

Norges miljø- og
biovitenskapelige
universitet

Bacheloroppgave 2020 15 stp

Fakultet for biovitenskap, Institutt for Husdyr- og Akvakultur Vitenskap (IHA)
Hanne Fjerdingby Olsen
Gunnar Klemetsdal
Dag Inge Våge

Sammenlikning av Genotypebaserte Diversitetsmål

Comparison of Genotypic Diversity measures

Nathalie Almaas Smogeli

Husdyr avl og genetikk, Fakultet for biovitenskap, Institutt for Husdyr- og
Akvakultur vitenskap (IHA)

Innholdsfortegnelse

Forord.....	2
Ordliste.....	3
1. Innledning.....	5
2. Bakgrunn	8
2.1. Hardy-Weinberg frekvenser	8
2.2. Fra stamtavle til genomiske mål.....	8
3. Genomiske mål på diversitet.....	10
3.1. Genomisk VanRaden	11
3.2. Molekylært slektskap (<i>SNP-by-SNP</i> Homozygoti)	12
3.3. Runs of Homozygosity	13
3.4. Felles genomiske segmenter	15
4. Konklusjon	17
Referanser	18

Forord

Jeg ønsker å takke mine veiledere Hanne Fjerdingby Olsen, Gunnar Klemetsdal og Dag Inge Våge for å ha kommet med tips om artikler og bøker å lese, og ikke minst veiledning på hvordan å skrive denne oppgaven. Takk til Peer Berg, og Simen Rødt Sandve for tips til fagbøker og artikler. Jeg ønsker også å takke NMBU for muligheten til å studere spennende og dagsaktuelle temaer.

Ordliste

Drift: Tilfeldig variasjon i allelfrekvens mellom generasjoner, større fenomen i små populasjoner (Klug, Cummings, Spencer & Palladino, 2017).

Dødelige ressesive alleler: Et mutert allel som i homozygot tilstand vil føre til at individet dør før en som ikke har mutasjonen (Klug et al., 2017).

Fiksering: Når seleksjonen blir så intens at en variant av et allel ikke lenger eksisterer i populasjonen, hvor det allelet som står igjen er det fikserte allelet (Klug et al., 2017).

Genetisk markør: er et definert punkt på genomet som viser variasjon (en polymorfisme). Noen baserer seg på identiteten av nukleotidene (alleler) (A, T, G, C) i en spesifikk posisjon (Austin, 2020).

Migrasjon: en forflytning av gener, til eller fra en populasjon som reduserer genetisk forskjell mellom populasjoner, men kan øke genetisk varians innad i populasjoner (Klug et al., 2017).

Referansepopulasjon: en populasjon fra fortid eller nåtid som brukes som et referansepunkt for sammenligning av diversitet, innavl eller andre mål (Klug et al., 2017).

Seleksjonsintensitet: basert på hvor mange individer som plukkes ut til å være foreldre til neste generasjon. Hvis det er mange individer som får reprodukere er intensiteten lav, og hvis det er få individer som får reprodukere er intensiteten høy (Klug et al., 2017).

SNP: Single nukleotid polymorfisme, er en posisjon på genomet som viser en enkelt-base-variasjon. De finnes både i ikke-kodende og kodende regioner, mellom kodende gener og benyttes ofte som genetiske markører (Fernández & Bennewitz, 2017).

IBD: Identitet ved arv (*'Identity by descent'*) brukes til å beskrive om et allel eller segment er identisk på grunn av innavl, og at det homozygote segmentet kommer fra en og samme stamfar (International Society of Genetic Genealogy, 2020).

IBS: Identitet av tilstand (*'Identity by state'*) brukes til å beskrive om et allel eller segment er identiske, men ikke nødvendigvis på grunn av innavl eller arv fra samme stamfar (International Society of Genetic Genealogy, 2017).

Innavlsdepresjon: Er det tapet i snittverdi til en gitt fenotype i en gitt populasjon (Falconer & Mackay, 1996), og innavlsdepresjon er avhengig av retningsseleksjon og allelfrekvensendringer.

Retningsdominans: Når flere gener viser dominans i en retning, noe som oppstår igjennom evolusjonen for egenskaper som er essensielle for en populasjons overlevelse (Johsi, Esko, Mattsson & Eklund, 2015), som for eksempel at gen A og B begge gir bedre fertilitet.

1. Innledning

Genetisk diversitet har alltid vært viktig for utviklingen av raser og arter, men i de siste årene kan en se hyppige miljøforandringer som en konsekvens av global oppvarming. Den genetiske diversiteten er forskjellen mellom- og innad- i raser og arter som ligger i deres arvemateriale (Collins, 2020). En reduksjon i genetisk diversitet fører til at en populasjon vil mislykkes i å overleve i et variert miljø, og ikke klarer å tilpasse seg over tid. Dette vil gi en lavere biologisk fitness for populasjonen (Yadav et al., 2019).

Helt siden husdyrenes domestisering har mennesket valgt ut dyr med spesielle egenskaper for å tilfredsstille egne behov. Dette kalles kunstig seleksjon og har resultert i et bredt spekter av raser innen arter. Rasene er ofte, eller har vært, spesialiserte på egenskaper som kjøttproduksjon, melk, fiber, eller en kombinasjon av flere (Brito et al., 2017). Når en har seleksjon på dyr, slik som ved kjæledyr og husdyr, vil allelfrekvenser endres. Konsekvensene av seleksjon er at det over tid kan føre til fiksering av allel, og med det også et tap av genetisk diversitet (Falconer & Mackay, 1996). Det er sannsynlig og logisk at seleksjon vil legge press på de regionene i genomet som kontrollerer egenskapene under seleksjon, men også andre egenskaper som tilpasningsevne til miljø, lynne og sykdoms- og parasittresistens.

