



## Genetische analyses voor Stekelrog en Blonde rog

Jan Jaap Poos, Timo Staeudle en Hendrik-Jan Megens



**WAGENINGEN**  
UNIVERSITY & RESEARCH



# Genetische analyses voor Stekelrog en Blonde rog

Jan Jaap Poos<sup>1</sup>, Timo Staedle<sup>1</sup> en Hendrik-Jan Megens<sup>2</sup>

1 Aquaculture and Fisheries Group, Wageningen University & Research

2 Animal Breeding and Genomics Group, Wageningen University & Research



Europees Fonds voor Maritieme Zaken en Visserij

Dit rapport is tot stand gekomen binnen het project OSW 2.1 Innorays, uitgevoerd door Wageningen University and Research.

Het rapport is gratis te downloaden via: <https://doi.org/10.18174/582408>

Cover foto: Reindert Nijland

Wageningen, December 2022.



Voor dit rapport geldt een Creative Commons Naamsvermelding 3.0 Nederland licentie.  
© Wageningen University & Research, 2022

De gebruiker mag het werk kopiëren, verspreiden en doorgeven en afgeleide werken maken. Materiaal van derden waarvan in het werk gebruik is gemaakt en waarop intellectuele eigendomsrechten berusten, mogen niet zonder voorafgaande toestemming van derden gebruikt worden. De gebruiker dient bij het werk de door de maker of de licentiegever aangegeven naam te vermelden, maar niet zodanig dat de indruk gewekt wordt dat zij daarmee instemmen met het werk van de gebruiker of het gebruik van het werk. De gebruiker mag het werk niet voor commerciële doeleinden gebruiken.

# Samenvatting

Verbeterd beheer van roggen in de Noordzee is urgent om twee redenen. Ten eerste vormen roggen verstikkingssoorten in het kader van de Europese aanlandplicht. Dit betekent dat de visserij op platvis stilgelegd wordt wanneer alle roggenvangsten moeten worden aangelegd en het roggencotum is opgevisst. Ten tweede vormen roggen als langlevende soort een wezenlijk onderdeel van het Noordzee ecosysteem en daaraan verbonden Europese natuurdoelstellingen. Zo is het verbeteren van de populatieomvang van elasmobranchen (kraakbeenvissen: haaien en roggen) in de Noordzee een belangrijke milieudoelstelling onder de EU Kaderrichtlijn Mariene Strategie.

Voor beter beheer van de roggencotums in de Noordzee zijn betere gegevens voor de bestandsschattingen nodig. Het OSW 2.1 Innorays project heeft als doel om de gegevensbasis voor de bestanden stekelrog (*Raja clavata*) en blonde rog (*Raja brachyura*) te verbeteren. Het project is gefinancierd uit de regeling Samenwerkingsprojecten Wetenschap en Visserij in het kader van het Nederlandse operationele programma voor het Europees Fonds voor Maritieme Zaken en Visserij.

Om de populatieomvang van roggen te schatten kan een nieuwe en innovatieve DNA-methode toegepast worden, genaamd Close-Kin Mark-Recapture. Bij deze techniek wordt met behulp van genotypering van een groot aantal samples de familierelatie vastgesteld binnen deze samples. Deze techniek is toegepast op stekelrog en blonde rog. Voor de genotypering zijn in een eerste stap de genoomsequenties van de twee soorten bepaald. Hiertoe wordt eerst een groot aantal stukken uit het DNA van een enkel individu bepaald, waarna deze "geassembleerd" dienen te worden. Dat wil zeggen, de correcte volgorde en samenhang van de verschillende stukken te bepalen. Voor stekelrog bleek er een goede assemblage van het genoom mogelijk te zijn.

Ons onderzoek geeft aan dat de CKMR techniek succesvol toegepast kan worden op elasmobranchen populaties in de Noordzee. Met behulp van de genotypering van individuen kon voor stekelrog een bestandsschatting gedaan worden. De monsters voor deze bestandsschatting werden verzameld met behulp van onderzoeksschepen en met behulp van de visserijsector ook aan boord van vissersschepen. Een belangrijke voorwaarde voor succesvolle toepassing bleek een robuuste genotypering. Voor stekelrog werd deze gevonden in een SNP ("Single-Nucleotide Polymorphism") array, die meer dan 5000 werkende SNPs opleverde voor de soort.

Voor blonde rog heeft de genotypering niet genoeg SNPs opgeleverd om tot een betrouwbare bestandsschatting te komen. Bij vervolgpiloten over deze soort zou idealiter een hogere dichtheid genotyping-by-sequencing gekozen moeten worden om tot een voldoende aantal SNPs te komen. De huidige gegevens zouden in de toekomst wel bijvoorbeeld gebruikt kunnen worden om populatiestructuren op Europese schaal waar te nemen.

Voor stekelrog heeft de genotypering daadwerkelijk een bestandsschatting opgeleverd. De schatting van de populatiegrootte voor stekelrog in de Noordzee op basis van de CKMR techniek is dat er tussen 4 honderdduizend en 6 miljoen ouderdieren zijn in de zuidelijke Noordzee. Deze bestandsschatting is nog onzeker doordat er een relatief klein aantal monsters genomen is, maar deze kan in de toekomst verbeterd worden door monsters toe te voegen. Daarnaast kan de CKMR methode, die nu in isolatie uitgevoerd is, in de toekomst geïntegreerd worden in bestaande bestandsschattingen, om zo tot een populatieschatting te komen.

Met behulp van de vangsten en de populatiegroottes kan een inschatting gemaakt worden van de effecten van visserij op de roggensorten in de Noordzee. Dit draagt bij aan betere bestandsschattingen en daarmee een beter advies over de vangstmogelijkheden en het verminderen van visserij-impact op roggen in het kader van een duurzaam beheer van de Noordzee. Indien succesvol kunnen deze technieken ook worden ingezet voor het verbeteren van de gegevensbasis van andere Noordzee visbestanden.

De Internationale Raad voor Onderzoek der Zee (ICES), die verantwoordelijk is voor de vangstadvisen aan de Europese Commissie, heeft in 2023 een zogenaamde 'benchmark' voor roggen gepland. Alle bestaande en nieuwe gegevens voor bestandsschattingen worden dan geëvalueerd. De resultaten van dit Innorays project zullen dan worden ingebracht.

## Summary

Improved management of skates and rays (elasmobranch stocks) in the North Sea is urgent for two reasons. First, skates and rays are choke species under the European landing obligation. This means that fishing for flatfish is halted when all ray catches must be landed and the ray quota has been fished. Second, as a long-lived species, rays are an essential part of the North Sea ecosystem and related European nature objectives. For example, improving the population size of elasmobranchs in the North Sea is an important environmental objective under the EU Marine Strategy Framework Directive.

Better management of the ray populations in the North Sea requires better data for stock estimates. The OSW 2.1 Innorays project aims to improve the knowledge base for the thornback ray (*Raja clavata*) and blonde ray (*Raja brachyura*) stocks. The project is financed from the Science and Fisheries Research Collaboration scheme under the Dutch operational program for the European Maritime and Fisheries Fund.

To estimate the population size of rays, a new and innovative DNA method called Close-Kin Mark-Recapture can be applied. This technique uses genotyping of a large number of samples to determine the family relationship within these samples. This technique was applied to thornback ray and blonde ray. For genotyping, the genome sequences of the two species were determined in a first step. For this purpose, a large number of pieces are first determined from the DNA of a single individual, after which they must be "assembled." That is, to determine the correct order and coherence of the various pieces. For thornback ray, proper genome assembly was found to be possible.

Our research indicates that the CKMR technique can be successfully applied to elasmobranch populations in the North Sea. Using the genotyping of individuals, a stock estimate could be done for thornback ray (*Raja clavata*). The samples for this stock estimation were collected using research vessels and with the help of the fishing industry also on board fishing vessels. An important prerequisite for successful application proved to be robust genotyping. For thornback ray, this was found in an SNP ("Single-Nucleotide Polymorphism") array, which yielded more than 5,000 working SNPs for the species.

For blonde ray (*Raja brachyura*), genotyping did not yield enough SNPs to arrive at a reliable stock estimate. Follow-up projects on this species should ideally choose a higher density genotyping-by-sequencing to arrive at a sufficient number of SNPs. The current data could, however, be used in the future, for example, to observe population structures on a European scale.

For thornback ray, genotyping actually provided a stock estimate. The population size estimate for thornback ray in the North Sea based on the CKMR technique is that there are between 4 hundred thousand and 6 million parent animals in the southern North Sea. This stock estimate is still uncertain due to that a relatively small number of samples were taken, but it can be improved in the future by adding samples. In addition, the CKMR method, which is now carried out in isolation, can be integrated into existing stock estimates in the future, in order to arrive at a population estimate.

Using the catches and population sizes, an estimate of the effects of fishing on skate species in the North Sea can be made. This contributes to better stock estimates and thus better advice on fishing opportunities and reducing fishery impacts on rays in the context of sustainable management of the North Sea. If successful, these techniques can also be used to improve the knowledge base of other North Sea fish stocks.

The International Council for the Exploration of the Sea (ICES), which is responsible for the catch advice to the European Commission, has planned a so-called 'benchmark' for rays in 2023. All existing and new data for stock estimates are then evaluated. The results of this Innorays project will then be presented.

# Inhoud

Samenvatting.....	4
Summary.....	6
Introductie.....	8
Doelstelling .....	8
Kennisvragen.....	8
Aanpak.....	10
Genoom assemblage.....	10
Monsternamen en genotypering weefselmonsters .....	10
Kwaliteitscontrole genotypes .....	11
Populatie-structuur .....	11
Verwantschapsanalyse .....	12
Populatiegrootte.....	13
Resultaten .....	14
Genoom assemblage.....	14
Stekelrog .....	14
Blonde rog.....	15
Oxford Nanopore sequencing.....	16
Monsternamen en genotypering weefselmonsters .....	16
Stekelrog .....	17
Blonde rog.....	18
Populatie-structuur .....	19
Verwantschapsanalyse .....	22
Stekelrog .....	22
Blonde rog.....	24
Populatiegrootte.....	27
Stekelrog .....	27
Blonde rog .....	28
Conclusies .....	30
Dankwoord .....	31
Literatuur .....	32
Appendix A .....	33
Appendix B .....	34

# Introductie

Het Innorays project waarbinnen dit onderzoek heeft plaatsgevonden heeft als doel gehad om de gegevensbasis voor het 'data limited stock' (DLS) bestand "roggen" te verbreden. Hierbij was de samenwerking tussen de visserijsector en de wetenschap cruciaal om tot goede gegevens en input voor de bestandsschattingen voor het visserijbeheer te komen. Verbeterd beheer is urgent om twee redenen. Ten eerste vormen roggen een verstikkingssoort in het kader van de Europese aanlandplicht. Dit betekent dat de visserij op platvis stilgelegd wordt wanneer alle roggenvangsten moeten worden aangelegd en het roggengotum is opgevisst. Ten tweede vormen roggen als langlevende soort een wezenlijk onderdeel van het Noordzee ecosysteem en daaraan verbonden Europese natuurdoelstellingen. Zo is het verbeteren van de populatieomvang van elasmobranchen in de Noordzee een belangrijke milieudoelstelling onder de EU Kaderrichtlijn Mariene Strategie.

Voor verschillende soorten roggen in de Noordzee werden gegevens verzameld over de vangsten. Dit gebeurde met behulp van video-opnames aan boord, waarbij geëvalueerd wordt of deze opnames in de toekomst ook gebruikt kunnen worden voor automatische beeldherkenning van roggen. De resultaten worden in een apart rapport beschreven. Daarnaast werd een nieuwe en innovatieve DNA-methode toegepast om populatiegroottes te schatten voor twee roggensoorten. Deze methode heet "Close-Kin Mark-Recapture" en maakt gebruik van het ontdekken van familierelaties binnen genotypes van bemonsterde individuen. Deze genotypes beschrijven het unieke DNA-profiel van individuen, waarbij individuen binnen een familie qua genotype meer op elkaar lijken dan individuen die geen familie van elkaar zijn. Het aantal gevonden familierelaties binnen een steekproef van een populatie geeft een indicatie van de populatieomvang. Over dit deel van het onderzoek gaat dit rapport.

Met behulp van de vangsten en de populatieomvang kan een inschatting gemaakt worden van de effecten van visserij op de roggensoorten in de Noordzee. Dit draagt bij aan betere bestandsschattingen en daarmee een beter advies over de vangstmogelijkheden en het verminderen van visserij-impact op roggen in het kader van een duurzaam beheer van de Noordzee. Indien succesvol kunnen deze technieken ook worden ingezet voor het verbeteren van de gegevensbasis van andere Noordzee visbestanden.

Het project besloeg een periode van vier jaar en werd gefinancierd uit de regeling Samenwerkingsprojecten Wetenschap en Visserij in het kader van het Nederlandse operationele programma voor het Europees Fonds voor Maritieme Zaken en Visserij.

## Doelstelling

De doelstelling van het onderzoek in dit rapport is het verbeteren van de gegevensbasis voor het 'Data Limited Stock' (DLS) bestand "roggen in de Noordzee", door toepassing van Close-Kin Mark-Recapture (CKMR), een nieuwe en innovatieve DNA-methode om populatiegroottes te schatten voor twee roggensoorten. Daarnaast is met de genotypes die verzameld zijn in het project een schatting gemaakt van de populatiestructuur in Noordwest Europa. Specifiek is er onderzoek gedaan naar twee soorten: Stekelrog (*Raja clavata*) en Blonde rog (*Raja brachyura*).

## Kennisvragen

In het project zijn in totaal 5 kennisvragen opgenomen, verdeeld over twee werkpakketten: werkpakket D en E. In werkpakket D werd het antwoord gezocht op de volgende kennisvragen:

1. Wat is het genoom van de stekelrog en blonde rog? Hiertoe werden sequences genomen van de twee genomen, die vervolgens geassembleerd werden. Deze genomen helpen om de genetische merkers uit kennisvraag 2 op het genoom te kunnen plaatsen.
2. Wat zijn de genotypes van 3000 monsters (2000 stekelroggen en 1000 blonde roggen). Deze genotypes worden gebruikt om de familie relaties te schatten tussen individuen later in het project

In werkpakket E werden de genotypes die verzameld werden in werkpakket D : bestandsschatting

3. Welke populatie-structuur is er in de genotypes van blonde rog en stekelrog, zowel in de Noordzee als daarbuiten?



4. Hoeveel familierelaties worden er gevonden in de bemonsterde genotypes in de Noordzee?
5. Wat is de census populatiegrootte?. Deze kan met behulp van de familierelaties bepaald worden, indien de genotypes van voldoende kwaliteit zijn én er voldoende familierelaties gevonden worden.

