

Rapport

«Artssammensetning og lokal populasjonsstruktur av sil i Tanafjorden»

Personell involvert i oppdraget: Førsteamanuensis Kim Præbel, PhD (prosjektleder),
Forskningsassistenter MSc Adam Andrews og MSc Filipe Figueiredo, Bioinformatikker Dr. Shripathi
Bhat.

Professor Peter R. Møller og Bioinformatikker Filipe G. Vieira, Zoologisk Museum, Københavns
Universitet, Danmark bidro også til prosjektet med hhv. referanse materiale/røntgen analyser og
bioinformatisk støtte.

Resyme av konklusjon

Dette prosjektet hadde to hovedmål: 1. å undersøke hvilke arter av sil finnes i munningen av Tanaelven og 2. å undersøke om sil har lokal populasjonsstruktur i munningen av Tanaelven. Det ble brukt tradisjonell Sanger sekvensering av mitokondrie DNA og nestegenerasjons sekvensering av kjerne DNA til besvarelse av målene.

Samlet sett, konkluderer vi at **sil i munningen av Tanaelven *A. marinus***, men at *A. tobianus* og *H. lanceolatus* også kan forekomme i området. **Sil i munningen av Tanaelven er genetisk ulik sil fra Altafjorden**, og det er tegn på at sil i munningen av Tanaelven også er genetisk strukturert. Resultatene viser også at de genetiske forskjellene mellom Tana og Altafjorden formentlig er drevet av lokale tilpasninger.

Bakgrunn

Dette dokumentet danner grunnlaget for avrapporteringen av de populasjons genetiske analysene i forbindelse med prosjektet "Artssammensetning og lokal populasjonsstruktur av sil i Tanafjorden". Det ble ønsket å få svar på to spørsmål; 1. hvilke arter av sil finnes i munningen av Tanaelven? og 2. Danner sil lokal populasjonsstruktur i munningen av Tanaelven?

Det ble i tillegg d. 16/2-18 fremsendt følgende spørsmål per email vedr. genetikkundersøkelsene:

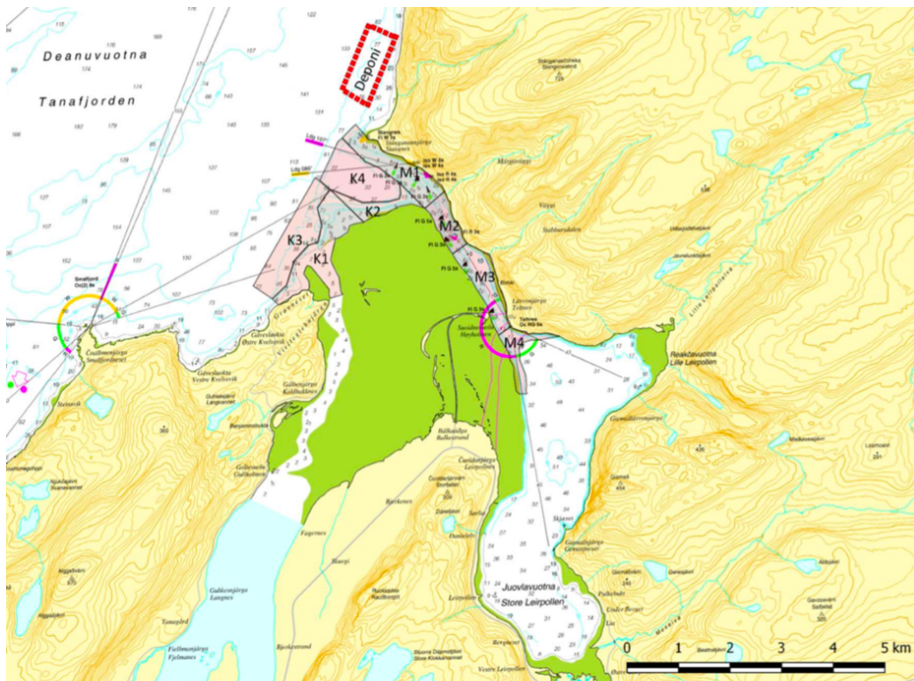
- a. *Hva betyr disse genetiske forskjellene?*
- b. *Er disse forskjellene spesielle, forvaltningsrelevante og/eller unike?*
- c. *Hvor lang tid tar det å evolvere slike forskjeller?*
- d. *Kan det tenkes at ny segregering kan oppstå innen få generasjoner?*
- e. *Hva er total og effektiv populasjonsstørrelse (N , N_e)?*
- f. *Hvor stor genetisk avstand er det mellom populasjonene (F_{ST})?*
- g. *Trengs flere undersøkelser for å besvare disse spørsmålene? Kan dere spesifisere omfang og type undersøkelser?*

Denne rapporten vil, til dels, besvare a., b., f., og g., men ikke c.-e., da dette er spørsmål som ikke umiddelbart kan besvares innenfor rammen av dette prosjektet. Her vil vi også påpeke at de populasjonsgenetiske analyser som skal belyse om sil fra Tana er en unik lokal populasjon (spørsmål 2.), bare er basert på sammenligning med en enkelt annen lokalitet (Alta) og ikke med prøver fra andre lokaliteter i Finnmark. Vi påpekte ved prosjektstart at slik geografisk dekning var nødvendig for et klart svar på om sil fra Tana er en unik lokal populasjon. Det betyr at de populasjonsgenetiske resultatene må tolkes med forsiktighet.

Materiale og Metode

Biologisk materiale

Sil ble innsamlet av Naturrestaurering AS i 2017 i munningen av Tana Elven (heretter Tana) og i Alta fjorden (heretter Alta) og vevsprøver på 96% etanol ble stillet til rådighet for prosjektet (Tabell 1).



Figur 1. Kart over områdene hvor det ble innsamlet prøver av sil til de genetiske undersøkelsene (Kart fra Kirkemoen og Colman, 2017)

Prøvene fra Tana ble innsamlet i de geografiske delområdene M1-4 og K2 (Figur 1). Sil fra Neiden (N = 2) og vest i Tana fjorden (N = 6) ble også fremsendt. Individene fra vest i Tana fjorden ble inkludert i undersøkelsen av arter i Tanamunningen. Verken sil fra denne lokalitet eller fra Neiden ble inkludert i de populasjonsgenetiske analysene, da N var for lav for meningsfulle analyser. Sil funnet i magene til huse fanget i Tufjord på Magerøy, Finnmark, ble også forsøkt inkludert i undersøkelsen. Men selv om vi gjorde flere forsøk med ulike ekstraksjonsmetoder, lykkes det ikke å isolere brukbart DNA fra disse prøver og de ble derfor ikke inkludert i prosjektet. Det ble også sendt prøver fra hhv. Mandal og Oslofjorden, innsamlet av hhv. Multiconsult AS og Naturrestaurering AS (Tabell 1). Disse prøvene ble bare brukt i analysene for artsbestemmelse, da den geografiske avstanden mellom Tana og disse lokaliteter tilsier at det vil være betydelige populasjonsgenetiske forskjeller mellom disse prøver og materialet fra Tana. Videre viste Genetikk Delrapport I (Appendiks 1), at sil fra disse lokaliteter var ulike arter (*Hyperoplus lanceolatus* og *Ammodytes tobianus*) enn dem som forekommer i Tana (*A. marinus*). Inkludering av dette materialet i de populasjonsgenetiske analysene ville derfor ha påvirket analysene i en slik grad at resultatene ikke ville være meningsfulle. Vi fremskaffet selv et referansemateriale bestående av vevsprøver fra to individer fra hver av fem verifiserte arter av sil fra det Zoologiske museet i København (ZMUC), til bruk for besvarelse av spørsmål 1 (Tabell 1).

Det ble ikke inkludert fenotypisk data (for eksempel; lengde, vekt, modenhet, og alder) for de innsamlede sil i denne undersøkelsen, da slik data ikke var tilgjengelig for dette prosjektet. Fenotypisk data ville ha støttet opp om de genetiske analysene. For eksempel i form av å understøtte om det var juvenil eller gyteklar/gytende sil som danner den genetiske strukturen vi fant eller om strukturen var en signatur på ulike økologiske typer av sil (fra lengde og vekt) og dermed ha bidratt til en dypere tolkning av resultatene.

Tabell 1. Oversikt over sil materialet brukt til påvisning av artssammensetningen og populasjonsstruktur av sil i Tanaelvens munning. Sted, område/voucher nummer til Københavns Zoologiske Museum (ZMUC), Art som undersøkelsen bestemte, antall individer hhv. Totalt (N), hvor det kunne isoleres brukbart DNA fra (N_{DNA}), antall individer brukt i artsbestemmelse med COI/Rho (N_{Art}), og antall individer med brukbar genomisk DNA og data til artsbestemmelse og populasjonsgenetikk (N_{Pop}).

Lokalitet	Område/ZMUC nr.	Art	N	N_{DNA}	N_{Art}	N_{Pop}
Tana	3M1	<i>Ammodytes marinus</i>	59	59	36	47
Tana	3M2	<i>Ammodytes marinus</i>	89	89	40	71
Tana	3M3	<i>Ammodytes marinus</i>	16	16	16	13
Tana	3M4	<i>Ammodytes marinus</i>	10	10	10	8
Tana	3K2	<i>Ammodytes marinus</i>	4	4	4	4
Tana	Vest Tana	Formentlig <i>A. tobianus</i> og <i>H. lanceolatus</i> *	6	6	-	4
Neiden	-	Ukjent*	2	2	-	-
Tufjord	Magerøy	Ukjent*	>70	0	-	-
Alta	Alta Fjorden	<i>Ammodytes marinus</i> *	27	25	-	25
Syd norge ¹	Mandal	<i>Ammodytes tobianus</i>	50	34	2	11
Oslo fjorden ¹	Gåsøya	<i>H. lanceolatus</i> , <i>A. marinus</i> , <i>A. tobianus</i>	28	28	7	13
Referanse	10534 & 61996	Småsil (<i>Ammodytes tobianus</i>)	2	2	2	2
Referanse	10515 & 10519	Havsil (<i>Ammodytes marinus</i>)	2	2	2	1
Referanse	61045 & 611254	Storsil (<i>Hyperoplus lanceolatus</i>)	2	2	2	2
Referanse	611253 & 611254	Uplettet storsil (<i>Hyperoplus immaculatus</i>)	2	2	2	1
Referanse	611007 & 611015	Glattsil (<i>Gymnammodytes semisquamatus</i>)	2	2	2	0
		Total	371	283	125	202

*disse prøver ble mottatt etter artsbestemmelse med COI/Rho var utført (Delrapport 1), men de genomiske analyser viste at sil fra Vest Tana formentlig var *A. tobianus* og *H. lanceolatus* og sil fra Altafjorden var *A. marinus*.

¹Resultatene for disse to lokaliteter følger denne rapporten og ikke Del-rapport I (som var preliminær)

DNA ekstraksjon, kvalifisering og kvantifisering

I prosjektet isolerte vi DNA fra ca. 370 individer ved hjelp av et kommersielt DNA ekstraksjons kit (DNaeasy Blood and Tissue, Qiagen). Vi inkluderte replikat prøver for 10 individer. DNA isolasjon og analyse av disse prøver ble utført uavhengig av de originale prøvene og de ble inkludert i både sanger og nestegenerasjons sekvenseringen. Kvalitet og kvantitet av det isolerte DNA ble bestemt med henholdsvis elektroforese og spektrofotometrisk. Vi brukte Quant-iT™ PicoGreen™ dsDNA Assay Kit (ThermoFisher) til den spektrofotometriske bestemmelsen. 283 individer hadde god nok DNA kvalitet og kvantitet til videre analyse (Tabell 1).

