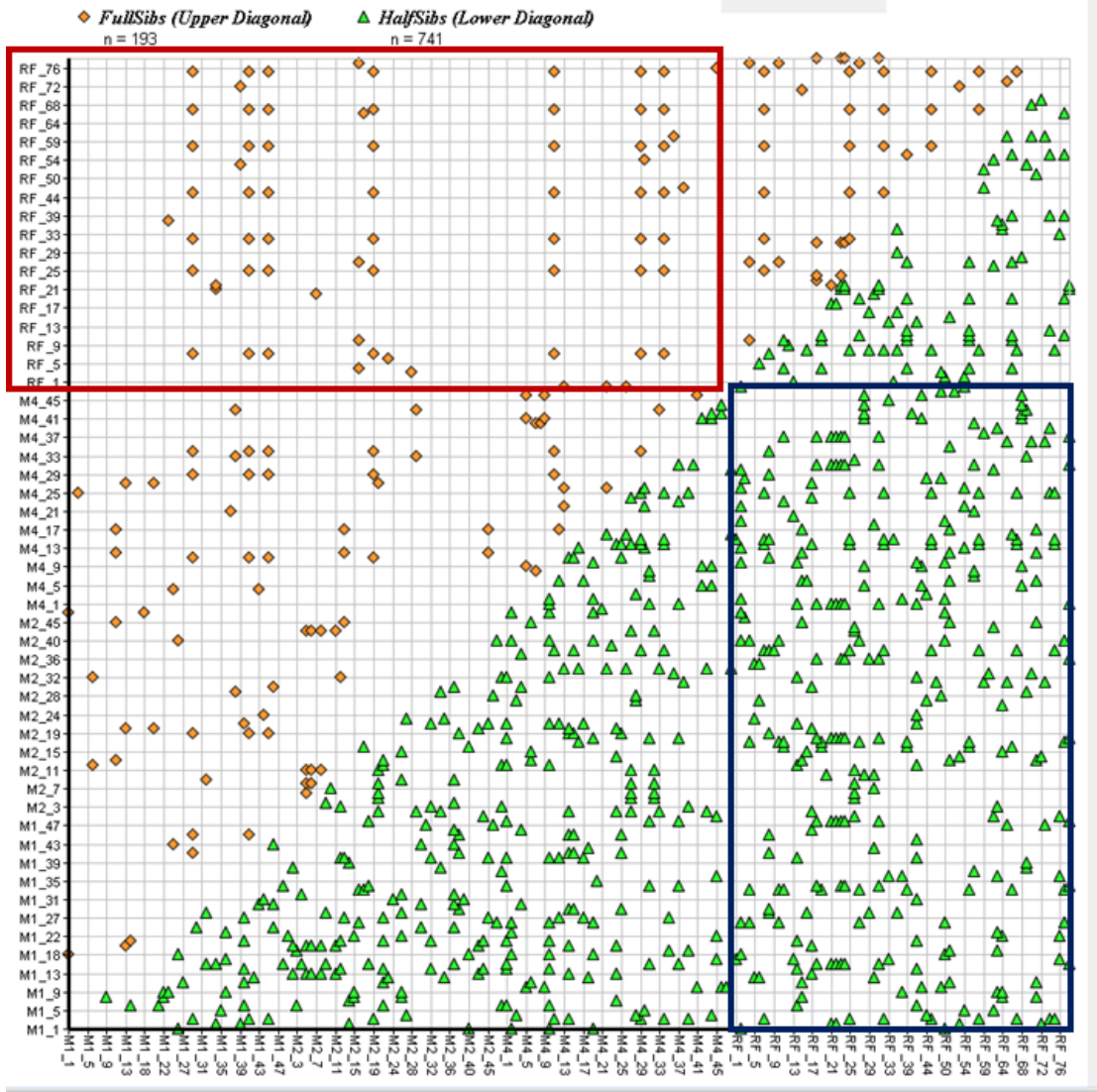


Fig 1. Plott som viser resultatet av Colony analysene. Grønne punkter i nedre diagonalfelt skildrer halvøskenpar og oransje punkter i øvre diagonalfelt viser helsøskenpar. Røde rektangel viser mulige helsøskenpar mellom rømt fisk og anlegg mens mørke blå rektangel viser mulige halvøskenpar mellom rømt fisk og anlegg.