# Aanpak

Voor de vissers en andere belanghebbenden werd een infographic ontwikkeld waarin de onderstaande beschrijving van de aanpak van een bestandsschatting met behulp van DNA sterk versimpeld is uitgelegd (Appendix B).

## Genoom assemblage

Onder werkpakket D valt het sequencen en assembleren van de genomen van *Raja clavata* en *Raja brachyura*. Deze zijn noodzakelijk om de merkers in kennisvraag 2 op het genoom te kunnen plaatsen. Voor het sequencen en assembleren van het genoom waren een beperkt aantal proefdieren nodig. Het genoom wordt namelijk bepaald met behulp van DNA in het bloed van één enkel dier. De bloedmonsters voor de assemblage zijn in 2018 verzameld. Voor het bepalen van de DNA sequenties zijn deze monsters voor analyse opgestuurd in 2019. Voor stekelrog (*Raja clavata*) is het bepalen van de DNA sequenties voor het genoom gedaan om basis van Pacific Biosciences (PacBio) "Single Molecule Real Time Sequencing" (SMRT) technologie en met behulp van "Oxford Nanopore" (ON). Daarbij zijn met behulp van PacBio in totaal 215 miljard baseparen bepaald in 17,1 miljoen fragmenten totaal en met behulp van ON zijn 80 miljard baseparen bepaald in 4,8 miljoen fragmenten.

Gezien het hoge aantal sequence fouten die in deze technologieën gangbaar zijn werd ook een toegevoegde dataset gegenereerd om die fouten te kunnen corrigeren, in de vorm van 'Illumina short-read' sequences. In totaal werd ruim 75 miljard 'Illumina short-read' sequences data toegevoegd.

*Tabel 1 Statistieken van de PacBio en ON data acquisitie voor stekelrog (Raja clavata).*

Statistiek	Stekelrog (Raja Clavata)		
	Aantal basen PacBio	Aantal basen Illumina	Aantal basen ON
Mean read length	12,578	NA*	16,613
Median read length	9,983	NA*	10,089
Number of reads	17,122,907	2,163,857,074	4,851,879
Read length N50	12,626	NA*	30,588
Total bases	215,366,732,877	324,578,561,100	80,606,720,591

\*Illumina sequencing heeft een vaste leeslengte, in dit geval 150 basen.

*Tabel 2 Statistieken van de 10X Genomics en ON data acquisitie voor blonde rog (Raja brachyura).*

Statistiek	Blonde rog (Raja brachyura)	
	Aantal basen 10X/Illumina	Aantal basen ON
Mean read length	NA*	12,165
Median read length	NA*	7,527
Number of reads	3,030,890,820	593,292
Read length N50	NA*	22,069
Total bases	454,633,623,000	6,294,687,292

\*Illumina sequencing heeft een vaste leeslengte, in dit geval 150 basen.

De blonde rog (*R. brachyura*) is gekarakteriseerd met de zogenaamde '10X Genomics pseudo-long read technologie' (10X Genomics) en met behulp van Oxford nanopore. 10x Genomics is een techniek om op basis van een goedkopere sequence technologie (Illumina sequencing), die probeert de voordelen te bewerkstelligen van een 'long read sequencing' technologie.

## Monsternamen en genotypering weefselmonsters

De monsternamen van de weefsels die nodig waren voor de genotypering werd gedaan op verschillende visafslagen in Nederland, tijdens surveys met onderzoeksschepen, aan boord van reizen waar discards verzameld werden, en met behulp van schippers die de vangsten van roggen nauwkeurig registreerden en deze aan land brachten. Bij deze laatste monsternamen werden roggen heel naar Wageningen Marine Research (WMR) gebracht en werden de weefselmonsters dan genomen door WMR personeel.

Om de monsternamen goed te laten verlopen zijn er korte protocollen gemaakt die meegegeven konden worden aan onderzoekers (Appendix A). Met behulp van deze protocollen werden de weefselmonsters zo consistent mogelijk genomen. Naast deze protocollen kregen de bemonsteraars ook voorgevulde formulieren mee waarop de positie, het geslacht, de lengte, en het gewicht van het monster genoteerd

kon worden. Daarnaast kregen de bemonsteraars 200 à 300 Eppendorf buisjes mee, gevuld met 96% alcohol. Hierin werden de weefselmonsters verzameld en opgeslagen bij -20 °C.

Als tweede categorie weefsels is er nog een groep samples die door de Universiteit van Groningen beschikbaar gesteld zijn. Dit waren weefselmonsters van stekelrog die voor onderzoek verzameld zijn in de periode 2000-2010. Deze samples gaven inzicht in de veranderingen van de genetische samenstelling van de roggenpopulaties, binnen en buiten de Noordzee en deze weefselmonsters zijn daarom opgestuurd om de genotypes te laten bepalen.

De genotypering voor stekelrog is gedaan met behulp van een Infinium® XT iSelect96 SNP-array. Deze SNP ("Single-nucleotide Polymorphism") array is door Labogena ([www.labogena.fr](http://www.labogena.fr)) ontwikkeld voor de genotypering van stekelrog. Op de SNP-array bevinden zich 9120 potentiële SNPs voor stekelrog. Deze SNPs zijn bepaald met behulp van RADseq (Restriction site Associated DNA sequencing) voor finclips van 225 individuen uit kustzones van de Noordoost Atlantische Oceaan en de Middellandse Zee, beschreven in Marandel et al. (2020).

De genotypering voor blonde rog is gedaan met behulp van Genotyping by sequencing (GBS; Elshire et al. 2011). De genotypering werd gedaan door Diversity Arrays Technology ([www.diversityarrays.com](http://www.diversityarrays.com)). In het project is er gekozen voor een lage dichtheid (LD) sequencing, gebruik makend van DArTseq™ zoals beschreven in Kilian et al. (2012). Monsters werden gesequenced (single read) met behulp van een Illumina® HiSeq® 2500, waarbij voor elk individu ongeveer 0.5 miljoen sequenties bepaald werden. Deze lage dichtheid GBS techniek was ongeveer 50% goedkoper dan de SNP array die bij stekelrog gekozen werd. Dit is gedaan om te zien of er kosten bespaard konden worden.

Voor beide soorten werden ongeveer 1% van het totaal individuen als duplicaat opgestuurd ter genotypering. Een vergelijking van de *duplicate* genotypering gaf inzicht in de consistentie van de genotyperingsmethode.

### *Kwaliteitscontrole genotypes*

De gevonden genotypes werden onderworpen aan een kwaliteitscontrole. Daarin werden stapsgewijs SNPs en monsters verwijderd die niet voldeden aan een aantal criteria. Ten eerste moesten de SNPs bepaald zijn in minimaal 95% van de genotypes. Vervolgens werden de incorrect geïdentificeerde soorten verwijderd uit de dataset op basis van hun extreem lage heterozygotie. Vervolgens werden die SNPs verwijderd met een minimum *allele frequency* < 0.01.

Het aantal SNPs dat gevonden werd in de genotypes na kwaliteitscontrole was van belang voor de verwantschapsanalyse: Om tot een betrouwbare meting van verwantschap te komen zijn er minstens 5000 werkende SNPs nodig. Voor de SNP array waren er 9120 potentiële SNPs, maar niet al deze SNPs gaven voldeden aan onze kwaliteitscontrole criteria. Voor de GBS was op voorhand niet bekend hoeveel SNPs deze op zouden leveren, maar gezien ervaringen elders was de verwachting dat er meer dan 5000 SNPs gevonden zouden worden na kwaliteitscontrole.

### Populatie-structuur

Om te bepalen wat de populatiestructuur voor stekelrog is in heel West-Europa is er naast de genotypes die verzameld zijn in het INNORAYS-project ook gebruik gemaakt van genotypes die bepaald zijn met behulp van weefselmonsters die verzameld door IFREMER in Frankrijk. Deze genotypes van de IFREMER monsters zijn bepaald met behulp van dezelfde SNP array als die gebruikt zijn voor het project. Daarom konden deze direct gebruikt worden voor het bepalen van de populatiestructuur in een groter verspreidingsgebied dan alleen de Noordzee.

De populatiestructuur in deze monsters op basis is bepaald door deze monsters in te delen op basis van hun geografische herkomst. Daarvoor zijn 6 gebieden gebruikt, gebaseerd op gebieden die ook in ICES gebruikt worden (Figuur 5): Centrale Noordzee, zuidelijke Noordzee, Engelse kanaal, Ierse Zee, Bristol Channel & Keltische Zee, en Golf van Biskaje. De populatiestructuur is vervolgens bepaald met behulp van twee methodes: met behulp van *admixture* modellen door middel van fastSTRUCTURE (Raj et al., 2014), en met behulp van paarsgewijze  $\theta_{ST}$  (Weir & Cockerham, 1984) en  $F_{ST}$  waardes (Nei, 1987).

## Verwantschapsanalyse

Voor de analyses van de familierelaties zijn in ieder geval de volgende vormen van verwantschap van belang: (i) Parent-offspring pairs (POPs), (ii) Full-sibling pairs (FSPs), (iii) Half-Sibling pairs (HSPs). Deze verwantschap kan op twee manieren bepaald worden: door Mendeliaanse exclusie (zie bijvoorbeeld Bravington et al. 2016), of met behulp van een maximum likelihood model (zie bijvoorbeeld Hillary et al. 2019).

Voor de verwantschapsanalyse moeten de SNPs die gebruikt worden in Hardy-Weinberg evenwicht zijn (Hillary et al., 2019). Dat betekent dat de allel en genotype frequenties in de populatie constant blijven van generatie op generatie. SNPs die niet in Hardy Weinberg equilibrium waren werden verwijderd van de SNP-dataset.

Verwantschapsbepaling door exclusie kan alleen gebruikt worden voor POPs. Daarin wordt aan de hand van genetische afstanden in het homozygote deel van de genotypes bepaald of beesten POPs zijn: bij foutloze genotypering moet de genetische afstand voor deze homozygote delen nul zijn, terwijl de genetische afstand voor de gehele genomen groter moet zijn dan nul. Afhankelijk van de grootte van de fouten in de genotypering neemt de afstand in het homozygote deel toe. Om POP paren te identificeren werden daarom figuren gemaakt waarbij voor elke combinatie van individuen de genetische afstanden voor homozygote delen van genotypes op de x-as geplaatst werden en genetische afstanden voor de gehele genotypes op de y-as.

De verwantschapsanalyse voor FSP en HSPs is gedaan op basis van Hillary et al. (2019). De methode in Hillary et al. (2019) werkt op basis van de Pseudo Likelihood werd toegepast op de resterende genotypes. Bij deze methode worden alle monsters vergeleken. Bij deze methode voor FSP wordt voor elk paar en voor elke SNP de waarschijnlijkheid van het geobserveerde genotype berekend voor twee type verwantschap: FSP of geen verwantschap (UP; *unrelated pair*). Bij deze methode voor HSP wordt voor elk paar en voor elke SNP de waarschijnlijkheid van het geobserveerde genotype berekend voor twee type verwantschap: HSP of geen verwantschap. Vervolgens wordt de log-waarde van de ratio van de twee waarschijnlijkheden (in het Engels "*Likelihood ratio*") berekend. Dit wordt voor alle SNPs binnen een vergelijking van twee individuen gedaan, en hiervan wordt het gemiddelde berekend om tot een "PLOD" statistiek te komen voor elk paar individuen. Als de PLOD waardes negatief zijn dan is dat een sterke aanwijzing tegen FSP verwantschap, en dus dat het *Pair Unrelated* (UP) is. Als de PLOD waardes positief zijn dan is dat een aanwijzing voor FSP verwantschap. De Pseudo-Likelihood methode kan vervolgens op een zelfde manier herhaald worden voor HSP verwantschap.

*Tabel 3. Identity-by-descent waarschijnlijkheden voor verschillende aantallen allelen:  $K_0$  voor 0 gedeelde allelen,  $K_1$  voor 1 gedeeld allel, en  $K_2$  voor 2 gedeelde allelen.*

Type verwantschap	Afkorting	IBD waarschijnlijkheid		
		$K_0$	$K_1$	$K_2$
Zelf, homozygote tweeling	S	0	0	1
Parent-offspring	POP	0	1	0
Full sibling pair	FSP	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$
Half-sibling Pair	HSP	$\frac{1}{2}$	$\frac{1}{2}$	0
Grootouder-kleinkind		$\frac{1}{2}$	$\frac{1}{2}$	0
Tante-neef		$\frac{1}{2}$	$\frac{1}{2}$	0

De PLOD statistiek voor het bepalen van verwantschap maakt gebruik van het feit dat de waarschijnlijkheid dat allelen verkregen worden door "identity-by-descent" anders is voor verschillende typen verwantschap. Deze zijn weergegeven voor verschillende typen verwantschap in Tabel 3. Zoals te zien in deze tabel is de verwantschap tussen *half-siblings* en grootouders met hun kleinkinderen of tantes/ooms met hun neven/nichten genetisch niet te onderscheiden. Daarom is aanvullende informatie nodig over het geboortjaar van de gevonden paren. Indien de verschillen in geboorte jaar kleiner zijn dan de leeftijd waarop de roggen geslachtsrijp worden, zijn alternatieve verwantschappen anders HSPs met grote zekerheid uit te sluiten.

## Populatiegrootte

De populatiegrootte kan in het Close-Kin Mark-Recapture raamwerk bepaald worden met behulp van POP paren of HSP paren. Naast de aanwezigheid van deze paren is ook het geboortjaar van de individuen in de bemonsterde populatie van belang. Deze is nodig om alternatieve familierelaties van HSPs uit te sluiten, maar ook om in de bestandsschatting een nauwkeurigere schatting te maken van de populatiegrootte (Hillary et al. 2019).

De leeftijden van de individuen in de HSP paren kan bepaald worden met behulp van een Von Bertalanffy groei vergelijking:

$$L_a = L_\infty (1 - e^{-K(t-t_0)}).$$

Deze vergelijking beschrijft de lengte  $L_a$  van een individu met leeftijd  $a$  met behulp van drie parameters: de asymptotische lengte  $L_\infty$ , de groeicoëfficiënt  $K$  en  $t_0$ , een correctie voor de lengte bij leeftijd 0. De inverse van de Von Bertalanffy groei vergelijking kan gebruikt worden om leeftijd te schatten aan de hand van de lengte van individuen:

$$a = -1/K(\log(1 - L_a/L_\infty) + t_0).$$

Voor stekelrog geldt dat de vrouwtjes en mannetjes andere parameters kennen voor de Von Bertalanffy groeicurve.