Genomiske analyser, laboratorie del

Vi brukte bioinformatiske metoder til å designe 18247 RNA prober ut ifra eksisterende genomisk data fra sil som vi har adgang til via samarbeidspartnere. Probene ble anskaffet fra MyBaits, Mycarray, MI, og anvendt til å "fange" informative DNA sekvenser (100 bp) hos hvert individ. Det ble fremstilt genomiske biblioteker fra hvert individ ved hjelp av NEBnext Ultra DNA Library prep kit (NEB) for

Illumina og DNAet fra alle individer ble påsatt en kjent unik DNA barkode (strekkode). Otte individ biblioteker ble poollet i ens mengde og probene ble anvendt til å fange DNA. Etter vask av probene, ble 12 slike pools slått sammen (totalt 96 prøver) og sekvensert i tre baner paired-end med 150 basepar leselengde på en Illumina HiSeq 4000 sekvenseringsplattform (Novogene, Hongkong). Det ble inkludert 10 replikat prøver som ble fordelt over alle sekvenseringsbaner. I alt ble 288 prøver sekvensert, hvorav det var 30 replikat prøver. For ytterligere informasjon om arbeidsgang, de genomiske biblioteker og filtrering og kvalitetssjekk av den genomiske dataen, se Appendiks 2.

Artsbestemmelse ved hjelp av hhv. mitokondrie og nukleær DNA barkoder og enkelt nukleotid polymorfismer

Artsbestemmelsen av sil i Tana munningen ble utført med to ulike tilnærminger; 1. analyse av variasjon i et mitokondrielt DNA fragment i Cytochrome oxidase I (COI) genet og et fragment av Rhodopsin (Rho) genet som finnes i det nukleære genomet jvf. protokollen i Thiel & Knebelberger (2016) (Rapportert i Del-Rapport I, Appendiks 1), og 2. variasjon ved 11.519 enkelt nukleotid polymorfismer (SNPs) i det nukleære genomet for Tana/Referanse og alle prøver/referanse (Tabell 1). For metodiske detalier om de to tilnærminger henvises til henholdsvis Appendiks 1 og 2.

Det ble også sendt individer til en ekspert i morfologisk artsbestemmelse av sil (Prof. Peter R. Møller, Københavns Zoologiske museum, Danmark), for å verifisere disse resultater. Denne bestemmelse ble utført med en kombinasjon av røntgenbilder av hele fisk (for bestemmelse av antall ryggvirvler og finnestråler) og visuell identifikasjon av morfologiske nøkkeltrekk.

Populasjonsgenomiske analyser

For å undersøke om sil i tanamunningen er en egen populasjon, ble populasjonsstrukturen for sil innsamlet i Tanamunningen og Alta fjorden bestemt ved hjelp av flere ulike metoder som anvender/ikke anvender genetiske modeller i algoritmene (med modell; NGS admix; Admixture [Skotte et al. 2013; Alexander et al. 2009])/uten modell; DAPC [Jobart et al. 2010]). Vi anvendte ulike metoder, for å oppnå resultater som avspeiler biologisk relevant struktur. For eksempel er det velkjent at algoritmer som bruker genetiske modeller ofte underestimerer antall genetiske klynger i et datasett og dermed utlader biologisk relevant variasjon (se f.eks. Janes et al. 2017). I alle tilfeller ble det mest sannsynlige antallet av populasjoner/klynger (K) i datasettet bestemt uten noen støtte informasjon (for eksempel lokalitet/miljøvariabler). Men vi estimerte også populasjonsstrukturen basert på geografisk opprinnelse av individene, noe som etter vår erfaring gir et mer biologisk relevant resultat. Det ble anvendt en delvis hierarkisk tilnærming, hvor «rene» populasjoner ble fjernet fra analysen underveis, da dette er vist at gi mer statistisk styrke til analysene og et mer biologisk relevant resultat (Väha & Primmer 2006; Väha et al. 2007; Warnock et al. 2010; Janes et al. 2017).

Vi estimerte genetisk differensiering (F_{ST} , Weir & Cockerham, 1984) mellom sil fra Tanamunningen og Altafjorden, mellom de ulike genetiske klyngene identifisert med Admixture, og mellom sil fanget på

lokalitetene i Tanamunningen. F_{ST} – verdiene ble estimert, og testet for signifikans med 10.000 permutasjoner, i Arlequin 3.5 (Excoffier & Lischer 2010).

Vi undersøkte også om det kunne identifiseres SNPs som var utsatt for retningsbestemt naturlig seleksjon, siden slike SNPs vitner om lokale adaptasjoner. Vi brukte FDIST2 modulen (Beaumont & Nichols 1996) i Arlequin 3.5 (Excoffier & Lischer 2010) til å teste hvilke SNPs som var henholdsvis nøytrale og utsatt for retningsbestemt seleksjon. Det bemerkes at resultatene fra alle slike tester skal tolkes med forsiktighet da de er utsatt for beregningsusikkerheter (se for eksempel Narum & Hess 2011). Vi brukte derfor en strikt p-verdi ($p < 0.01$) til differensiering mellom nøytrale SNPs og SNPs utsatt for retningsbestemt seleksjon for å minimere usikkerheten (unngå falske positive). Seleksjonstesten ble også kun utført på sammenligningen mellom sil fra Tanamunningen og Altafjorden, da de genetiske forskjellene mellom sil fanget på lokalitetene i Tanamunningen var for lave til at få meningsfulle resultater fra en slik analyse.

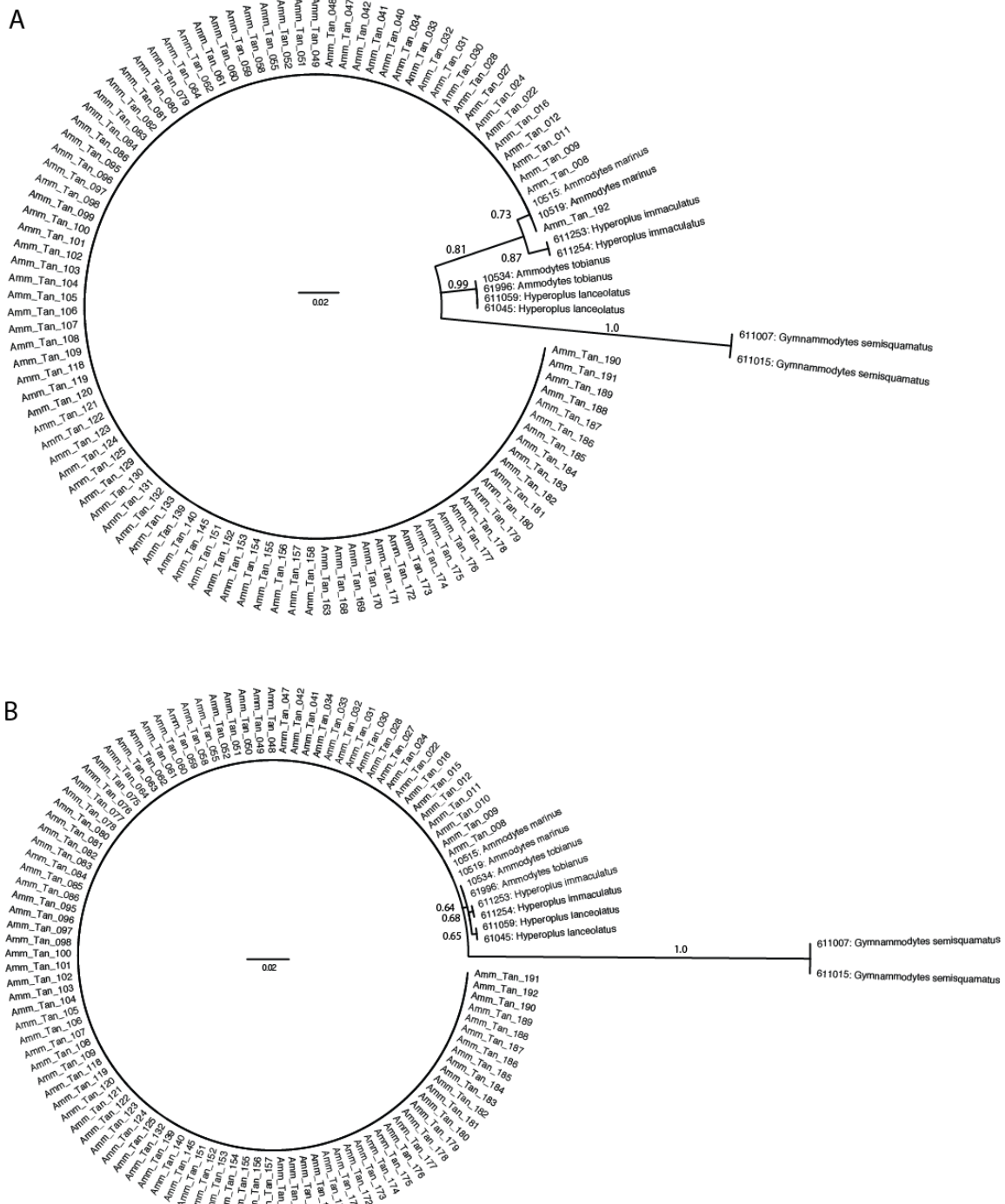
RESULTATER OG DISKUSJON

Hvilken art er av sil finnes i Tana munningen (Oppgave 1)

Artsbestemmelse av sil i Tana munningen ble utført med to ulike metoder: 1. variasjon i et mitokondrielt DNA fragment i Cytochrome oxidase I (COI) genot og et fragment av Rhodopsin (Rho) genot som finnes i det nukleære genomet jvf. protokollen i Thiel & Knebelsberger (2016) og 2. variasjon ved 11.521 SNPs enkelt nukleotid polymorfismer i det nukleære genomet for Tana vs referanse og alle lokaliteter vs. referanse. For detaljer om metode 1. og 2. henvises til Appendiks 1 og 2.

Artsbestemmelse av sil ved hjelp av mitokondrie og nukleære DNA fragmenter

De fylogenetiske analysene av COI sekvensene viste, at alle individene fra Tanaelvens munning ikke var signifikant forskjellige fra referanse individene for *A. marinus* (Figur 2A). De fire andre referansearter grupperte seg der i mot som signifikant forskjellige fra *A. marinus*. Samme grupperingen ble observert i den fylogenetiske analysen for Rho sekvensene (Figur 2B). De fylogenetiske analysene fra denne metoden tilsier at sil prøvene fra Tanaelvens munning er *A. marinus*. Manuell gjennomgang av nukleotidvariasjonene for Rhodopsin genot og sammenligning med resultatene i Thiel & Knebelsberger (2016) bekreftet dette resultatet (Appendiks 1). Resultatet støttes også når sil innsamlet i Oslofjorden og Mandal inkluderes i fylogeniene (Appendiks 1; Bilag 1). Resultatet fra denne tilnærmingen viste altså at sil i Tanamunningen er *A. marinus*, men siden nye data tyder på at sil familien må revideres (enda ikke publisert), var ytterligere analyser nødvendige.