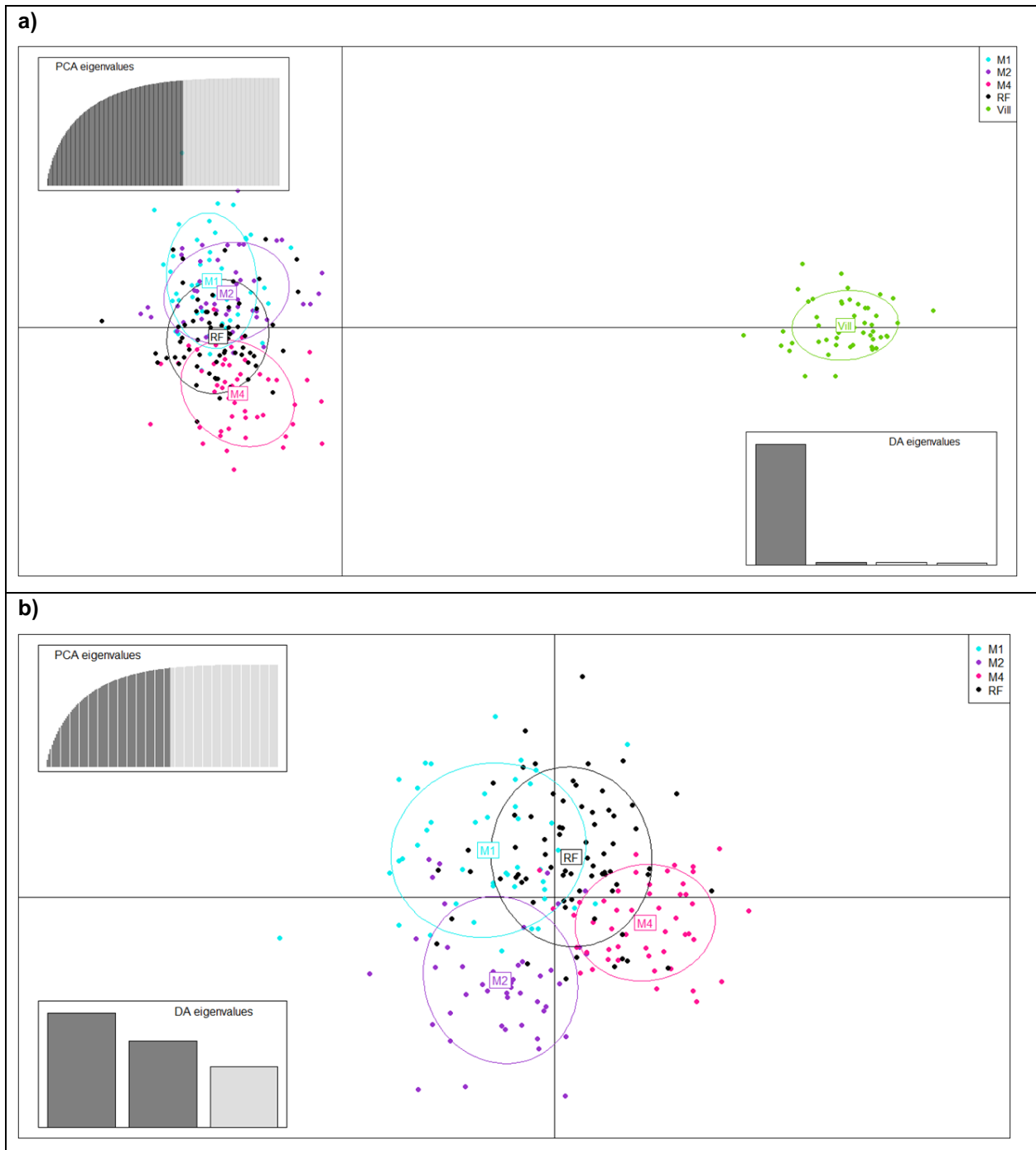


Fig 2. Diskriminantanalyse av Hovedkomponenter (DAPC). Figuren a) inneholder rømte fisk (RF), tre anlegg (M1, M2, M3) og vill fisk. Ett hundre og tjue hovedkomponenter (PCs) ble beholdt og første aksens (PC1), som forklarer den 94.5% av variasjonen, skiller villfisk tydelig fra oppdrett og rømt fisk. Andre aksens (PC2) bare forklarer 2.2% av variasjonen og kan ikke gjøre forskjell mellom oppdrett og rømt torsk. I nedre figuren (b), ble vill fisk ekskludert og 80 PCs står for 43.8 av variasjonen på den første aksens og 32.9% på den andre. Stor overlapping er funnet blant treghetsellipsene til de fire prøvene.