De parameters die nodig zijn om met behulp van een Von Bertalanffy groeicurve de leeftijd te bepalen met behulp van lengte zijn bepaald aan de hand van lengtegegevens uit visserijreizen en surveys met onderzoeksschepen (Schuitemaker, 2020), met behulp van de ELEFAN techniek (Mildenberger et al. 2017). Voor mannetjes zijn de waardes 0.23, 109.3, en 0.094, voor respectievelijk  $K$ ,  $L_\infty$ , en  $t_0$ . Voor vrouwtjes zijn deze waardes 0.24, 120 en 0.0576. Daarmee kunnen de leeftijden van de individuen in de verwante paren bepaald worden.

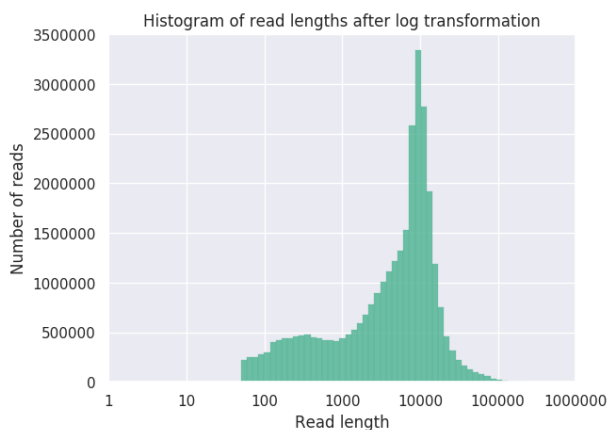
De populatiegrootte is met behulp van de methode Close-Kin Mark-Recapture methode beschreven in Hillary et al. (2018) op basis van HSP paren. Dit is een methode die de likelihood (in het Nederlands ook wel "aannemelijkheid" genoemd) van de omvang van de populatie volwassen dieren schat op basis van het aantal gevonden paren en het leeftijdsverschil tussen de paren. Behalve de Half-Sib paren zijn voor deze methode ook een geschatte populatiegroei en een jaarlijkse overlevingskans van de moeder benodigd. Voor de populatiegroei werd aangenomen dat de populatie stabiel was en niet groeide in de periode 2013-2019. Voor de jaarlijkse overlevingskans van de moeder werd aangenomen dat 90% van de dieren jaarlijks overleeft. Als er meer data beschikbaar komt in de toekomst dan hoeven de populatiegroei en sterfte van de moeders niet aangenomen te worden, maar kunnen deze ook geschat worden in het model

# Resultaten

## Genoom assemblage

### *Stekelrog*

De sequence data die terugkwam van de sequence provider voor *Raja clavata* toonde een aantal eigenschappen die aangaven dat het proces niet optimaal is gelukt, maar tegelijkertijd ook de beperkingen aantonen van de stand van de technologie op het moment van data acquisitie. De voornaamste parameter is de lengte van de DNA moleculen die zijn gesequenced (zie Figuur 1; Tabel 1). De mediaan ligt rond de 9 duizend basen. Dit is een acceptabel, maar geen geweldig resultaat. Positief was het aantal verkregen basen, en bovendien bleek achteraf onze schatting van de grootte van het genoom redelijk in lijn met het behaalde resultaat. De hoeveelheid behaalde data was daarmee ruim voldoende, hoewel gehoopt werd dat de gemiddelde lengte van de sequences (de 'sequence reads') hoger zou zijn. Echter, met de stand van de PacBio technologie op moment van DNA sequencing was dit resultaat 'acceptabel'. Gezien het hoge aantal sequence fouten die in deze technologie gangbaar zijn werd ook een toegevoegde dataset gegenereerd om die fouten te kunnen corrigeren, in de vorm van 'Illumina short-read' sequences. In totaal werd ruim 75Gb (giga-basen) van deze data toegevoegd.



Figuur 1 Histogram van de read lengths van de PacBio sequence voor stekelrog (*Raja clavata*).

De genoom assemblage van *Raja clavata* bleek uiteindelijk redelijk succesvol, tenminste in sommige statistieken. De eerste statistiek is de totale grootte van de assemblage: 2,474,273,273 basen. Dat is iets kleiner dan de ongeveer 3 miljard basen die we verwachtten. Een genoom assemblage resulteert normaal gesproken in een groter of kleiner aantal fragmenten die samen het totale genoom vertegenwoordigen. De N50 statistiek, die aangeeft hoe continue de geassembleerde stukjes zijn, is 884,727 basen. Dat is een wat teleurstellend resultaat, omdat dit aangeeft dat we nog erg ver weg zijn van het bij elkaar puzzelen van het genoom tot chromosoom-armen (het best denkbare resultaat).

De reden voor deze gefragmenteerde assemblage is deels duidelijk: maar liefst 2/3 van het geassembleerde genoom blijkt van het type 'interspersed repeat' te zijn, ofwel zogenaamde 'transposable elements' – DNA-sequenties die random door het DNA springen en zich zo kopiëren (ter vergelijking: in het menselijk genoom is dit 45%). Omdat we een niet erg optimaal zoekbeeld hebben voor deze transposons is dit ongetwijfeld nog een onderschatting. Daarbij komt dat de 'simple repeats' erg slecht te kwantificeren zijn als ze geconcentreerd zijn, bijvoorbeeld in het centromeer, omdat ze in de genoom assemblage te veel op elkaar lijken. Een ruwe schatting leert dus dat we wellicht rekening dienen te houden met >80% repeats en repeat elements (ter vergelijking, dit is 50% in de mens). Dit verklaart veel van de problemen bij de genoom sequencing en assemblage.

Het goede nieuws is echter dat we wel bijna alle genen van de rog hebben kunnen karakteriseren. Een functionele analyse (zogenaamde BUSCO-score) leerde dat ongeveer 86% van de genen compleet aanwezig is, en maar liefst 94% van de genen tenminste gedeeltelijk. Dit leert ons tevens dat het grootste deel van de transposons zich heeft genesteld tussen genen en niet in de genen (bijvoorbeeld in intronen). Dit blijkt onder andere ook uit de betrekkelijk korte lengte van genen (ongeveer 31 duizend basen,

vergeleken met de mens rond de 66 duizend basen, terwijl het genoom vergelijkbaar is in grootte). Dit betekent dus dat we wel gedetailleerd functioneel onderzoek kunnen doen op basis van de genomassemblage.

Een andere indicatie dat we wel degelijk het overgrote deel van het genoom in kaart hebben gebracht is dat we maar liefst 8,580 van de ongeveer 9 duizend SNPs van het SNP panel op het genoom hebben kunnen plaatsen. Dit betekent dus dat we daarmee de mogelijkheid hebben om populatie data, bijvoorbeeld indicaties van selectie, te gaan koppelen aan individuele genen. Maar ook hebben we daarmee een beeld van hoe SNPs bij elkaar horen omdat ze dicht bij elkaar in het genoom liggen, wat het mogelijk maakt om te komen tot meer gedetailleerde populatie genetische analyses in de toekomst.

*Tabel 4 Resultaten repeat analyse van het genoom voor Raja clavata.*

	Number of elements	Length occupied	Percentage of sequence
SINEs:	6,392	532,763 bp	0.02 %
ALUs	6	1,678 bp	0.00 %
MIRs	1	29 bp	0.00 %
LINEs:	4,132,522	945,692,440 bp	38.22 %
LINE1	44,341	49,545,971 bp	2.00 %
LINE2	12,971	5,745,819 bp	0.23 %
L3/CR1	2,867,461	654,002,253 bp	26.43 %
LTR elements:	242,058	150,679,006 bp	6.09 %
ERV1	222	47,861 bp	0.00 %
ERV1-MaLRs	17	759 bp	0.00 %
ERV classI	12,094	7,242,455 bp	0.29 %
ERV classII	1,056	76,998 bp	0.00 %
DNA elements:	148,210	32,396,953 bp	1.31 %
hAT-Charlie	2,206	567,419 bp	0.02 %
TcMar-Tigger	2,296	181,954 bp	0.01 %
Unclassified:	2,782,612	490,847,586 bp	19.84 %
Total interspersed repeats:		1,620,148,748 bp	65.48 %
Small RNA:	630	45,837 bp	0.00 %
Satellites:	4,237	838,863 bp	0.03 %
Simple repeats:	290,129	24,199,428 bp	0.98 %
Low complexity:	48,255	5,077,750 bp	0.21 %

### *Blonde rog*

Het karakteriseren van dit type genomen is niet gelukt met 10X genomics. Het *Raja brachyura* genoom bleek zich slecht te lenen voor deze technologie. In totaal wisten we minder dan 2 Gb van het genoom te karakteriseren met 10 Genomics, en dan nog enkel in kleine fragmenten. de overgrote meerderheid in stukjes van minder dan 50 duizend basen. Ook het percentage teruggevonden genen was zeer teleurstellend. In totaal zijn 2,513,055,107 basen van het genoom opgelost – ongeveer 60 tot 70% van het verwachte genoom (contig N50: 43,315). Met de opgedane ervaring voor *R. clavata*, en onze kennis van andere haaien en roggen genomen, is het aan te nemen dat ook genomen van *R. clavata* een exceptionele hoeveelheid repeats bevatten, waardoor de 10X Genomics methode niet goed werkt. Bovendien blijkt uit onze analyses van de ruwe sequence data dat het genoom van *R. brachyura* wellicht substantieel groter is dan van *R. clavata*. Mogelijk in de orde van 4.5 Gb, dus anderhalf keer groter. Op voorhand hadden we geen goede schatting van het *R. brachyura* genoom. Echter, het is bekend dat haaien en roggen binnen hetzelfde genus, van soort tot soort, significant in genoom grootte kunnen verschillen. Het is te verwachten dat deze vergroting van het genoom gepaard gaat met een verhoging van de complexiteit (bijv. de expansie van families van repeat elementen). Het *Raja brachyura* genoom is daarmee waarschijnlijk nog een flinke slag ingewikkelder om te assembleren in vergelijking tot *R. clavata*. Echter, delen die zijn geassembleerd geven een eerste karakterisering van het *R. brachyura* genoom en dit zal helpen om de Genotyping-By-Sequencing data voor deze soort in de genoom-context te plaatsen.

## *Oxford Nanopore sequencing*

Het betrouwbaar karakteriseren van genomen met zoveel transposons als de beide *Raja* soorten is volledig afhankelijk van twee factoren: 1) de lengte van de gelezen DNA-moleculen en 2) de betrouwbaarheid waarmee dat gebeurt. Met de ontwikkeling van nieuwe sequentie technologie kwamen gedurende het project nieuwe kansen. Met name de Oxford Nanopore sequentie bepaling heeft een stormachtige ontwikkeling gemaakt. Met deze technologie is het veel eenvoudiger om lange DNA-moleculen te lezen vergeleken met PacBio.

Hoewel de gemiddelde molecuul lengte van 30 duizend basen inderdaad gehaald wordt met behulp van Oxford Nanopore, blijkt dat het DNA van *R. brachyura* tot snelle degradatie leidt van de "pores" in de flowcells die gebruikt worden bij de Oxford Nanopore techniek. Dit heeft gevolgen voor de te behalen hoeveelheid gelezen DNA basen door elke flowcell. De oorzaak van de degradatie van de cellen ligt waarschijnlijk in de aanwezigheid van lange segmenten in het DNA met korte herhalingen. Het blijkt dat roggen niet de enige diersoorten zijn waar deze snelle degradatie van de flowcell is te zien; bijvoorbeeld ook vogels, die ook specifieke 'repeat elements' laten zien, vertonen iets vergelijkbaars, hoewel minder ernstig.

Intensief overleg met de leverancier, Oxford Nanopore, heeft geen definitieve oplossing opgeleverd, maar wel een kleine optimalisatie – door het breken van de langste moleculen ('shearing') wordt deels voorkomen dat deze lange moleculen gedurende hun aflezing de poriën waar ze doorheen migreren verstoppen. Voorts dient elke 8 tot 12 uur de run – die normaal gesproken twee dagen loopt – gestopt te worden. Tijdens elke pauze wordt de flowcell dan gespoeld om de poriën te ontstoppen, en moet vervolgens opnieuw geprepareerd DNA worden toegevoegd. Een zeer bewerkelijk procedé.

Met veel inzet – in totaal waren zo'n 20 spoelingen nodig over 6 flowcells - is derhalve tot een geoptimaliseerde procedure gekomen. Niet perfect, want het probleem van het verstopt raken van de poriën is maar deels opgelost. Echter, al met al is nu rond de 80 Gigabasen van *Raja clavata* gelezen (Tabel 1) op deze manier. De moleculen zijn veel langer dan bij de PacBio sequencing (read N50 van 31 duizend basen).

De assemblage op basis van deze data laat een veel beter resultaat zien – de contig N50 is 4.01 miljoen basen, met een totaal geassembleerd genoom van 2.9 miljard basen. Bovendien is het zeer bemoedigend dat het grootste fragment wat is gekarakteriseerd zo'n 50 miljoen basen is – dat gaat richting wat we verwachten van de kleinere chromosoom armen.

## Monsternamen en genotypering weefselmonsters

De monsters voor beide soorten zijn verzameld in het verspreidingsgebied van de twee roggensoorten: in de Zuidelijke Noordzee en het Oostelijke deel van het Engelse Kanaal (Figuur 2).

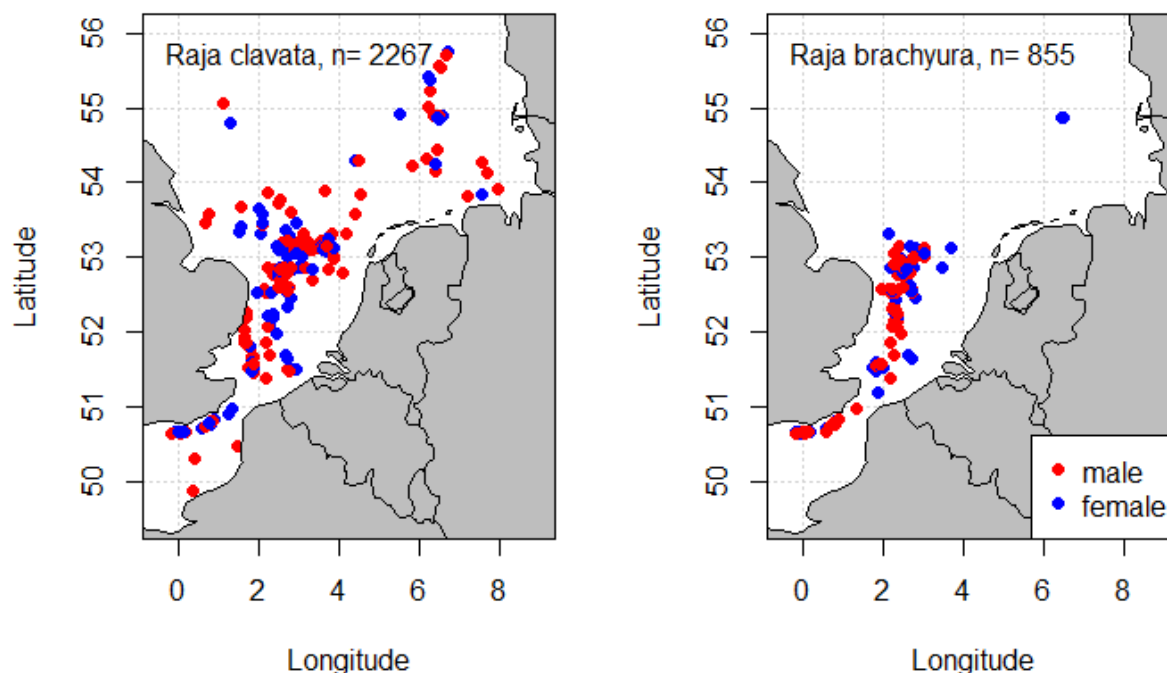
Aan het einde van het project zijn er 3122 monsters verzameld. Dat was meer dan de 3000 monsters die er bij aanvang van het project beoogd waren. Voor stekelrog zijn er 2267 monsters verzameld, terwijl er 2000 monsters beoogd waren. Dit grotere aantal monsters komt doordat het moeilijk bleek om exact te voorspellen hoeveel monsters er per monsternamen verzameld zouden worden. Uiteindelijk zijn er hierdoor dus ~ 10% meer monsters voor stekelrog verzameld dan beoogd.