I et avlsarbeid foretas det bevisst seleksjon av en foreldrepopulasjon med formål om forbedring i egenskaper hos fremtidige generasjoner. Formålet kan være økonomisk rettet, som for eksempel å øke ytelsen på populasjonen, mindre forbruk per kilo tilvekst, eller mindre sykdom i en populasjon. Det kan også være funksjonelt rettet som for eksempel mot lynne, eller holdbarhet. Seleksjon kan ha som hensikt å bevare en eller flere egenskaper, og ofte ønsker en flere egenskaper i avlsmålet samtidig og et samlet avlsmål utfra dette. Det kan også være et mål å ivareta små populasjoner som gamle raser eller ville populasjoner, eller ivareta genetisk diversitet i en populasjon. Avlsprogram i dag har som mål å vedlikeholde den genetiske diversiteten og samtidig få høy gevinst på en eller flere egenskaper. Dette kan oppnås ved å minimere innavlsraten (ΔF).

Hos husdyr- og i ville populasjoner over hele verden er innavl et uunngåelig fenomen. Innavl i et individ kan beskrives som sannsynligheten for at to alleler på samme locus er IBD (Wright, 1922), som også kan beskrives som paring mellom beslektede individer (Falconer & Mackay, 1996). En innavlsgrad viser til forventet tap av genetisk varians relativt

til en referansepopulasjon, med dette blir forvaltning av genetisk diversitet en kontroll på innavl, og innavlsraten (ΔF) (Meuwissen, Sonesson, Gebregiwergis & Wooliams, 2020). For et gitt allel vil innavl øke andelen homozygote individer, og i teorien vil en fullstendig innavlet populasjon bare bestå av homozygote genotyper. Fitness-relaterte egenskaper knyttet mot overlevelse, reproduksjon og resistens har en tendens til å være mer sensitive til innavl (Falconer & Mackay, 1996) enn andre egenskaper fordi disse egenskaper har en større grad av retnings-dominans enn andre egenskaper (Baes et al., 2019). Det er generelt akseptert at det er viktig å kontrollere innavlsgraden for å unngå innavlsdepresjon og flere studier har rapportert innavl på forskjellige husdyr som storfe (Makina et al., 2015; Zhao, McParland, Kearney, Du & Berry, 2015), fjørfe (Stainton et al., 2015) og svin (Ai et al., 2014; Ai, Huang & Ren, 2013). Yadav et al., (2019) viser en oversikt over studier over noen dyr og effekter i sin rapport. Flere studier viser negative konsekvenser med innavl, men ved seleksjon for forbedring av egenskaper er opphopningen av homozygoter på fordelaktige allel en viktig del av målet. Dette vil ha konsekvenser for genetisk diversitet og seleksjonsrespons i senere tid. Akkumulering av innavl er et uønsket resultat av seleksjon i avlsprogrammer (Maltecca, Tietzzi, Cole & Baes, 2020).

Det er viktig med data på genetisk diversitet for å utvikle forståelse av evolusjon og egenskaper som har hatt en intens seleksjon (Qanbari & Simianer, 2014). Noen genomiske mål er mer egnet til visse formål enn andre. Det reiser spørsmål om hvilke av de genomiske redskapene en står med i dag som er best egnet til en gitt populasjon. I populasjoner med seleksjon er det da forventet at allelfrekvenser vil vike fra den klassiske Hardy-Weinberg forventningen (Meuwissen et al., 2020). Det er derfor viktig å dokumentere nivåer og variasjoner i genetisk diversitet, innad og mellom populasjoner. Forskjellige diversitetsmål kan brukes til å oppdatere eksisterende, eller gi nye seleksjonsprogrammer (Brito et al., 2017).

Genetisk diversitet, og evaluering av diversitet innad og mellom populasjoner, er i dag et aktuelt og diskutert tema i forhold til husdyr, og ville populasjoner. Dette for å øke fitness i en tid hvor miljøforandringer er hyppige. Det er kommet flere forskjellige metoder for hvordan beregne diversitet i populasjoner. Metodene kan virke veldig like, eller vidt forskjellige avhengig av hvilke forutsetninger som er tatt og hvordan metodene regnes ut. Det er ikke en del av denne oppgaven å finne det beste diversitet målet, men å se på

forskjeller mellom noen metoder for å senere kunne bygge videre på denne kunnskapen. Formålet med oppgaven er å gi en oversikt over ulike måter å beregne genetisk diversitet med genetisk informasjon, og se nærmere på mål som er hyppig brukt i senere tid av forskere. Det ønskes også å se på hvilke styrker og svakheter de utvalgte målene har, som vil gi et kunnskapsgrunnlag til en masteroppgave.

2. Bakgrunn

2.1. Hardy-Weinberg frekvenser

En del av populasjonsgenetikk er å beregne allel- og genotypefrekvenser i en populasjon, og observere hvordan frekvensene kan skifte mellom generasjoner i en ideell populasjon. En ideell populasjon er en uendelig stor populasjon med tilfeldig paring, hvor det ikke forekommer migrasjon, mutasjon eller drift (Klug et al., 2017). Allel- og genotypefrekvenser og forholdet mellom dem, beskrives av Hardy-Weinbergs matematiske modell, med noen forutsetninger. Blant annet blir det antatt at alle individer i en gitt populasjon har likt genetisk bidrag og lik sannsynlighet for å overleve. Hardy-Weinbergs forventninger ganger to allelfrekvenser for å gi sannsynligheten for en genotype. Forventet homozygoti er den andelen homozygote individer i en populasjon med de samme allelfrekvensene, og kan regnes som

$$EH = 1 - \sum p^2 ,$$

hvor EH er forventet homozygoti («*Expected homozygosity*»), og p^2 er den kvadrerte allelfrekvensen. Disse forventningsestimaterne er nyttige for å se hvordan en populasjon viker fra en ideell populasjon, som grunnlag for å finne ut hvorfor de viker.

