

Genetiske markører til adskillelse af vilde og opdrættede gråænder (*Anas platyrhynchos*)

Notat fra DCE - Nationalt Center for Miljø og Energi

Dato: 22. maj 2019

Liselotte Wesley Andersen¹, Lars-Erik Holm² & Frank Panitz²

¹Institut for Bioscience & ²Institut for Molekylær Biologi og Genetik

Antal sider: 7

Faglig kommentering:
Aksel Bo Madsen
Kvalitetssikring, centret:
Jesper Fredshavn



AARHUS
UNIVERSITET

DCE - NATIONALT CENTER FOR MILJØ OG ENERGI

Tel.: +45 8715 0000
E-mail: dce@au.dk
<http://dce.au.dk>

Indhold

Baggrund	3
SNP identifikation	3
Differentieringsgrad og antal SNP markører	5
Konklusion	6
Referencer	7

Baggrund

Gråand (*Anas platyrhynchos*) er opdrættet i fangenskab igennem flere årtier, hvilket bl.a. har medført at opdrættede gråænder er genetisk forskellige fra den vilde gråand (Søderquist m.fl. 2017). Dette skyldes at der foregår en tilpasning til opdrætsmiljøet gennem udvælgelsen af de fugle, der benyttes som avlsmateriale både i forhold til de egenskaber opdrætterne ønsker (selektion) (Frankham 2008), og dels gennem det faktum at det er et begrænset antal avlsfugle (genpulje), der bidrager til næste generation (genetisk drift). På denne baggrund er formålet med projektet at udvikle genetiske markører der kan skelne mellem vilde og udsatte gråænder.

Med henblik på at udvikle genetiske markører i form af single nucleotide polymorphisms (SNP), der adskiller opdrættede og vilde gråænder, er der indsamlet prøver fra henholdsvis opdrættede gråænder samt kendte og formodede vilde gråænder. Fra fire gråandeopdræt, der bidrager med en væsentlig del af de ænder, der bliver udsat i Danmark, er der indsamlet 30 prøver fra hver. Disse repræsenterer opdrætsfugle. Dernæst er der indsamlet 30 vilde gråænder fra Norge via Norges Jeger og Fiskerforbund, hvor vildkonsulent Webjørn Svendsen har stået for indsamlingen, og endelig er der indsamlet 40 gråænder fra den formodede vilde danske gråandebestand ved Vadehavet. Der er sekventeret 30 individer fra hver af de fire opdræt og de to vilde populationer ved anvendelse af Next Generation Sequencing teknologi. DNA fra 30 individer inden for én population (dvs opdræt eller vilde) er poolet til én prøve, som så er blevet sekventeret, da denne strategi giver information om tilstedeværelsen af SNP's inden for den enkelte population og samtidig giver et godt estimat af frekvenserne af de forskellige SNP's inden for den enkelte population. Analysen af de opnåede sekvenser fra de seks populationer foretages ved hjælp af værktøjet Popoolation2 (Kofler m.fl. 2011) som er specielt udviklet til at analysere poolseq data. Denne analyse vil give information om den genetiske differentieringsgrad (F_{ST} estimator), som giver et udtryk for forskellen mellem to populationer. F_{ST} estimatorerne bliver beregnet parvis mellem alle seks populationer for hver enkelt af de identificerede SNP markører. Vi benyttede disse seks populationer til at identificere genetiske markører (SNP), der tydeligt differentierer opdrættede og vilde gråænder, dvs har en høj detektions evne. Samtidigt blev de seks populationer opdelt i hanner og hunner, således at der kan studeres genetiske forskelle mellem kønnene i senere analyser.