Voor de blonde rog zijn er 853 van de 1000 monsters verzameld. Het verschil in aantallen komt voornamelijk omdat de abundantie van stekelrog in de Noordzee groter is dan die van blonde rog (Amelot et al. 2021). Daarom is het makkelijker om grote aantallen stekelroggen te verzamelen dan om blonde roggen te verzamelen.

De monsters voor beide soorten zijn op verschillende manieren verzameld: vanaf visafslagen, vanaf discardsreizen door observers, en direct aan boord door schippers, waarbij de samples in de visafslag genomen werd (Tabel 3). Daarnaast was er beschikking over de weefselmonsters uit Groningen niet alleen verzameld in de Noordzee en het Engels kanaal, maar ook in bijvoorbeeld de Ierse zee. Deze monsters



kunnen gebruikt worden voor de bepaling van de populatiestructuur buiten de Noordzee en het Oostelijk deel van het Engels kanaal.



*Figuur 2. Ruimtelijke verdeling van de monsters voor Stekelrog (Raja clavata, linker paneel) en Blonde rog (Raja brachyura, rechterpaneel). De totale aantallen samples staan in de figuren vermeld. Voor stekelrog zijn er meer individuen verzameld (2267) dan dat er individuen opgestuurd zijn voor genotypering (1994).*

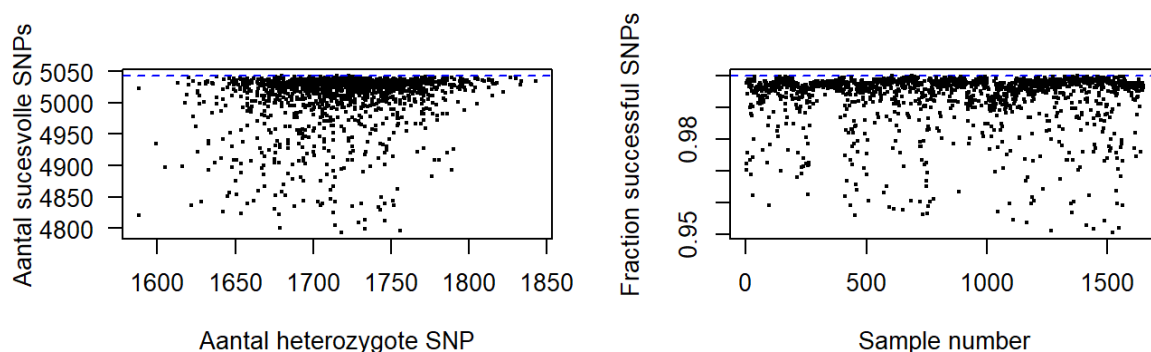
Het merendeel van de samples is verzameld aan boord van vissersschepen waarbij de bemanning roggan aangeland heeft met daarbij de exacte vangstlocatie (categorie "Vessel") en aan boord van verschillende onderzoeksreizen aan boord van onderzoeksschepen (categorie "Survey") (Tabel 3). Daar waar de monsters direct aan boord gevangen zijn (categorieën "survey" en "vessel") was de verwachting dat de kwaliteit dan de genotypes beter was: immers, de monsters werden van levende dieren verzameld en konden aan boord direct ingevroren worden. Uit analyses van de genotypes bleek echter dat er geen verschil in kwaliteit was tussen de genotypes die bepaald werden met behulp van weefselmonsters die direct van levende dieren verzameld werden of van dode dieren die werden aangevoerd en in de haven verzameld.

*Tabel 5. Herkomst van de monsters die voor genotypering zijn opgestuurd voor de twee soorten in het project. Deze aantallen zijn kleiner dan de totale aantallen die verzameld zijn in het project.*

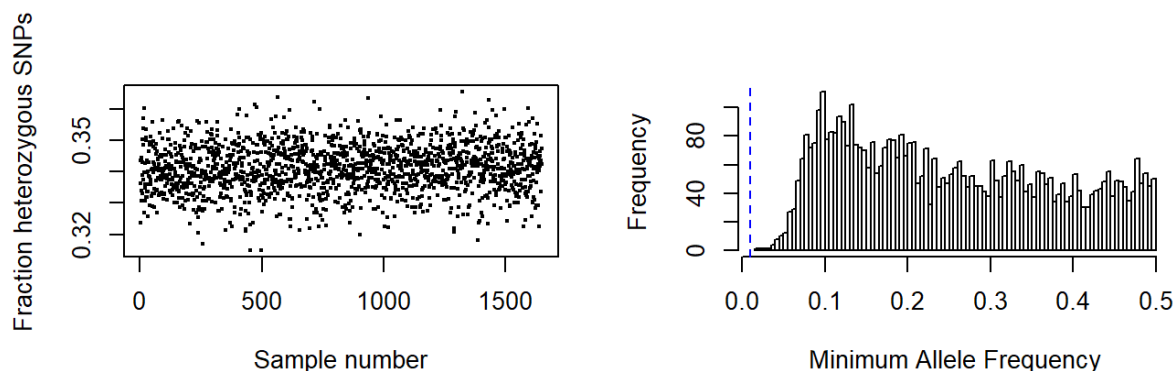
Soort	Herkomst monster					Totaal
	Auction	Discard trips	Groningen archive	Survey	Vessel	
Raja clavata	273	106	61	544	1283	1994
Raja brachyura	270	65	0	43	477	855

### *Stekelrog*

Van de stekelrog weefselmonsters zijn 1994 individuen opgestuurd voor genotypering. Omdat er een aantal individuen in duplo gebruikt zijn om de consistentie van de genotypering te testen, zijn er voor deze 1994 individuen 2016 genotypes bepaald. Uit analyse van de duplo's bleek dat dat de fouten in de genotypering van erg gering waren (gemiddeld <1 basepaar verschil in de volledige SNP set). Daarom kunnen we het exclusie principe toepassen voor de POPs in de verwantschapsanalyse.



*Figuur 3. Overzicht van de genotypes na kwaliteitscontrole. Succesgraad en heterozygotie van de 1650 monsters waarvoor genotypes beschikbaar zijn voor (paneel links). Elk punt is het genotype van één individu. Het maximaal aantal succesvolle SNPs in één individu is aangegeven met een blauwe horizontale lijn. Fractie succesvolle SNPs per genotype (paneel rechtsboven). In dit paneel staat het aantal resterende individuen/genotypes op de x-as en is de maximaal gevonden waarde aangegeven met een horizontale blauwe lijn.*



*Figuur 4. Fractie heterozygote SNPs per sample (paneel links). Verdeling van de minimum allel frequentie per SNP, met daarbij de drempelwaarde voor de minimum allel frequentie als verticale lijn (paneel rechts).*

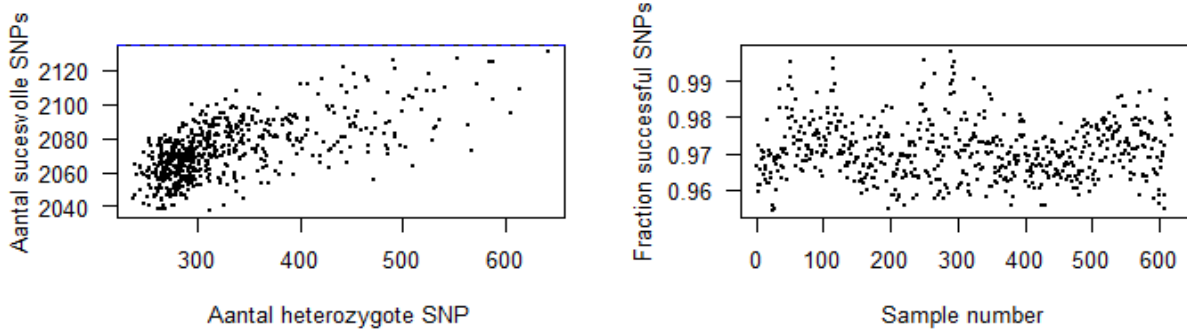
In de Niet alle SNPs op de chip hebben gewerkt voor de samples, en voor een aantal monsters hebben niet alle SNPs gewerkt. In de kwaliteitscontrole stap werden SNPs verwijderd die voor minder dan 95% van de monsters gewerkt hebben, of waarvoor de minimum allel frequentie lager was dan 0.01 (Figuur 3 & 4).

Uiteindelijk waren er na deze kwaliteitscontrole stap 1650 individuen over, met 5927 SNPs voor gebruik in de populatiestructuur analyse en de verwantschapsanalyse.

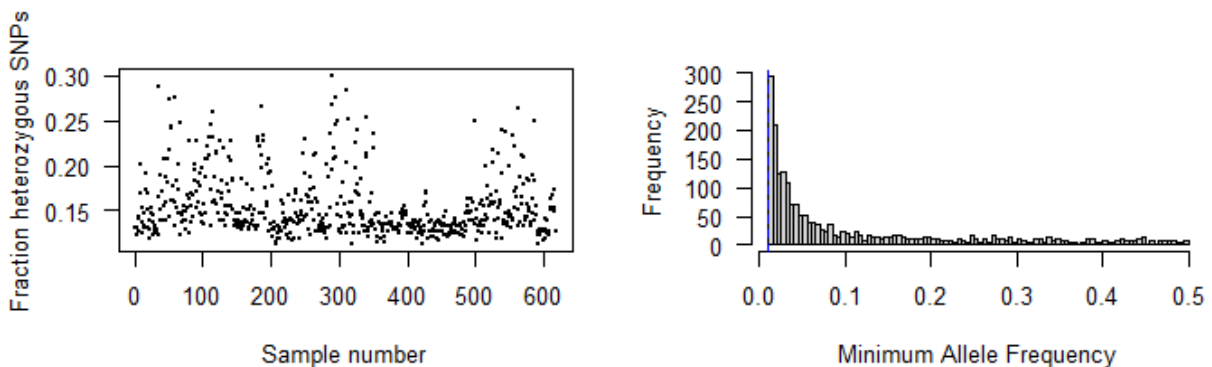
### **Blonde rog**

Er zijn in totaal 855 weefselmonsters verzameld. Er zijn 934 genotypes bepaald. Er zijn dus meer genotyperingen uitgevoerd dan weefselmonsters verzameld zijn voor de genotypering omdat er (1) een aantal individuen meerdere keren gegenotypeerd zijn om de consistentie van de genotypering te testen en (2) er monsters van andere soorten gegenotypeerd zijn met deze methode zodat genotypes van individuen die per ongeluk fout geïdentificeerd zijn verwijderd konden worden.

Met de GBS-techniek die we gekozen hebben zijn er ongeveer  $\sim 2100$  SNPs na kwaliteitscontrole. In deze kwaliteitscontrole worden SNPs verwijderd die voor minder dan 95% van de monsters gewerkt hebben, of waarvoor de minimum allel frequentie lager is dan 0.01 (Figuur 5 & Figuur 6).



Figuur 5. Overzicht van de genotypes na kwaliteitscontrole. Succesgraad en heterozygotie van de monsters waarvoor genotypes beschikbaar zijn voor (paneel links). Elk punt is het genotype van één individu. Het maximaal aantal succesvolle SNPs in één individu is aangegeven met een blauwe horizontale lijn. Fractie succesvolle SNPs per genotype (paneel rechtsboven). In dit paneel staat het aantal resterende individuen/genotypes op de x-as en is de maximaal gevonden waarde aangegeven met een horizontale blauwe lijn.

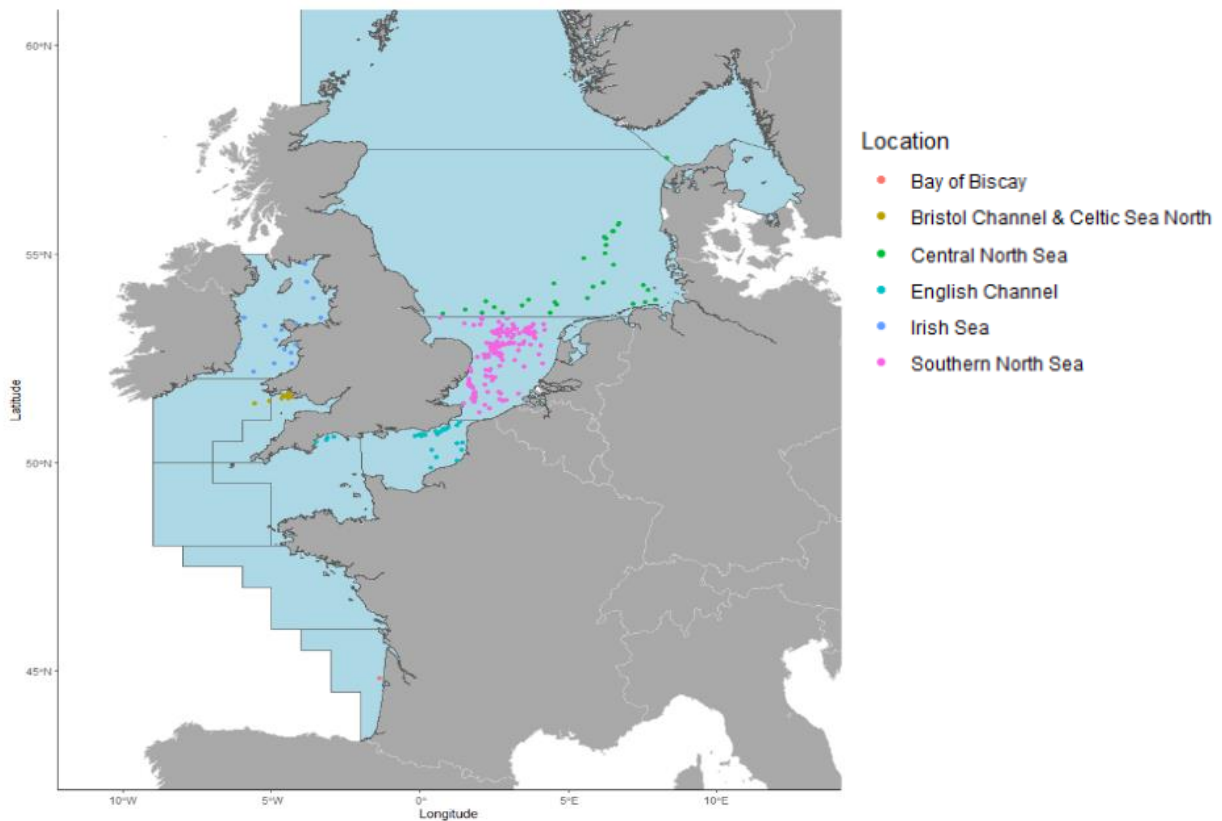


Figuur 6. Fractie heterozygote SNPs per sample (paneel links). Verdeling van de minimum allel frequentie per SNP, met daarbij de drempel waarde voor de minimum allele frequency als verticale lijn (paneel rechts).