2.2. Fra stamtavle til genomiske mål

Før genomiske data ble tatt i bruk for estimering av innavl har slektskap vært beregnet fra stamtavleinformasjon. Dette er en relativt utbredt metode for å beregne innavlsgraden på individer eller en populasjon. Så lenge det føres opp hvilke dyr som er i slekt med hverandre kan innavlskoeffisienten med stamtavleinformasjon gi et relativt presist estimat. Beregning av innavlskoeffisient basert på stamtavleinformasjon kan gjøres som foreslått av Meuwissen og Luo (1992) der de anvender en slektskapsmatrise. Det er derimot en del faktorer som kan gjøre beregning av innavlskoeffisient basert på stamtavleinformasjon usikker og upresis. Eksempler på dette er dersom en forelder ikke er med i stamtavlen, eller stamtavlen inneholder ukjente individer eller feilregistreringer, så vil innavlskoeffisienten estimeres feil. Det er i tillegg antatt at basepopulasjonen i et individs stamtavle ikke er beslektet, en konsekvens av dette blir underestimering av innavlskoeffisienten. En økende tilgjengelighet av flere molekylære markører for ulike arter tillater derimot beregning av slektskap uten

stamtavleinformasjon. Genomiske mål har vist seg å være minst like gode på beregning av innavlskoeffisient, da de er beregnet ut fra faktisk slektskap og ikke forventet slektskap. Det kan argumenteres for at forventningene om ukoblede nøytrale loci, ikke er reelle for loci som er koblet med varianter av kvantitative egenskaper (*QTL*). Stamtavleinformasjon alene blir for simpelt til å beregne IBD når allel- og genotypfrekvenser kan variere som konsekvens av seleksjon (Meuwissen et al., 2020). Underestimert innavlgrad ved bruk av stamtavleinformasjon kan forklares ved at det er et forventningsestimat, og at variasjon tilhører forventingen ikke er med i estimatet (Keller, Wisscher & Goddard, 2011). Den sanne innavlsverdi for et locus er forventet å være innavlsgraden med variansen av innavlsgraden (Garcia-Cortés, Legarra, Chevalet & Toro, 2013).

Flere metoder har blitt foreslått for å kontrollere innavl gjennom tidene. En metode som har vist seg mer effektiv over lengre tid er Seleksjon for Optimale Bidrag (*Optimal Contribution Selection; OCS*) (Meuwissen T. H., 1997). OCS er en mye brukt metode som ved å minske gjennomsnittlig slektskap på individene som selekteres, og maksimere genetisk gevinst i populasjonen, kan effektivt konservere alleler. OCS fordeler bidrag til alle mulige foreldre for å minske det genetiske bidraget til hver forelder (Maltecca et al, 2020). Med den genomiske teknologien som brukes i dag trekkes det konklusjoner om at genetisk seleksjon behøver en genomisk kontroll over innavl. Det foreslås Genomisk Seleksjon for Optimale Bidrag (*Genomic Optimal Contribution; GOC*) (Sonesson, Wooliams & Meuwissen, 2012), hvor A-matrisen erstattes av målt genomisk slektskap (G-matrise) i samme formel som ved OCS (Meuwissen et al., 2020).

3. Genomiske mål på diversitet

Genomisk seleksjon har vist seg til å være et kraftig verktøy for avlsprogrammer, og med dette trengs det også et kraftig verktøy til å vurdere den genomiske diversiteten til en populasjon. Implementeringen av genomisk seleksjon på storfe har hatt store fordeler som for eksempel å halve generasjonsintervallene (Austin, 2020), og med dette få en økt genetisk gevinst på egenskapene som selekteres for. På den ene siden vil innavlsraten per generasjon minke, mens på den andre siden vil innavlsraten per år øke fordi seleksjonsintensiteten har økt (Daetwyler, Villanueva, Bijma & Wooliams, 2007). Makanjuola et al., (2019) fant i sin studie at innavl økte som en følge av implementeringen av genomisk seleksjon i Holstein og Jersey kyr, som viser at det er nødvendig med genomisk kontroll på innavl ved bruk av genomisk seleksjon.

Genomiske mål på diversitet og genomiske innavls mål er to sider av samme sak, og gjensidig utelukkende fenomen, da en ikke kan ha høy diversitet og innavl på en populasjon eller individ. Flere studier viser en korrelasjon mellom beregning av innavlskoeffisient basert på stamtavleinformasjon og genomiske mål generelt, som kan tyde på at genomiske mål på innavl er pålitelig (Li et al., 2011; Saura et al., 2013; Mastrangelo et al., 2014; Rodríguez-Ramilo et al., 2015). Hensikten med utviklingen av genomiske mål er å kunne beregne innavlsgraden i en populasjon mer presist enn en kan gjøre med stamtavleinformasjon (de Cara et al., 2011; Gómez-Romano et al., 2013; Clark et al., 2013; Toro et al., 2014).

Genomiske mål gir oss muligheten til å beskrive innavl på en måte som innavlskoeffisienten fra stamtavleinformasjon og Hardy-Weinbergs forventninger ikke kan.

Seleksjon kan som nevnt øke andel homozygote individer i en populasjon, og noen metoder bruker differansen mellom observert og forventet homozygoti som et mål på innavl (Marras et al., 2015; Bosse et al., 2012). Det kan også tenkes at i lengre homozygote segmenter finner vi de ressesive dødelige mutasjoner (Szpiech et al., 2013), som gir grunnlaget for metoder som måler homozygote segmenter på et individ eller en populasjon. For å nevne noen eksempler på genomiske diversitetsmål er genomisk VanRaden, felles genomiske sekvenser («*shared genomic sequences*»), *runs of homozygosity*, *SNP-by-SNP* homozygoti (Meuwissen et al., 2020). I tillegg finnes flere metoder som kan være lignende, eller vidt forskjellig fra de nevnte metodene over. Det er mange studier som bruker forskjellige genomiske innavls mål på ulike populasjoner. Innsamling av mye genomisk data, og

kartlegging av regionene som bidrar til innavlsdepresjon, vil gi bedre estimater for de risikoene som kommer med innavl (Kardos et al., 2016; Curik et al., 2017).

