FINAL REPORT (DEVELOPMENT AWARD)

AWARD CODE and TITLE

20088/328.16 2008/328.16 People development program: 2012 FRDC Visiting Expert Bursaries - Professor Sigbjorn Lien

AWARD RECIPIENT: Dr Sigbjørn Lien

ADDRESS: CIGENE, Norwegian University of Life Sciences, PO Box 5003, 1432 Aas, Norway

HOST ORGANISATION: CSIRO Animal, Food and Health Sciences

DATE: 19/08/2013

ACTIVITY UNDERTAKEN

- Visited the Queensland Bio-Precinct at St Lucia and the EcoSciences-Precinct at Dutton Park, Brisbane, QLD: met with staff from CMAR and CSIRO, gave one seminars on his work and participated in Galaxy workshop
- Visited the University of the Sunshine Coast: met with staff and students from the Aquaculture group, gave seminar on his work
- Visited Armidale, NSW: met with staff and gave seminar at CSIRO Animal, Food and Health Science, the Animal Genetic and Breeding Unit and the Department of Animal Science at UNE, participated in Galaxy workshop at CSIRO
- Visited CSIRO Marine and Atmospheric Research at Hobart, TAS: gave seminar at CMAR on his work to staff members and SALTAS representatives, met with industry and CSIRO staff

OUTCOMES ACHIEVED TO DATE

Main outcomes of the visit were to exchange knowledge and experience within genomics and bioinformatics and discuss practical implementation of genomic information to advance aquaculture industry.

Acknowledgments

Dr. Sonja Dominik, CSIRO Animal, Food and Health Sciences, FD McMaster Laboratory, Armidale, is greatly acknowledged both for organising the visit and for being such a great host during the trip.

Background

Traditionally, the genetic improvement of economically important traits within agri- and aquaculture has been accomplished without identifying the underlying genes involved. Instead it has been based on estimated breeding values calculated from phenotypic records and pedigrees. Although this has led to tremendous improvements in the efficiency of food production over the last 50 years, the process is inefficient for traits that can be measured only in one sex, after harvest, late in life, and can be expensive to measure (e.g. feed

efficiency or disease resistance). Furthermore, some species have long generation times and life histories, which greatly limit the selection process. As a consequence, genomic approaches that forecast long term changes, and where possible allow faster and more precise genetic selection, are highly desired by breeders.

Today the field of genomics is offering new ways to enhance bioproduction and address research questions for a number of production species. The critical ingredients available for these species are (i) a reference genome, (ii) larger collections of molecular markers (iii) high quality phenotypic records and biobank material, and (iv) bioinformatics and statistical tools for analyzing and implementing data. These reference genomes, together with genomic data derived by re-sequencing individuals in breeding populations using next-generation sequencing (NGS) technologies allows an unprecedented large-scale discovery of genetic markers (Single Nucleotide Polymorphisms – SNPs) for breeding purposes. Correlating SNPs, haplotype information and sequence data with high resolution phenotype data enables industry to implement marker assisted selection (MAS) or genomic selection (GS) in their breeding programs, while researchers can reveal the biological systems and genes underlying disease resistance and trait development. GS in particular has the potential to revolutionize the structure of breeding schemes and reduce selection costs and generation interval. Two important features of this approach are that it; (1) increases the genetic gain (potentially up to 80%) because of a reduced generation interval, and (2) utilizes phenotype records from a reference population rather than the breeding candidates which makes it easier to include multiple new traits.

The Centre of Integrative Genetics, CIGENE, (www.cigene.no) was established at the Norwegian University of Life Sciences (UMB) in 2003 as a national core facility in the FUGE program responsible for detection, genotyping and interpretation of SNPs. Since then, CIGENE has become Norway's foremost SNP genotyping facility, and has built a strong network and gained experience in handling sequencing data, performing SNP discovery, developing and validating large-scale SNP arrays, performing data quality control, and finally analysis of genetic variation data on a wide range of species. Significant bioinformatics and computational resources to deal with the data have been established. CIGENE has been involved in large-scale sequencing projects since 2007. Various bioinformatics activities are ongoing with a focus on *de novo* assembly, among these International Collaboration to Sequence the Atlantic Salmon Genome (ICSASG) and SNP discovery. From these efforts genotyping arrays have been made available for Atlantic salmon, Atlantic cod, sea louse, brown trout and additional products are in development.

CIGENE has created close collaboration with key breeding industries in Norway (both agri- and aquaculture) with the aim of implementing genomics with agri- and aquaculture. Key elements in this collaboration are development of genomics, bioinformatics and computational resources to secure development, transfer and application of genome-based selection in the breeding industry. One important component of the bioinformatics pipeline is the implementation of the online software resource, Galaxy (http://main.g2.bx.psu.edu/). Galaxy enables scientists with less programming skills to handle data of large volume efficiently, extract subsets and analyse data. Implementation of local version of Galaxy has made it possible to provide access to in-house developed software tools through a standard web-browser. By arrangement, collaborators and users will be given access to their data using secure ftp or via a password protected web server.