Na de kwaliteitscontrole zijn er 618 samples over met 1588 SNPs (Figuur 6). Dit is niet voldoende om tot een nauwkeurige verwantschapsanalyse te komen. Het feit dat de lage dichtheid GBS niet voldoende SNPs heeft opgeleverd om tot een goede verwantschapsanalyse te komen terwijl dit voor andere soorten wel gelukt heeft ligt waarschijnlijk aan de grootte van het genoom: het genoom van blonde rog is aanzienlijk groter dan voor de blauwvintonijn. Het is daarom moeilijker om betrouwbare SNPs te vinden met behulp van GBS met lage dichtheid.

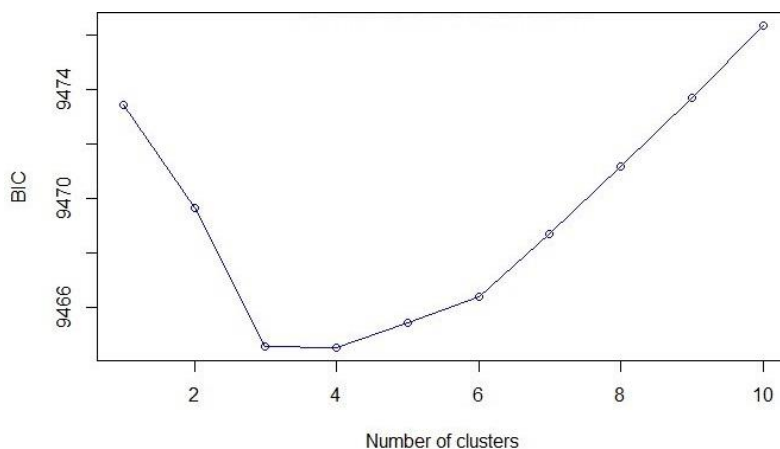
## Populatie-structuur

Om te bepalen wat de populatiestructuur voor stekelrog is in West-Europa is er gebruik gemaakt van alle genotypes die beschikbaar waren: de weefselmonsters die verzameld zijn gedurende het project (inclusief de monsters uit het weefselarchief in Groningen), en de genotypes van weefselmonsters die verzameld zijn door IFREMER in Frankrijk. Daarom kon de populatiestructuur in een groter verspreidingsgebied dan alleen de Noordzee bepaald worden (Figuur .7).

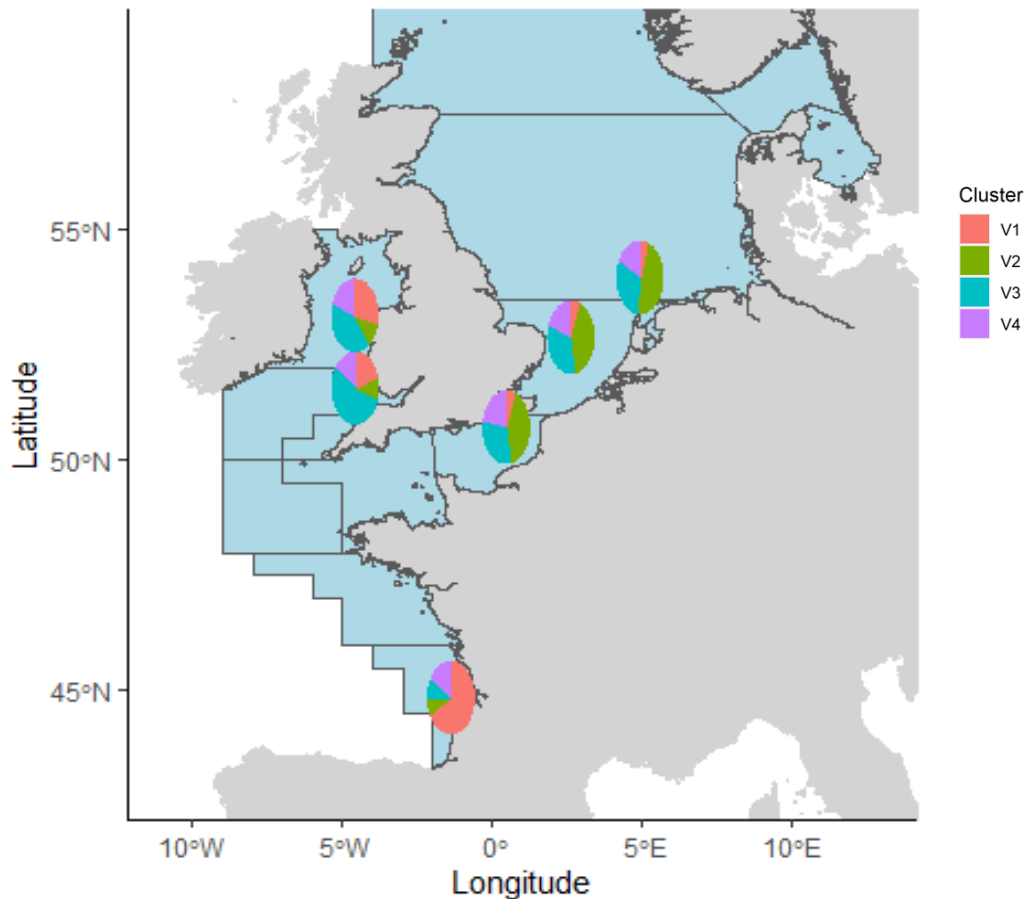


*Figuur 7. Locaties van alle genotypes die beschikbaar waren voor populatiestructuur van stekelrog, ook die verzameld zijn buiten de Noordzee en het Oostelijk deel van het Engelse kanaal. Elk punt in deze figuur is een locatie waar één of meerdere genotypes beschikbaar waren. De legenda geeft de classificatie van deze locaties in grotere gebieden weer, gebaseerd op ICES-regio's.*

De *admixture* modellen werden geanalyseerd voor verschillende clusterwaardes, waarbij de uitkomsten voor 3 of 4 clusters als beste modellen geëvalueerd werden op basis van hun BIC waarden (Figuur 8). De verdeling van de vier clusters in het desbetreffende model geeft aan dat de Golf van Biskaje duidelijk anders was dan de andere gebieden, met een dominantie van cluster 1. De Noordzee en het Engelse kanaal daarentegen werden gedomineerd door cluster 2, terwijl er binnen deze gebieden minder duidelijke verschillen waren in clustersamenstelling. De *admixture* modellen gaven daarmee een duidelijke eerste indicatie dat er wel een structuur is binnen de verschillende stekelrog populaties binnen Noordwest Europa, maar minder binnen de Noordzee en het Engelse kanaal (Figuur 9).



*Figuur 8. BIC waarden (Y-as) mogelijke aantallen clusters in de data (X-as). De laagste waarden voor de BIC geeft het beste model voor de gegevens. In het geval van stekelrog ( $X = 3-4$ ): Hier ligt het breekpunt in de balans tussen model complexiteit en verklarende waarde.*

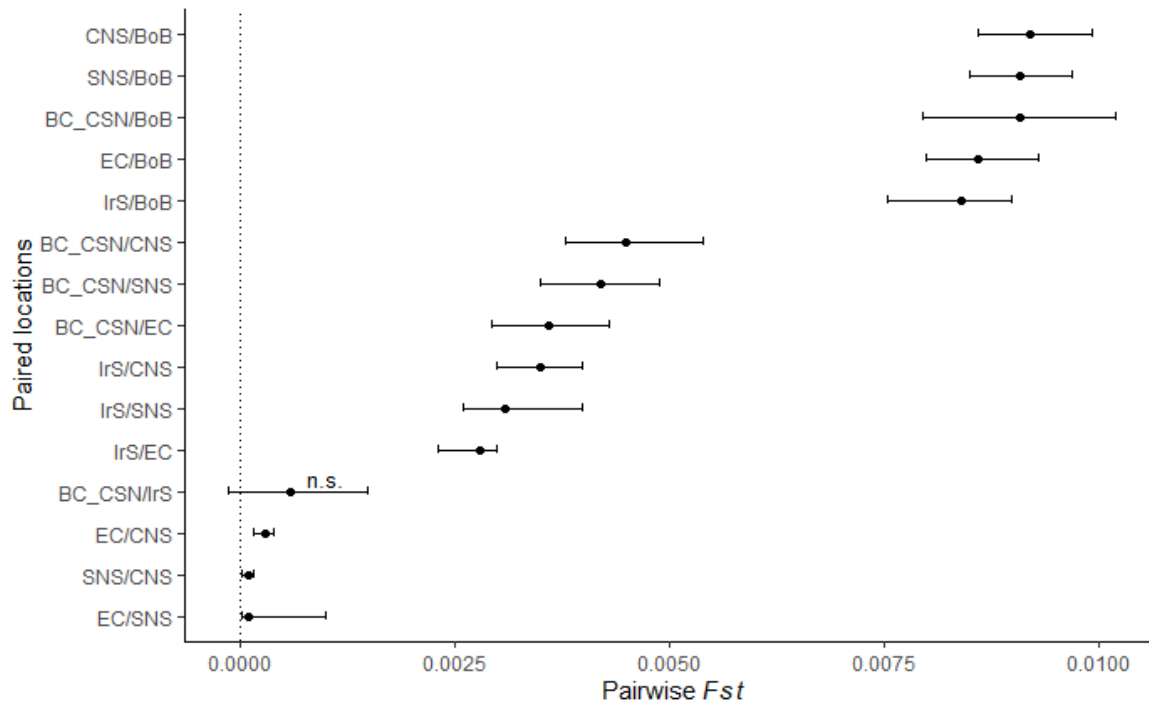


Figuur 9. Faststructure resultaten: de verdeling van de verschillende clusters over de gebieden voor de resultaten met vier clusters (V1-V4).

Paarsgewijze  $F_{ST}$  waarden zijn berekend tussen de monsterlocaties om de genetische differentiatie tussen de gebieden schatten. Zowel Weir & Cockerham's (1984)  $\theta_{ST}$  en Nei's (1987)  $F_{ST}$  waarden zijn geschat (Tabel 5). In deze tabel is te zien dat de  $F_{ST}$  waarden voor de Golf van Biskaje met andere gebieden hoger zijn dan voor de andere gebieden onderling. Figuur 10 laat dit beter zien: de  $F_{ST}$  waarden voor alle gebieden wijken significant af van 0, met uitzondering van het paar Bristol Channel - Ierse zee. De waarden voor gepaarde gebieden binnen het grotere Noordzee ecosysteem (d.w.z. Centrale Noordzee, Zuidelijke Noordzee en Engels kanaal) zijn echter substantieel lager dan de andere waarden voor de andere paren. Dit geeft aan dat er binnen de Noordzee relatief weinig genetische verschillen te vinden zijn tussen de subgebieden.

Tabel 6. Matrix van paarsgewijze Weir & Cockerham's (1984) ( $\theta_{ST}$ ) and Nei's (1987)  $F_{ST}$  waarden voor de verschillende monsterlocaties:  $\theta_{ST}$  waarden staan beneden de diagonaal,  $F_{ST}$  waarden staan boven de diagonaal. Afkortingen: BC\_CSN: Bristol Channel & Celtic Sea North; IrS: Ierse zee; BoB: Golf van Biskaje; EC: Engels kanaal; SNS: zuidelijke Noordzee; CNS: Centrale Noordzee.

	BC_CSN	IrS	EC	SNS	CNS	BoB
BC_CSN		0.0006	0.0036	0.0042	0.0045	0.0091
IrS	0.0007		0.0028	0.0031	0.0035	0.0084
EC	0.0036	0.0028		0.0001	0.0003	0.0086
SNS	0.0042	0.0031	0.0001		0.0001	0.0091
CNS	0.0045	0.0035	0.0003	0.0001		0.0092
BoB	0.009	0.0083	0.0087	0.0091	0.0093	

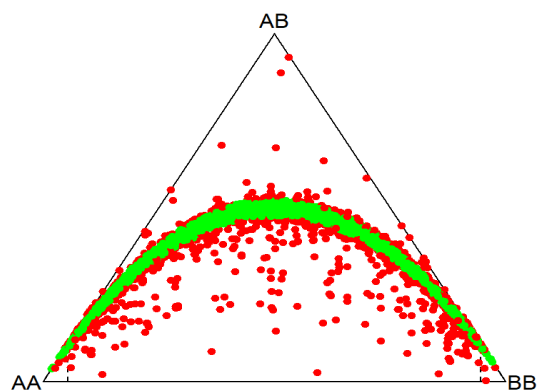


Figuur 10. Nei's (1987)  $F_{ST}$  waarden voor gepaarde gebieden, met bijbehorende 95% betrouwbaarheidsintervallen. Voor betekenis van afkortingen zie Tabel 2. Punten geven de geschatte  $F_{ST}$  waarde per paar. De horizontale lijnen geven de 95% betrouwbaarheidsintervallen aan. De verticale lijn geeft een  $F_{ST}$  waarde van 0. N.s.: niet significant afwijkend van 0.