3.1. Genomisk VanRaden

Det er foreslått varianter av VanRaden (2008) for å kunne skille mellom IBD og IBS i individer. Metoden gjør dette med en kvadrert, vektet differanse mellom den observerte og forventede homozygoti (Caballero, Villanueva & Druet, 2020). Mikrosatellitt-markører brukes (Córtes, Eusebi, Dunner, Sevane & Cañón, 2019) for å innhente informasjon på genomet til et individ. Mikrosatellitt-markører er korte repeterende sekvenser, ofte av to til sju nukleotid repetisjoner, som for eksempel baserekkefølgen G-A-T-A repetert flere ganger etter hverandre (Kumar, 2017). Genomisk VanRaden blir lik som A-matrisen fra stamtavleinformasjon, men matrisen byttes ut med det observerte genomiske slektskap i en G-matrise hvor formelen er;

$$F_{GRM} = \frac{1}{n} \sum_{i=1}^n \frac{(x_i - 2p_i)^2}{h_i} - 1$$

hvor x_i og p_i er antall kopier og frekvensen av referanseallelet for *SNP* i , og h_i er $2p_i(1 - 2p_i)$, og totalt antall *SNP*er er n . Matrisen er symmetrisk, med like mange antall rader og kolonner som individer tatt med for undersøkelsen (Wiley, 2016) og de ikke-diagonale elementene i matrisen korresponderer med andelen gener som er IBD mellom individene. De diagonale elementer blir et individs slektskap med seg selv (innavl), og innavlskoeffisienten kan regnes ut i fra diagonalen ved;

$$F_{G_j} = \sum_{j=1}^n \left(\frac{z_i z_i'}{2p_j(1-p_j)} - 1 \right)$$

hvor F_{G_j} er den genomiske innavl av individ j ; z_i er vektor med element $2-2p_j$, $1-2p_j$, $0-2p_j$ for genotypene AA, Aa, og aa. Z_i' er den transponerte z_i , og p_j er den andre allelfrekvens av individ j (Baes et al., 2019). Som nevnt er det flere forskjellige metoder av VanRaden, og hvis man ser litt mer på to av de kan en lettere forstå hva som gjør de forskjellige. Genomisk VanRaden metode 1 (G_{vr1}) bruker et vektet snitt, og vil ved å se på et enkelt loci gi identisk slektskapsmatrise som Genomisk VanRaden Metode 2 (G_{vr2}). Ser en derimot på flere loci vil

G_{vr1} bruke ett vektete snitt, mens G_{vr2} bruker et snitt av et estimat av en loci og vekter alle loci likt (Meuwissen et al., 2020).

Fordelen med variantene av disse metodene er at datasettet man har på individene, som avkomsgransking og stamtavleinformasjonen, forblir intakt. Det er derimot begrensninger i metodene, slik som at datasettene kan være mangelfulle som tidligere beskrevet i beregning av innavlskoeffisient basert på stamtavleinformasjon, og at det ikke kan sammenlignes med sanne verdier (Wiley, 2016). Resultatene blir ikke betydelig mer presis ved å bruke flere markører, som ved bruk av høytetthets *SNP*, men antall genotypede individer inkludert i G-matrisen ser ut til å ha størst effekt på presisjonen (Wiley, 2016). Flere studier på storfe viser at genomiske evalueringer er på sitt beste like presis som foreldrenes snittverdi hvis mindre enn 1000 okser er med i studie (VanRaden et al., 2009).

3.2. Molekylært slektskap (*SNP-by-SNP* Homozygoti)

Ved flere studier er molekylært slektskap utregnet på *SNP-by-SNP* basis, det vil si at en ser på mange markører, om de er homozygote eller ikke. Etter Malecot (1948, i Curik, Ferenčaković, Sölkner, 2014) sin teori er denne typen slektskap det dobbelte av sannsynligheten for at to alleler valgt ut tilfeldig er identiske. For individet selv, vil slektskapet være 1 pluss sannsynligheten for at to allel er identiske med hverandre. Metoden går ut på å estimere antall alleler som er IBS for hver enkelt *SNP*, og i teorien fanger den opp slektskap som kommer av like stamfedre tilbake til basepopulasjon (Rodríguez-Ramilo, Elsen & Legarra, 2019). Den er basert på en skalert differanse mellom observert og forventet homozygoti, og kan beregnes slik;

$$F_{HOM} = 1 - \frac{\sum_{k=1}^S x_k(2-x_k)}{\sum_{k=1}^S 2p_k(1-p_k)},$$

hvor s er antall markører, X_k er antall *minor allele* (0, 1 eller 2), og p_k er frekvensen av *minor allele* i populasjonen.

Metoden er noe upresis. Ved ett simulasjonsstudie *SNP-by-SNP* vises det at homozygoti innavls mål estimerer mer presist innavl på populasjoner med stor effektiv populasjonsstørrelse og intens seleksjon, mens den gir dårligere estimerer ved små populasjoner med lav seleksjonsintensitet (Caballero et al., 2020). Dette kan komme av at

metoden ikke kan skille mellom IBD og IBS, hvor den generelt kan overestimere innavlen i en populasjon (Rodríguez-Ramilo et al., 2019). Dette kan spille seg ut mer dramatisk i en liten populasjon med lav seleksjonsintensitet, med en forventet lav innavlsgrad. Ved beregning av effektiv populasjonsstørrelse fra molekulære slektskapsdata på små populasjoner viser Rodríguez-Ramilo et al., (2019) estimater som ligger langt over (ca. $N_e=100$) de andre metodene. Dette kan gi store konsekvenser for forvaltning av populasjoner, som naturlige populasjoner eller gamle raser, og understreker at metoden ikke kan brukes til å vurdere genetisk diversitet alene. Metoden har som et formål å opprettholde alle alleler, som i utgangspunktet kan tenkes er bra for genetisk diversitet, men som også viser seg ugunstig da det opprettholder en frekvens av dødelige mutasjoner med OCS (de Cara, Villanueva, Toro & Fernández, 2013).

