NORWEGIAN SEQUENCING CENTRE

Organization and synteny of hemoglobin gene cluster in codfishes: our experience with targeted sequence capture

Ave Tooming-Klunderud, NSC and CEES at UoOslo,
ave.tooming-klunderud@ibv.uio.no
www.sequencing.uio.no

Leiden, 02/05/2017

Oslo universitetssykehus
The Research Council of Norway
CEES Centre for Ecological and Evolutionary Synthesis
HELSE SØR-ØST
UNIVERSITAS OSLOENSIS MDCCCXI

Norwegian Sequencing Centre

- Two nodes:
  - CEES at Department of Biosciences, University of Oslo
  - Department of Medical Genetics, Oslo University Hospital
- Funding:
  - Host institutions
  - Research Council of Norway
    - NSC - INFRAstructure program 2009-2013
    - NSC Phase II – INFRAstructure 2014 – 2017
    - NorSeq (National consortium for sequencing and personalized medicine) – INFRAstructure 2015 – 2018

Research of Atlantic cod at NSC

LETTER

doi:10.1038/nature10342

The genome sequence of Atlantic cod reveals a unique immune system

Bastiaan Star[1], Alexander J. Nederbragt[1], Sissel Jentoft[1], Unni Grim

www.nature.com/scientificreports

SCIENTIFIC REPORTS

OPEN

Genomic characterization of the Atlantic cod sex-locus

OPEN ACCESS Freely available online

Unraveling the Evolution of the Atlant

Tørresen *et al. BMC Genomics* (2017) 18:95
DOI 10.1186/s12864-016-3448-x

RESEARCH ARTICLE

An improved genome assembly ur prolific tandem repeats in Atlantic

Ole K. Tørresen[1*], Bastiaan Star[1], Sissel Jentoft[1,2], William B. Reinar[1], Harald Grov
Brian P. Walenz[5], James Knight[6], Jenny M. Ekholm[7], Paul Peluso[7], Rolf B. Edvard
Ave Tooming-Klunderud[1], Morten Skage[1], Sigbjørn Lien[3], Kjetill S. Jakobsen[1]
and Alexander J. Nederbragt[1,9*]

ow Salinity Promotes Genomic Divergence

www.nature.com/scientificreport

SCIENTIFIC REPORTS

www.nature.com/scientificdata

SCIENTIFIC DATA

OPEN

Data Descriptor: Whole genome sequencing data and *de novo* draft assemblies for 66 teleost species

Martin Malmstrøm[1], Michael Matschiner[1], Ole K. Tørresen[1], Kjetill S. Jakobsen[1] & Sissel Jentoft[1,2]

Kjetill S Jakobsen[1] & Sissel Jentoft[1,6]

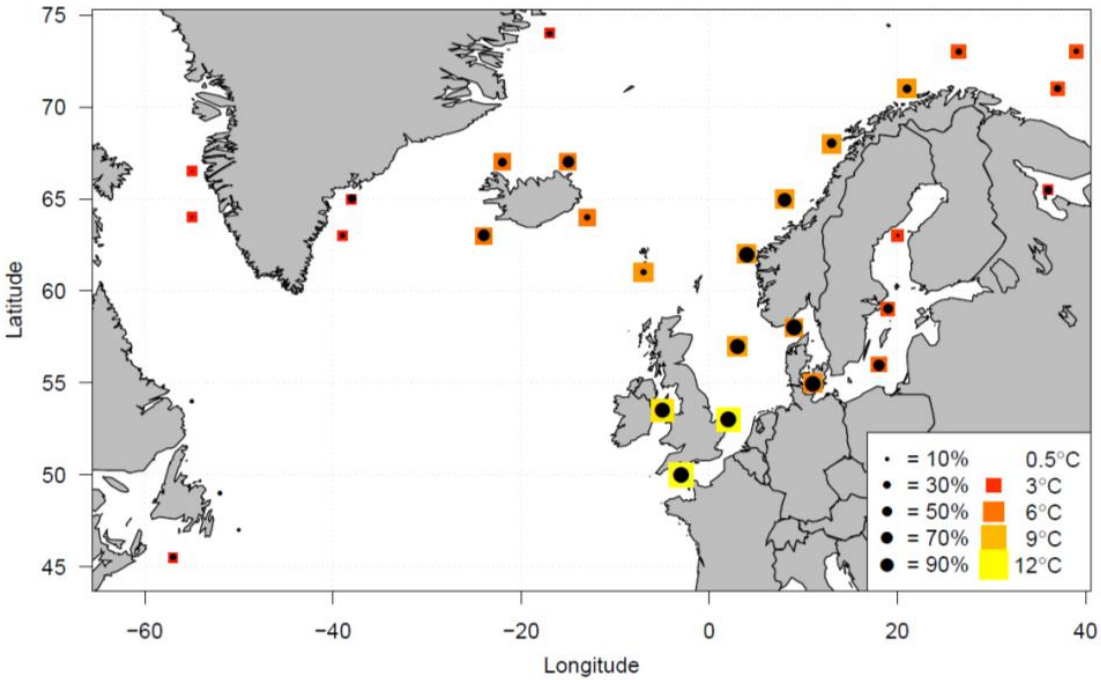
EVOLUTION OF HEMOGLOBIN GENES IN CODFISHES

Adaption to different environments

- Hemoglobins ability to bind O_2 is influenced by environmental factors
  - Temperature
  - Hydrostatic pressure

→ Molecular adaptation

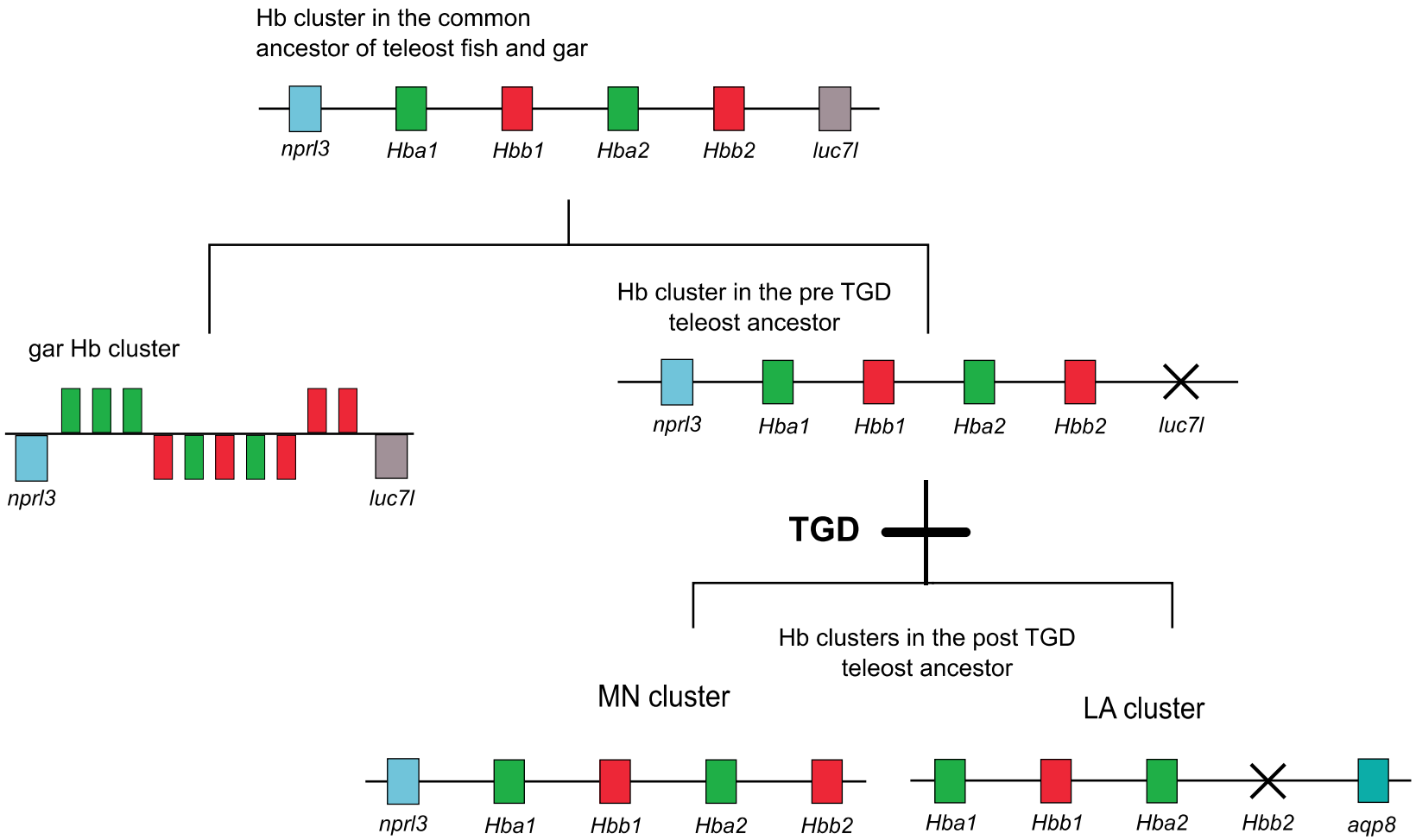
- Temperature adaptation of hemoglobin in Atlantic Cod - polymorphism in Hb gene β1