Sound industry breeding programs have been developed for Australian aquaculture (like Atlantic salmon and black-tiger prawns). Genetic selection in the programs is mainly based on traditional estimated breeding values calculated from phenotypic records and pedigrees. The next potential step is to integrate molecular information in breeding programs, in particular for difficult to measure traits (e.g. health or product quality) or traits of low heritability (e.g. reproduction). To achieve there is an urgent need for developing innovative, cost effective and less computationally intensive approaches to translate genomics data into highly efficient breeding schemes.

Main objectives of the visit were to;

1) review and assess the needs in SNP data handling, processing and analysis in Atlantic salmon and black tiger prawn breeding programs
2) transfer knowledge to researchers on the development a bioinformatics pipeline in Galaxy for SNP data in Atlantic salmon and black tiger prawn
3) provide industry with an overview of the use of molecular technologies in aquaculture breeding programs in Europe

Table 1. Meetings were arranged with the following people:

Person	**Organisation**	**Location**
Nigel Preston	CMAR	Brisbane
Melony Sellars	CMAR	Brisbane
Mat Cook	CMAR	Brisbane
James Kijas	CAFHS	Brisbane
Ross Tellam	CAFHS	Brisbane
Bill Barendse	CAFHS	Brisbane
Wayne Knipp	USC	Sunshine Coast
Ngyuen Hong Ngyuen	USC	Sunshine Coast
Scott Cummins	USC	Sunshine Coast
Technical staff	USC	Sunshine Coast
Postgraduate students	USC	Sunshine Coast
Peter Hunt	CAFHS	Armidale
Ian Purvis	CAFHS	Armidale
Julius van der Werf	UNE	Armidale
Brian Kinghorn	UNE	Armidale
Cecile Massault	UNE	Armidale
Heather Burrow	CAFHS	Armidale
Kim Bunter	AGBU	Armidale
Nick Elliott	CMAR	Hobart
Peter Kube	CMAR	Hobart
Richard Taylor	CMAR	Hobart
David Mitchell	Huon Aquaculture Company	Hobart
Yvonne Sheehan	SALTAS	Hobart
Linda Sams	TASSAL	Hobart

Methods
Dr Lien travelled to Brisbane, the Sunshine Coast, Armidale and Hobart during his three week visit. He talked to staff, students and industry representatives at

- CSIRO Animal, Health and Food Sciences (CAFHS) at Queensland Bio-Precinct (QBP)
- CSIRO Marine and Atmospheric Research (CMAR) at the Queensland Bio-Precinct (QBP)
- EcoSciences Precinct at Dutton Park
- University of the Sunshine Coast
- CAFHS in Armidale
- Animal Genetics and Breeding Unit in Armidale
- University of New England in Armidale
- CMAR in Hobart

Dr Lien gave five seminars at various locations to staff, students and industry representatives (Table 1). The seminar covered the developments in the Salmon Genome project and developments in molecular technologies by the Center of Integrative Genetics (CIGENE) in Norway.

Results/Discussion
Today traditional breeding is increasingly being supplemented with, or even replaced, by genome based approaches. Critical ingredients for such an implementation are; (i) a reference genome for the species of interest, (ii) larger collections of molecular markers (iii) high quality phenotypic records and biobank material, and (iv) bioinformatics and statistical tools for analyzing and implementing data. Species are differently advanced with respect to the 'genomics toolbox' available which also affects strategies for practical implementation into breeding schemes. Despite this there is a strong overlap in the data processing and analysis needs for the different species including bioinformatics pipelines addressing genotype quality issues, pedigree checking procedures and data analysis, as well as computational infrastructure dealing with the ever-increasing amount of data.

The FRDC support made it possible for Dr. Lien to visit advances teams of CSIRO researchers in Brisbane, Armidale and Hobart, consisting of geneticists, biologists and staff with industry expertise. Seminars and meetings were organised with researchers and industry (IMB at UQ, CSIRO, DEEDI, UNE, AGBU, SALTAS, TASSAL, Huon Aquaculture Company) to exchange ideas and discuss ongoing and prospecting research within genomics and breeding. A number of opportunities for interaction and future collaborations were pointed out.

In the short term, direct collaboration it seems to be most realistic for Atlantic salmon. CIGENE and collaborators have recently developed a new high density SNP panel (~220K) for Atlantic salmon. There is great interest at CSIRO and its industry partner, SALTAS, for using use this SNP-chip in genome based selection against amoebic gill disease (AGD), and a scoping study on the applicability of the chip in the Tasmanian Atlantic salmon population has recently been funded.

Benefits and Adoption

Seminars were arranged with other research groups and industry to learn about the prospects of genome based breeding in Norwegian industries (in Brisbane, Armidale and Hobart). Strong foundations for similar developments exist for Australian aquaculture and strategies for knowledge transfer and tool developments were pointed out and discussed (in particular for prawns and salmon). A particular focus was put on cost-effective and integrated genomics tools and streamlined bioinformatics pipelines. Such tools are rather generic in nature but highly demanded both in research and industry related application.