3.3. Runs of Homozygosity

Broman & Weber (1999) har en teori om at Runs of Homozygosity (ROH) oppstår når et segment som kommer fra samme stamfar nedarves til et individ gjennom sine foreldre, og vil derfor være IBD. Korte homozygote segmenter vil vise seg når det er gammel innavl i en populasjon, og lengre segmenter finnes på autosomet når det har vært nylig innavl i populasjonen (Purfield, Berry & Bradley, 2012). Dette gjør at ROH er IBD basert innavlsmål når en ser på lengre segmenter, men det vil også være et mål på homozygoti på haplotyper ved korte segmenter (de Cara et al., 2013). Med ROH er det som nevnt tenkt at et homozygot segment indikerer et segment som er IBD. Summen av lengdene på homozygote segmenter over lengden på hele genomet (minus kjønnskromosomer, i.e. autosomet) vil gi et estimat om hvor stor del av autosomet som er IBD (Meuwissen et al., 2020). Det observeres i flere studier en relativt høy korrelasjon mellom ROH og innavlsberegning basert på stamtavleinformasjon. Dette bekrefter at det er en mulig metode for å se på avstamning ved manglende stamtavleinformasjon på individer (Purfield et al., 2012). Beregning av ROH gjøres ved å beregne lengden av genomet som er dekket av ROH, delt på hele lengden av genomet av autosomet (Oliveira et al., 2020). Formelen er;

$$F_{ROH_j} = \frac{\sum L_{ROH_j}}{L},$$

hvor F_{ROHj} er den genomiske innavlen av individ j ved bruk av ROH, L_{ROH} er total lengde av ROH i individ j , og L er total lengde av autosomet (Baes et al., 2019).

En av de sterke sidene til ROH er at den vil hovedsakelig oppdage IBD segmenter.

Sannsynligheten for at et individ bærer to identiske, lange haplotyper er liten, hvis det ikke er nedarvet fra samme stamfar. Det er også flere studier som kan indikere at ROH rekombinasjons-mønstre kan bidra til forståelse av populasjonshistorie (Purfield et al., 2012; Purfield et al., 2017; Bjelland et al., 2013). Det er for eksempel blitt sett på genomet til en ulvepopulasjon, hvor det ble funnet mange individer med ROH over hele kromosomer. Konklusjonen i studiet er at store deler av innavlen i populasjonen kom fra en felles stamfar for mindre enn ti generasjoner siden (Kardos et al., 2018).

Presisjonen på ROH kan virke høy hvis en vet hva en ser etter, og bruker forskjellige lengder av ROH. Et eksempel på dette er et studie på sau hvor de fant at et tilfeldig utvalg på 150 individer ga relativt gode estimater sammenlignet med hvis man brukte alle individer i populasjonen (Oliveira et al., 2020). ROH har også vært sammenlignet med innavlsberegning med bruk av stamtavleinformasjon og genomiske slektskapsmatriser, hvor ROH viser seg å gi et bedre estimat for autozygoti og detektering av innavlsdepresjon (Keller et al., 2011). Caballero et al. (2020) viser i sin studie at ROH er mer presis en *SNP-by-SNP* slektskap i populasjoner med liten effektiv populasjonsstørrelse, og relativt liten seleksjonsintensitet.

Det er vanskelig å sammenligne ROH resultater fra forskjellige studier da det for hver studie er forskjellige faktorer som påvirker identifiseringen av ROH. Det å ha en nøyaktig definisjon av ROH-kriterier er vanskelig da kriteriene varierer i forskjellige studier på bakgrunn av begrensninger i tilleggsutstyr som kan påvirke resultatet man får. Blant annet varierer lengden av ROH målt i; markører, cM, markørtetthet. I noen studier er det en tillatelse av heterozygote genotyper som følge av genotypefeil (Oliveira et al., 2020). Kvalitetskontrollen (Albrechtsen, Nielsen & Nielsen, 2010), antallet heterozygote genotyper tillat i et ROH segment (Purfield et al., 2012), gjør det vanskelig å definere en basepopulasjon (Keller et al., 2011). Det er også forskjellige terskelverdier som brukes under sekvensanalyser (Howrigan, Simonson & Keller, 2011), hvor alle disse variablene understreker Rodriguez-Ramilo et al. (2019) sin konklusjon om et sterkt behov for å etablere konsistente kriterier for å identifisere og kvantifisere ROH.

I større populasjoner med intens seleksjon viser ROH seg mindre presis enn *SNP-by-SNP* slektskap (Caballero et al., 2020). Dette vises også fra en studie på sau hvor antallet genotypedede individer påvirket hvor mange ROH-segmenter metoden oppdaget (Oliveira et al., 2020). Metoden vil ikke se på homozygote segmenter mindre enn en minimumslengde som er satt og kan derfor overse relevant innavl. Det vises at innavlsgraden man får med å bruke ROH, er svært sensitiv på lengden av ROH-segmentene som også kan være forskjellig mellom populasjoner (Rodríguez-Ramilo et al., 2019).

3.4. Felles genomiske segmenter

For å kunne skille ut de dødelige mutasjonene er det blitt foreslått en metode som baserer seg på felles genomiske segmenter (de Cara et al., 2013; Bosse et al., 2015; Rodríguez-Ramilo et al., 2015; Gómez-Romano et al., 2016). Med denne metoden sammenligner en haplotyper mellom to individer og regner ut de delene på deres genom som er felles mellom individene (de Cara et al., 2013). Metodens mål er å minimere antall og lengden av de felles segmentene (Olsen, Tenhunen, Dolvis, Våge & Klemetsdal, 2020). Dette beregnes ved;

$$f_{seg(i,j)} = \frac{\sum_k \sum_{a_i=1}^2 \sum_{b_j=1}^2 (L_{seg_k}(a_i, b_j))}{4L_{auto}},$$

hvor $L_{seg_k}(a_i, b_j)$ er lengden av det k-ente IBD felles segment seg_k , over homolog a for individ i og homolog b for individ j, og L_{auto} er lengden av det autosomale genom (Fernández & Bennewitz, 2017).