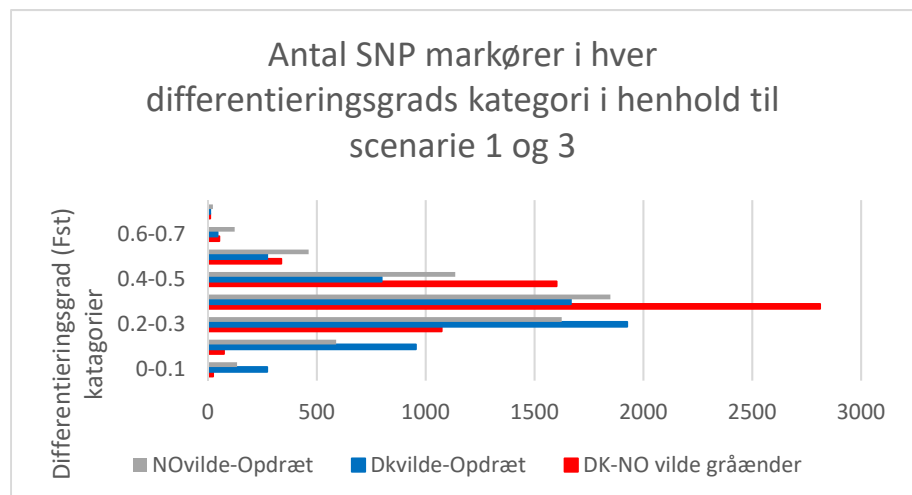
SNP identifikation

For at identificere SNP-markører, dvs. hvor der er genetisk variation i form af enkelt nukleotid forskelle (mutationer) i de opnåede sekvenser fra opdrættede og vilde gråænder, skal de opnåede sekvenser sammenlignes med det nyeste reference-genom (april 2019), der findes offentligt tilgængeligt for gråand på databasen Ensembl (version 96; <http://www.ensembl.org>). Før disse analyser er rådata-sekvenserne filtreret i forhold til forskellige kvalitetsparametre for at sikre et høj kvalitetsdatasæt. Selve identifikationen af SNP-markørernes placering på kromosomerne blev foretaget på dette datasæt og udført ved at benytte samme procedure og software som beskrevet af Kofler et al (2011; PoPoolation2). Dernæst er det muligt at estimere SNP-markørernes allelfrekvens forskelle mellem to populationer (i dette tilfælde parvist mellem de seks forskellige sekventerede pools), samt den genetiske differentieringsgrad (F_{ST}) parvist mellem de seks pools samt antallet af hyppigt forekomne alleler og mindre hyppigt forekommende alleler i de seks pools.

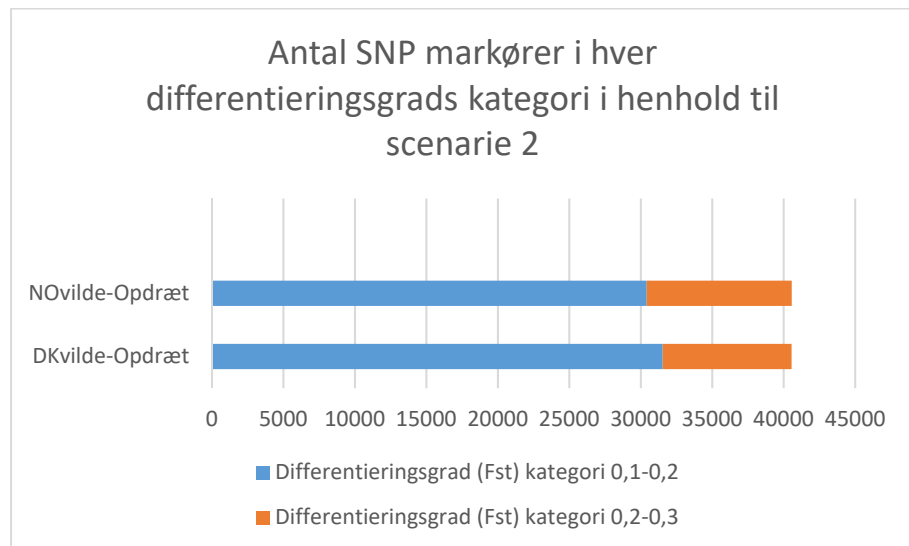
Til selve udvælgelsen af SNP-markører kan anvendes forskellige scenarier:

- 1) SNP's med høj genetisk differentieringsgrad (F_{ST}) mellem alle fire opdræt og de to vilde gråande-datasæt. Det er muligt at finde flere SNP's i et sekvensområde. Den enkelte SNP i et bestemt sekvens-område på kromosomerne, der viste det højeste F_{ST} estimat i sekvensen mellem de fire opdræt og de to vilde gråandedatasæt, blev valgt, mens resten, der lå inden for en minimum afstand på 100.000 bp ikke blev benyttet. Denne metode blev valgt for at mindske chancen for, at to SNP's er koblede, dvs. de repræsenterer samme sekvensområde og derfor ikke er uafhængige observationer. I praksis beregnes dette ved den højeste gennemsnitlige F_{ST} værdi af de 4x2 (opdræt - vild) sammenligninger mellem opdrættede og vilde gråænder. SNP's med disse høje F_{ST} værdier reflekterer sandsynligvis selektion for specielle træk i de opdrættede gråænder. Dette identificerede 5942 SNP (Figur 1).
- 2) SNP's der har en differentieringsgrad (F_{ST} værdi) mellem 0,1 og 0,3 mellem de fire opdræt til begge vilde gråande-datasæt. Det vil sige alle otte (4x2) F_{ST} estimater ligger i dette interval. Dette scenarie vælges, da F_{ST} værdier i intervallet sandsynligvis reflekterer genetisk drift, dvs. de genetiske ændringer, der sker over tid, som følge af tilfældighed. Dette identificerede 40650 SNP (Figur 2).
- 3) SNP's med den højeste differentieringsgrad (F_{ST}) mellem de to vilde gråande-datasæt og som ligger med en afstand på minimum 100.000 bp til den næste SNP blev valgt. Dette vælges, da disse formodentlig vil kunne identificere forskellige vilde populationer af gråænder. Dette identificerede 5980 SNP (Figur 1).