<u>β1 locus</u>
β1-1(Met55Lys62)
β1-2 (Val55Ala62)

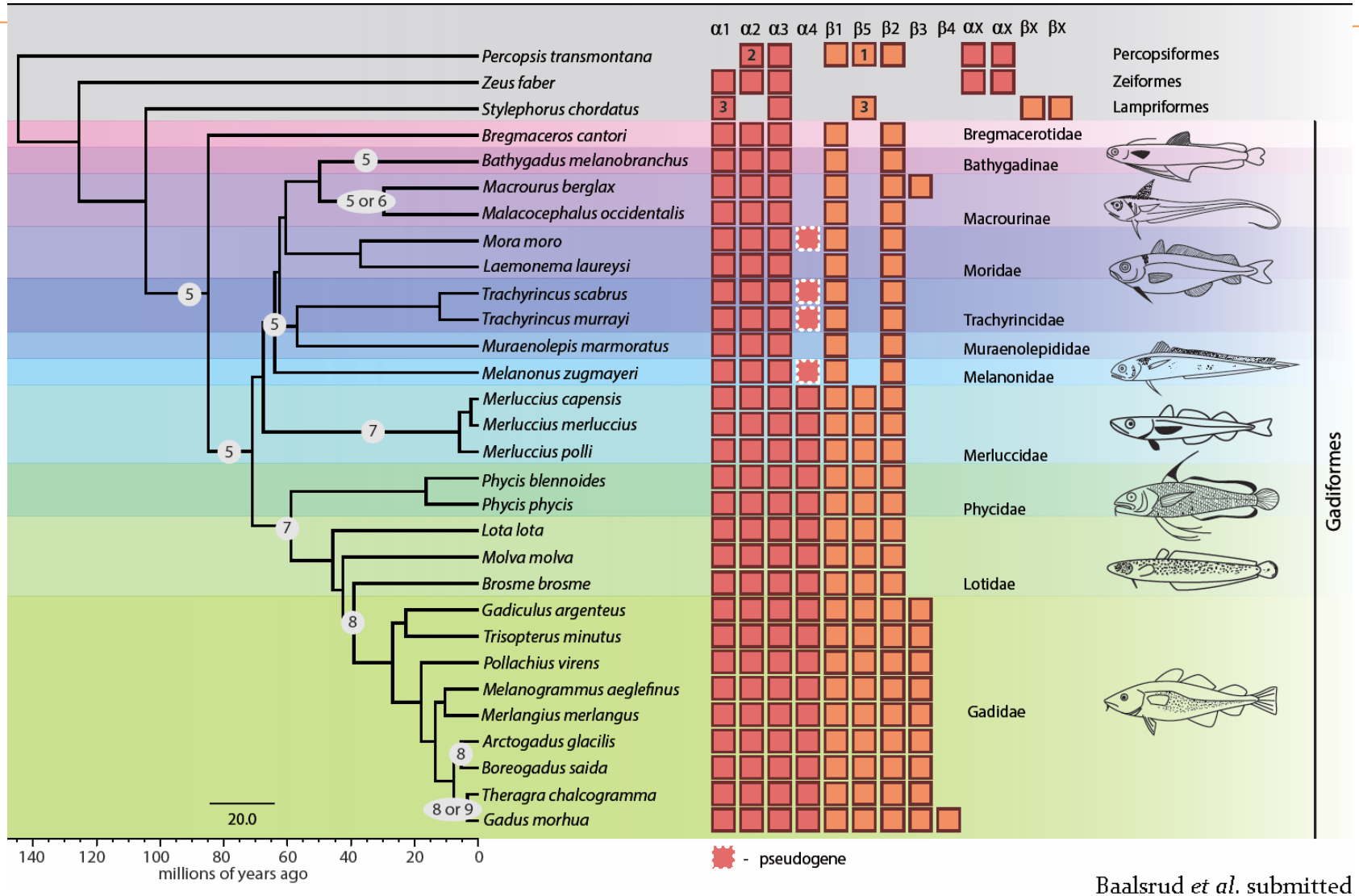
Hemoglobin β1-1 frequencies of cod in the North Atlantic.
Ross *et al*. 2013

Hemoglobin genes in fish



Opazo et al. 2012

Hemoglobin genes in codfishes



Baalsrud *et al.* submitted

What about synteny?

Two alternatives:

- Generate high quality reference genomes
- Sequence a subset of the genome by targeted sequence capture

PACIFIC BIOSCIENCES®

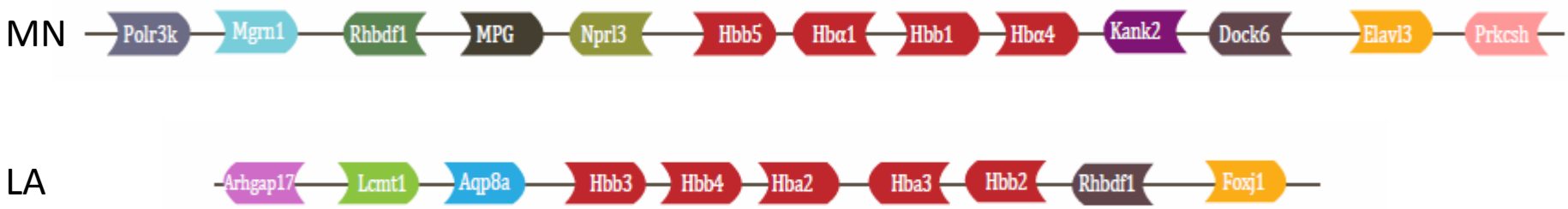
PacBio SampleNet – Shared Protocol

Please note: the shared protocols described herein may not have been validated by Pacific Biosciences and are provided as-is and without any warranty. Use of these protocols is offered to those customers who understand and accept the associated terms and conditions and wish to take advantage of their potential to help prepare samples for analysis using the PacBio® system. If any of these protocols are to be used in a production environment, it is the responsibility of the end user to perform the required validation.