Det er vist at fitness øker i populasjoner når en bruker segmentbasert slektskap sammen med OCS (Bosse et al., 2015). Dette bekrefter at metoden har sin fordel når en ikke vet hvor på genomet en dødelig mutasjon befinner seg, siden metoden har som mål å minimere de felles segmentene mellom individer. Olsen et al. (2020) fant i en studie på to populasjoner av fjordhest at korte segmenter vil påvirke resultatet på en segmentbasert innavlskoeffisient. Studiet konkluderte med at gammel innavl er grunnen til forskjellen mellom segmentbasert innavlskoeffisient og innavlskoeffisient basert på stamtavleinformasjon, på tross av gode stamtavledata. Dette viser at stamtavleinformasjon ikke er presist nok for å se hvor mye innavl det faktisk er i en populasjon, selv om stamtavledataen er dyptgående og gode.

Noen utfordringer med metoden er at den ikke kan skille mellom IBD og IBS (Meuwissen et al., 2020), og at det kan se ut som presisjonen er følsom for hvor mange individer som er med i undersøkelsen. Det observeres at metoden gir mye mindre presise tall på effektiv populasjonsstørrelse når en liten populasjon blir brukt. Innavlskoeffisienten med felles genomiske segmenter ble nesten dobbelt av innavlskoeffisient på stamtavleinformasjon på den samme populasjonen (Olsen et al., 2020). Denne sensitiviteten gjør at en ikke vil kunne stole på målet ved bruk i populasjoner som gamle raser og naturlige populasjoner, og bør brukes med andre genomiske mål for å vurdere den faktiske diversiteten.

4. Konklusjon

Seleksjon på individer som gir en populasjon større varians kan være nøkkelen for overlevelse i flere populasjoner. Det vises at de genotypebaserte diversitetsmål tatt med i denne sammenligningen (ROH, Genomisk slektskapsmatrise, Felles genomiske sekvenser, og *SNP-by-SNP* homozygoti) har alle sine styrker og svakheter, hvor de presterer ulikt i forskjellige situasjoner.

De fleste metoder i dette studiet ser ut til å være mer eller mindre upresise i sin estimering av innavlsgrad, og kan gi avvik i målt innavlsgrad i populasjoner. Det kan se ut som at flere av målene er mindre presise ved mindre populasjoner, men ROH har vist seg å ha en fordel med å kunne estimere relativt gode tall på en populasjon med kun et lite utvalg av individer. Felles genomiske segmenter og *SNP-by-SNP* homozygoti er begge upresise i sine estimeringer når de brukes på små populasjoner. Begge metodenes estimater ser ut til å påvirkes av seleksjonsintensiteten i populasjonen. Der *SNP-by-SNP* homozygoti kan tendere til å overestimere innavl i populasjoner, kan ROH også gjøre dette, med falske positive ved bruk av lav-tetthets *SNP chip*, og vil identifisere en ROH der hvor det ikke er det. Derimot kan også ROH overse relevant innavl, da en bare ser på segmentstørrelser etter forhåndsbestemte kriterier, og alt under minimumskriteriet på segmentlengden vil ikke tas med i beregningen. Dette forekommer også med felles genomiske segmenter hvor man kan overestimere hvis en tar små segmenter med i beregningen, eller underestimere hvis en overser de små segmentene. Genomisk VanRaden ser ut til å ligge nærmere innavlsberegning på stamtavleinformasjon enn de andre målene. Den øker ikke sin presisjon signifikant ved flere markører på genomet, men kun hvis en tar med flere genotypede individer med i beregning. Det må i tillegg være et stort antall genotypede individer for at metoden skal bli mer presis en estimattall som foreldrenes snittverdi.

Det er ikke noe godt svar på hvilke metode som er best, derimot kan det heller virke som om anbefalingen av metode kan avhenge av hvordan populasjonen ser ut med tanke på populasjonsstørrelse, innavlsgraden, forventet seleksjonsintensitet i populasjonen og utstyret en har tilgjengelig.

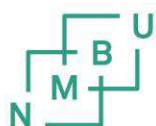
Referanser

- Ai, H., Huang, L., & Ren, J. (2013). Genetic diversity, linkage equilibrium and selection signatures in Chinese and Western pigs revealed by genome-wide SNP markers. *PLoS One*, *8*;2, e56001.
- Ai, H., Yang, B., Li, J., Xie, X., Chen, H., & Ren, J. (2014). Population history and genomic signatures for high-altitude adaptation in Tibetan pigs. *BMC genomics*, *15*;1, 1.
- Albrechtsen, A., Nielsen, F. C., & Nielsen, R. (2010). Ascertainment biases in SNP chips affect measures of population divergence. *Mol. Biol. Evol.* *27*, 2534-2547.
- Austin, C. P. (2020, 25.11). *Geno*. Retrieved from Geno: <https://www.genome.gov/genetics-glossary/Marker>
- Baes, C. F., Makanjuola, B. O., Miglior, F., Marras, G., Howard, J. T., Fleming, A., & Maltecca, C. (2019). Symposium review: The genomic architecture of inbreeding: How homozygosity affects health and performance. *J. Dairy Sci.*, *102*, 2807-2817.
- Bjelland, D. W., Weigel, K. A., Vukasinovic, N., & Nkrumah, J. D. (2013). Evaluation of inbreeding depression in Holstein cattle using whole-genome SNP markers and alternative measures of genomic inbreeding. *J. Dairy Sci*, *96*, 4697-4706.
- Bosse, M., Megens, H., Madsen, O., Crooijmans, R., Ryder, O. A., & de Cara, M. (2015). Using genome-wide measures of coancestry to maintain diversity and fitness in endangered and domestic pig populations. *Genome research* *25* (7), 970-981.
- Bosse, M., Megens, H.-J., Madsen, O., Paudel, Y., Frantz, L., Schook, L., . . . Groenen, M. (2012). Regions of homozygosity in porcine genome: consequence of demography and the recombination landscape. *PLoS Genet* *8*(11), :e 1003100.
- Brito, L. F., Kijas, J. W., Ventura, R. V., Sargolzaei, M., Porto-Nero, L. R., Cànovas, A., . . . Schenkel, F. (2017). Genetic diversity and signatures of selection in various goat breeds revealed by genome-wide SNP markers. *BMC Genomics*, *18*:229, 1-18.
- Broman, K. W., & Weber, J. L. (1999). Long homozygous chromosomal segments in reference families from the Centre d'Etude du Polymorfisme Humain. *American Journal of Human Genetics* *65*(6), 1496-1500.
- Caballero, A., Villanueva, B., & Druet, T. (2020). On the estimation of inbreeding depression using different measures of inbreeding from molecular markers. *Evolutionary Applications*, *00*;, 1-13.
- Clark, S. A., Kinghorn, B., Hickey, J. M., & van der Werf, J. H. (2013). The effect of genomic information on optimal contribution selection in livestock breeding programs. *Genet. Sel. Evol.* *45* (1), 44.
- Collins, F. S. (2020, 25.11). *Genome.gov*. Retrieved from Genetic glossary: <https://www.genome.gov/genetics-glossary/Genetic-Variation>