Figur 2. Fylogenetiske treer for COI (A) og Rho (B) for sil innsamlet i Tanaelvens munning (merket Amm_Tan_). Referanseprøver av verifiserte arter av *Ammodytes marinus*, *Ammodytes tobianus*, *Hyperoplus immaculatus*, *Hyperoplus lanceolatus* og *Gymnammodytes semisquamatus* ble inkludert i analysen for å oppnå en mer sikker artsbestemmelse. Figuren viser at sil i Tanaelvens munning formentlig er *A. marinus*, siden det er ikke er signifikant (>50%) forskjell mellom referanse prøvene og sil fra Tana og i tillegg samsvar mellom resultatene for COI og Rho. Figur fra Præbel et al. (2017).

Artsbestemmelse av sil ved hjelp av SNPs

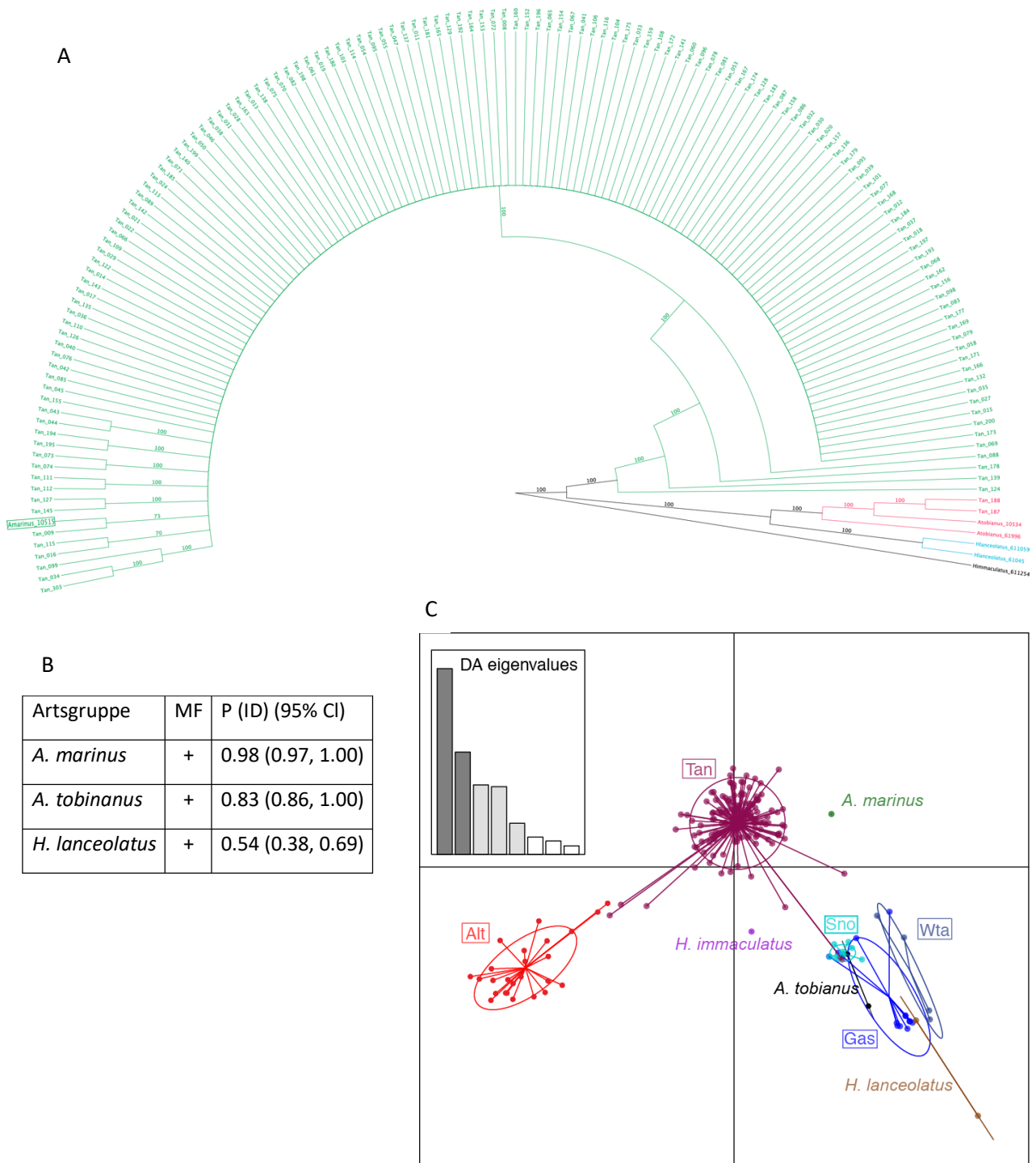
Den fylogenetiske artsbestemmelse av sil fra Tanaelvens munning med 11.521 SNPs grupperte 99% (141 av 143) av individene sammen med referanse individet for *A. marinus* (Figur 3A). Denne gruppen var monofyletisk og signifikant forskjellig fra de andre gruppene i treet (Figur 3B). De to siste individene fra Tanaelvens munning grupperte seg med *A. tobianus*, også denne gruppe var monofyletisk (Figur 3B). Grupperingen ble også testet med en DAPC (Discriminant Analysis of Principal Components) som ikke er avhengig av genetiske modeller. Denne analysen grupperte også sil fra Tanaelvens munning nærmere *A. marinus* enn noen av de andre referansearter inkludert i undersøkelsen (Figur 3C). Dette resultatet ble også støttet når sil fra vest i Tanafjorden, Altafjorden, Oslofjorden og Mandal ble inkludert i analysene (Appendiks 3). Her grupperte 99% (141 av 143) av silen fra Tanaelvens munning, 100 % av silen fra Altafjorden, 50% (2 av 4) av sil fra vest i Tana fjorden, og 23 % (3 av 13) av sil individene fra Oslofjorden, monofyletisk med referansen for *A. marinus*. Samme mønster ses for disse lokalitetene når dataen analyseres med DAPC (Figur 3C).

For å sikre disse analyser fikk vi utført en detaljert morfologisk undersøkelse av tre sil individer innsamlet i Tanamunningen. De undersøkte individer hadde 72-73 ryggvirvler, 62 finnestråler i dorsal finnen, og 32-34 stråler i analfinnen, hvilket tilsier at individene er *A. marinus* (Tabell 2).

Når det kommer til sil innsamlet fra andre geografiske lokaliteter, viste analysene at prøvene fra vest i Tanafjorden var både *A. tobianus* (N=2) og *H. lanceolatus* (N=2) (Figur 3C; Appendiks 3). All sil fra Mandal ble tilordnet til *A. tobianus*. Prøvene av sil fra Oslofjorden viste seg å være både *H. lanceolatus* (N=9), *A. tobianus* (N=1) og *A. marinus* (N=3). Artssammensetningen på de sørlige lokalitetene reflekterer hva som tidligere er rapportert for disse områdene (f.eks. Thiel & Knebelberger 2016; Fietz et al. 2018). Nævdal og Thorkildsen 2002 analyserte artssammensetningen av sil i norsk farvann og konkluderte med at *A. marinus* dominere norsk farvann, med mindre forekomst av *H. lanceolatus*. Studien konkluderte videre at *Gymnammodytes semisquamatus*, *H. immaculatus* og *A. tobianus* ikke forekommer hyppig eller er fraværende fra norsk farvann. Våre resultater viser at sil bestanden formentlig er langt mer artsrik enn hva tidligere er trodd og det bør fortas en nasjonal undersøkelse, med høyoppløste genomiske metoder, for å kartlegge dette. Samlet sett viser alle våre analyser at det primært er *A. marinus* som finnes i Tana munningen. Vi konkluderer derfor: **at *A. marinus* dominerer silbestanden i Tanaelvens munning på denne årstid**, men at andre arter av sil (*A. tobianus* og *H. lanceolatus*) også kan forekomme i området.

Tabell 2. Morfologisk karakteristikk av tre stykk sil individer. REF angir grenseverdier for morfologiske trekk for tre ulike arter av sil (Kilde: P. R. Møller, Københavns Zoologiske Museum, Danmark).

Individ/Referanse	Antall ryggvirvler	Stråler i dorsal finne	Stråler i analfinne
TANA - AMTAN 17008	73	-	34
TANA - AMTAN 17151	72	62	32
TANA - AMTAN 17093	72	62	33
REF – <i>A. marinus</i>	65-75	56-63	29-34
REF – <i>A. tobianus</i>	60-65	51-56	26-31
REF – <i>H. lanceolatus</i>	66-68	52-57	25-29



Figur 3. Fylogenetisk cladogram (A) og tilhørende analyse data for avgrensning av arter (B) av sil fra Tanamunningen. Inkludert er også DAPC graf (C) av sil fra Tana munningen (Tan), Altafjorden (Alt), vest Tana (Wta), Mandal (Sno), og Oslofjorden (Gas). Alle analyser er basert på variasjon ved 11.521 SNPs og det er brukt referanse individer (ZMUC) fra *A. marinus*, *A. tobianus*, *H. Immaculatus* og *H. lanceolatus* til å assistere artsbestemmelsen. I cladogrammet (A) ble *H. lanceolatus* anvendt som utgruppe og styrken på grenene ble testet med 1000 bootstraps. Alle forgreninger hadde en statistisk styrke på 50% eller mer. I (B) viser +/- MF om grupperingen er monofyletisk og P(ID) viser sannsynligheten for rett tilordning av individer med ukjent opphav til denne grupperingen innenfor et 95% konfidensintervall. Et fylogenetisk cladogram av datasettet i DAPC grafen (C) finnes i Appendiks 3.

Populasjonsstruktur (Oppgave 2)

Summering av datagrunnlag

For å undersøke om sil i tanamunningen er en egen populasjon, og eventuelt har genetisk struktur innen Tanamunningen, ble den populasjons genomiske strukturen undersøkt ved hjelp av neste generasjons sekvensering av genomiske målrettede RNA prober. Sekvenseringen av de tre HiSeq4000 baner resulterte i 599.145.913 pair-end reads med 150 basepar leselengde, hvor 582.329.094 pair-end reads (tilsvarer $87,4 \times 10^9$ nukleotider) passerte vår kvalitetsfiltrering. Vi fikk gjennomsnittlig 1.052.068 – 1.551.467 kvalitets DNA lesninger for de 18.247 prober per individ/geografisk lokalitet med <10% manglende data (Tabell 2). Det ble identifisert 27.733 SNPs som var felles for alle individer og etter filtrering og kvalitetsanalyser ble det etablert et markørsett på 11.521 SNPs som ble brukt for videre analyse. Basert på dette markørsett, filtrerte vi også ut et markørsett på 2.293 SNPs, som viste høyere evne til genetisk differensiering ($F_{ST} > 0$) mellom sil fra Tana og Alta, enn markørsettet på 11.521 SNPs. Begge markørsett ble anvendt i analysene for sammenligningen mellom Tana og Alta. For analyser innen Tana elvens munning, ble bare markørsettet på 11.521 SNPs anvendt, da det var for lav genetisk differensiering mellom lokalitetene til å filtrerer et slikt datasett. Videre kan settet fra Tana-Alta sammenligningen ikke anvendes innen Tana, da vi kan tilføye systematisk feiler i resultatet grunnet variasjon i geografiske adskilte populasjoner (ascertainment bias).