## Verwantschapsanalyse

### Stekelrog

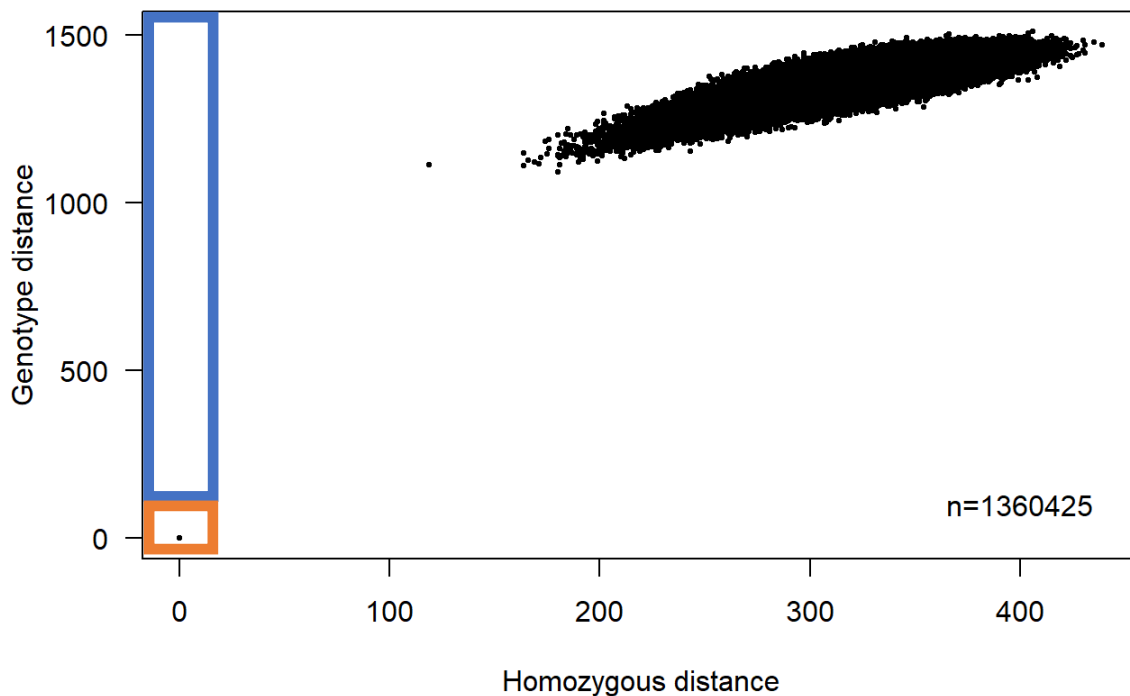
De meeste van de SNPs waren in Hardy-Weinberg evenwicht, maar er was ook een deel wat significant afweek (Figuur 11). Na verwijdering van de SNPs die niet Hardy-Weinberg evenwicht waren bleven er 1650 samples over met 5043 SNPs.



Figuur 11. Ternary plot van de genotype frequenties van de beschikbare SNPs in de samples van Stekelrog. Groene punten geven SNPs in Hardy-Weinberg evenwicht (HWE) weer, rode punten geven SNPs weer die significant afwijken van HWE. Er zijn in totaal 267 SNPs waar de frequentie van de heterozygoten is significant hoger zijn dan verwacht en 617 waar de frequentie van de heterozygoten lager is dan verwacht.

De POP paren zijn te zien in Figuur 12. Die zowel de totale genotype afstand, als de genotype afstand voor de homozygoten weergeeft. De duplicaten in de dataset die per ongeluk ontstaan zijn tijdens de verwerking van de monsters zijn te zien in Figuur 12, dicht bij de oorsprong van de grafiek. Voor deze duplicaten is zowel de afstand voor de volledige SNP set, als voor de subset van homozygoten in beide individuen nagenoeg nul. Inderdaad zaten er 3 duplicaten in de set. Deze duplicaten zijn ontstaan door fouten tijdens

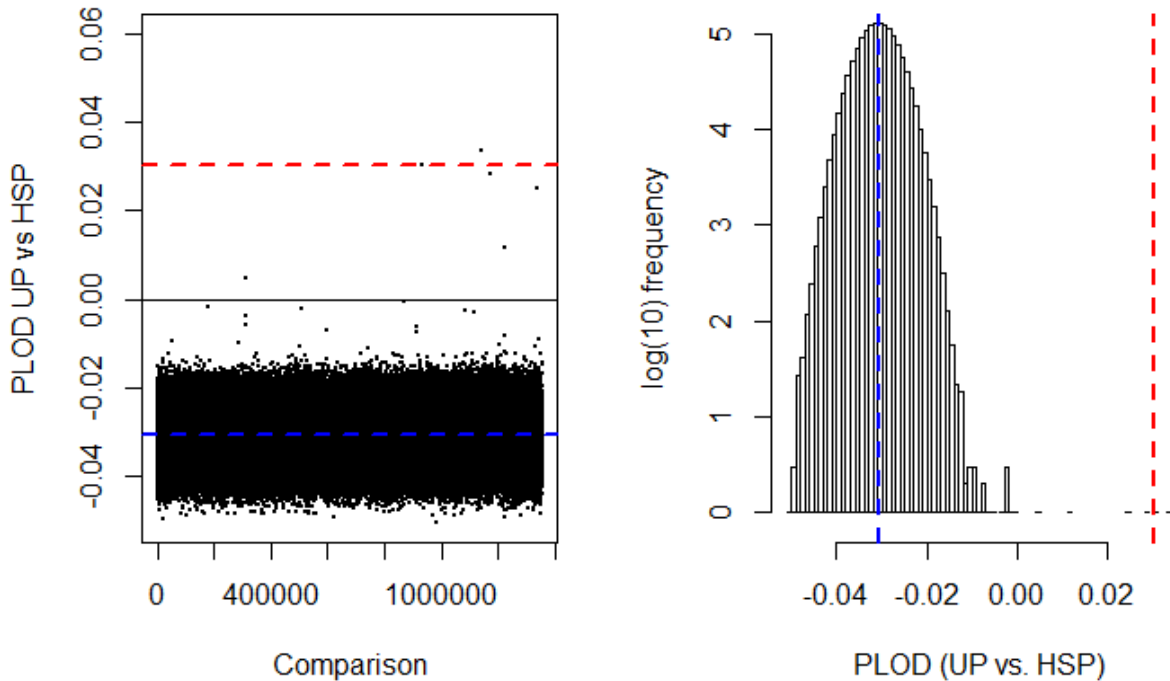
de bemonstering. In Figuur 12 verwachten we POPs op de lijn  $x=0$ , waarbij  $y>1$ . Zoals te zien is, is er in onze 1650 monsters nog geen POP paar gevonden.



*Figuur 12. Genotype afstanden tussen monsters voor de stekelrog genotypes. Afstand is bepaald als de som van de verschillen voor elke SNP, waarbij de afstand tussen heterozygoot en homozygoot gelijk is aan 0.5, afstand tussen twee ongelijke homozygoten gelijk is aan 1. Voor alle andere combinaties is de afstand 0. Op de x-as de afstanden waarbij alleen de SNPs vergeleken zijn die homozygoot zijn voor beide monsters. Op de y-as de afstanden voor het volledige genotype. Identieke samples zullen in deze figuur in de buurt van het punt (0,0) liggen, in het rode vak in de figuur. POPs zullen op de lijn (0,.) liggen, in het blauwe vak in de figuur. Bij 1650 monsters zijn er in totaal  $(1650*1649)/2$  afstanden te bepalen.*

Voordat de Pseudo-Likelihood methode toegepast werd om de FSP en HSP paren te bepalen moesten eerst de duplicaten uit de data gehaald worden. Deze werden, vanwege hun overeenkomsten in het genotype anders geclassificeerd als FSP verwante individuen. Om deze duplicaten te verwijderen zijn de resultaten in Figuur 12 gebruikt. Van de 1650 individuen zijn er dan 1647 over. Er zijn geen FSP paren in de data gevonden, maar wel 4 HSP paren in de data gevonden (Figuur 13). De verwachte PLOD score voor UP versus HSP is ongeveer 0.03. Rond deze lijn zijn vier paren te vinden. De overige waarnemingen worden gevonden rond de lijn die de verwachte PLOD score voor zulke unrelated pairs weergeeft.

Tabel 6 geeft de 8 individuen in de dataset die samen de 4 gevonden HSP paren vormen. Van de 1647 individuen zijn er dus 8 individuen die een HSP paar vormen, terwijl er geen POP of FSP paren gevonden zijn. Deze informatie, samen met de leeftijd van de bemonsterde dieren werd gebruikt in de populatieschatting voor stekelrog.



Figuur 13. PLOD schattingen voor Half-Sib pairs (HSPs) voor individuele schattingen (paneel links) en als histogram (paneel rechts). In het linker paneel geeft elk zwart symbool geeft een vergelijking tussen twee individuen. De blauwe lijnen geven de verwachting voor niet verwante individuen en de rode lijnen geven de verwachting voor HSP paren.

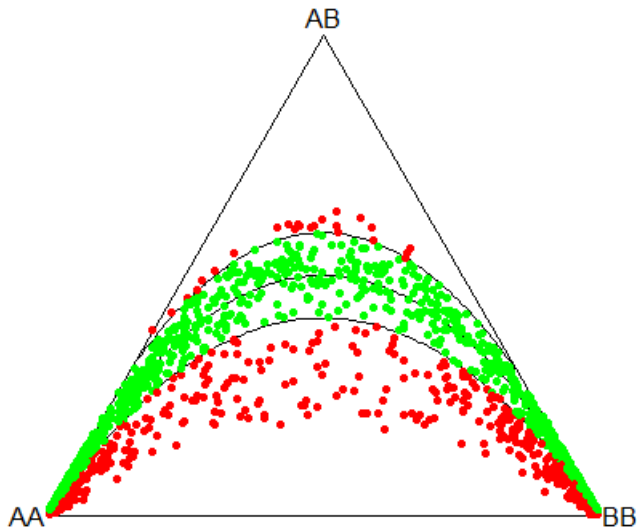
Tabel 7. Genotypes met binnen paren met PLOD UP vs. HSP schattingen  $> 0.02$ , die duiden op Half-Sibling verwantschap. Het aantal succesvolle SNPs en, het jaar van vangst zijn toegevoegd.

Genotype	Vis 1			Vis 2			PLOD UP vs HSP
	Aantal SNPs	succes	Jaar	Aantal SNPs	succes	Jaar	
EY0048505		5038	2020	WURH211	5019	2020	0.030
EY0054150		5022	2020	WURF723	4999	2020	0.034
EY0054209		5039	2019	WURG429	4967	2020	0.028
WURH410		5030	2020	WURH443	5030	2020	0.025

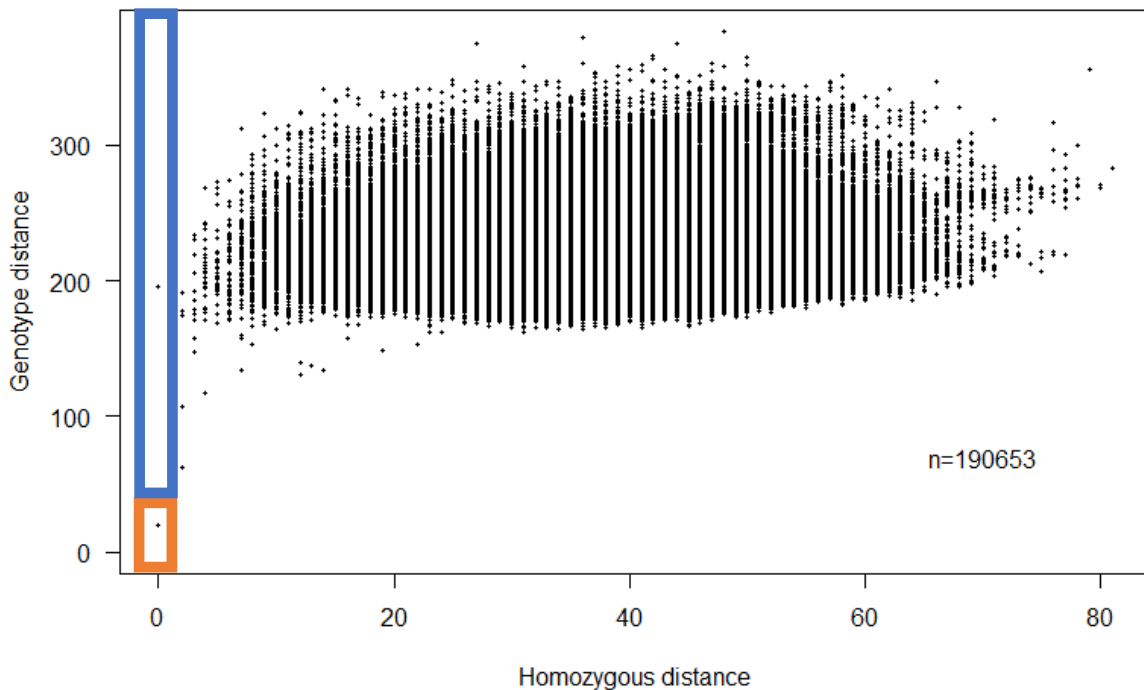
### Blonde rog

Voor de verwantschapsanalyse van Full-Sib en Half-Sib pairs moeten de SNPs gebruikt worden die in Hardy-Weinberg evenwicht zijn. Een eerste verkenning van de resultaten voor de stekelrog geeft aan dat de meeste van de SNPs inderdaad in Hardy-Weinberg evenwicht zijn, maar dat er ook een deel significant afwijkt (Figuur 14). In totaal zijn er 2006 SNPs die gebruikt konden worden voor de verwantschapsanalyse. Dat is aanzienlijk minder dan de  $\sim 5000$  SNPs die normaliter nodig zijn voor goede verwantschapsanalyse. Desondanks zijn de analyses wel uitgevoerd, en de resultaten worden hieronder besproken.





Figuur 14. Ternary plot van de genotype frequenties van de beschikbare SNPs in de samples van blonde rog. Groene punten geven SNPs in Hardy-Weinberg evenwicht (HWE) weer, rode punten geven SNPs weer die significant afwijken van HWE. Er zijn in totaal 2006 SNPs waar de frequentie van de heterozygoten binnen de verwachtingswaarde van het Hardy-Weinberg equilibrium ligt.