Further Development

Atlantic salmon and black-tiger prawns are at different stages of 'genomics toolbox' developments. Also the application of genomics information in breeding programs of these two species is quite different (genomic selection in salmon and pedigree assignment and improvement of management in prawns). Still there is a strong overlap in the data processing and analysis needs with respect to quality control, pedigree checking procedures and links to external software for data analysis and pedigree assignment. Thus both industries, as well as other aquaculture industries in the future, would highly benefit by the development of efficient bioinformatics pipelines facilitating implementation of genome information in breeding programs.

Appendices

The Powerpoint presentation of seminars send as a separate document

Atlantic salmon genomics to advance aquaculture

Sigbjørn Lien
Centre for Integrative Genetics (CIGENE)
Department of Animal and Aquacultural Sciences
Norwegian University of Life Sciences
Ås, Norway

CIGENE
Centre for Integrative Genetics

NORWEGIAN UNIVERSITY OF LIFE SCIENCES
MDCCCLIX

CIGENE
Centre for Integrative Genetics

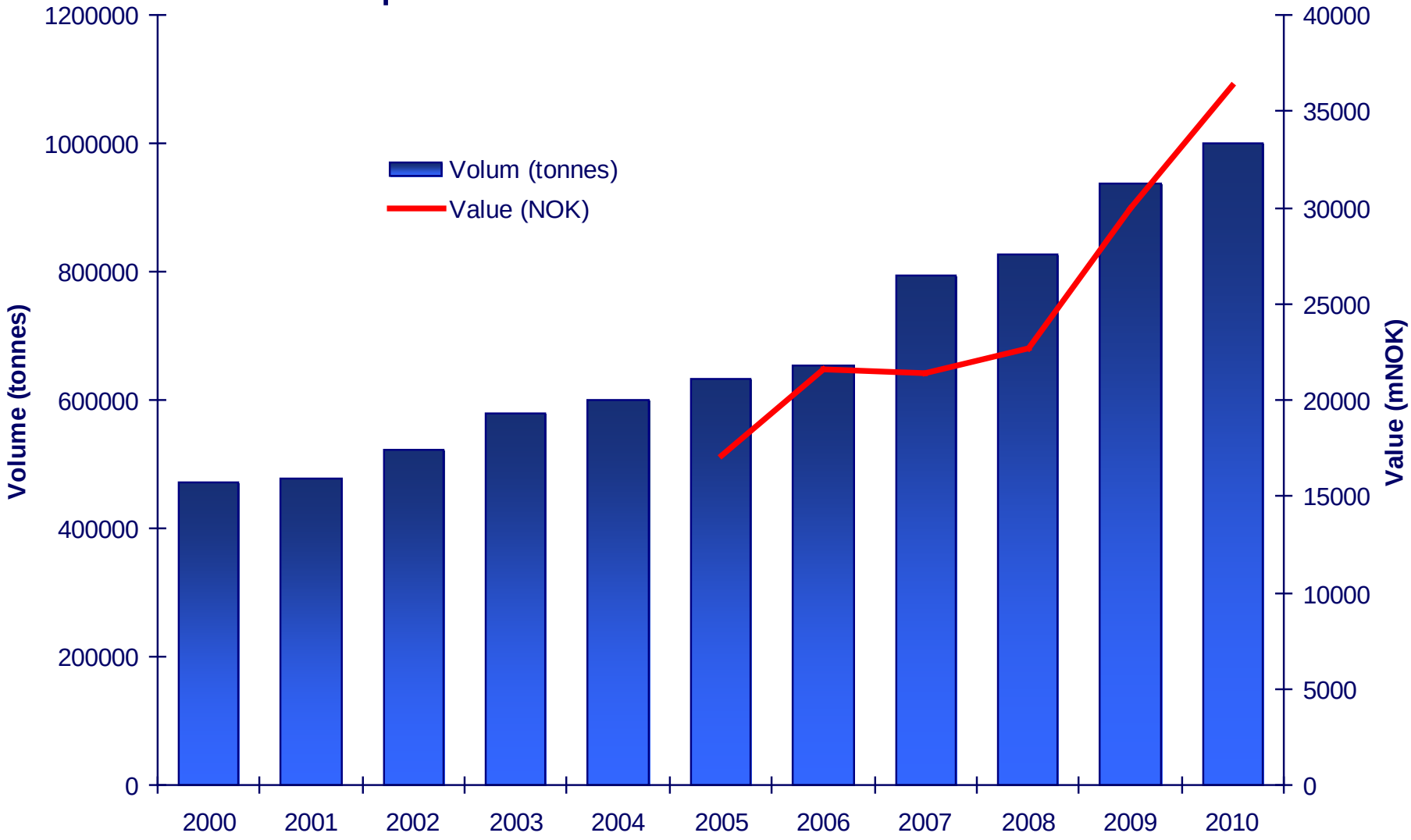
- Established in 2003 and funded by the Norwegian Government to serve as a national SNP genotyping facility
- Research focus on production biology species
- Close collaborations with livestock and aquaculture industries