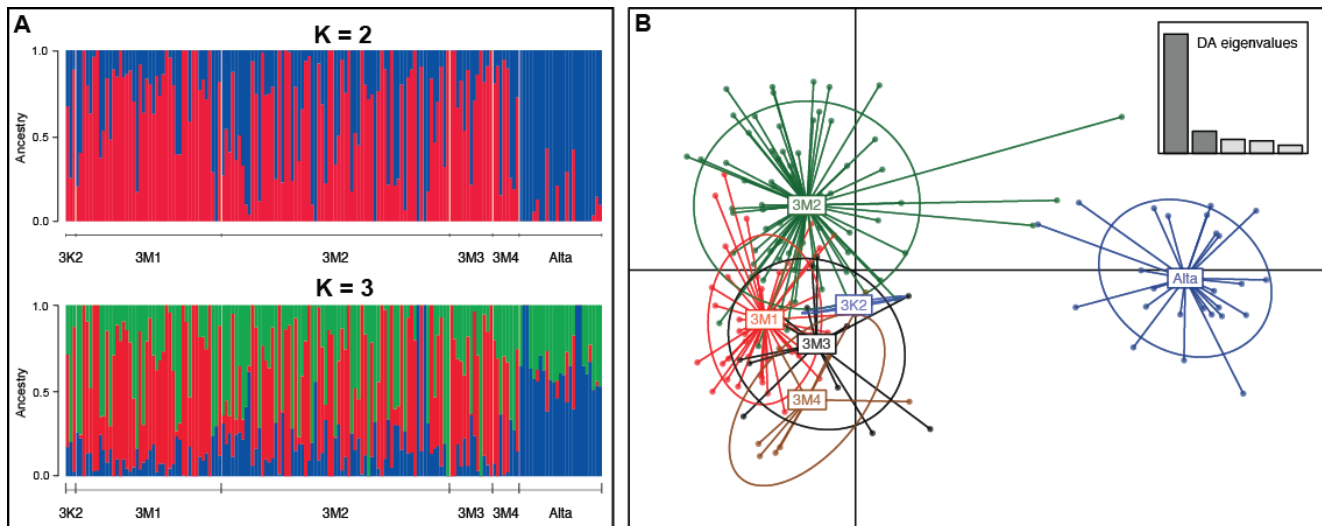
Den genetiske variasjonen (observert [H_o] og forventet [H_e] heterozygositet) var generelt høyere hos sil fra Tana elvens munning (gjennomsnittlig $H_o/H_e = 0.356/0.328$) enn hos sil fra Alta ($H_o/H_e = 0.313/0.293$) (Tabell 2). Det ses også variasjon i den genetiske variasjonen innen sil fra ulike lokaliteter fra Tana. Dette kan stamme fra variasjon i prøvestørrelser, men også være et resultat av at det er prøve tatt sil fra ulike populasjoner på hver lokalitet.

Genetisk sammenligning av sil fra Tana og Alta

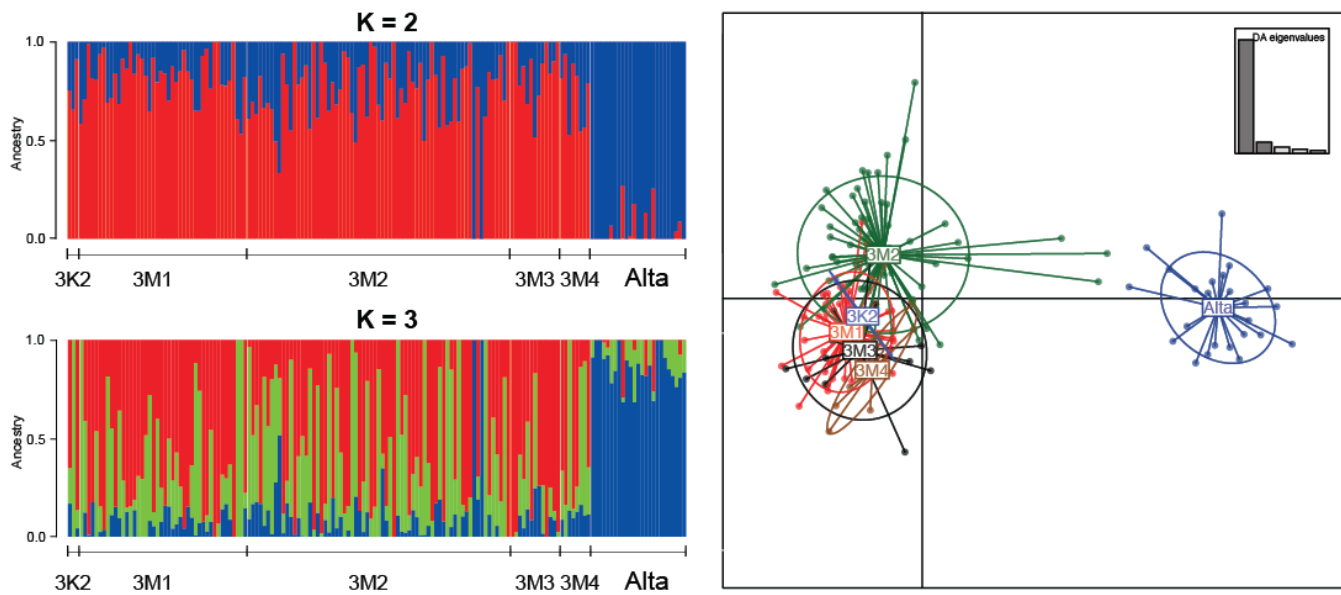
Den populasjonsgenomiske sammenligningen av sil fra Tana og Altafjorden viste at datasettet med 11521 SNPs, til dels, kunne differensiere sil fra Tana fra Altafjorden (Figur 4), men den genetiske forskjellen kunne bare påvises ved statistisk søkning for den ene av de to analyse metoder (Figur 4B, Appendiks 4A, Appendiks 5). Dette er en vanlig observasjon i genomiske og genetiske studier av marine fisk og andre organismer (se for eksempel, Benestan et al. 2015; Rodríguez-Ezpeleta et al. 2016). Disse studier har også vist at DAPC metoden gir mest biologisk relevant informasjon og struktur sammenlignet med andre analyse metoder.

Tabell 2. Oversikt over antall prøver initialt sekvensert (N), gjennomsnittlig antall lesninger (reads) per lokalitet, antall prøver per lokalitet som passerte kvalitetsvalideringen (N_{qseq}) og dermed ble inkludert i analysen, og observert (H_o) og forventet (H_e) heterozygositet per lokalitet.

Sted	Kode	N	Gjennomsnittlig antall reads +/- SD	% manglende Data	N_{qseq}	H_o	H_e
Tana 3M1	3M1	59	1327990 +/- 666146	<10%	47	0.311	0.277
Tana 3M2	3M2	89	1488356 +/- 635994	<10%	71	0.267	0.257
Tana 3M3	3M3	16	1117848 +/- 638077	<10%	13	0.331	0.307
Tana 3M4	3M4	10	1184071 +/- 327441	<10%	8	0.485	0.456
Tana 3K2	3K2	4	1052068 +/- 592566	<10%	4	0.387	0.343
Alta	Alta	27	1551467 +/- 698357	<10%	25	0.313	0.293



Figur 4. Analyse av populasjonsstrukturen til sil med 11521 SNPs. Figur A viser Admixture analyse av sil fra Tanamunningen og Alta fjorden, i tilfellet hvor individene stammer fra hhv. to ($K=2$) og tre ($K=3$) populasjoner. I figur A er hver vertikal bjelke et individ og andelen av farge viser genetisk tilhørighet til de ulike populasjoner, slik at individer som viser >0.9 tilhørighet til en populasjon tilhører denne populasjonen. Den statistiske søkningen for antall populasjoner i datasettet (cross validation) viste at denne tilnærmingen sannsynligvis finner $K=1$ (Appendiks 4A). DAPC (Discriminant Analysis of Principal Components), Figur B, viste en mer klar separasjon av populasjonene og for denne analysemetoden viste den statistiske søkningen for antall populasjoner at datasettet sannsynligvis inneholder $K=2$ populasjoner. Forkortelser av populasjonsnavn jvf. Tabell 1.



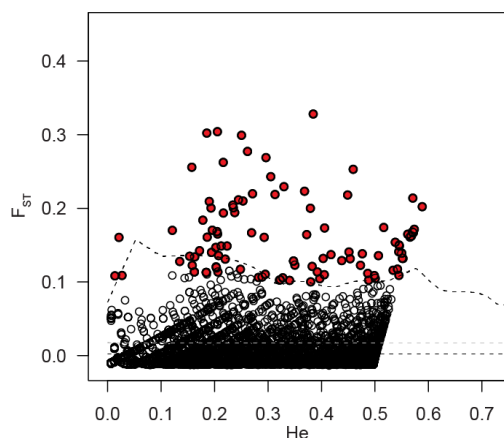
Figur 5. Analyse av populasjonsstrukturen til sil med 2293 SNPs med høy evne til genetisk differensiering. Figur A viser admixture analyse av sil fra Tanamunningen og Alta fjorden, i tilfellet hvor individene stammer fra hhv. to ($K=2$) og tre ($K=3$) populasjoner. I figur A er hver vertikal bjelke et individ og andelen av farge viser genetisk tilhørighet til de ulike populasjoner, slik at individer som viser >0.9 tilhørighet til en populasjon tilhører denne populasjonen. Den statistiske søkningen for antall populasjoner i datasettet (CV, se Appendiks 4B) viste at denne tilnærmingen sannsynligvis finner $K=2$ populasjoner. DAPC (Discriminant Analysis of Principal Components), Figur B, viste også en klar separasjon av sil fra Tana og Alta. Også for denne analysemetoden viste den statistiske søkningen for antall populasjoner at datasettet sannsynligvis inneholder $K=2$ populasjoner. Forkortelser av populasjonsnavn jvf. Tabell 1.

Når vi replikerte analysen med datasettet med 2293 SNPs, med høy genetisk differensieringsevne, ble denne struktur signifikant støttet av begge de analytiske metodene (Figur 5, Appendiks 4B). Her ble sil fra Tana klart genetisk differensiert fra sil fra Altafjorden. Det er også klart, at det finnes genetisk rene individer ($q > 0.9$) i Tana elvens munning, samt individer som har noe ($q < 0.9$ & > 0.5) innblanding av genetisk materiale fra Alta individene. Om denne innblandingen er nylig eller er signaturer fra tidspunktet for adskillelsen av de to populasjonene kan ikke fastslås innen denne rapporten. Det skal også bemerkes at både for datasettet med 11521 og med 2293 SNPs, fant Admixture analysen genetiske rene individer for en tredje populasjon innen Tana elvens munning (Figur 4 og 5). Selv om det ikke er statistisk støtte for tre populasjoner i datasettet, er det vår og andres erfaring at slik struktur ofte har biologisk relevans (For eksempel Janes et al. 2017).

Vi estimerte heretter den genetiske forskjellen mellom sil fra Tana elvens munning og sil fra Alta Fjorden og fikk en F_{ST} – verdi på 0.01 som var høy signifikant ($P < 0.001$). Hvis den genetiske grupperingen som ble funnet i Admixture analysen ($K=2$) (og ikke lokalitet) ble lagt til grunn for estimeringen av F_{ST} – verdi økte denne til 0.07 ($P < 0.001$). Denne genetiske forskjellen er nok så høy for marine arter, når det sammenlignes over små geografiske avstander. At den genetiske forskjellen øker når den baseres på genetisk gruppering ($F_{ST} = 0.07$), og ikke lokaliteter ($F_{ST} = 0.01$), støtter at det finnes migranter og/eller litt genflytt mellom de to lokaliteter. Men dette kan også være en genetisk signatur som stammer fra oppsplittelsen av de to sil populasjonene. Hvilke av disse mekanismer som er biologisk relevant, vil kreve ytterligere analyser av dataen og inkludering av et større datamateriale.