Figur 1. Antal observerede SNP's sorteret efter deres differentieringsgrad ved den parvise sammenligning mellem opdræt og vilde gråænder samt mellem de formodede vilde gråænder fra Danmark og vilde gråænder fra Norge (scenarie 1 og 3).



Figur 2. Antal observerede SNP's sorteret efter deres differentieringsgrad (0,1-0,2 og 0,2-0,3) ved den parvise sammenligning mellem opdræt og vilde gråænder samt mellem de formodede vilde gråænder fra Danmark og vilde gråænder fra Norge (scenarie 2).



Differentieringsgrad og antal SNP markører

Er det muligt at identificere vilde fra opdrættede gråænder?

Til sammenligning med de ovenfor beskrevne antal SNP-markører givet de forskellige kriterier i de tre scenarier til identifikation af vilde og udsatte ænder har fx Nussberger m.fl. (2013) benyttet 200 SNP's til at identificere vildkatte fra tamkatte. Disse markører blev udvalgt på baggrund af genetiske data fra 45 formodede tamkatte og 33 formodede vildkatte, hvor allelerne fra SNP-markøren kun fandtes i den ene eller anden gruppe. SNP-markørerne blev udvalgt ud fra tre kriterier - 1) SNP-markøren var mindst sekventeret med en 10 % coverage i alle prøverne, 2) SNP-markøren havde en allel der kun fandtes i enten vild- eller tamkatten, 3) SNP's blev udvalgt så de fandtes på forskellige kromosomer eller med en afstand på mindst 10.000 bp og kunne antages at være u-koblede. Dernæst udvalgte de SNP's blandt de 200 SNP's, der udviste en differentieringsgrad $F_{ST} > 0,8$ mellem vild- og tamkatte, hvilket resulterede i 48 SNP's der kunne benyttes til at identificere vildkatte, tamkatte og krydsninger (hybrider) mellem disse. I gråande-datasættet fandt vi 5942 SNP's i scenarie 1) hvor vi har udvalgt SNP's med den højeste differentieringsgrad mellem vilde og opdrættede ænder. Disse er ligeledes fordelt over de forskellige kromosomer med en afstand på minimum 100.000bp, hvorfor de også kan antages at være u-koblede. Blandt disse SNP's er der også SNP's med en $F_{ST} > 0,7 - 0,8$ som var kriteriet for udvælgelsen af SNP's til identifikation af vilde og tamkatte. Vi vurderer derfor, at det er muligt på baggrund af det eksisterende gråande-datasæt, der har et højere antal SNP's med en høj differentieringsgrad at udvælge et sæt af SNP-markører, der kan skelne mellem vilde og opdrættede gråænder.

Et andet studie (Ozerov m.fl. 2016) der ligeledes udviklede SNP's til identifikation af vilde og udsatte individer, denne gang af laks, benyttede man også prøver fra kendte vilde og udsatte laks i udviklingen. Her benyttedes også pool-sekventering. Man fandt 5568 SNP's i alt, hvoraf 3928 havde to alleler som forventet. Disse 3928 SNP's blev benyttet og yderligere kvalitetschecket og sammen med et sidste kriterie baseret på at sjældne alleler, dvs alleler der forekommer i mindre end 5 % af prøverne ($MAF < 0,05$), blev frasorteret, resulterede dette i 1986 SNP's. Dernæst for at sikre så optimalt et estimat for krydsning mellem vilde og udsatte individer blev et udsnit af disse SNP's (596 SNP's) udvalgt på baggrund af den maksimale genetiske differentiering (F_{ST})

mellem vilde og udsatte laks. Til sammenligning med poolsekventeringen af gråænder, hvor vi har benyttet lignende filtrerings- og udvælgelseskriterier til at identificere SNP-markørerne, fandt vi flere SNP-markører (5942 SNP's) med en høj genetisk differentieringsgrad mellem opdræt og vilde sammenlignet med 596 SNP's for laks. Dette understreger ligeledes, at det er muligt at skelne vilde fra opdrættede gråænder genetisk med de identificerede SNP's fra udviklingsprojektet.