Target Sequence Capture Using Roche NimbleGen SeqCap EZ Library

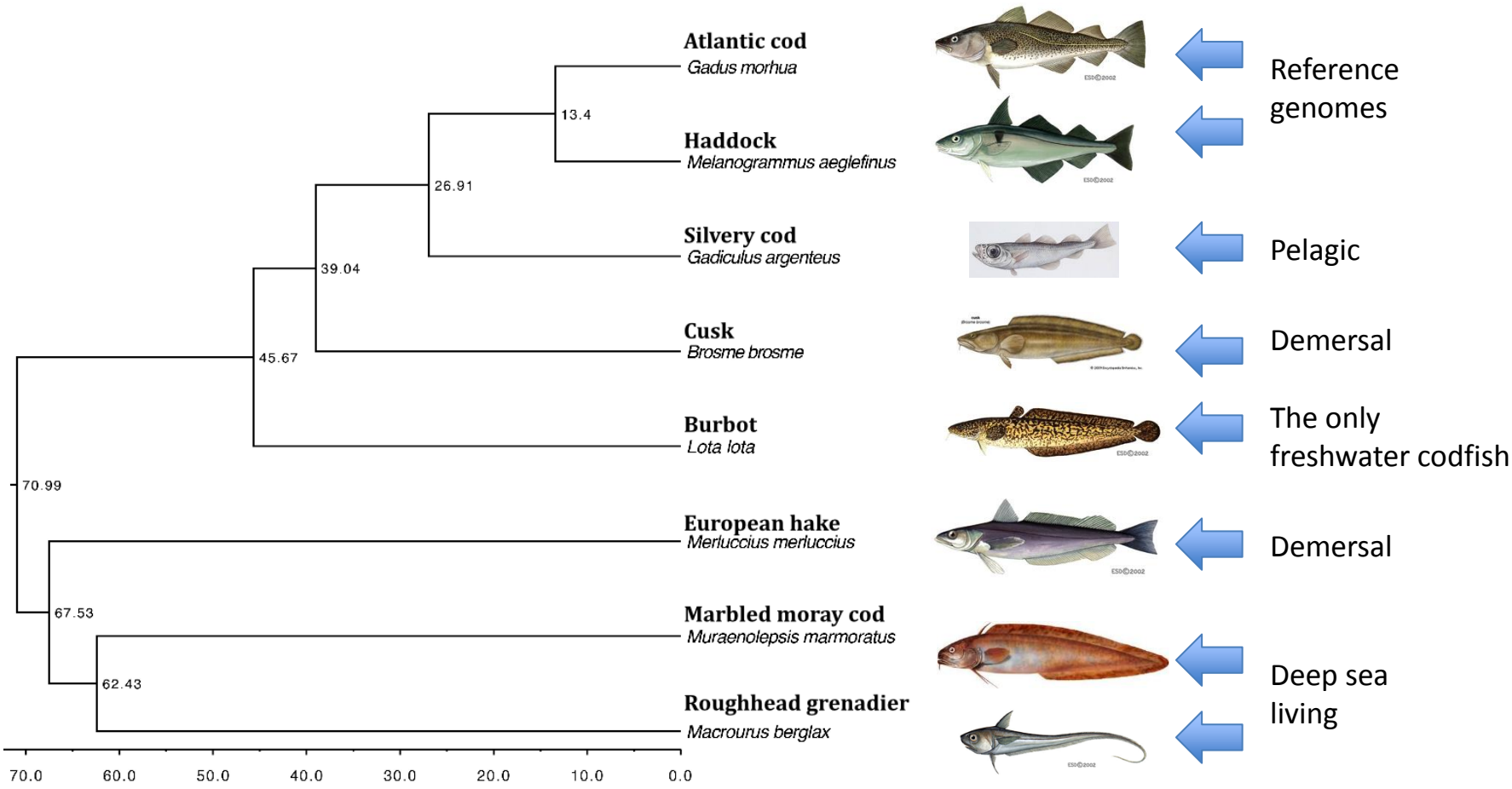
Custom probe design – Roche Nimblegen

- Sequences used for the probe design
  - The target region (MN and LA clusters) in Atlantic cod reference genome



  - Sequences of Hemoglobin and flanking genes
  - Partially assembled genomes for additional species – only regions orthologous to the target region of Atlantic cod

Species included in the study



Phylogenetic relationship of the eight species of codfishes included in the study. Time since divergence given at each node in millions of years. Subset from Malmstrøm *et al.* 2016

Pre capture multiplexing

- 16 bp PacBio barcode

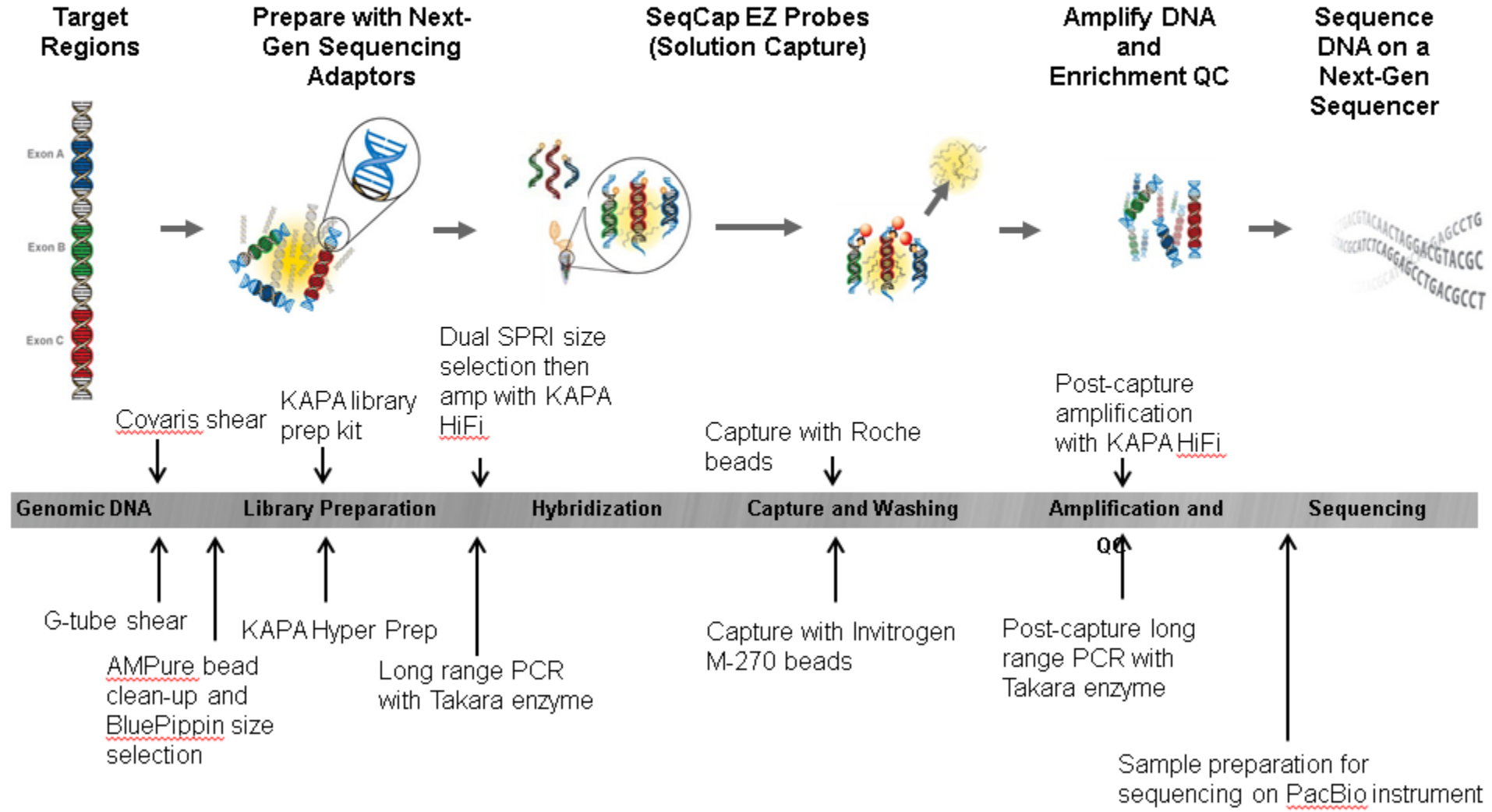
```
SeqCap Oligo1:                                AATGATACGGCGACCACCGAGA
PreCap-2 fwd:      GTCAGACGATGCGTCATAATGATACGGCGACCACCGAGA……
Block-2 fwd:       GTCAGACGATGCGTCAT   ------HE blocking from kit----------
PostCap-2 fwd:     GTCAGACGATGCGTCATAATGA

PreCap-2 rev:      CGCGATCTATGCACACGCAAGCAGAAGACGGCATACGAG…..
SeqCap  Oligo2 :                              CAAGCAGAAGACGGCATACGAG
Block-2 rev:       CGCGATCTATGCACACG  ------HE blocking from kit----------
PostCap-2 rev:     CGCGATCTATGCACACGCAAGC
```