Aqua Gen at a glance

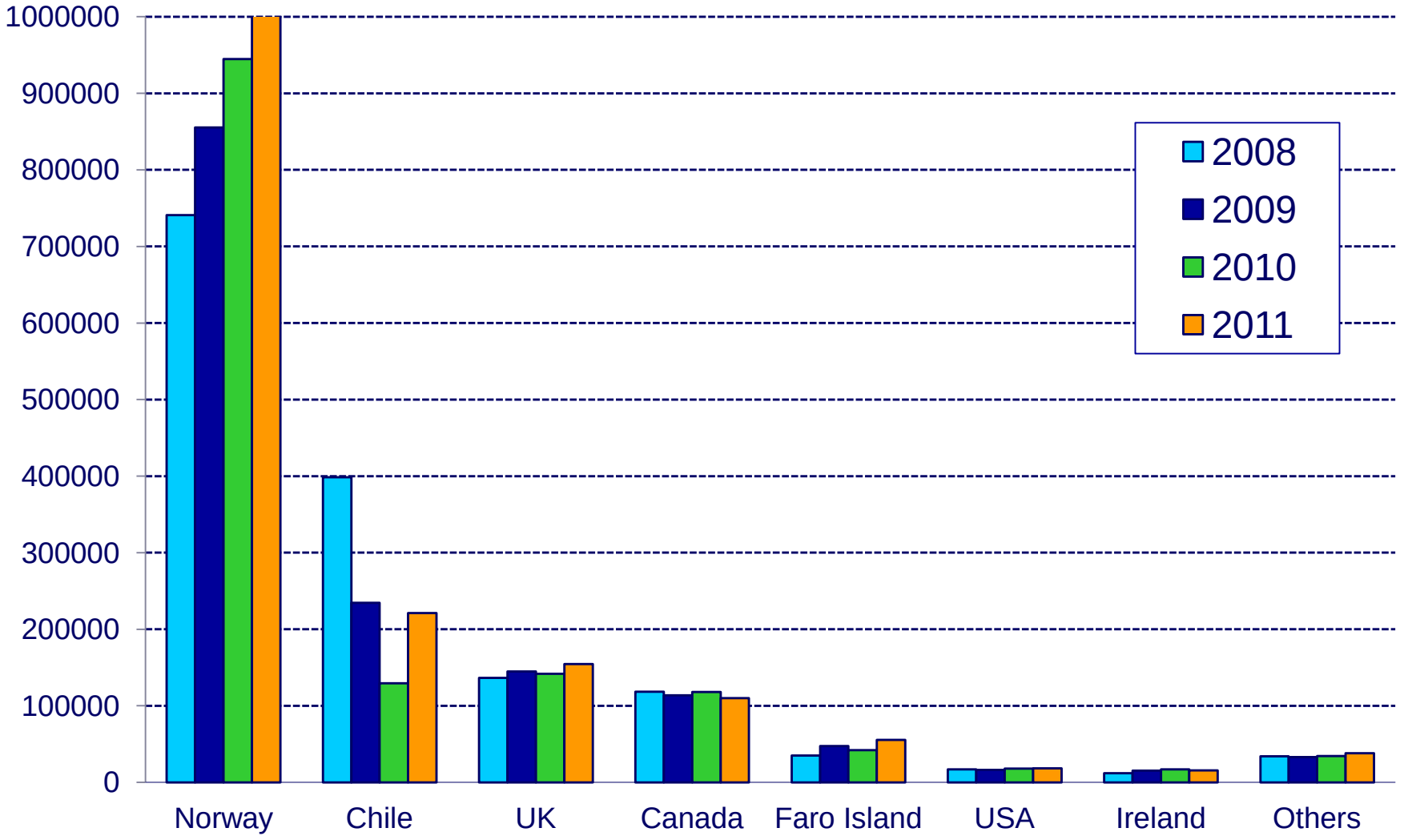
- ✓ Worlds first Atlantic salmon family breeding programme established in 1971 – 1974 by Akvaforsk/AUN
- ✓ Base population collected from more than 40 different Norwegian salmon rivers
- ✓ Selective breeding through 11 salmon and 13 trout generations
- ✓ Sale of 350-400 million salmonid eggs to 14 countries
- ✓ 170 employees
- ✓ Turnover of 330 mNOK (2010)

AquaGen™

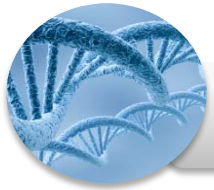
Norwegian Salmon industry -
production and value creation



Global salmon production (tonnes) 2008-2011



Evolution of breeding programs in salmon



Whole genome selection?