- Cortés, O., Eusebi, P., Dunner, S., Sevane, N., & Cañón, J. (2019). Comparison of diversity parameters from SNP, microsattellites and pedigree records in the Lida cattle breed. *Livestock Science* 219, 80-85.
- Curik, I., Ferenčaković, M., & Sölkner, J. (2014). Inbreeding and runs of homozygosity: A possible solution. *Livestock Science*, 166, pp. 26-34.
- Curik, I., Ferenčaković, M., & Sölkner, J. (2017). Genomic dissection of inbreeding depression: a gate to new opportunities. *Brazilian Journal of Animal Science*, 46:9, 773-782.
- Daetwyler, H. D., Villanueva, B., Bijma, P., & Woolliams, J. A. (2007). Inbreeding in genome-wide selection.
- de Cara, M. A., Fernández, J., Toro, M. A., & Villanueva, B. (2011). Using Genome-wide information to minimize loss of genetic diversity in conservation programmes. *Animal Breeding Genetics* 128 (6), 456-464.
- de Cara, M. A., Villanueva, B., Toro, M. A., & Fernández, J. (2013). Using genomic tools to maintain diversity and fitness in conservation programmes. *Molecular ecology*, 22(24), 6091-6099.
- Falconer, D. S., & Mackay, T. F. (1996). *Introduction to quantitative genetics*. Harlow: Pearson Education Limited.
- Fernández, J., & Bennewitz, B. (2017). Defining genetic diversity based on genomic tools. In K. Oldenbroek, *Genomic management of animal genetic diversity* (pp. 55-76). Wageningen, Nederland: Wageningen Academic Publishers.
- García-Cortés, L. A., Legarra, C. A., Chevalet, C., & Toro, M. A. (2013). Variance and Covariance of Actual Relationships between relatives at one locus. *PLoS One* 8:e57003.
- Gómez-Romano, F., Villanueva, B., de Cara, M. A., & Fernández, J. (2013). Maintaining genetic diversity using molecular coancestry: the effect of marker density and effective population size. *Genet. Sel. Evol.* 45(1), 38.
- Gómez-Romano, F., Villanueva, B., Sölkner, J., de Cara, M. A., Mészáros, G., Pérez O'Brien, A. M., & Fernández, J. (2016). The use of coancestry based on shared segments for maintaining genetic diversity. *Journal of Animal Breeding and Genetics* 133 (5), 357-365.
- Howrigan, D. P., Simonson, M. A., & Keller, M. C. (2011). Detecting autozygosity through runs of homozygosity: a comparison of three autozygosity detection algorithms. *BMC Genomics*, 12:460.
- International Society of Genetic Genealogy*. (2017, 01.31). Retrieved from https://isogg.org/wiki/Identical_by_state
- International Society of Genetic Genealogy*. (2020, 06.11). Retrieved from https://isogg.org/wiki/Identical_by_descent
- Johsi, P. K., Esko, T., Mattsson, H., & Eklund, N. (2015). Directional dominance on stature and cognition in diverse human populations. *Macmillan publishers Limited*, 523; 459-461.
- Kardos, M., Åkersson, M., Fountain, T., Flagstad, Ø., Liberg, O., Olason, P., . . . Ellegren, H. (2018). Genomic consequences of intensive inbreeding in an isolated wolf population. *Nature of ecology and evolution*, 2, 124-131.