- 6 bp Illumina barcode

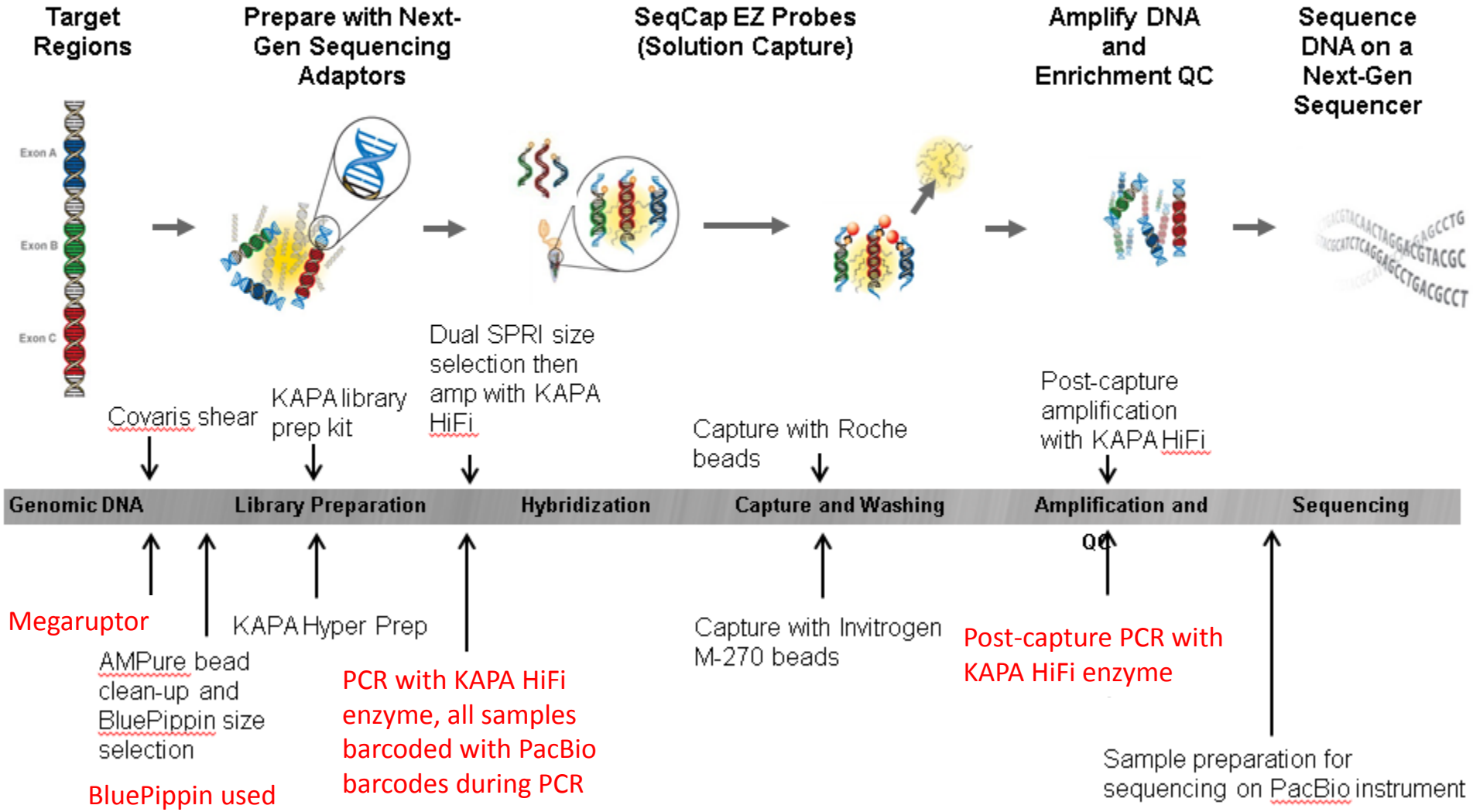
SeqCap EZ Enrichment System

Workflow for long fragments



SeqCap EZ Enrichment System

Workflow for long fragments

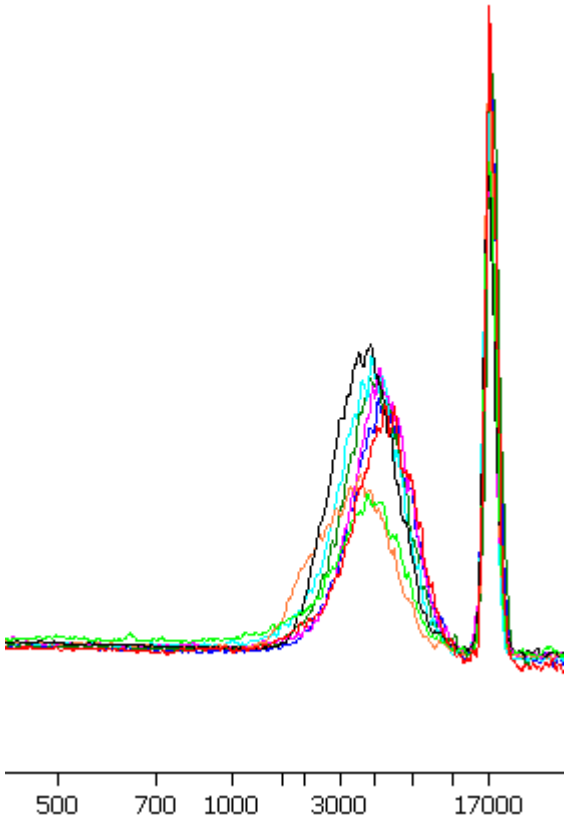


Modifications of the protocol

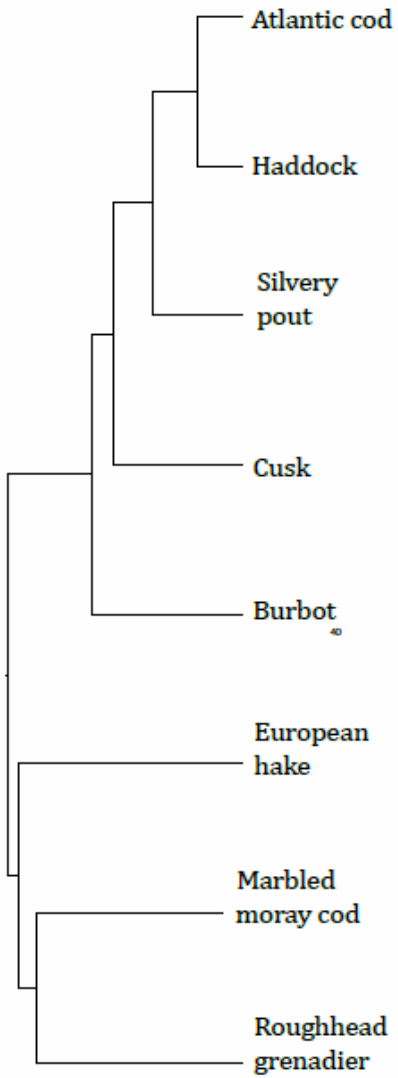
Library prep and capture

- 4 samples pooled before capture
- Individual captures for 6 of the samples, similarly sized post capture PCR products pooled into two final libraries