Genomic selection

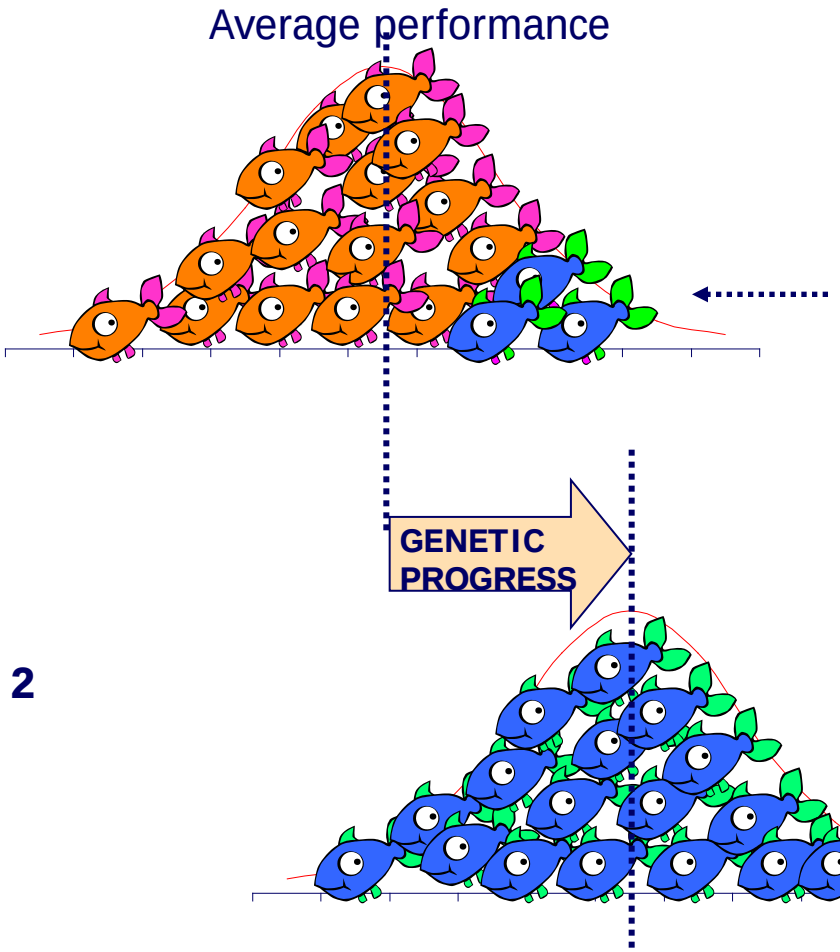
Within family selection

MAS

Traditional breeding

Breeding & Genetics

How to utilise genetic variation to generate genetic progress?

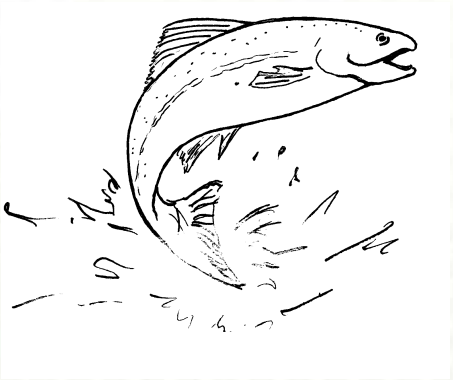


4-years test-program to identify the best performing families/individuals

- Growth/feed efficiency
- Product qualities
- Challenge tests for diseases

Most important traits in breeding goal

- **Growth and product quality**:
  - Weight
  - Fillet-weight and thickness
  - Fillet- colour and lipid content
  - Fillet texture