Samlet sett viste populasjonsstruktur analysene, **at sil i Tana elvens munningen er signifikant genetisk forskjellig sil fra Altafjorden**. Dette resultatet ble understøttet av tre ulike genetiske analyse metoder.

Vi testet i tillegg om det var genetiske markører som var utsatt for retningsbestemt seleksjon i sammenligningen mellom sil fra Altafjorden og Tana elvens munning. Ut av de 11.521 genetiske markører, ble 103 av markørene estimert til å være signifikant ($P < 0.01$) påvirket av retningsbestemt seleksjon (Figur 6). Flere av disse markørene hadde F_{ST} - verdier høyere enn 0.2, noe som er høyt for marine arter. Fremtidige undersøkelser bør identifisere hvilke fysiologiske og/eller fenotypiske egenskaper som er knyttet til de 103 markører, da slik informasjon kan avdekke hvilke faktorer i miljøet som er viktige for opprettholdelse av den genetiske strukturen. Denne test viste at det høyst sannsynlig er lokale tilpasninger som driver den genetiske strukturen mellom de to lokalitetene.



Figur 6. Identifikasjon av SNP markører utsatt for retningsbestemt naturlig seleksjon i sil fra sammenligningen mellom Tana og Alta. Testen ble utført med FDIST2 modulen (Beaumont & Nichols 1996) i Arlequin 3.5 (Excoffier & Lischer 2010). 103 SNPs ble funnet signifikant ($P < 0.01$) påvirket av retningsbestemt seleksjon (røde sirkler) og 11418 SNPs ble funnet å være nøytrale eller under balanserende seleksjon (svarte åpne sirkler).

Genetisk sammenligning av sil fra ulike områder i Tana elvens munning

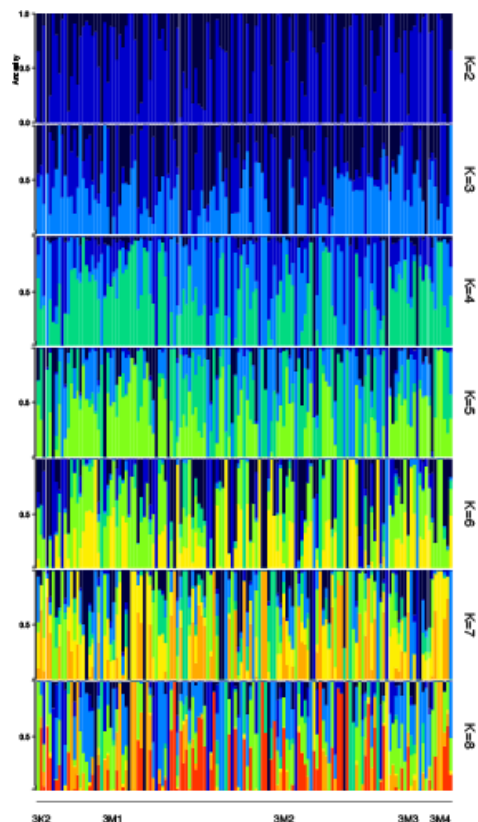
Den genetiske strukturen innen de ulike innsamlingsområdene i Tanamunningen ble også undersøkt. Den statistiske søkingen etter struktur viste at det området kan betraktes som en populasjon (data ikke vist). Men, litteraturen har også vist, at i datasett med lav genetisk differensiering mellom populasjoner (F_{ST} – verdier < 0.01), feiler denne tilnærmingen i å vise relevant biologisk struktur (for eksempel, Banestan et al. 2015). Vi undersøkte derfor populasjonsstrukturen med NGS admix som ikke inkluderer statistisk søking etter antall populasjoner, men visualiserer biologisk relevant struktur. Denne analysen viste at det formentlig er utbredt genetisk struktur hos sil i Tanamunningen, men at strukturen er genetisk svak (Figur 7). For eksempel ses det «genetisk rene» individer ($q > 0.9$) alle mulige situasjoner hvor datasettet inneholder 2 til 8 populasjoner (Figur 7). Vi undersøkte også de genetiske forskjellene mellom prøvetakningslokalitetene i Tanamunningen. Denne analysen viste at sil fra område 3M3 og 3M4 var signifikant ($P < 0.05$) forskjellige, med en F_{ST} – verdi på 0.001 (Tabell 3). Selv om dette er en lille genetisk forskjell kan den være biologisk relevant, hvilket er vist i flere andre studier av relevante marine arter (for eksempel, Drinan et al. 2018; Fietz et al. 2018). Fremtidige undersøkelser bør også inkludere sil innsamlet fra ulike lokaliteter i Tana fjorden og fra nabo fjorder, for å fastslå om denne strukturen er en signatur på svak genetisk struktur i Tanafjorden, altså om sil i fjorden utgjør en, eller flere, populasjoner.

Samlet sett viser disse analysene, **at det må forventes at det er genetisk struktur for sil i de ulike områdene innen Tana elvens munning**, men dette må verifiseres med ytterligere analyser og større prøvestørrelser. Det bør også undersøkes om denne strukturen er koblet til miljø og biologiske egenskaper til sil, for eksempel ulike gytetidspunkter.

Tabell 3. F_{ST} - verdier for parvise sammenligninger av genetiske forskjeller mellom sil innsamlet i ulike geografiske områder i Tana elvens munning. Områdene er vist i Figur 1. Bare en sammenligning var signifikant (* = $P < 0.05$) etter korreksjon for paralleltester.

	3M1	3M2	3M3	3K2	3M4
3M1	-				
3M2	0.000	-			
3M3	-0.002	-0.001	-		
3K2	-0.008	-0.004	-0.007	-	
3M4	-0.002	0.001*	-0.004	-0.009	-

Figur 7. NGS admix analyse av sil fra Tana elvens munning basert på datasettet med 11521 SNPs. I figuren er hver vertikal bjelke et individ og andelen av farge viser genetisk tilhørighet til de ulike populasjoner (K), slik at individer som viser > 0.9 tilhørighet til en populasjon tilhører denne populasjonen. Denne tilnærmingen inkluderer ikke statistiske søkingen for antall populasjoner, men er heller en biologisk relevant metode for å observere populasjonsstruktur i datasettet.



Sett i et forvaltningsmessig perspektiv, betyr resultatene fra analyse av populasjonsstrukturen (Oppgave 2), at sil i Tana elvens munning er særegen for dette området (sammenlignet med sil fra Altafjorden), og resultatene støtter at sil fra disse to lokaliteter formentlig ikke gyter i sammen. Dette resultatet hadde vært betydelig styrket hvis sil fra andre lokaliteter i Tana fjorden og nabo fjorder var blitt inkludert i analysen, siden vi ikke kan fastslå om sil i Tana elvens munningen er en del av en større populasjon av sil, for eksempel i Tana fjorden. Hvis sil i hele Tanafjorden utgjør en samlet populasjon, bør det forventes at sil fra fjorden vil kunne rekolonisere området som skal mudres, dersom den lokale silbestanden påvirkes. Resultatene viste også at naturlig retningsbestemt seleksjon har medvirket til den genetiske strukturen mellom sil fra Tana elvens munningen og Altafjorden. Dette resultatet er minst like viktig som selve den genetiske strukturen, siden retningsbestemt seleksjon er basisen for akkumulering av biologisk tilpasning til miljøet i en populasjon. Slike lokale tilpasninger kunne for sil i Tana elvens munningen for eksempel være salt og temperaturløselighet, timing av gyttidspunkt i forhold til klimatiske forhold og i forhold til predasjonspress fra for eksempel fugl, og biologiske prosesser knyttet til veksthastighet. Som nevnt overfor, bør eventuelle fremtidige undersøkelser analysere dataen som ble opparbeidet i denne undersøkelsen med henblikk på å identifisere hvilke egenskaper markørerne som var utsatt for seleksjon er knyttet til, da slik viten vil gi forvaltningen innblikk i hvilke biologiske og miljømessige egenskaper som er viktige for sil i Tana elvens munning.

Videre bør fremtidige undersøkelser inkludere sil innsamlet fra ulike lokaliteter i Tana fjorden og fra nabo fjorder, for å fastslå om sil Tanafjorden utgjør en, eller flere, populasjoner. Det bør også undersøkes i hvilken grad de ulike arter av sil benytter Tana elvens munning til gyting på ulike årstider og i hvilken grad de ulike arter interagerer økologisk. For eksempel, er det observert genetisk differensiering mellom vår og høst gytere av *A. tobianus* i Østersjøen (Fietz et al. 2018) og ynglene fra disse gytinger inngår i dietten til andre arter av i området (Pers. Com. P. R. Møller). Inkludering av fenotypisk informasjon (for eksempel, lengde fordeling og alder) fra individene som vi analyserte hadde derfor vært ønskelig.

Samlet sett, konkluderer vi at **sil i Tana elvens munningen er *A. marinus***, men *A. tobianus* og *H. lanceolatus* kan forekomme i området. **Sil fra Tana elvens munning er genetisk ulik sil fra Altafjorden**, og analysene tyder på at disse forskjeller er drevet av lokale tilpasninger. Siden lokale tilpasninger oppstår som en konsekvens av blant annet miljøfaktorer, fødespesialisering og habitatvalg er dette et forvaltningsmessig viktig resultat. Dette spesielt fordi det tar generasjoner å akkumulere slike genetiske tilpasninger i en populasjon. Det skal påpekes at de populasjonsgenetiske analyser som skulle belyse om Tana er en unik lokal populasjon, bare er basert på sammenligning med en enkelt annen lokalitet (Alta) og ikke med prøver fra andre lokaliteter i Finnmark. Vi mener dog ikke at det har betydning for analysen og tolkningen av lokale tilpasninger, siden tilpasningene er unike for denne sammenligningen av lokaliteter. Men det kan ikke konkluderes om tilpasningene for sil fra Tana elvens munning deles med sil fra f.eks. andre områder i Tanafjorden, siden dette materialet ikke er inkludert i undersøkelsen.

Resultatene tydet også på at det er genetisk struktur hos sil i Tanaelvens munning. Slik lokal struktur er normalt drevet av sterke biologiske prosesser (partnervalg, ulike gyttidspunkter, homing, etc), men også sterk naturlig seleksjon/tilpasninger til leveområde og predasjon kan danne slike strukturer. For å oppnå statistisk styrke i slike situasjoner, kreves et større antall individer og eventuelt bruk av

maskinlærings algoritmer. Selv om resultatet om lokal struktur er indikativt, er vår vurdering at det likevel ikke helt bør forkastes i den forvaltningsmessige vurderingen.