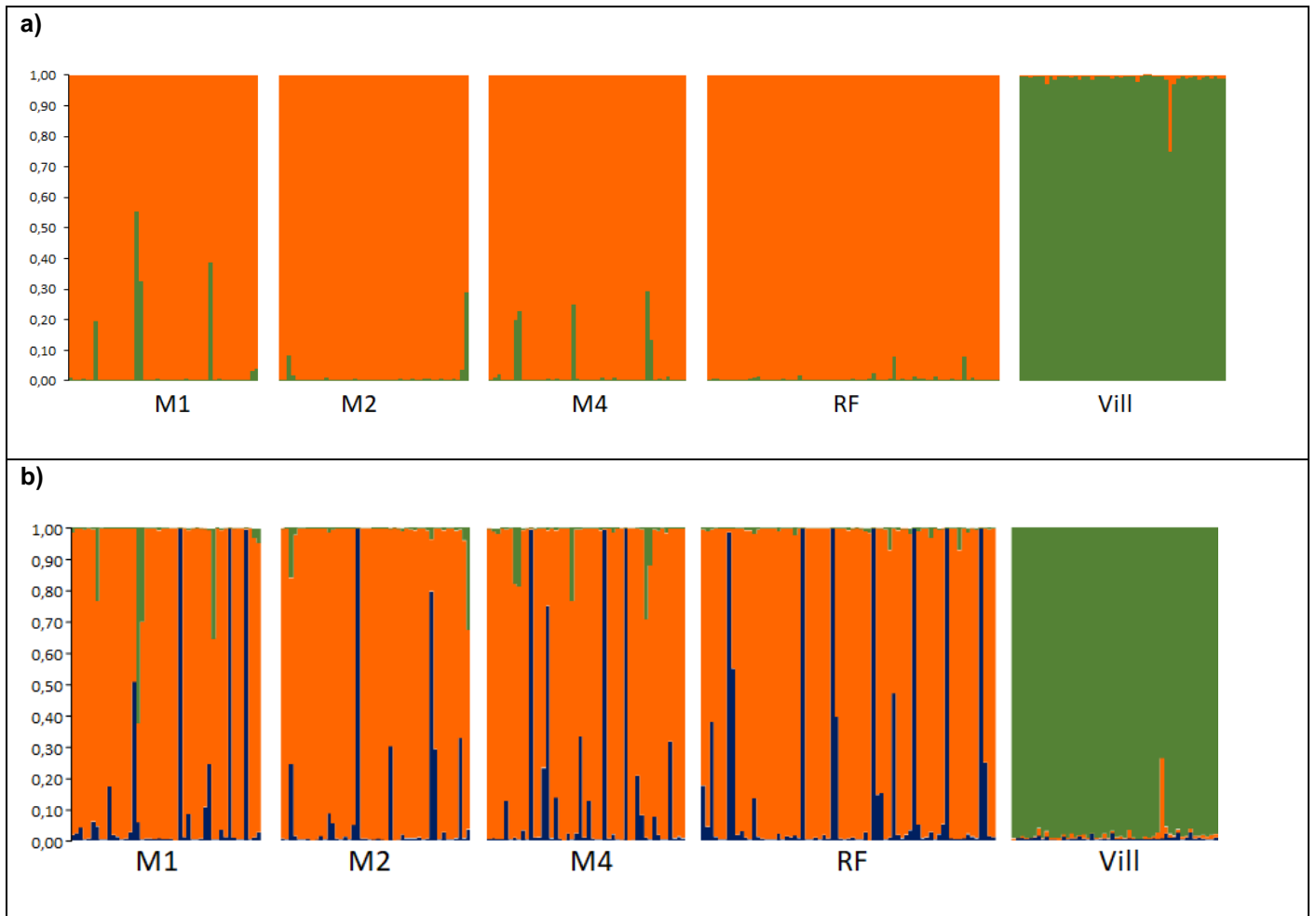


Fig 3. Strukturdiagram som viser hvordan prøvene fordeler seg i ulike genetiske grupper. Villfisk skiller seg tydelig fra oppdrett (M1, M2, M4) og rømte torsk (RF) både på K=2 clusterne (a) og på K=3 clusterne (b). Den nederste plott viser 14 hele mørkeblå linjer fordelt på oppdrett og rømte prøvene tilsvarende de medlemmer av samme helsøsken familie markert i Tabell 3. Familien inneholder individer M1_29, M1_41, M1_45, M2_19, M4_11, M4_29, M4_34, RF_7, RF_25, RF_32, RF_45, RF_58, RF_67 og RF_75.