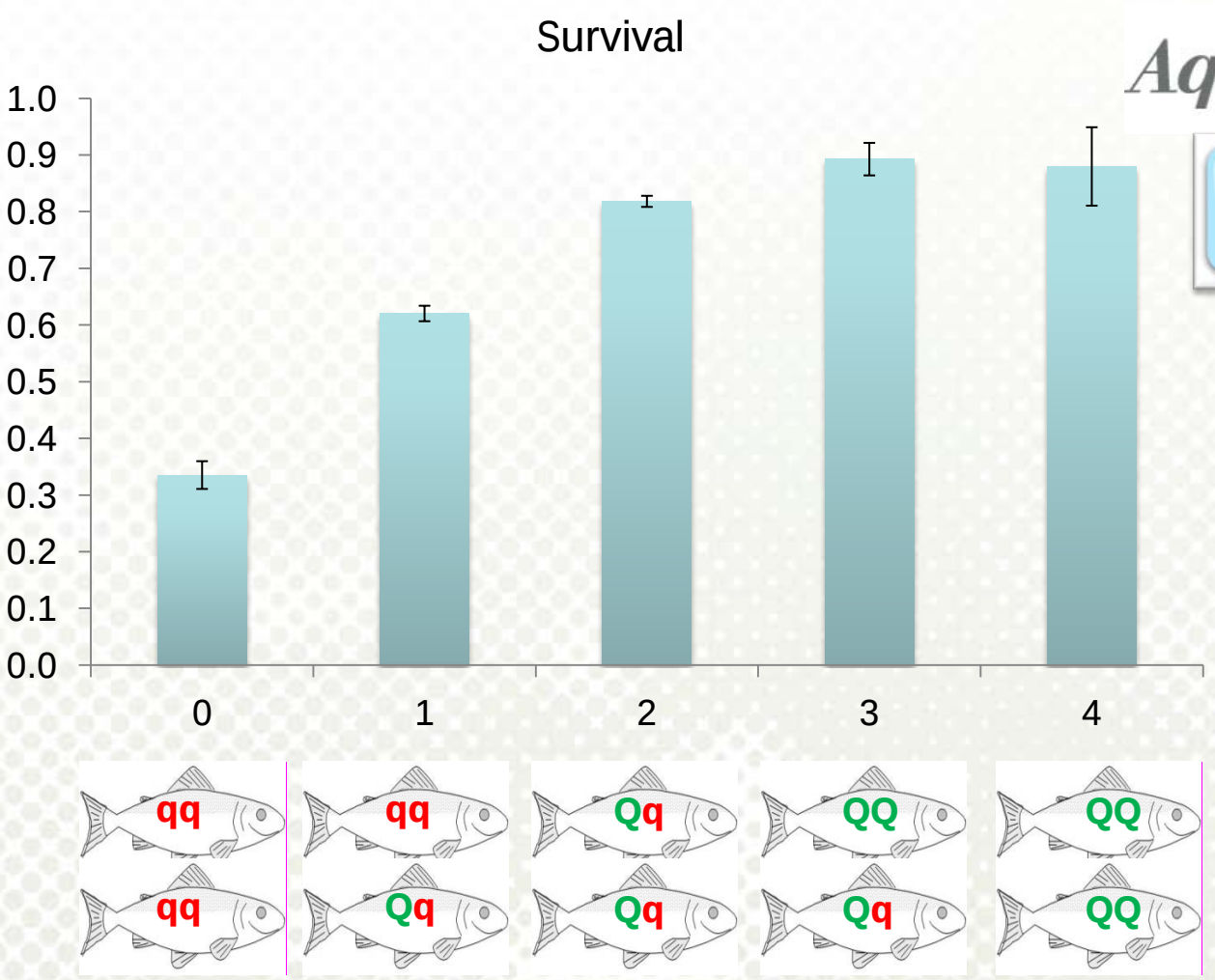
- **Diseases:**
  - Infectious pancreatic necrosis (IPN)
  - Infectious salmon anemia (ISA)
  - Pancreas Disease (PD)
  - Salmon Rickettsial Syndrome (SRS)
- **Sea lice resistance** (*Lepeophtheirus salmonis*)

Infectious pancreatic necrosis (IPN)

- Viral disease
- Average losses: 8 % (smolt producers), 5 % (grow-out)
- Part of Aqua Gen breeding goal since 2001 (family selection)
- Overall heritability: 0.31 (Wetten et al. 2007)
- Large QTL detected on chr. 26 by interval mapping
- QTL explains 80% of genetic variation
- QTL-selection at breeding nucleus and multiplier level