- Kardos, M., Taylor, H. R., Ellegren, H., Luikart, G., & Allendorf, F. W. (2016). Genomics advances the study of inbreeding depression in the wild. *Evolutionary Applications*, *9*, 1205-1219.
- Keller, M. C., Visscher, P. M., & Goddard, M. E. (2011). Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. *Genetics* *189*, 237-249.
- Klug, W. S., Cummings, M. R., Spencer, C. A., & Palladino, M. A. (2017). *Essentials of Genetics*. Harlow: Pearson Education Limited.
- Kumar, R. (2017, 27.10). *Springer Link*. Retrieved from Springer Link: https://link.springer.com/referenceworkentry/10.1007%2F978-3-319-47829-6_167-1
- Li, M.-H., Strandén, I., Tiieikka, T., Sevón-Aimonen, M.-J., & Kantanen, J. (2011). A Comparison of approaches to estimate the inbreeding coefficient and pairwise relatedness using genomic and pedigree data in sheep population. *PLoS One* *6*(11), e26256.
- Makanjuola, B. O., Miglior, F., Abdalla, E. A., Maltecca, C., Schenkel, F. S., & Baes, C. F. (2019). Effect of genomic selection on rate of inbreeding and coancestry and effective population size of Holstein and Jersey cattle populations. *J. Dairy Sci.*, *103*, 5183-5199.
- Makina, S., Muchadeyi, F., van Marle-Köster, E., Taylor, J. F., Makgahlela, M. L., & Maiwashe, A. (2015). Genome-wide scan for selection signatures in six cattle breeds in South Africa. *Genet. Sel. Evol.* *47*;1, 1-14.
- Maltecca, C., Tiezzi, F., Cole, J. B., & Baes, C. (2020). Symposium review: Exploiting homozygosity in the era of genomics: Selection, inbreeding, and mating programs. *Journal of Dairy Science*, *103*;6.
- Marras, G., Gaspa, G., Sorbolini, S., Dimauro, C., Ajmone-Marsan, P., Valentini, A., . . . Macciotta, N. (2015). Analysis of runs of homozygosity and their relationship with inbreeding in five cattlebreeds farmed in Italy. *Animal Genetics* *46*(2), 110-121.
- Mastrangelo, S., Saura, M., Tolone, M., Salces-Ortiz, J., Di Gerlando, R., Bertolini, F., . . . Portolano, B. (2014). The genome-wide structure of two economically important indigenous Sicilian cattle breeds. *Journal of Animal Science* *92* (11), 4833-4842.
- Meuwissen, T. H. (1997). Maximizing the response of selection with a predefined rate of inbreeding. *J. Anim. Sci.*, *75*, 934-940.
- Meuwissen, T. H., & Luo, Z. (1992). Computing Inbreeding Coefficients in Large Populations. *Genet Sel Evol*, *24*, 305-313.
- Meuwissen, T. H., Sonesson, A. K., Gebregiwegis, G., & Wooliams, J. A. (2020). Management of Genetic Diversity in the Era of Genomics. *Frontiers in Genetics*, *11*:880, 1-14.
- Oliveira, H. R., McEwan, J. C., Jakobsen, J., Blichfeldt, T., Meuwissen, T., Pickering, N., . . . Brito, L. F. (2020). Genetic connectedness between Norwegian White Sheep and New Zealand Composite Sheep Populations With Similar Development History. *Frontiers of Genetics*, *11*:371, 1-12.
- Olsen, H. F., Tenhunen, S., Dolvis, N. I., Våge, D. I., & Klemetsdal, G. (2020). Segment-based coancestry, additive relationship and genetic variance within and between the Norwegian

- and the Swedish Fjord horse populations. *Acta Agriculturae Scandinavica, Section A- Animal Science*, 69:1-2, 118-126.
- Purfield, D. C., Berry, D. P., & Bradley, D. G. (2012). Runs of homozygosity and population history in cattle. *BMC Genet.* 13:70.
- Purfield, D. C., McParland, S., Wall, E., & Berry, D. P. (2017). The distribution of runs of homozygosity and selection signatures in six commercial sheep breeds. *PLoS One*.
- Qanbari, S., & Simianer, H. (2014). Mapping signatures of positive selection in the genome of livestock. *Livestock Science* 166, 133-143.
- Rodríguez-Ramilo, S. T., Elsen, J. M., & Legarra, A. (2019). Inbreeding and effective population size in French dairy sheep: comparison between genomic and pedigree estimates. *J. Dairy Sci.* 10, 4227-4237.
- Rodríguez-Ramilo, S. T., Fernández, J., Toro, M. A., Hernández, D., & Villanueva, B. (2015). Genome-wide estimates of coancestry, inbreeding and effective population size in the Spanish Holstein population. *PLoS One*, 10 (4), e0124157.
- Saura, M., Fernández, A., Rodríguez, M. C., Toro, M. A., Barragán, C., & Fernández, A. I. (2013). Genome-wide estimates of coancestry and inbreeding in a closed herd of ancient Iberian pigs. *PLoS One* 8 (10), 1-7, e78314.
- Sonesson, A. K., Wooliams, J. A., & Meuwissen, T. E. (2012). Genomic selection requires genomic control of inbreeding. *Genet. Sel. Evol.* 44:27.
- Stainton, J., Haley, C. S., Charlesworth, B., Kranis, A., Watson, K., & Wiener, P. (2015). Detecting signatures of selection in nine distinct lines of broiler chickens. *Anim. Genet.* 46;1, 37-49.
- Szpiech, Z. A., Xu, J., Pemberton, T. J., Peng, W., Zöllner, S., Rosenberg, N. A., & Li, J. Z. (2013). Long runs of homozygosity are enriched for deleterious variation. *The American Journal of Human Genetics* 93 (1), 90-102.
- Toro, M. A., Villanueva, B., & Fernández, J. (2014). Genomics applied to management strategies in conservation programmes. *Livestock Science*, 166, 48-53.
- VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91, 4414-4423.
- VanRaden, P. M., Van Tassell, C. P., Wiggans, G. R., Sonstegard, T. S., Schnabel, R. D., Taylor, J. F., & Schenkel, F. S. (2009). Invited review: Reliability of genomic predictions for North American Holstein bulls. *J. Dairy Sci.*, 92:, 16-24.
- Wiley, J. I. (2016). *Genomic selection in animals*. Hoboken, New Jersey: John Wiley & sons, Inc.
- Wright, S. (1922). Coefficients of inbreeding and relationships. *Amer. Nat.* 56, 330-338.
- Yadav, A., Jain, A., Sahu, J., Dubey, A., Gadpayle, R., Barwa, K. D., & Kumar, V. (2019). A review on the concept of inbreeding and its impact on livestock. *International Journal of Fauna and Biological Studies*, 6 (5), 23-30.
- Zhao, F., McParland, S., Kearney, F., Du, L., & Berry, D. P. (2015). Detection of selection signatures in dairy and beef cattle using high-density genomic information. *Genet. Sel. Evol.* 47:1, 1-12.



Norges miljø- og biovitenskapelige universitet
Noregs miljø- og biovitenskapelige universitet
Norwegian University of Life Sciences

Postboks 5003
NO-1432 Ås
Norway