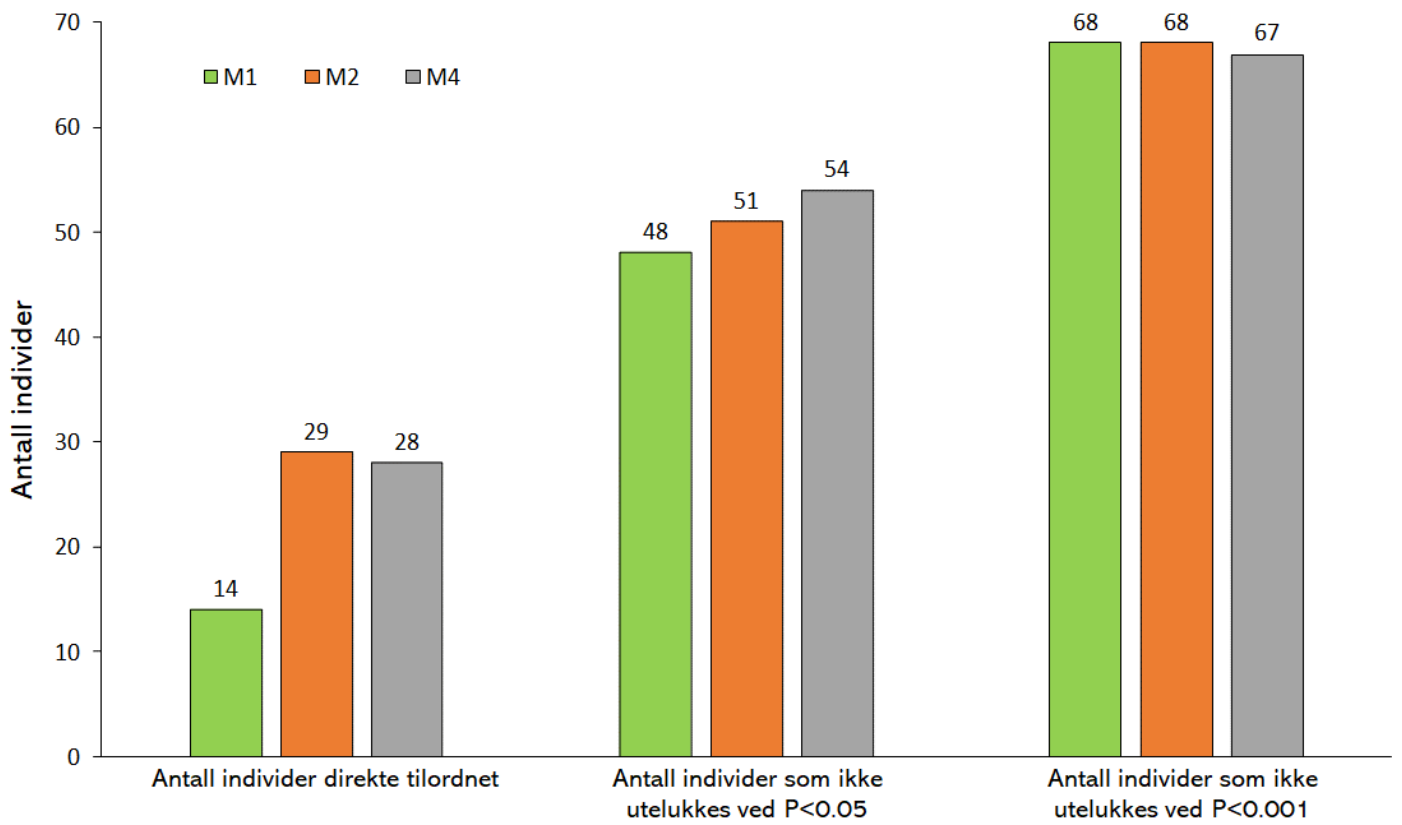


Fig 4. Direkte tilordning av de 71 rømte individer til anlegg. Antall rømte fisk som ikke blir ekskludert fra prøven ved 0.05 og 0.001 signifikansnivå henholdsvis. Rømte individer ble direkte tilordnet til prøver i relativt like proporsjoner på grunn av lav differensiering mellom prøvene.