- All 8 samples pooled before capture, 250 ng of each



Sequencing results



	Pool of 8 samples	Total amount of data produced		
	% of reads	#reads	#bases	av. length
Atlantic cod	12.6%	84 304	258 Mb	3.06 kb
Haddock	5.4%	42 530	132 Mb	3.1 kb
Silvery pout	17.1%	86 067	268 Mb	3.1 kb
Cusk	12.2%	66 450	214 Mb	3.2 kb
Burbot	7.1%	64 472	195 Mb	3.01 kb
European hake	21.5%	76 604	215 Mb	2.8 kb
Marbled moray cod	9%	64 202	186 Mb	2.9 kb
Roughhead grenadier	15.2%	54 391	155 Mb	2.85 kb

«Jumping» PacBio barcodes

Due to our barcode design:

SeqCap Oligo1:	AATGATACGGCGACCACCGAGA
PreCap-2 fwd:	GTCAGACGATGCGTCATAATGATACGGCGACCACCGAGA……
Block-2 fwd:	GTCAGACGATGCGTCAT ------HE blocking from kit----------
PostCap-2 fwd:	GTCAGACGATGCGTCATAATGA
PreCap-2 rev:	CGCGATCTATGCACACGCAAGCAGAAGACGGCATACGAG…..
SeqCap Oligo2 :	CAAGCAGAAGACGGCATACGAG
Block-2 rev:	CGCGATCTATGCACACG ------HE blocking from kit----------
PostCap-2 rev:	CGCGATCTATGCACACGCAAGC

some degree of barcode «jumping» was expected

«Jumping» PacBio barcodes

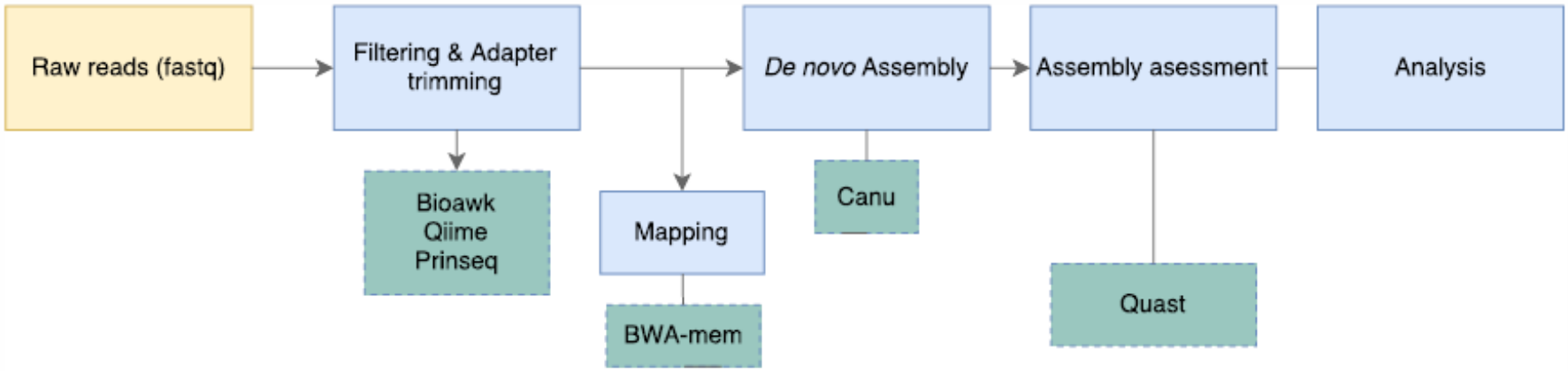
1486	CGCAAGCAGAAGACGGCATACGAGATACATCGGTGACTGGAGTTCAGACG	
84	GCAAGCAGAAGACGGCATACGAGATACATCGGTGACTGGAGTTCAGACGT	
57	ACGCAAGCAGAAGACGGCATACGAGATACATCGGTGACTGGAGTTCAGAC	
51	CGCAAGCAGAAGACGGCATACGAGATACGGAACTGTGACTGGAGTTCAGA	A14
29	CGCAAGCAGAAGACGGCATACGAGATACATCGTGACTGGAGTTCAGACGT	
28	TGCAAGCAGAAGACGGCATACGAGATTGGTCAGTGACTGGAGTTCAGACG	A4
28	TCCAAGCAGAAGACGGCATACGAGATTACAAGGTGACTGGAGTTCAGACG	A12
23	ACCAAGCAGAAGACGGCATACGAGATCGTTTCACGTGACTGGAGTTCAGA	A19
22	GCCAAGCAGAAGACGGCATACGAGATGATCTGGTGACTGGAGTTCAGACG	A7
16	TCCAAGCAGAAGACGGCATACGAGATCGGGACGGGTGACTGGAGTTCAGA	A16
12	CGCAAGCAAGCACAGATGTAGACATGGACGCACACACACCAGCAAACCCA	A16
11	CTCAAGCAGAAGACGGCATACGAGATCACTGTGTGACTGGAGTTCAGACG	A5

9.8% reads with wrong adapter

% of reads with wrong adapter for other samples:

0.4% 0.8% 1.5% 2.4% 2.5% 3.4% 5.0%

Data analysis workflow

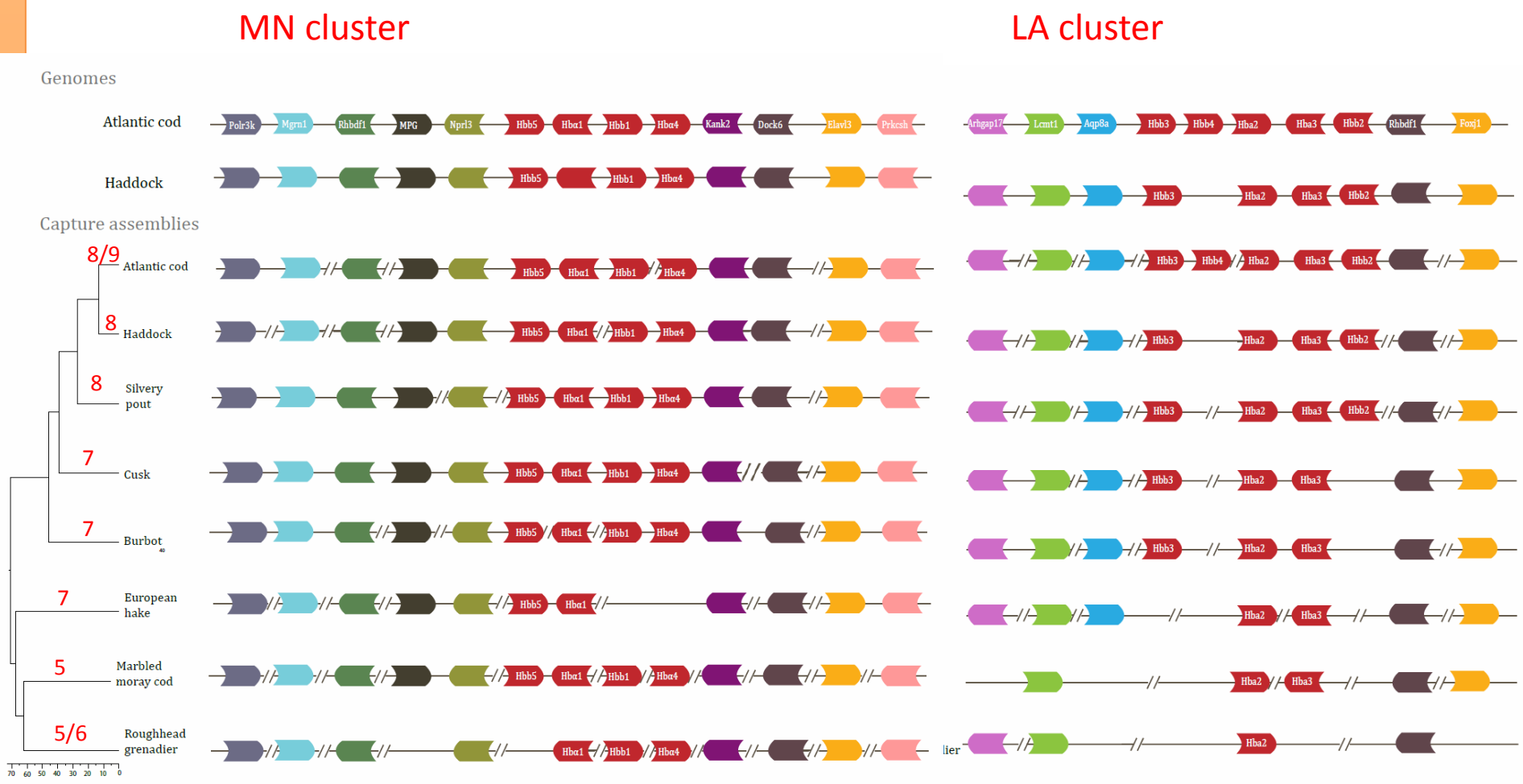


Rawreads = reads filtered and demultiplexed using RS_Reads of Insert.1 pipeline

De novo assembly

	Total length (>= 1000 bp)	Total number unitigs	Largest unitigs (bp)	GC (%)	N50	N75
Atlantic Cod	1757471	215	41343	44.91	8575	5750
Haddock	971706	116	44949	45.20	8776	6169
Cusk	2232327	248	117090	42.62	9260	6405
Silvery cod	1713941	210	41689	45.41	8649	5879
Burbot	943952	103	58914	44.23	10360	6386
European Hake	1410818	197	36393	46.01	7037	5400
Roughhead grenadier	1407647	182	42357	48.44	8192	5655
Marbled moray Cod	1552724	206	38697	45.02	7963	5539

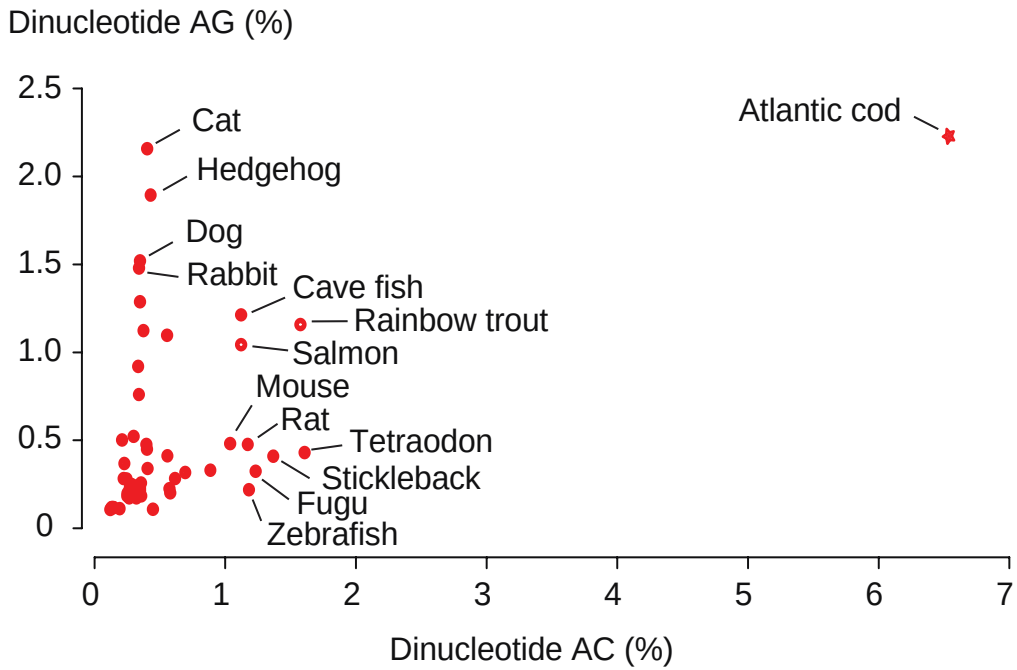
Synteny of the hemoglobin clusters



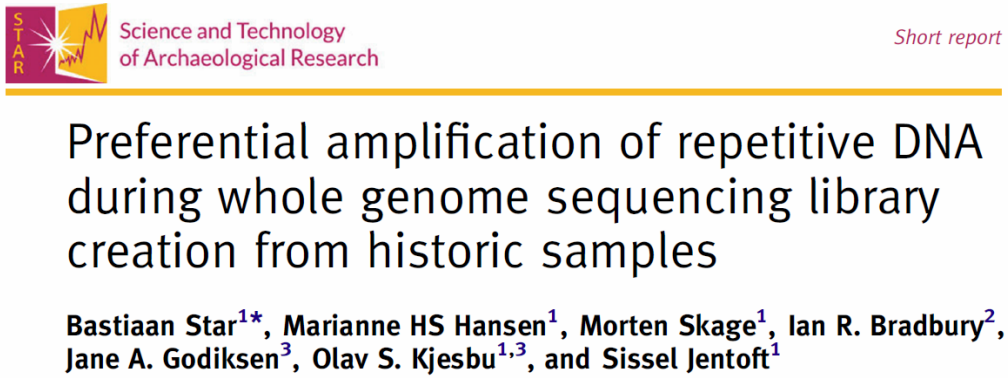
Number of genes detected from draft assemblies indicated on tree