Med vennlig hilsen,

Tromsø 14/5-2019,  (På vegne av prosjektgruppen)

Kim Præbel, PhD
Norges fiskerihøgskole, UiT Norges Arktiske Universitet, 9037-Tromsø
Mobil: 45 43 64 15, e-post: kim.praebel@uit.no

Referanser

- Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* 19:1655–1664
- Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society London Series B* 263:1619-1626
- Benestan L, Gosselin T, Perrier C, Sainte-Marie B, Rochette R, Bernatchez L (2015) RAD genotyping reveals fine-scale genetic structuring and provides powerful population assignment in a widely distributed marine species, the American lobster (*Homarus americanus*). *Molecular Ecology* 24:3299-3315
- Drinan, DP, Gruenthal, KM, Canino, MF, Lowry, D, Fisher, MC, Hauser, L (2018) Population assignment and local adaptation along an isolation-by-distance gradient in Pacific cod (*Gadus macrocephalus*). *Evolutionary Applications* 11:1448-1464
- Excoffier L, Lischer HEL (2010) Arlequin suite ver. 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* 10:564-567
- Fietz K, Hintze COR, Skovrind M, Nielsen TK, Limborg MT, Krag M, Palsboll PJ, Hansen LH, Møller PR, Gilbert MTP (2018) Mind the gut: genomic insights to population divergence and gut microbial composition of two marine keystone species. *Microbiome* 6:82
- Janes JK, Miller JM, Dupuis JR, et al. (2017) The K = 2 conundrum. *Molecular Ecology* 26:3594-3602
- Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics* 11:94
- Kirkemoen O, Colman JE (2017) Feltnotat - sampling av sil innenfor og utenfor Leirpollen, Tanafjord. Data fra seks tokt mars-juli 2017. NRAS-notat 2017-08-15
- Korneliussen TS, Albrechtsen A, Nielsen R (2014) Angsd: analysis of next generation sequencing data. *BMC Bioinformatics* 15:1-13
- Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis version 7.0. *Molecular Biology and Evolution* 33:1870-1874
- Narum SR, Hess JE (2011) Comparison of F_{ST} outlier tests for SNP loci under selection. *Molecular Ecology Resources* 11:184-94
- Nævdal G, Thorkildsen S (2002) Genetic studies on species composition and population structure of sand eels (Genera : *Ammodytes*, *Hyperoplus* and *Gymnammodytes*) in Norwegian waters. *Journal of Applied Ichthyology* 18:124-126
- Præbel K, Andrews A, Bhat S (2017) Del-rapport I: Artssammensetning og lokal populasjonsstruktur av sil i Tanafjorden. Norges fiskerihøgskole, UiT Norges Arktiske Universitet, 9037-Tromsø. (Se Appendiks I).
- Rodríguez-Ezpeleta N, Bradbury IR, Mendibil I, Álvarez P, Cotano U, Irigoien X (2016) Population structure of Atlantic mackerel inferred from RAD-seq-derived SNP markers: effects of sequence clustering parameters and hierarchical SNP selection. *Molecular Ecology Resources* 16:991-1001

Skotte L, Korneliussen TS, Albrechtsen A (2013) Estimating individual admixture proportions from next generation sequencing data. *Genetics* 195:693-702

Thiel R, Knebelsberger T (2016) How reliably can northeast Atlantic sand lances of the genera *Ammodytes* and *Hyperoplus* be distinguished? A comparative application of morphological and molecular methods. *ZooKeys* 617: 139-164

Väha JP, Primmer CR (2006) Efficiency of model-based Bayesian methods for detecting hybrid individuals under different hybridization scenarios and with different numbers of loci. *Molecular Ecology* 15:63-72

Väha JP, Erkinaro J, Niemelä E, Primmer CR (2007) Life-history and habitat features influence the within-river genetic structure of Atlantic salmon. *Molecular Ecology* 16:2638-2654

Warnock WG, Rasmussen JB, Taylor EB (2010) Genetic clustering methods reveal bull trout (*Salvelinus confluentus*) fine-scale population structure as a spatially nested hierarchy. *Conservation Genetics* 11:1421-1433

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population-structure. *Evolution* 38:1358-1370

Appendiks

Appendiks 1:

Del-rapport I

«Artssammensetning og lokal populasjonsstruktur av sil i Tanafjorden»

Personell involvert i oppdraget: Førsteamanuensis Kim Præbel, PhD (prosjektleder),
Forskningsassistent MSc Adam Andrews, Bioinformatikker Shripathi Bhat, PhD

I forbindelse med prosjektet "Artssammensetning og lokal populasjonsstruktur av sil i Tanafjorden", ble det ønsket å få en preliminær avklaring på hvilke arter av sil som finnes i munningen av Tanaelven (oppgave 1 i prosjektet). Det presiseres, at resultatene som presenteres her er utført med en metode som har lavere oppløselighet enn den genomiske metoden som ble beskrevet i tilbudsbeskrivelsen for prosjektet. Dette skyldes, at vi ikke kan gjennomføre det genomiske arbeidet før vi har det komplette komparative materialet av sil fra Neiden, Alta og andre nærliggende lokaliteter som skal inngå i prosjektet (Oppgave 1+2).

For å imøtekomme dette ønsket har vi analysert sil innsamlet av Naturrestaurering AS i 2017. Et referansemateriale bestående av fem verifiserte arter av sil ble stilt til rådighet for prosjektet av Københavns Zoologiske museum (ZMUC) (Tabell 1). Det ble fra Multiconsult AS og Naturrestaurering AS side stilt sil fra hhv. Mandal og Oslofjorden til rådighet til sammenligning.

Vi isolerte DNA fra 125 individer til denne undersøkelse (Tabell 1). DNA kvaliteten var generelt god og DNA analysene ble gjennomført vha. protokollen i Thiel & Knebelberger (2016). Det ble oppkopiert og sekvensert to DNA fragmenter; det mitokondrielle Cytochrome oxidase I (COI) og et fragment av Rhodopsin (Rho) genet som finnes i det nukleære genomet. Sekvensering ble utført i genetikklaboratoriet, NFH, UiT, og på avdeling for medisinsk genetikk, UNN, Tromsø.

Sekvensene ble trimmet og kvaliteten verifisert i MEGA 7 (Kumar et al. 2016). Sekvensene ble sammenlignet ved bruk av MUSCLE algoritmen og Tamura-Nei modellen implementert i MEGA 7. Fylogenetiske trær ble konstruert ved hjelp av Maximum-Likelihood tilnærmingen med inkludering av komplement-slett alternativet for manglende data. Støtte for divergens mellom arter ble funnet ved hjelp av 1000 permuteringer og *G. semisquamatus* ble brukt som utgruppe for å styrke fylogeniene. Alle enkelt nukleotid polymorfismer i Rho genet ble manuelt verifisert. Analysene med utført med og uten sil materialet fra Oslofjord og Mandal.

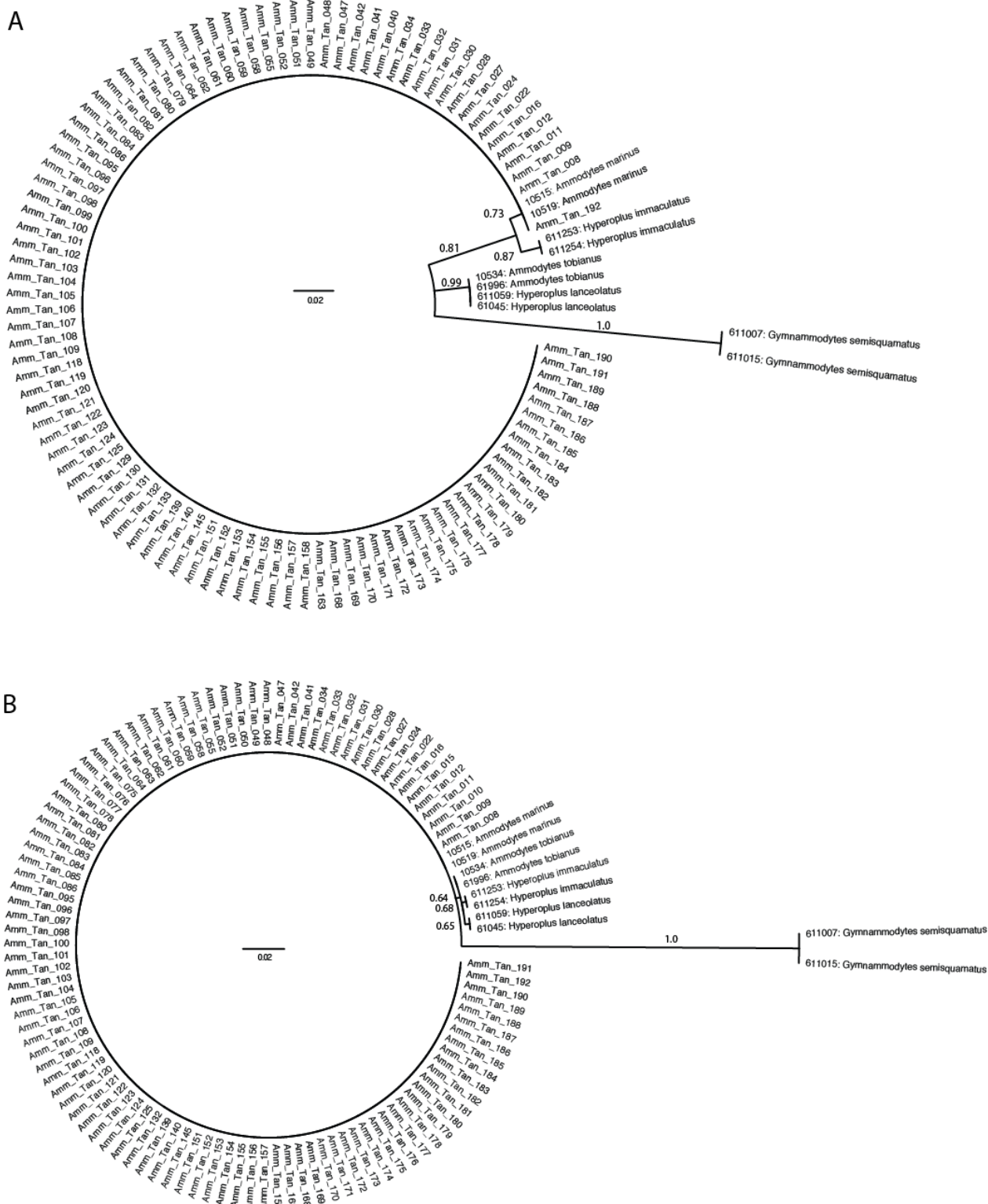
Tabell 1. Oversikt over referanse materiale og viltfanget sil brukt til påvisning av artssammensetningen av sil i Tanaelvens munning.

Art	ZMUC nummer eller område	N
Storsil (<i>Hyperoplus lanceolatus</i>)	61045 & 611254	2
Uplettet storsil (<i>Hyperoplus immaculatus</i>)	611253 & 611254	2
Havsil (<i>Ammodytes marinus</i>)	10515 & 10519	2
Småsil (<i>Ammodytes tobianus</i>)	10534 & 61996	2
Glattsil (<i>Gymnammodytes semisquamatus</i>)	611007 & 611015	2
Ukjent	Tana, 3M1	36
Ukjent	Tana, 3M2	40
Ukjent	Tana, 3M3	16
Ukjent	Tana, 3M4	10
Ukjent	Tana, 3K2	4
Ukjent	Mandal	2
Ukjent	Oslo fjorden	7

De fylogenetiske analysene av COI sekvensene viste, at alle individene fra Tanaelvens munning ikke var signifikant forskjellige fra referanse individene for *A. marinus* (Figur 1A). De fire andre referansearter grupperte seg der i mot som signifikant forskjellige fra *A. marinus*. Samme grupperingen ble observert i den fylogenetiske analysen for Rho sekvensene (Figur 1B). De fylogenetiske analysene tilsier at sil prøvene fra Tanaelvens munning er *A. marinus*. Manuell gjennomgang av nukleotidvariasjonene for Rhodopsin genet og sammenligning med resultatene i Thiel & Knebelsberger (2016) bekreftet dette resultatet (Tabell 2). Resultatet støttes også når sil innsamlet i Oslofjorden og Mandal inkluderes i fylogeniene (**Bilag 1**). Det er derfor sannsynlig at *A. marinus* dominerer silbestanden i Tanaelvens munning på denne årstid, og at sil fra Oslofjorden og Mandal sannsynligvis var henholdsvis *A. tobianus* og *H. lanceolatus* (Appendiks 1, Tabell 3).