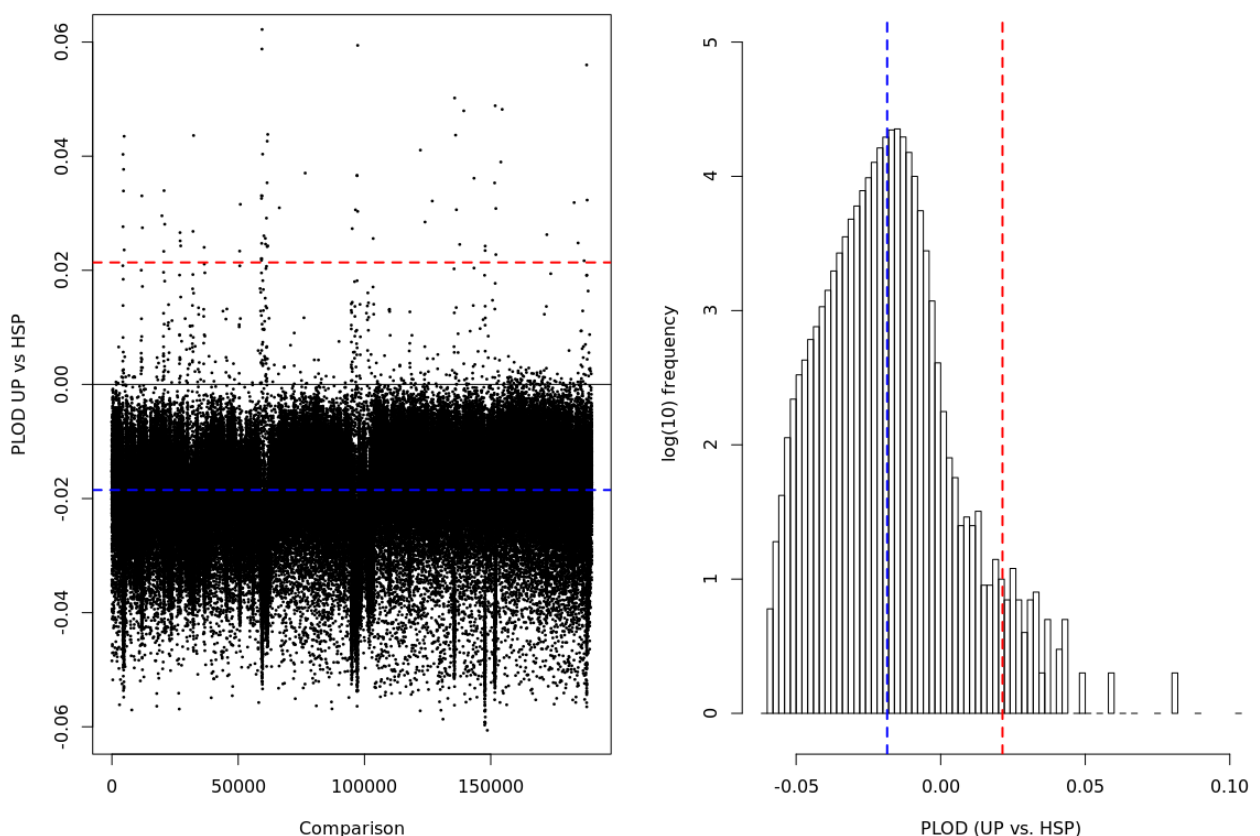


Figuur 16. Genotype afstanden tussen monsters voor de blonde rog genotypes. Afstand is bepaald als de som van de verschillen voor elke SNP, waarbij de afstand tussen heterozygoot en homozygoot gelijk is aan 0.5, afstand tussen twee ongelijke homozygoten gelijk is aan 1. Voor alle andere combinaties is de afstand 0. Op de x-as de afstanden waarbij alleen de SNPs vergeleken zijn die homozygoot zijn voor beide monsters. Op de y-as de afstanden voor het volledige genotype. Identieke samples zullen in deze figuur in de buurt van het punt (0,0) liggen, in het rode vak in de figuur. POPs liggen op de lijn (0,.), in het blauwe vak in de figuur. Bij 618 monsters zijn er in totaal  $(618*617)/2$  afstanden te bepalen.

Omdat er een aantal monsters in duplicaat verstuurd zijn weten we inmiddels dat de fouten in de genotypering gering zijn ( gemiddeld  $<10$ ). Daarom kunnen we het exclusie principe toepassen voor de POPs. Er is in Figuur 16 één combinaties van genotypes gevonden dicht bij de oorsprong van de grafiek. Voor deze duplicaten is zowel de afstand voor de volledige SNP set, als voor de subset van homozygoten in beide individuen nagenoeg nul. Dit duplicaat werd verwijderd uit de dataset.

In Figuur 16 verwachten we POPs op de lijn  $x=0$ , waarbij  $y>1$ . Zoals te zien is, is er in onze 618 monsters nog één POP paar gevonden. Echter, zoals te zien is in Figuur 16 ligt de puntenwolk die de niet POP paren beschrijft dicht bij de  $y$ -as. Dit komt omdat er relatief weinig SNPs gevonden zijn en omdat de genotyperingsmethode niet zo robuust is als de SNP array die gebruikt is voor de stekelrog. Bij het aanwijzen van dit POP paar is er dus een kans dat deze een "false-positive" is: een paar dat aangewezen wordt, terwijl de dit in werkelijkheid geen paar is.

Vervolgens is er een analyse gemaakt van de half-sib paren. De resultaten van deze analyse wordt getoond in Figuur 17. Daar wordt nogmaals duidelijk wat het effect is van de lage dichtheid aan SNPs. Er worden wel resultaten waargenomen die rond de verwachte PLOD score voor Half-Sib paren liggen, maar er is geen duidelijke scheiding tussen de verwachte waarnemingen van Unrelated pairs versus Half-sib pairs. Dit wordt duidelijk als Figuur 17 vergeleken wordt met Figuur 13, waar de resultaten voor de stekelrog in staan. Er zijn dus geen duidelijke Half-Sib pairs aan te wijzen: elke aangewezen paar heeft een grote kans om een "false positive" te zijn, die eigenlijk in de verdeling rond de Unrelated pairs valt.



*Figuur 17: PLOD schattingen voor Half-Sib pairs (HSPs) voor individuele schattingen (paneel links) en als histogram (paneel rechts) voor blonde rog. In het linkerpaneel geeft elk zwart symbool geeft een vergelijking tussen twee individuen. De blauwe lijnen geven de verwachting voor niet verwante individuen en de rode lijnen geven de verwachting voor HSP paren.*

## Populatiegrootte

### Stekelrog

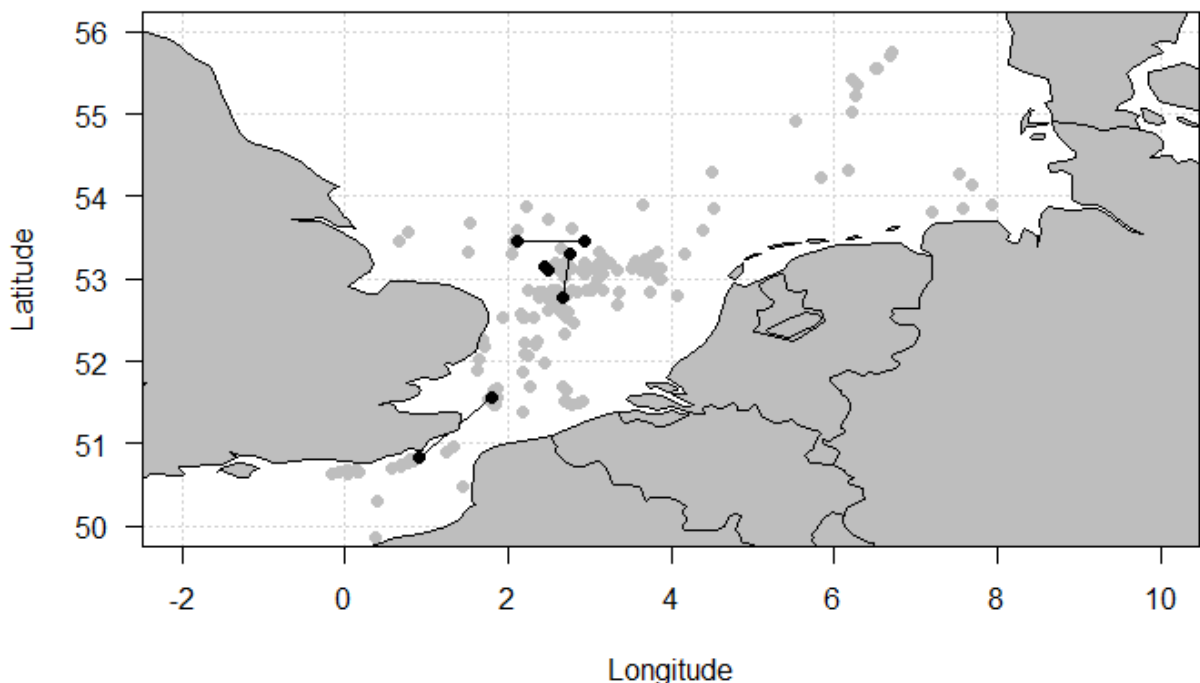
Voor een schatting van de populatiegrootte zijn leeftijdsschattingen nodig zijn voor de bemonsterde dieren. Deze leeftijden zijn nodig om alternatieve verwantschappen voor HSPs uit te sluiten, maar ook omdat de leeftijd cruciale informatie bevat die nodig is in de bestandsschatting. Voor de schattingen van de leeftijden van de HSPs zijn daarom hun lengtes nodig die omgezet zijn naar leeftijden. Voor de vissen binnen de Half-Sib paren zijn deze weergegeven in Tabel 5.

Omdat de individuen in de gevonden HSP paren relatief klein zijn (het grootste individu is 75 cm), is de leeftijdsbepaling relatief goed: beneden 75 cm is de Von Bertalanffy groeicurve nog relatief steil, zodat kleine verschillen in lengte veroorzaakt worden door relatief grote verschillen in leeftijd. Bovendien is de grootste lengte van 75 cm in de ongeveer gelijk aan de lengte waarbij deze roggen geslachtsrijp worden. Het is daarom onwaarschijnlijk dat de paren die als HSP worden aangemerkt grootouders - kleinkinderen zijn, of bijvoorbeeld tante/oom - neef/nicht paren.

Tabel 8. Genotypes met binnen paren met PLOD UP vs HSP schattingen > 0.02, die duiden op Half-Sibling verwantschap. De lengte van de individuen, het geslacht, en de geschatte leeftijd zijn toegevoegd.

Genotype	Vis 1				Vis 2				
	Lengte (cm)	Geslacht	Leeftijd (jaar)	Cohort	Genotype	Lengte (cm)	Geslacht	Leeftijd (jaar)	Cohort
EY0054150	59.8	F	2.93	2018	WURF723	57.9	M	3.39	2017
EY0054209	55.2	M	3.16	2018	WURG429	44.9	M	2.40	2018
EY0048505	57.0	F	2.74	2017	WURH211	71	F	3.79	2016
WURH410	74.7	F	4.12	2016	WURH443	51	M	2.84	2018

De locaties van de geschatte HSP paren zijn weergegeven in Figuur 18. Zoals te zien is zijn de Halfsib paren veelal geografisch dicht bij elkaar gevonden, behalve voor één paar, welke de Zuidelijke Noordzee met het Engelse Kanaal verbindt.



Figuur 18. Ruimtelijke verdeling van de gevonden HSPs (zwarte symbolen verbonden door lijnen), ten opzichte van de locaties van alle bemonsterde individuen (grijze symbolen).

Op basis van de 8 half siblings die er gevonden zijn in onze sample van ~1650 individuen, en hun leeftijd kan de populatiegrootte geschat worden met behulp van de CKMR methode. De populatiegrootte wordt daarmee geschat op ~ 1 miljoen volwassen dieren. De schatting is echter omgeven door een grote onzekerheid: de 95% betrouwbaarheid intervallen worden geschat op 4 honderdduizend – 6 miljoen (Figuur 19).

Om onze schatting in perspectief te plaatsen kan deze vergeleken worden met een populatieschatting die recent gedaan is door Amelot et al. (2021). Daar wordt een totale biomassa geschat voor de populatie van 75 duizend ton. Deze getallen zijn niet direct te vergelijken: Ten eerste is de schatting van Amelot et al. uitgedrukt in gewicht en is de schatting uit de CKMR analyse in aantallen dieren. Ten tweede is de schatting van Amelot et al. een schatting voor het totale aantal dieren. Om toch een vergelijking te kunnen maken moet de schatting van het aantal geslachtsrijpe ouderdieren omgezet worden naar de totale gewichten van de populatie.

De gemiddelde lengte waarbij een vrouwelijke stekelrog geslachtsrijp wordt is ongeveer 74 cm (McCully et al., 2012). De lengte-gewicht relatie van stekelrog kan beschreven worden als:

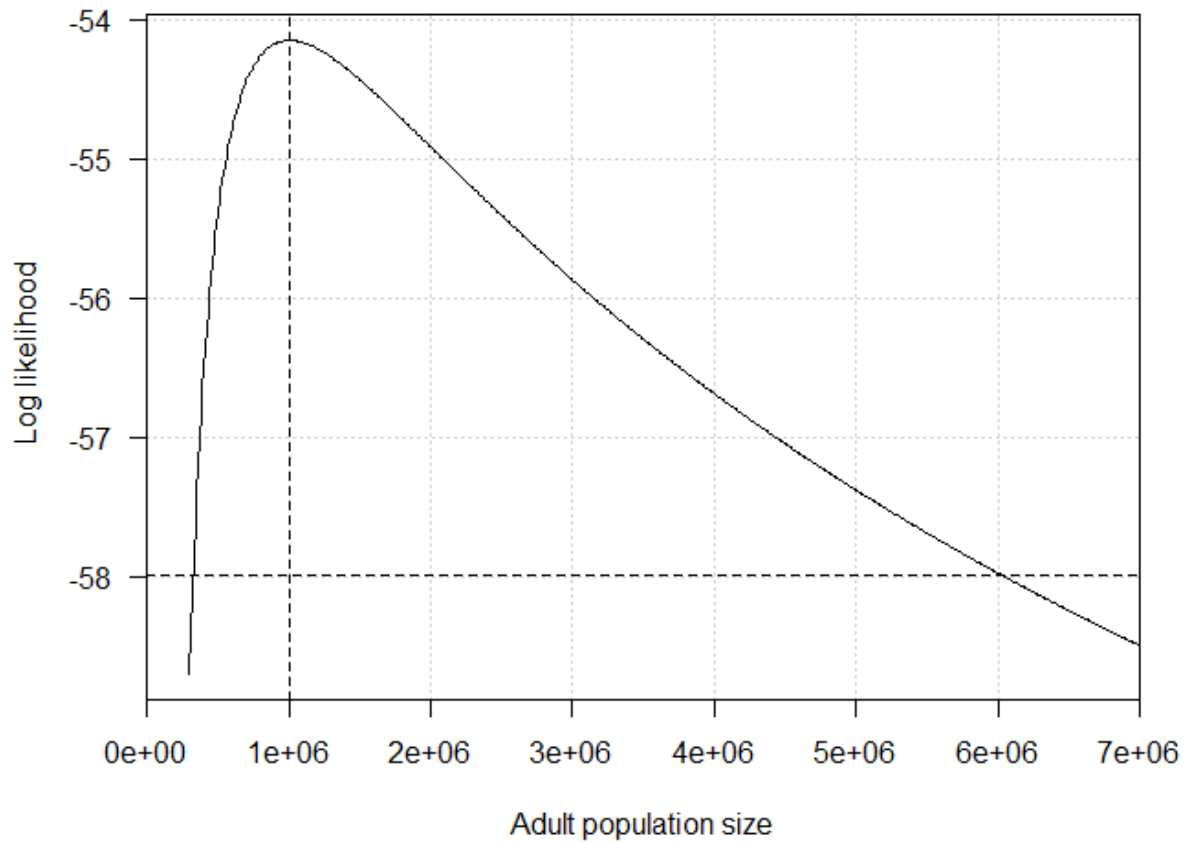
$$W = aL^b,$$

Met parameters  $a=0.0045$  en  $b=3.0686$ , waarbij de lengte gegeven is in cm en het gewicht in gram (Silva et al. 2013). Een vrouwelijke stekelrog weegt dus als deze geslachtsrijp wordt ongeveer 2.5 kg. Vervolgens is de verhouding tussen juvenielen en volwassen beesten nodig. Uit een studie van Parmentier (2019) blijkt dat de verhouding tussen adulte en juveniele beesten ongeveer 10% is. Dat zou betekenen dat er ongeveer tussen 4 miljoen en 60 miljoen jonge beesten zijn. De gemiddelde lengte van deze juveniele beesten kan aangenomen worden op 40 cm., met een bijbehorend gewicht van 0.370 kg.