AquaGen™ Nofima CIGENE Centre for Integrative Genetics

Strong correlation between deduced QTL genotypes and survival in challenge tests



Market share – 2 years after launch

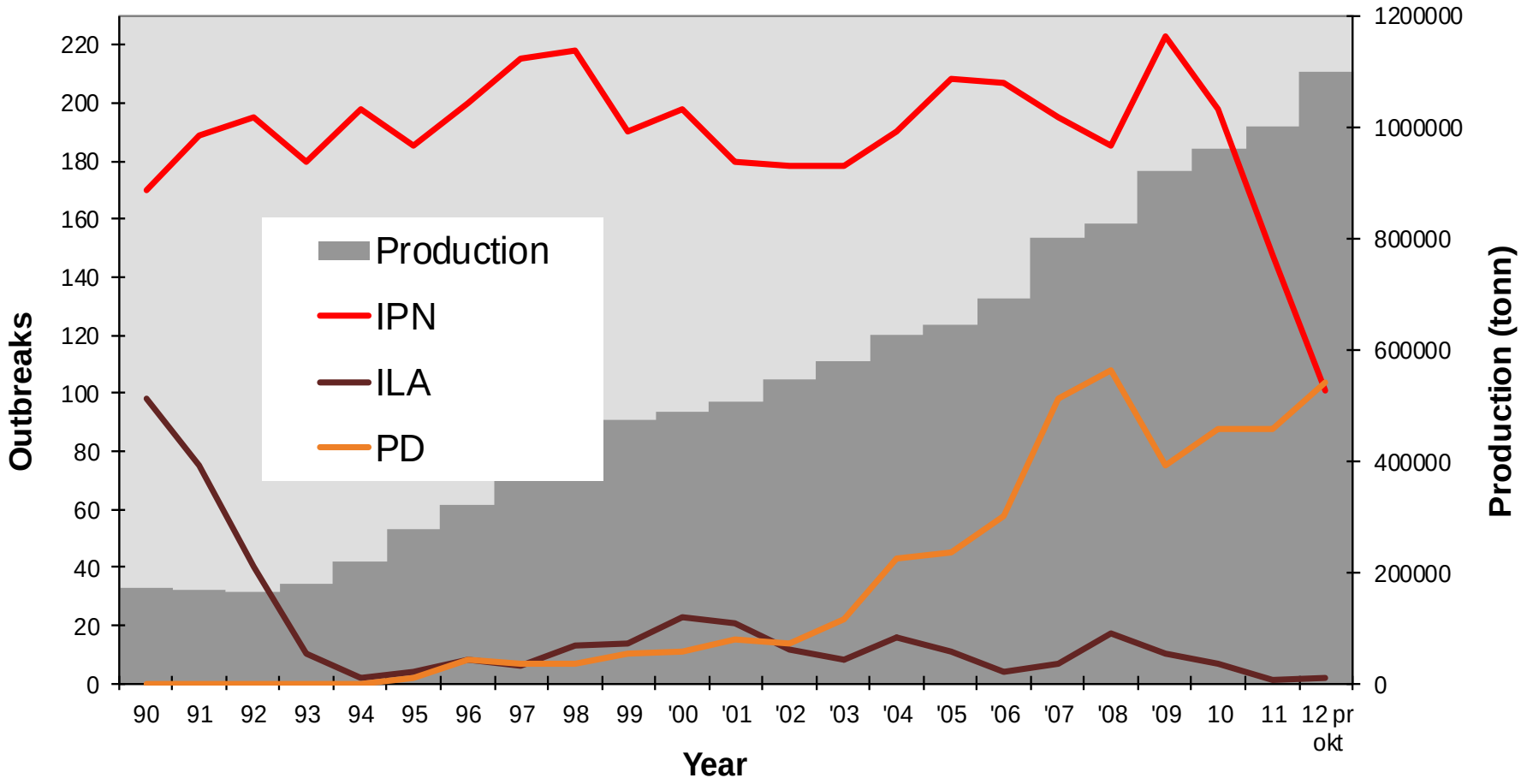
Europe

- 2009/10 – 30%
- 2010/11 – 65%
- 2011/12 – 95 %

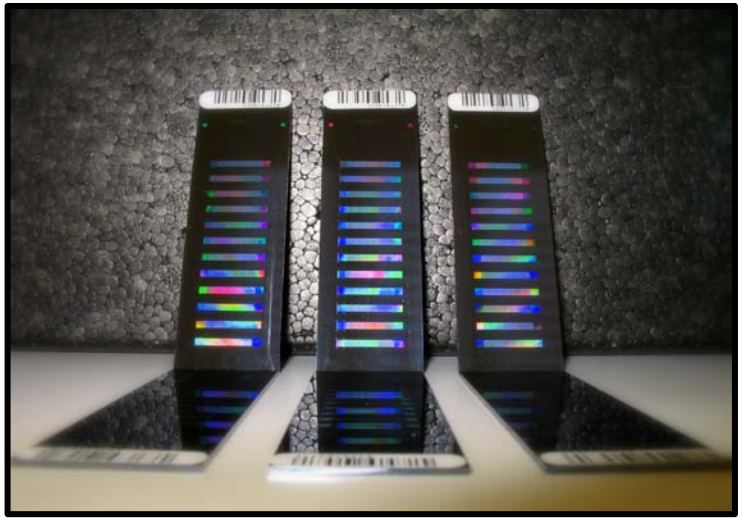
Chile

- 2010 – 34%
- 2011 – 50%

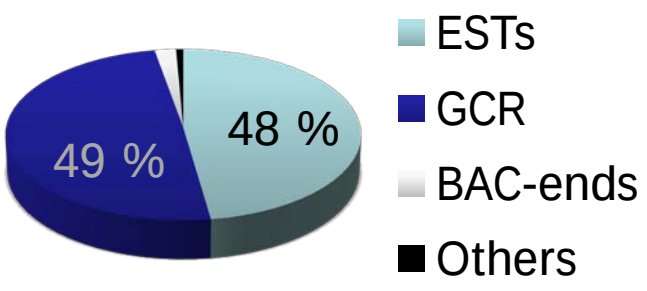
Outbreaks of viral diseases in Norwegian aquaculture