Evaluation of capture - Atlantic cod

- Probes do not cover entire sequence due to repeats



Star et al. 2016

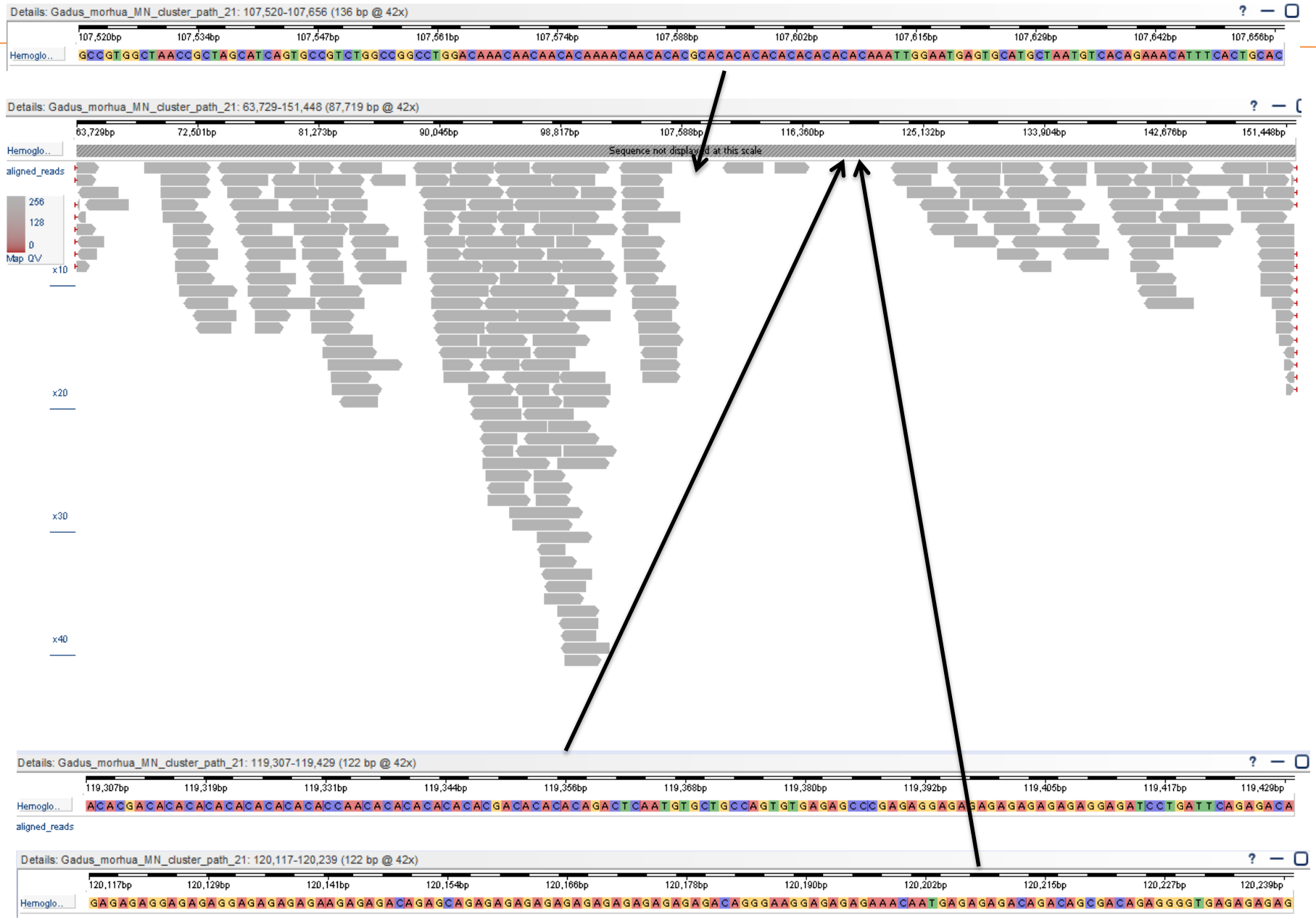

Science and Technology of Archaeological Research

Short report

Preferential amplification of repetitive DNA during whole genome sequencing library creation from historic samples

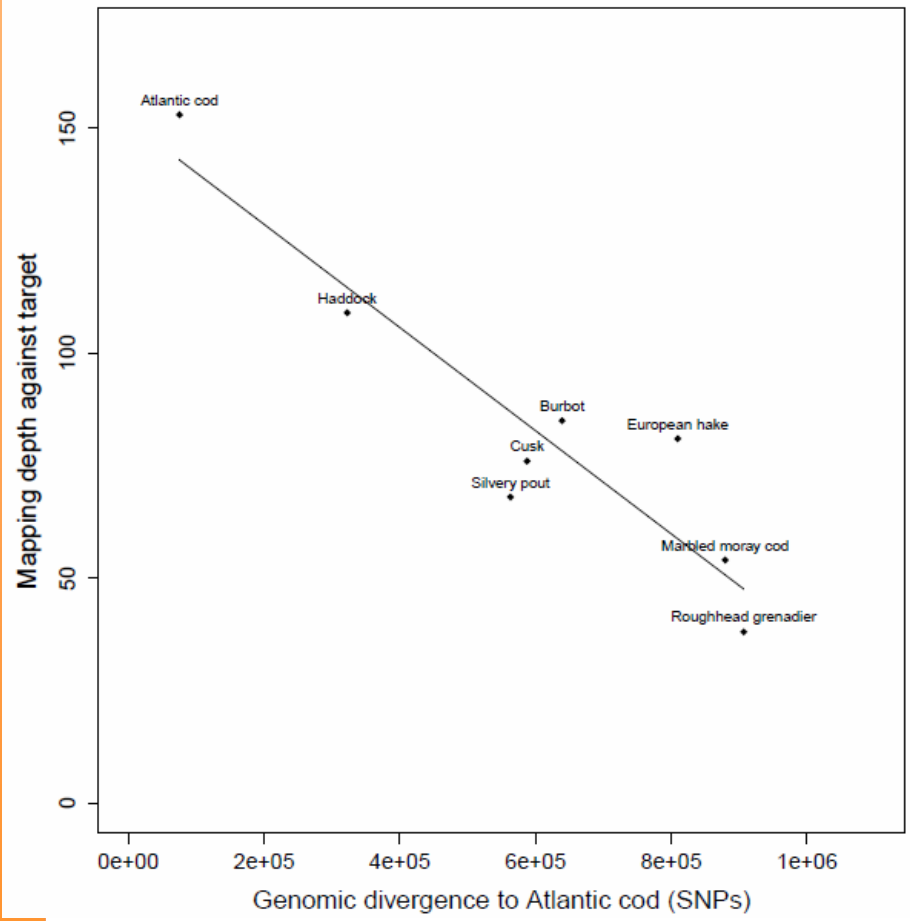
Bastiaan Star[1]*, Marianne HS Hansen[1], Morten Skage[1], Ian R. Bradbury[2], Jane A. Godiksen[3], Olav S. Kjesbu[1,3], and Sissel Jentoft[1]

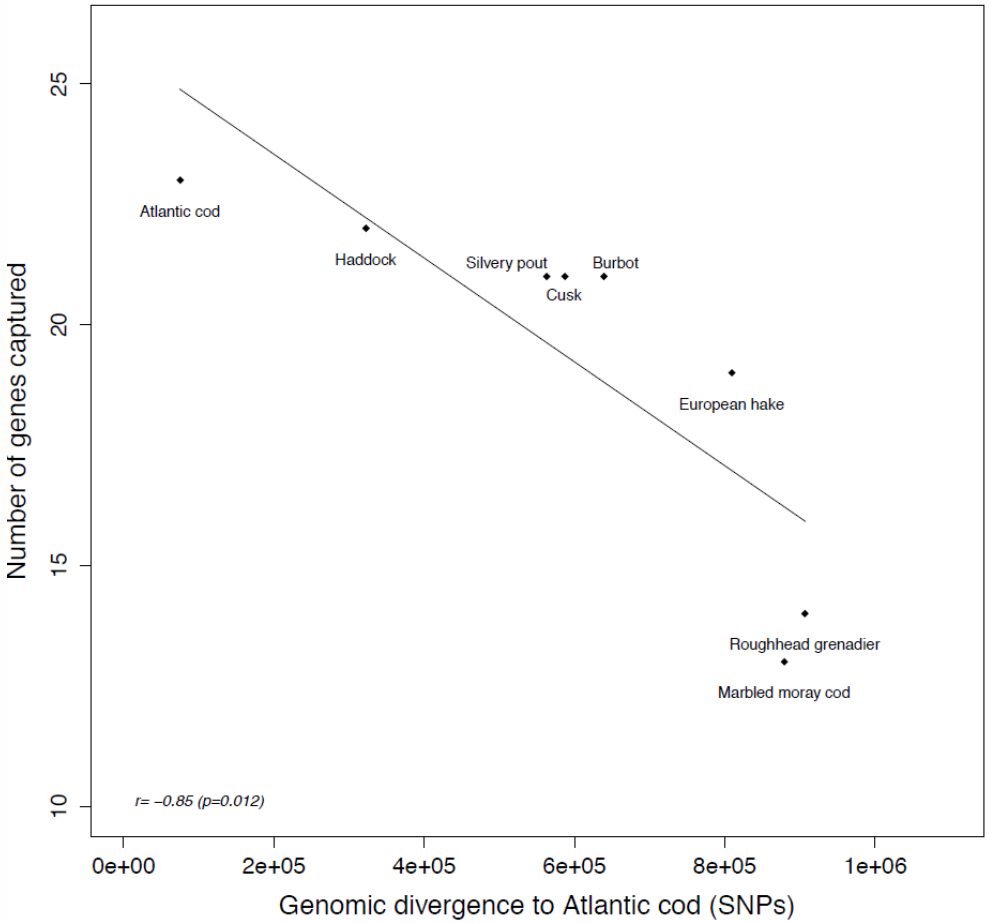
Repeat content is 20.29% in LA and 10.7% in MN cluster

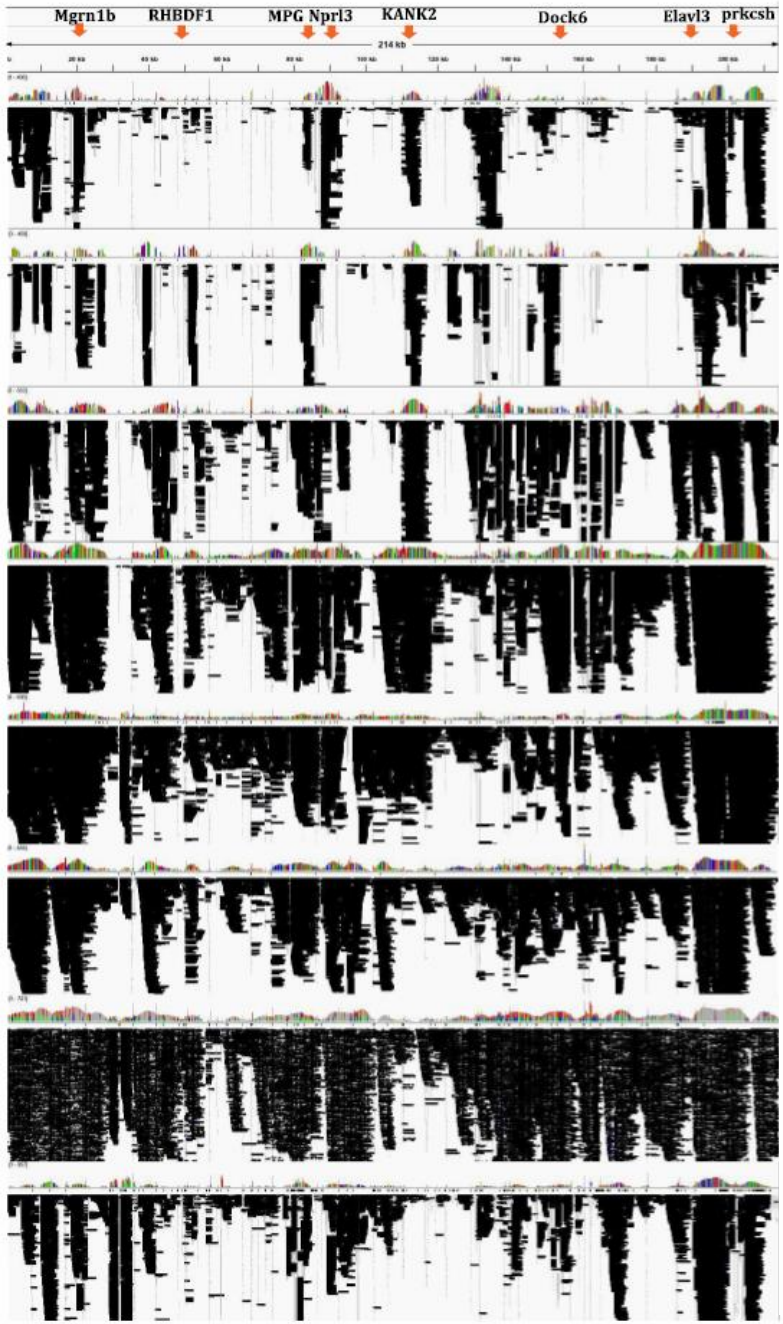


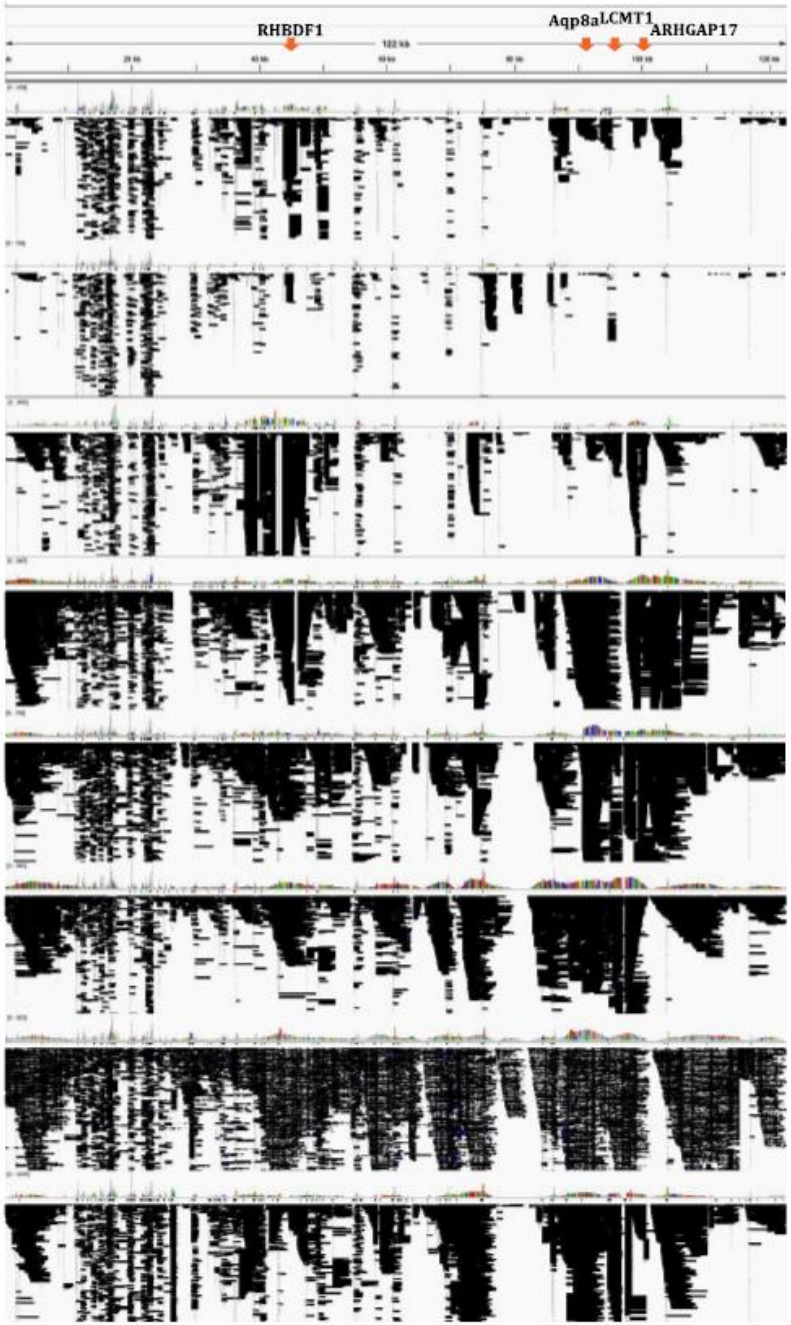
Evaluation of capture – rest of the fishes

- Capture efficiency decreases with increased evolutionary distance from Atlantic cod









Marbled moray cod

Roughhead grenadier

European hake

Burbot

Cusk

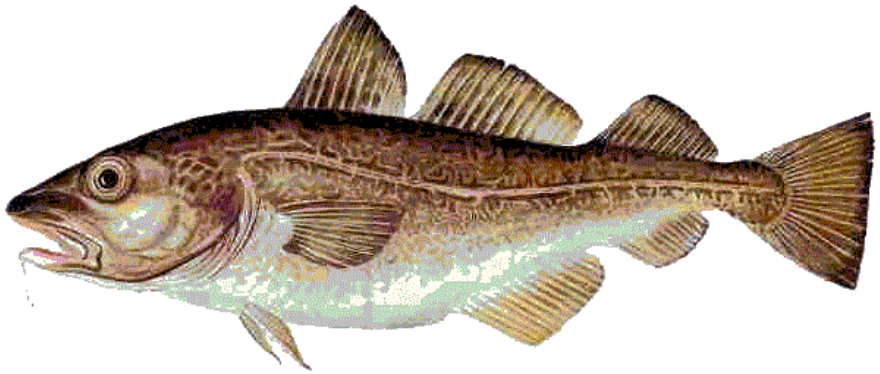
Haddock

Atlantic cod

Silvery cod

MN cluster

LA cluster

Acknowledgments

- Cod-group
  - Siv Nam Khang Hoff
  - Helle Baalsrud
  - William B.Reinar
  - Sissel Jentoft
  - Kjetill S. Jakobsen
  - Ole Kristian Tørressen
  - Martin Malmstrøm
- NSC
  - Morten Skage
  - Marianne H.S. Hansen
  - Lex Nederbragt
- Roche/Nimblegen
  - Reza Shirzadi
  - Gregor Obernosterer
  - Todd Richmond