Met behulp van deze aannames kan dan een ruwe schatting van de totale biomassa van de populatie gemaakt worden die hoort bij de Close-Kin Mark-Recapture schatting. Deze is 2.5 duizend ton – 37.2 duizend ton. Dat is dus lager dan de schatting in Amelot. Eén van de mogelijke verklaringen voor dit verschil is dat de schatting van Amelot de volledige vangsten en survey informatie uit de zuidelijke Noordzee en het Engelse Kanaal gebruikt heeft, terwijl de schatting van de CKMR gefocust is op de zuidelijke Noordzee (zie Figuur 1).

## Blonde rog

Omdat er geen duidelijke verwante dieren gevonden zijn is er voor de blonde rog geen bestandsschatting gemaakt. Zoals eerder vermeld komt dit voornamelijk door het lage aantal SNPs dat er gevonden is met behulp van de low-density Genotyping-by-sequencing techniek.



*Figuur 19: Eén-dimensionaal likelihood profiel van de volwassen populatiegrootte. De maximum likelihood schatting voor de populatiegrootte is ongeveer 1 miljoen volwassen individuen, met 95% betrouwbaarheid intervallen die aangegeven worden met behulp van de gestreepte horizontale lijn. De verticale gestreepte lijn geeft de maximum likelihood schatting aan.*

## Conclusies

Voor stekelrog bleek er een goede assemblage van het genoom mogelijk te zijn met behulp van het materiaal wat er binnen het project verzameld is. In eerste instantie werd hierbij gekozen voor sw PacBio techniek, later aangevuld door Oxford Nanopore sequencing. Voor blonde rog was de sequentie informatie met behulp van 10X Genomics niet voldoende voor een goede assemblage van het genoom. Het toevoegen van een beperkt aantal baseparen met behulp van de Oxford Nanopore techniek heeft dit uiteindelijk niet kunnen verhelpen. Het feit dat het sequencen van de roggen lastig bleek komt waarschijnlijk door de aanwezigheid van lange segmenten in het DNA met korte herhalingen, gecombineerd met hun redelijk grote genoom.

De bemonstering voor genotypering van een groot aantal dieren was succesvol: van beide soorten is het aantal individuen dat beoogd was voor monsternamen behaald, waarbij de meeste individuen bemonsterd zijn in de Noordzee en het Engelse kanaal. Dit is tevens het voornaamste verspreidingsgebied van deze roggen bestanden. Na kwaliteitscontrole bleek dat de genotypering voor stekelrog voor 1650 van de 1994 (~ 83%) gegenotypeerde individuen in voldoende mate geslaagd was. Voor blonde rog was ~72% van de genotypes van de monsters van voldoende kwaliteit. Dit was bij aanvang van het project niet voorzien: De verwachting was dat de er een hogere mate van succesvolle genotypering zou zijn, zeker omdat de gebruikte techniek voor stekelrog bekend staat als robuust. Uiteindelijk was het aantal SNPs wat uit de analyses voor stekelrog kwam wel voldoende voor gebruik in een schatting van de populatiegrootte. Voor Blonde rog was de Low Density GBS techniek niet in staat om voldoende SNPs op te leveren voor een robuuste verwantschapsanalyse.

De populatiestructuur voor stekelrog kon bepaald worden voor een groot geografisch gebied: van de Golf van Biskaje tot aan de centrale Noordzee. Daarbij bleken wel duidelijke structuren tussen de Golf van Biskaje en de andere gebieden, maar geen duidelijke structuren in de Noordzee en het Engelse Kanaal.

De CKMR-techniek kon succesvol toegepast worden op elasmobranchen populaties. Met behulp van de genotypering van individuen kon voor stekelrog een bestandsschatting gedaan worden. De monsters voor deze bestandsschatting werden verzameld met behulp van onderzoeksschepen en met behulp van de visserijsector ook aan boord van vissersschepen. Een belangrijke voorwaarde voor succesvolle toepassing bleek een robuuste genotypering. Voor stekelrog werd deze gevonden in een SNP-array, die meer dan 5000 werkende SNPs opleverde voor de soort. Met behulp van deze SNP werden 4 half-sibling paren gevonden, en daarmee 8 dieren die aan elkaar verwant waren. Voor blond rog, waar uiteindelijk voor een goedkoper alternatief gekozen is om zo te onderzoeken of de methode ook met lagere kosten uitgevoerd kon worden, bleek dit alternatief niet genoeg SNPs op te leveren.

Voor blonde rog heeft de genotypering dus niet genoeg SNPs opgeleverd om tot een betrouwbare bestandsschatting te komen. Bij vervolgprouwen over deze soort zou idealiter een hogere dichtheid genotyping-by-sequencing gekozen moeten worden om tot een voldoende aantal SNPs te komen. Om tot een hoger aantal SNPs te komen zou een hogere dichtheid GBS gekozen kunnen worden, maar daarmee zijn ook hogere kosten gemoeid. De huidige gegevens zouden in de toekomst wel bijvoorbeeld gebruikt kunnen worden om populatiestructuren op Europese schaal waar te nemen.

Voor stekelrog heeft de genotypering daadwerkelijk een bestandsschatting opgeleverd. De schatting van de populatiegrootte voor stekelrog in de Noordzee op basis van de CKMR techniek is dat er tussen 4 honderdduizend en 6 miljoen ouderdieren zijn in de zuidelijke Noordzee. Deze bestandsschatting is nog onzeker doordat er een relatief klein aantal samples genomen is, maar deze kan in de toekomst verbeterd worden door samples toe te voegen. Daarnaast kan de CKMR-methode, die nu in isolatie uitgevoerd is, in de toekomst geïntegreerd worden in bestaande bestandsschattingen, om zo tot een populatieschatting te komen. De resultaten worden ingebracht in de 2023 "benchmarkmeeting" voor de roggenbestanden van de Internationale Raad voor Onderzoek der Zee (ICES), waarin alle bestaande en nieuwe gegevens voor bestandsschattingen worden geëvalueerd.

## Dankwoord

Deze studie zou niet mogelijk geweest zijn zonder de inspanning van talloze onderzoekers en vissers, die aan boord van hun schepen monsters verzameld hebben, of toestemming gegeven hebben om monsters te verzamelen. Daarvoor zijn wij hen zeer erkentelijk. Ook willen wij onze dank uitspreken aan Michelle van Leijssen en Joey Portier, die bijgedragen hebben aan de analyse van de populatiestructuur van stekelrog.

# Literatuur

- Amelot, M., Batsleer, J., Foucher, E., et al. (2021) Evidence of difference in landings and discards patterns in the English Channel and North Sea Rajidae complex fishery. *Fisheries Research* 242,106028.
- Bravington, M.V., Grewe, P.M., Davies, C.R. (2016) Absolute abundance of southern bluefin tuna estimated by close-kin mark-recapture. *Nature Communications* 7,13162.
- Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J., Kawamoto, K., Buckler, E.S., Mitchell, S.E. (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* 6(5),e19379.
- Hillary, R.M., Bravington, M.V., Patterson, T.A., et al. (2018) Genetic relatedness reveals total population size of white sharks in eastern Australia and New Zealand. *Scientific Reports* 8(1),2661.
- Kilian, A., Wenzl, P., Huttner, E., Carling, J., Xia, L., Blois, H., Caig, V., Heller-Uszynska, K., Jaccoud, D., Hopper, C., Aschenbrenner-Kilian, M., Evers, M., Peng, K., Cayla, C., Hok, P., & Uszynski, G. (2012) Diversity Arrays Technology: A generic genome profiling technology on open platforms. *Methods in Molecular Biology*, 888, 67– 89.
- Mildenberger, T.K., Taylor, M.H., Wolff, M. (2017) TropFishR: an R package for fisheries analysis with length-frequency data. *Methods in Ecology and Evolution* 8(11), pp. 1520-1527.
- Marandel F., Charrier G., Lamy J.-B., Le Cam S., Lorange P., Trenkel V. M. (2020) Estimating effective population size using RADseq: effects of SNP selection and sample size. *Ecology and Evolution*, 10: 1929–1937.
- McCully, S.R., Scott, F., and Ellis, J.R., (2012) Lengths at maturity and conversion factors for skates (Rajidae) around the British Isles, with an analysis of data in the literature. *ICES Journal of Marine Science*, 69(10): 1812-1822.
- Nei, M. (1987). *Molecular Evolutionary Genetics*. Columbia University Press. <https://doi.org/doi:10.7312/nei-92038>
- Parmentier, B. (2019) Thornback ray (*Raja clavata*) distribution and abundances over the past decades in the North Sea. MSc thesis. Aquaculture and Fisheries Group. Wageningen University and Research:35 pp.
- Raj, A., Stephens, M., Pritchard, J.K. (2014) FastSTRUCTURE: Variational inference of population structure in large SNP data sets. *Genetics* 197(2), pp. 573-589.
- Silva J. F., Ellis J. R. and Ayers R. A. (2013) Length-weight relationships of marine fish collected from around the British Isles. *Sci. Ser. Tech. Rep., Cefas Lowestoft*, 150: 109pp.
- Schuitemaker, A. (2020) Age and growth determination from length frequencies of two North Sea ray species (*Raja clavata* and *Raja brachyura*). MSc Thesis nr .2092. Wageningen University and Research. 26 pp.
- Trenkel, V.M., Charrier, G., Lorange, P., Bravington, M.V. (2022) Close-kin mark-recapture abundance estimation: Practical insights and lessons learned. *ICES Journal of Marine Science* 79(2), pp. 413-422.
- Weir, B. S., Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, 1358–1370.



# Appendix\_A

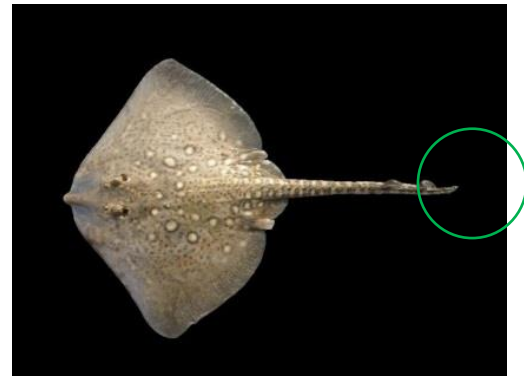
## **Sampling Protocol finclips rays and skates**

### **Material:**

Measuring board  
Scale (not compulsory)  
Gloves  
2 ml tubes filled with 70% pure ethanol  
Scissors  
Tweezers

### **Method:**

- 1) Provide geographic location of catch, as precisely as possible
- 2) Determine of the sex of the fish
- 3) Measure Total Length (TL) to the lower nearest cm
- 4) Where possible, measure the weight of the fish
- 5) Additional (not compulsory data): determine maturity stage



*Figuur A1 Location of dorsal fins*

### **Sampling**

Gloves may be used but are not mandatory. Rinse tweezers and scissors with tap or sea water between each fish(1). Cut a small piece of the dorsal fins(2) using tweezers and scissors. Store the samples in a pre-filled tube (photo). Tubes are pre-labeled and corresponding data are to be recorded on the provided form.

(1) The sampling needs to be done in domestic clean condition, e.g. like for cooking, but not strictly like "white lab" or "clean lab" conditions.

(2) Rays can be released alive, we already had ray sampled in this way, that were tagged and recaptured.

### **Storage of samples**

Store samples at -20 °C and transport in cooler with ice.



Thank

*Figuur A2 Example of 2 ml tube containing a sample of dorsal fin*

# Innorays: bestandsonderzoek met DNA

## Hoe en waarom?

In dit project onderzoeken we of we DNA-technieken kunnen gebruiken om een beter beeld te krijgen van de rogbestanden in de Noordzee. Er zijn nu nog te weinig gegevens over de rogbestanden voor een goede toestandsbeoordeling. Blonde rog en stekelrog zijn de soorten waar we deze technieken het eerst op testen.

**Vangstadvies**  
ICES\* geeft jaarlijks vangstadviezen voor commerciële visbestanden op basis van wetenschappelijk onderzoek. Hierop baseert de EU de vangstquota. Voor bestanden met weinig gegevens wordt een voorzichtig vangstadvies gegeven. Met meer gegevens zullen de adviezen beter aansluiten bij de praktijk.

\* De Internationale Raad voor Onderzoek der Zee (ICES) is een internationale wetenschappelijke organisatie gespecialiseerd in duurzaam gebruik van zeeën en oceanen.

1

Als eerste brengen we de gehele DNA-code (het genoom) van de blonde rog en de stekelrog in kaart. Het genoom is het "genetische bouwplan". Dit bouwplan is uniek voor elk individu, maar bij familie komt het bouwplan sterker overeen.

2

Als je weet hoe het bouwplan van de stekelrog en blonde rog er uit ziet, kun je bouwplannen van individuele roggen gaan vergelijken.

3

Onderzoekers knippen een stukje van uit de staart en slaan dit op in een buisje met alcohol. Deze DNA-monsters worden aan boord van kotters en op de visafslag verzameld.

4

Daarna vergelijken onderzoekers de DNA-monsters met elkaar. Als je veel familiebanden vindt, dan is de populatie waarschijnlijk klein. Als je weinig of geen familiebanden tussen de roggen vindt, dan is de populatie waarschijnlijk groot.

5

Een computermodel maakt met de DNA-gegevens en informatie over de leeftijd en de vangsten de bestandschatting.

### Onderzoeksamenwerking

#### Deelproject: innorays

De visserijsector, onderzoekers en maatschappelijke organisaties werken samen in onderzoeksprojecten aan duurzaam visserijbeheer. Wij bedanken deelnemende vissers voor hun bijdrage.

Dit onderzoek wordt gefinancierd door:

