



N I F E S

NATIONAL INSTITUTE
OF NUTRITION AND
SEAFOOD RESEARCH

Rapport

2016

21. mars

Nysekvensering av berggyltegenomet

Øystein Sæle¹, Sissel Jentoft², Ole Kristian Tørresen², Ave Tooming-
Klunderud², Lex Nederbragt² og Kai K. Lie¹

1: Nasjonalt institutt for ernærings- og
sjømatforskning (NIFES)

2: CEES, Institutt for biovitenskap, UiO



INNHALDSFORTEGNELSE

Sammendrag	3
Innledning	4
Problemstilling og formål	4
Prosjektgjennomføring	6
Oppnådde resultater	6
Referanser	7

Sammendrag

Genomet til en heldig utvalgt berggylte fra Marine Harvest sitt oppdrettsanlegg i Øygarden ble sekvensert. En kombinasjon av to ulike sekvenseringsteknologier ble benyttet: Illumina som gir korte sekvenser med veldig få feil og PacBio som gir lange sekvenser, men større sjans for feil. Denne sekvenseringsstrategien i kombinasjon med assemblyprogrammet canu og assemblyforbedringsprogrammet Pilon resulterte i et godt sammensatt genom med lange sammenhengende nukleotidesequenser (contigs: N50 704 kilobaser) og høy andel komplette gener funnet med valideringsverktøyet BUSCO (95 %). Det sammensatte berggyltegenomassemblyet er 805 megabaser (Mb), hvilket plasserer berggyltegenomet i samme størrelsesorden som genomene til torsk, medaka og tilapia. Til sammenligning er laksens genom estimert til å være ca 3000 Mb. Dette prosjektet har også demonstrert at det er mulig å frembringe et ferdig sammensatt og annotert genom fra en fisk i løpet av relativt kort tid og med stadig lavere kostnader.

Innledning

Å få bukt med lakselusproblemet på en bærekraftig måte er en stor utfordring, der rensefisk har allerede vist seg å være en del av løsningen. Bruk av rensefisk er blitt så populært at etterspørselen stadig stiger, og det er ikke lenger bærekraftig å kun satse på villfanget fisk. Løsningen kan bli oppdrett av de to mest lovende artene berggylt og rognkjeks. Skal berggylte være en del av løsningene på "Effektiv kontroll av lakselus med minst mulig medikamentbruk", en av FHF sine hovedmålsettinger, må produksjonen av berggylte øke. Takket være intensiv forskning og erfaring fra oppdrett av "nye" arter som kveite og torsk, og det nære samspillet mellom forskningsmiljøer og industri, har vi allerede nå en betydelig produksjon av både berggylte og etterhvert rognkjeks.

Prosjektet hadde oppstart 1. august 2015 og ble nylig avsluttet 1. mars 2016. Det totale budsjettet har vært på 427 000,- NOK.

Prosjektet har vært et samarbeid mellom Nifes og CEES/NSC ved universitetet i Oslo. Styring av prosjektet har lagt hos Øystein Sæle ved Nifes og styringsgruppen har bestått av Espen Grøtan og Olav Breck ved Marine Harvest i tillegg til Halvar Hovland, Havlandet Marin Yngel. Ivar Hordvik ved BIO, Universitetet i Bergen stilte som referansegruppe.

Problemstilling og formål

Flere analyseverktøy er i dag basert på molekylære metoder, der vi ser på regulering av gener og hvilke "versjoner" av genene en art har. Vi har allerede benyttet oss av dypsekvensering av berggyltens transkriptom (genuttrykk) for å belyse funksjonaliteten av denne fiskens bemerkelsesverdige fordøyelsessystem i det FHF finansierte prosjektet "LeppeProd". Vi jobber videre med denne fiskens fordøyelsessystem, med hovedfokus på appetittstyring og fordøyelse i NRF prosjektet "noGutsnoGlory". Målet er at den skal spise mer og vokse raskere. Andre aktører jobber med vaksiner, og oppdretterne selv ønsker seg

avl på arten. Et felles verktøy som vil representere et stort fremskritt i utviklingen av verktøykassen for alle de nevnte fagområdene vil være kartlegging av berggyltens genom.