Kraus m.fl. (2013) og Søderquist m.fl. (2017) benyttede SNP-markører til at påvise de genetiske forskelle bl.a. hos vilde og opdrættede gråænder. Disse markører var dog ikke udviklet med henblik på at identificere opdrætsfugle fra vilde fugle, men til at finde genetiske markører (SNP's) der påviste en fælles nedarvet genom historie mellem de tre forskellige, kendte kilder – museum, vilde og opdrættede – af gråænder dvs. allelfrekvenserne på de udvalgte markører var omkring 0,5 (Kraus m.fl. 2013). Søderquist m.fl. (2017) benyttede disse 384 SNP-markører, udviklet af Kraus m.fl. (2013) nævnt tidligere, til at analysere den genetiske indkrydsning af opdrættede gråænder med vilde gråænder. Af de 384 blev 24 SNP-markører sorteret fra, grundet de ikke havde virket i 10 % af de analyserede prøver (656 totalt), og ligeledes blev de individer, hvor 10 % af SNP markørerne ikke havde virket, sorteret fra. Det resulterede i et datasæt på 591 individer analyseret for 360 SNP's. Resultatet viste genetiske forskelle mellem opdrættede og vilde gråænder og ligeledes var der tegn på indkrydsning mellem de to grupper, hvilket indikerer, at nogle af de udsatte gråænder sandsynligvis overlever og yngler med de vilde gråænder. Dette foregår formodentlig ikke med stor hastighed, da det er påvist at opdrættede udsatte gråænder har en dårligere overlevelse sammenlignet med vilde gråænder (Champagnon m.fl. 2012). Hvor stort omfanget af indkrydsning mellem opdrættede og vilde gråænder er i Danmark er uvist, men resultatet af dette udviklingsprojekt muliggør en analyse af dette. At Søderquist m.fl. (2017) kunne påvise genetiske forskelle mellem opdrættede og vilde gråænder på baggrund af 360 SNP markører, der var udvalgt på baggrund af en lavere genetisk differentieringsgrad sammenlignet med udviklingsprojektet, understreger, at 5942 SNP- markørerne har så stor en differentieringsgrad, at det vil være muligt at udarbejde et panel på 384 SNP, der stadig kan skelne opdrættede fra vilde gråænder og samtidig se på hvor stor indkrydsningsgrad, der er mellem disse.

Konklusion

Der blev identificeret 5942 SNP's baseret på en parvis høj genetisk differentieringsgrad (F_{ST}) mellem opdrættede og vilde gråænder ud fra hvilke et mindre sæt af SNP's kan udvælges til at være en del af et panel på 384 SNP's, der kan bruges videre til et projekt om identifikation af opdrættede og vilde gråænder. Yderligere er der i alt fundet 40650 SNP's, der udviser en lavere differentieringsgrad mellem opdrættede og vilde, som sandsynligvis er et resultat af markører, der ændres mere tilfældigt gennem tid (selektive neutrale). Blandt disse kan der ligeledes udvælges markører til et panel på 384 SNP's for at verificere antagelser om, at de første SNP's er under selektion. De sidste 5980 SNP's udviste en høj differentieringsgrad mellem de to vilde gråandedatasæt, hvilke kan skyldes enten at de er ændret genetisk over tid og derved påviser populationsstrukturer eller, at en af disse gråandedatasæt evt. har en højere indkrydsning af opdrættede gråænder. Her kan der også udvælges et sæt SNP's, der kan verificere en af de to hypoteser.

Referencer

Champagnon J, Guillemain M, Elmberg J, Massez G, Cavallo F, Gauthier-Clerc M. (2012) Low survival after release into the wild: assessing the burden of captivity on Mallard physiology and behaviour. *European Journal of Wildlife Research* 58(1): 255–267.

Frankham R. (2008): Genetic adaptation to captivity in species conservation programs. *Molecular Ecology* 17:325-333.

Kofler R, Pandey RV, Schlötterer C. (2011): PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics* 27(24):3435-6.

Kraus RHS, Van Hooft P, Megens H-J, Tsvey A, Fokin SY, Ydenberg RC, Prins HHT. (2013): Global lack of flyway structure in a cosmopolitan bird revealed by a genome wide survey of single nucleotide polymorphisms. *Molecular Ecology* 22: 41–55.

Nussberger B, Greminger MP, Grossen C, Keller LF, Wandeler P. (2013): Development of SNP markers identifying European wildcats, domestic cats, and their admixed progeny. *Molecular Ecology Resources* 13: 447–460.

Ozerov MY, Gross R, Bruneaux M, Vähä J-P, Burimski O, Pukk L, Vasemägi A. (2016): Genomewide introgressive hybridization patterns in wild Atlantic salmon influenced by inadvertent gene flow from hatchery releases. *Molecular Ecology* 25: 1275–1293.

Söderquist P, Elmberg J, Gunnarsson G, Thulin C-G, Champagnon J, Guillemain M, Kreisinger J, Prins HHT, Crooijmans RPMA, Kraus RHS. (2017): Admixture between released and wild game birds: a changing genetic landscape in European mallards *Anas platyrhynchos*. *European Journal of Wildlife Research* 63: 98.