Tabell 2. Enkelt nukleotid polymorfismer (SNPs) i Rhodopsin genet som ble brukt til å differensiere de ulike arter av sil. Som det ses av tabellen er hver haplotype (kombinasjon av SNPs) diagnostisk for de fire ulike arter av sil. Sekvensen for *Gymnammodytes semisquamatus* inneholdt for mange SNPs til å kunne inkluderes her.

Art	Posisjon 33	Posisjon 383	Posisjon 410
<i>Ammodytes marinus</i>	A	G	C
<i>Ammodytes tobianus</i>	A	G	A
<i>Hyperoplus lanceolatus</i>	T	G	A
<i>Hyperoplus immaculatus</i>	A	A	A



Figur 1. Phylogenetiske treer for COI (A) og Rho (B) for sil innsamlet i Tanaelvens munning. Referanse prøver av verifiserte arter av *Ammodytes marinus*, *Ammodytes tobianus*, *Hyperoplus immaculatus*, *Hyperoplus lanceolatus* og *Gynnammodytes semisquamatus* ble inkludert i analysen for å oppnå en mer sikker artsbestemmelse. Figuren viser at sil i Tanaelvens munning formentlig er *Ammodytes marinus*, siden det er ikke er signifikant (>50%) forskjell mellom referanse prøvene og sil fra Tana og i tillegg samsvar mellom resultatene for COI og Rho.

Tabell 3. Oppsummering av artsidentifikasjon av ukjent materiale. Se Figur 1 og Appendiks 1 for de fylogenetiske analyser som ligger til grunn for denne tabell.

Art	Område	Sekvenserings suksess (COI/Rho), % individer
Havsil (<i>Ammodytes marinus</i>)	Tana, 3M1	94/97
Havsil (<i>Ammodytes marinus</i>)	Tana, 3M2	95/95
Havsil (<i>Ammodytes marinus</i>)	Tana, 3M3	94/100
Havsil (<i>Ammodytes marinus</i>)	Tana, 3M4	100/100
Havsil (<i>Ammodytes marinus</i>)	Tana, 3K2	100/100
Storsil (<i>Hyperoplus lanceolatus</i>)	Mandal	66/100
Småsil (<i>Ammodytes tobianus</i>)	Oslo fjorden	100/86

Thiel & Knebelsberger (2016) viste, at ved å sammenholde nukleotid variasjonen i COI og Rho, kan inntil fire arter av sil identifiseres. Men materialet av sil inkludert i deres studie var begrenset og de fylogenetiske treer strukturerte derfor ikke artene med sikkerhet (ikke monofyletisk). Helgenomiske undersøkelser har senere vist at artene innen *Ammodytes* og *Hyperoplus* utgjør et artskompleks og at artene innen disse familiene bør revideres. Selv om resultatene presentert her for silen i Tanaelvns munning entydig peker på at arten er *A. marinus*, bør dette verifiseres med de avtalte genomiske undersøkelser.

I mangel av komparativt materiale av sil fra Neiden, Alta og andre nærliggende bestander av sil kan vi heller ikke si noe om denne bestanden er lokaltilpasset og eventuelt genetisk segregert fra andre bestander.

Med vennlig hilsen,

Tromsø 31/8-2017,  (På vegne av prosjektgruppen)

Kim Præbel, PhD.

Norges fiskerihøgskole, UiT Norges Arktiske Universitet, 9037-Tromsø

Mobil: 45 43 64 15, e-post: kim.praebel@uit.no

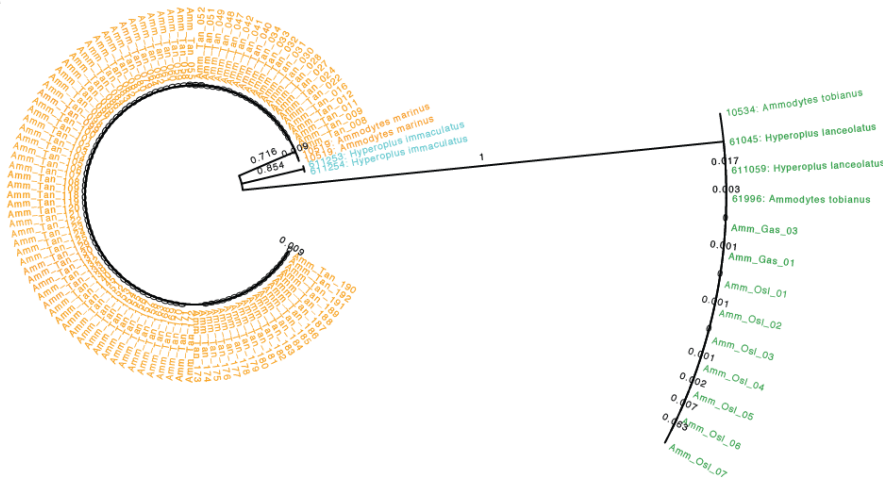
Referanser

Kumar S, Stecher G, Tamura K (2016). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0. *Molecular Biology and Evolution* 33:1870-1874

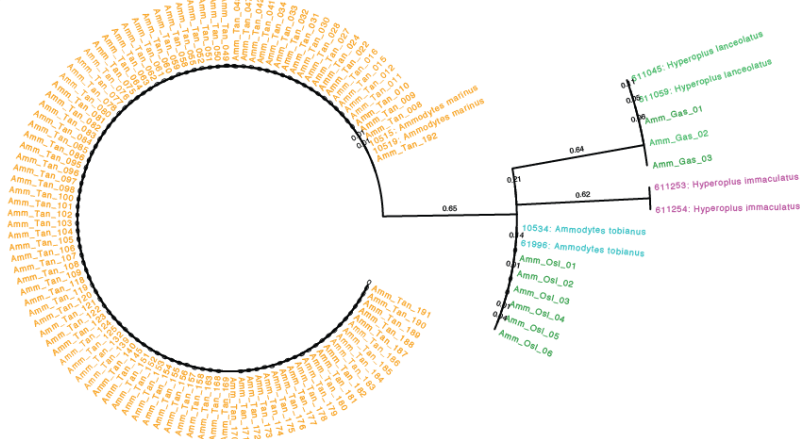
Thiel R, Knebelsberger T (2016). How reliably can northeast Atlantic sand lances of the genera *Ammodytes* and *Hyperoplus* be distinguished? A comparative application of morphological and molecular methods. *ZooKeys* 617: 139-164

Bilag 1. (Del-Rapport I). Treer for COI (A) og Rho (B) for sil innsamlet i Tanaelvens munning, Oslo fjorden og i Mandal. Referanse prøver av verifiserte arter av *Ammodytes marinus*, *Ammodytes tobianus*, *Hyperoplus immaculatus* og *Hyperoplus lanceolatus*, ble inkludert i artsbestemmelsen. Figuren viser at silen i Tanaelvens munning formentlig er *Ammodytes marinus*, siden det ikke er signifikant forskjell mellom sil fra denne lokalitet og referanse materialet for *A. marinus*. Dette støttes av at det er samsvar mellom analysene for COI og Rho. For COI (A) kunne sil fra Oslofjorden og Mandal ikke tilordnes art. Variasjon i Rho genet (B) tilsier at de 7 sil fra Oslo fjorden er småsil (*Ammodytes tobianus*) og at de 3 sil fra Mandal er Storsil (*Hyperoplus lanceolatus*).

A



B



Appendiks 2. Genomisk metode og arbejdsgang

Laboratory protocol

Illumina compatible next generation sequencing libraries were built using NEBnext Ultra DNA Library prep kit (NEB) for Illumina.

Probes for in solution sequence capture for targeted high-throughput sequencing (in short: target enrichment assay) were designed from in hand GBS data (CPH). 18247 RNA probes were produced by a commercial manufacturer (MyBaits, Mycroarray, MI). Pooled DNA libraries and probes were hybridized at 65 °C for 18-20 hours, following the Mybaits capture protocol Version 3,02. Captured libraries were sequenced paired-end in three flow cell lanes of a HiSeq 4000 at Novogene (Hongkong).

Quality and filtering of NGS data

De-multiplexed Illumina sequence data were processed using various bioinformatic tools (Fastqc, AfterQC, bbmap) to check and remove bad quality reads (presence of adapters, more than 10% Ns, low quality reads, Qscore less than 10). QC passed reads were further mapped against the probe set (18247 probes) with bwa-mem (bwa-0.7.17) using default settings. Mapped reads were further quality checked by filtering for mapping quality and PCR duplicates using samtools. Angsd (version: 0.921) (Korneliussen et al, 2014) was used to call the single nucleotide polymorphisms (SNPs) from the mapped reads using Minor Allele Frequency (MAF) of 10% and allowing only 20% missing data. Further, an in-house script was used to keep only one SNP per tag to avoid linkage disequilibrium between neighboring SNPs.

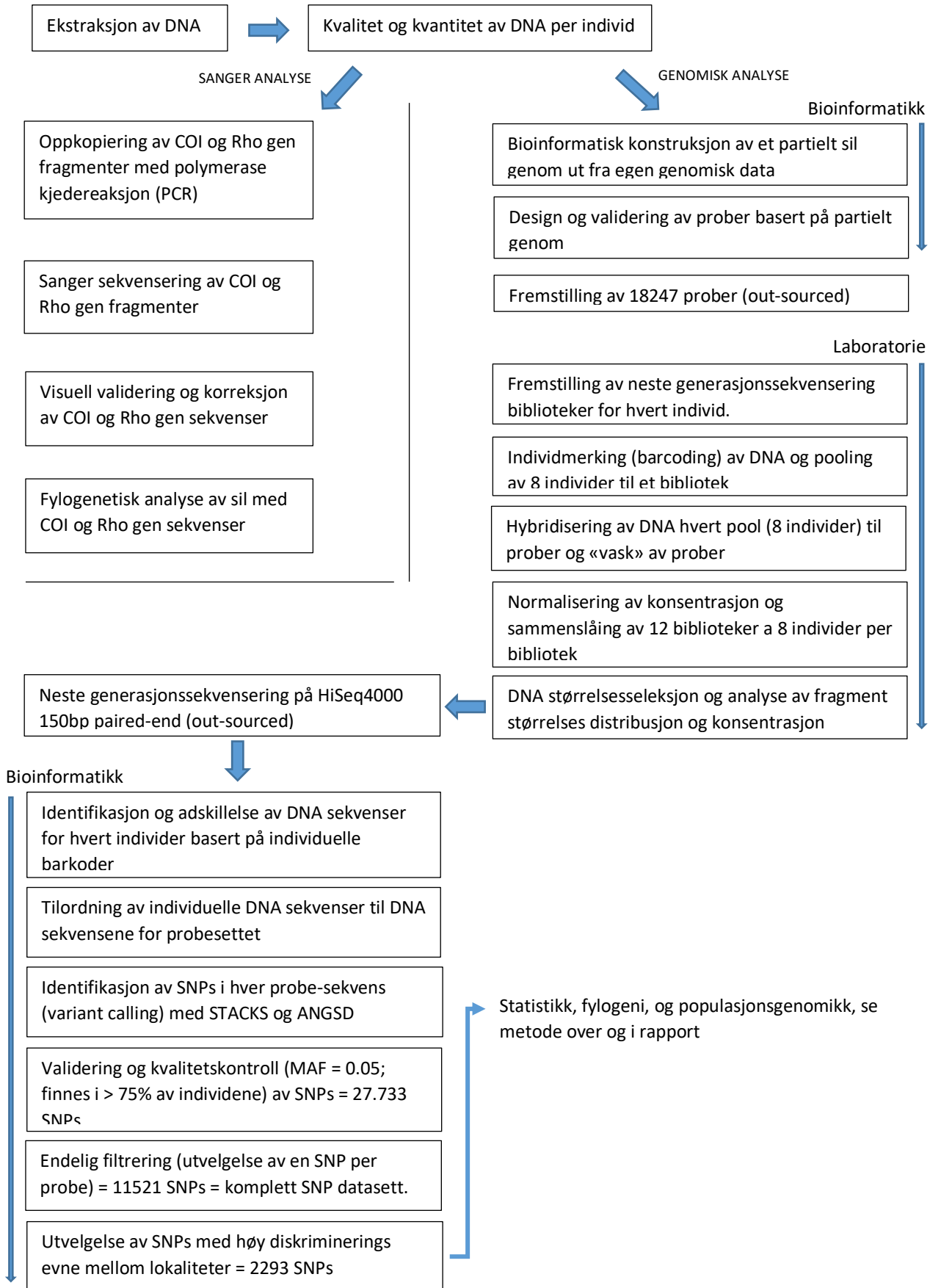
Species taxonomy

The 11521 SNPs were aligned as a single sequence file and imported into Geneious-Prime. We used neighbor-joining trees to infer taxonomic relationships and tested support for branches with 1000 bootstraps. We tested for monophyly using the Masters and Ross species delimitation plugin in Geneious.

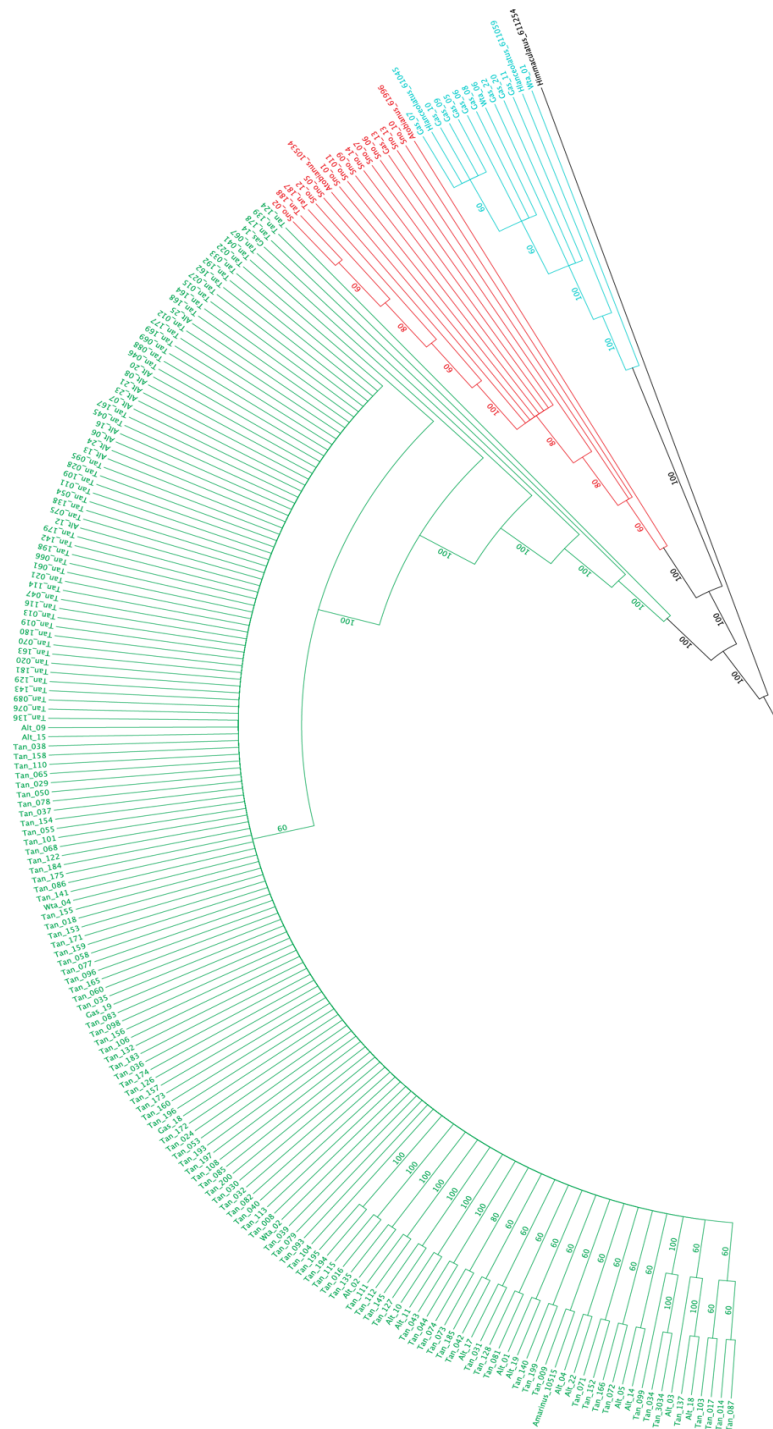
Population genomic analyses

We used a model based approach as implemented in Admixture (Alexander et al. 2009) and a model-free approach as implemented in DAPC (Jombart, 2010) to study population structure. As a relaxed approach for investigating population structure we used NGS admix (Skotte et al. 2013). We also estimated F_{ST} (Weir & Cockerham, 1984), and tested for significance using 10000 permutations, in Arlequin 3.5 (Excoffier & Lischer 2010). To infer SNPs influenced by directional natural selection we used the FDIST2 module (Beaumont & Nichols 1996) as implemented in Arlequin 3.5 (Excoffier & Lischer 2010).

Arbeidsgang – genomisk metode og genetiske analyser

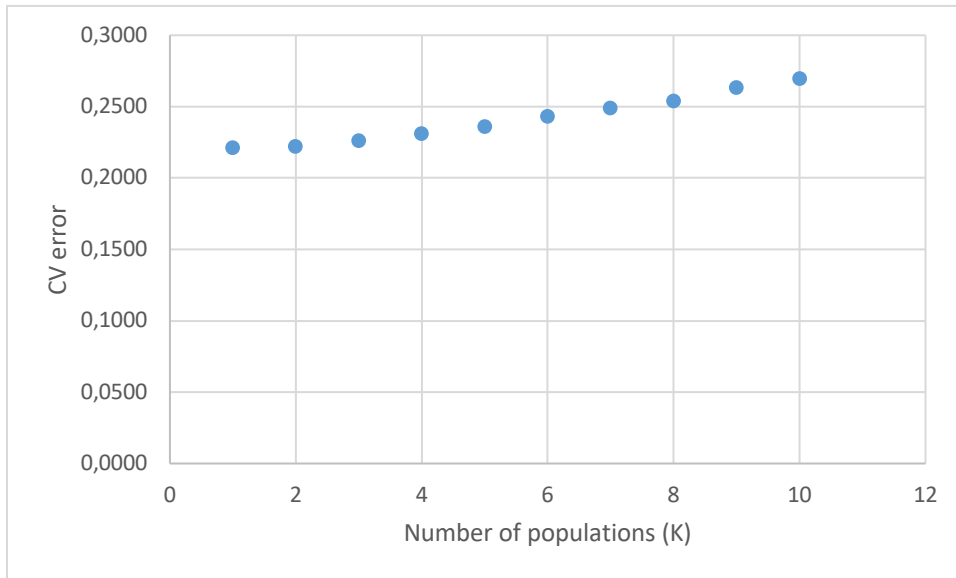


Appendiks 3. Fylogenetisk analyse av sil innsamlet i Tanaelvens munning (Tan), vest i Tanafjorden (Wta), Altafjorden (Alt), Oslofjorden (Gas) og i Mandal (Sno) basert på 11.521 SNPs. Referanse prøver av verifiserte arter av *Ammodytes marinus*, *Ammodytes tobianus*, *Hyperoplus immaculatus* og *Hyperoplus lanceolatus*, ble inkludert i artsbestemmelsen. *Hyperoplus immaculatus* ble brukt som utgruppe for å assistere grupperingen. Figuren viser at sil fra Tanaelvens munning og Altafjorden primært er *A. marinus*. For Tana støttes dette resultatet av at det er samsvar mellom denne analyse og analysene for COI og Rho. Det ble observert to individer av *A. tobianus* i materialet fra Tanaelvens munning. Sil fra vest i Tanafjorden grupperte seg med *H. lanceolatus* (N=2) og *A. tobianus* (N=2). All sil fra Mandal (Sno) ble tilordnet til *A. tobianus*. Prøvene av sil fra Oslofjorden (Gas) viste seg å være både *H. lanceolatus* (N=9), *A. tobianus* (N=1) og *A. marinus* (N=3).

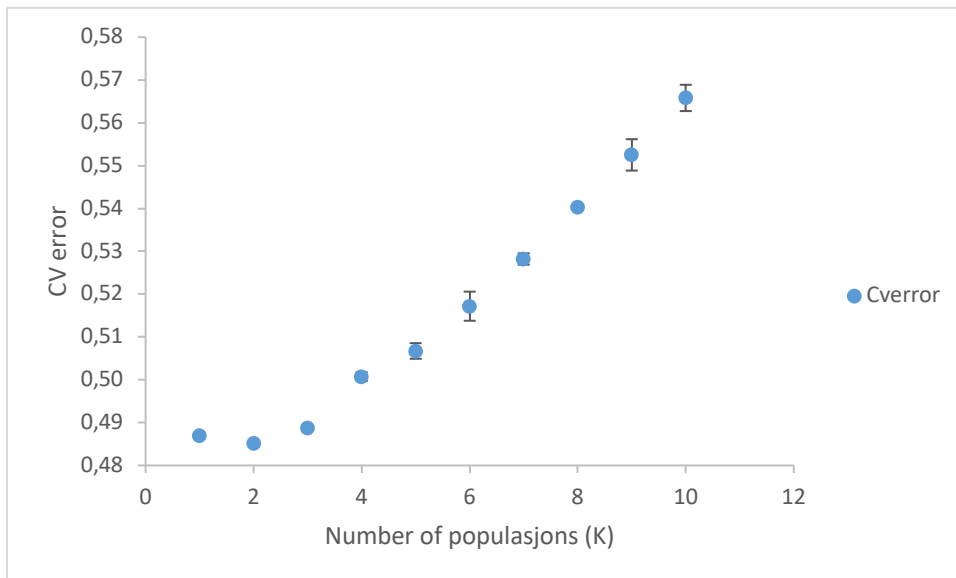


Appendiks 4. Kryssvaliderings verdier for estimering av det mest sannsynlige antallet av populasjoner (K) i Tana-Alta datasettet. Laveste CV verdi tilsier K.

A. 11521 SNPs



B. 2293 SNPs



Appendiks 5. NGS admix analyse av sil fra Tana og Altafjorden, basert på datasettet med 11521 SNPs. I figuren er hver vertikal bjelke et individ og andelen av farge viser genetisk tilhørighet til de ulike populasjoner, slik at individer som viser >0.9 tilhørighet til en populasjon tilhører denne populasjonen. Denne tilnærmingen inkluderer ikke statistiske søkningen for antall populasjoner, men er heller en biologisk relevant metode for å observere populasjonsstruktur i datasettet.