Formålet med prosjektet var å sekvensere og annotere berggyltegenomet slik at FoU på arten har et viktig verktøy tilgjengelig. Konkret bruk kan være:

- Kartlegging av immunsystemet og utvikling av diagnostiske verktøy.
- Kartlegging av immunrespons ved utprøving av vaksiner og sykdomsstudier
- Å gi svar på hvilke gener som finnes i arten, (Vi har for eksempel ikke funnet Grelin i dette genomet, et protein som i andre arter er viktig i regulering av sult og appetitt)
- Å gi svar på hvor mange varianter det finnes av et gen (vi måler ofte funksjon om mengde av ett gen, og det kan gi et feil bilde av virkerligheten i de tilfellene der det finnes flere varianter av samme gen og disse er ulikt regulert)
- Å gi svar på hvor lik genenes oppbygging er andre arters gener (vi kan regne ut sannsynligheten for hvor godt molekylære verktøy fra andre arter kan virke på "vår" art).
- Identifisere gener og variasjon av disse som styrer prosesser som:
 - Vekst
 - Kjønnsmodning og bytte
 - Sykdomsresistens
 - Temperaturtilvenning (kulde tilpasning?)

Tilgangen til et vertsgenom vil være nyttig på flere områder enn det vi selv ønsker å bruke det til. Som blant annet ved karakterisering av immunrespons ved bruk av RNA sekvensering (noe som kan være et viktig verktøy ved uttesting av nye vaksiner). I tillegg vil sekvensering og metagenomikk bli viktige diagnostiske verktøy også innenfor fiskehelse i fremtiden. I den anledning vil kjennskap til et vertsgenomet gjøre det mulig å skille vertsgenomet fra patogengenomet.

Ivar Hordvik (UIB) vil bruke dataene inn i sitt NFR prosjekt «Basic immunology studies and development of tools to monitor immune responses in cleaner fish».

Prosjektgjennomføring

En heldig utvalgt leppefisk fra Marine Harvest sitt berggylteanlegg i Øygarden ble sekvensert med en dekningsgrad på 144 ganger (144 x dekning). Til genomsekvenseringen ble det benyttet en kombinasjon av to ulike sekvenseringsteknologier: Illumina (111 x dekning) og PacBio (33 x dekning), som henholdsvis gir korte og lange sekvenser. Sekvens bitene ble satt sammen med assemblyprogrammet Canu, utviklet fra “Celera Assembler” (1) og forbedret med assemblyforbedringsprogrammet “Pilon” (2). Assemblyet ble deretter validert med verktøyet “BUSCO” (3).

Oppnådde resultater

Sekvenseringsstrategien i kombinasjon med assemblyprogrammet *Canu* og assemblyforbedringsprogrammet *Pilon* resulterte i et godt sammensatt genom med store contigs (N50 704 kilobaser) og høy andel komplette gener funnet med valideringsverktøyet BUSCO (95 %). Det sammensatte berggyltegenomassemblyet er 805 megabaser (Mb), hvilket er i samme størrelsesorden som genomene til torsk, medaka og tilapia. Til sammenligning er laksens genom estimert til å være ca 3000 Mb.

CEES er i ferd med å sette opp en genom-browser med hjelp av ELIXIR Norge

(<https://www.mn.uio.no/cees/english/research/about/infrastructure/genome-browser/>).

En genom-browser er et web-basert verktøy for å gi forskere og andre interesserte tilgang til genomisk informasjon. Browseren gjør det mulig å undersøke, og søke i, selve DNA sekvensen sett i sammenheng med andre typer genomiske data, som for eksempel annotasjonen (lokasjoner til genene), og eksperimentelle data som er plassert ift til DNA sekvens, for eks. individ variasjoner (SNPs). Brukeren kan også zoome inn og ut på områder av interesse, søke etter spesifikke gener og laste ned relevante deler av genomdata som vises i browseren. Eksempler på andre kjente genom-browsere er ENSEMBL (ensembl.org) og UCSC Genome Browser Home (<https://genome.ucsc.edu/>).

CEES skal snart gjøre en ny versjon av torskegenomet tilgjengelig via sin genom-browser. Deretter står berggyltegenomet for tur til å bli gjort tilgjengelig via samme browseren. Inntil dette er på plass vil vi ved Nifes være behjelpelig med data fra genomet på forespørsel fra interesserte.

Referanser

1. Myers EW, Sutton GG, Delcher AL, Dew IM, Fasulo DP, Flanigan MJ, et al. A whole-genome assembly of *Drosophila*. *Science*. 2000 Mar 24;287(5461):2196–204.
2. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. Wang J, editor. *PloS one*. Public Library of Science; 2014;9(11):e112963.
3. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 2015 Sep 21;31(19):3210–2.