Source: Veterinary Institute/FHL



Salmon 7K SNP array

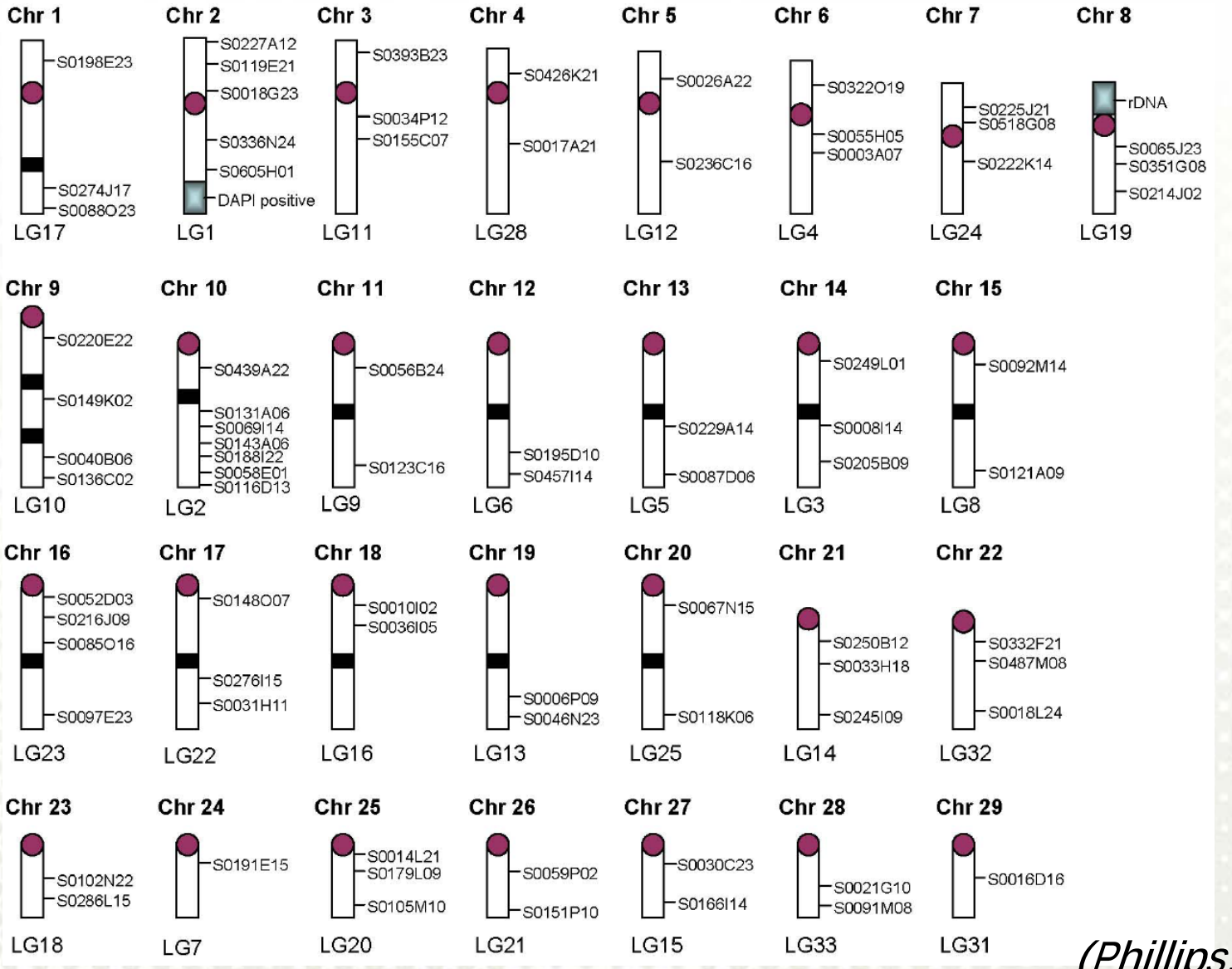


3297 salmon, 143 families
5919 SNPs mapped to chrs
5650 SNPs integrated in map

(Lien, 2011)

Chromosome	Number of SNPs	Female map	Male map	Female:male ratio
ssa01	386	135.3	130.1	1.04
ssa02	241	121.8	27	4.51
ssa03	291	115.4	61	1.89
ssa04	224	112.4	99.1	1.13
ssa05	255	116.6	54.9	2.12
ssa06	251	119.9	68.4	1.75
ssa07	158	114	72.2	1.58
ssa08	71	56.2	7.6	7.39
ssa09	311	106.8	79.1	1.35
ssa10	296	88.1	65.8	1.34
ssa11	233	85.1	58.4	1.46
ssa12	242	118.6	66.5	1.78
ssa13	285	89.7	84.1	1.07
ssa14	206	69.2	65.1	1.06
ssa15	215	80.1	82.3	0.97
ssa16	192	63	18.6	3.39
ssa17	166	69.3	11.3	6.13
ssa18	164	73.8	37.7	1.96
ssa19	157	66.3	71.6	0.93
ssa20	177	63.3	49.2	1.29
ssa21	107	53.9	66.7	0.81
ssa22	160	58.8	62.5	0.94
ssa23	126	51.4	58.1	0.88
ssa24	115	58.7	58.6	1.00
ssa25	117	55	56.2	0.98
ssa26	145	81.6	77.5	1.05
ssa27	162	53.6	57.7	0.93
ssa28	96	53.7	44.9	1.20
ssa29	101	71.4	60.7	1.18
	5650	**2403**	**1752.9**	**1.37**

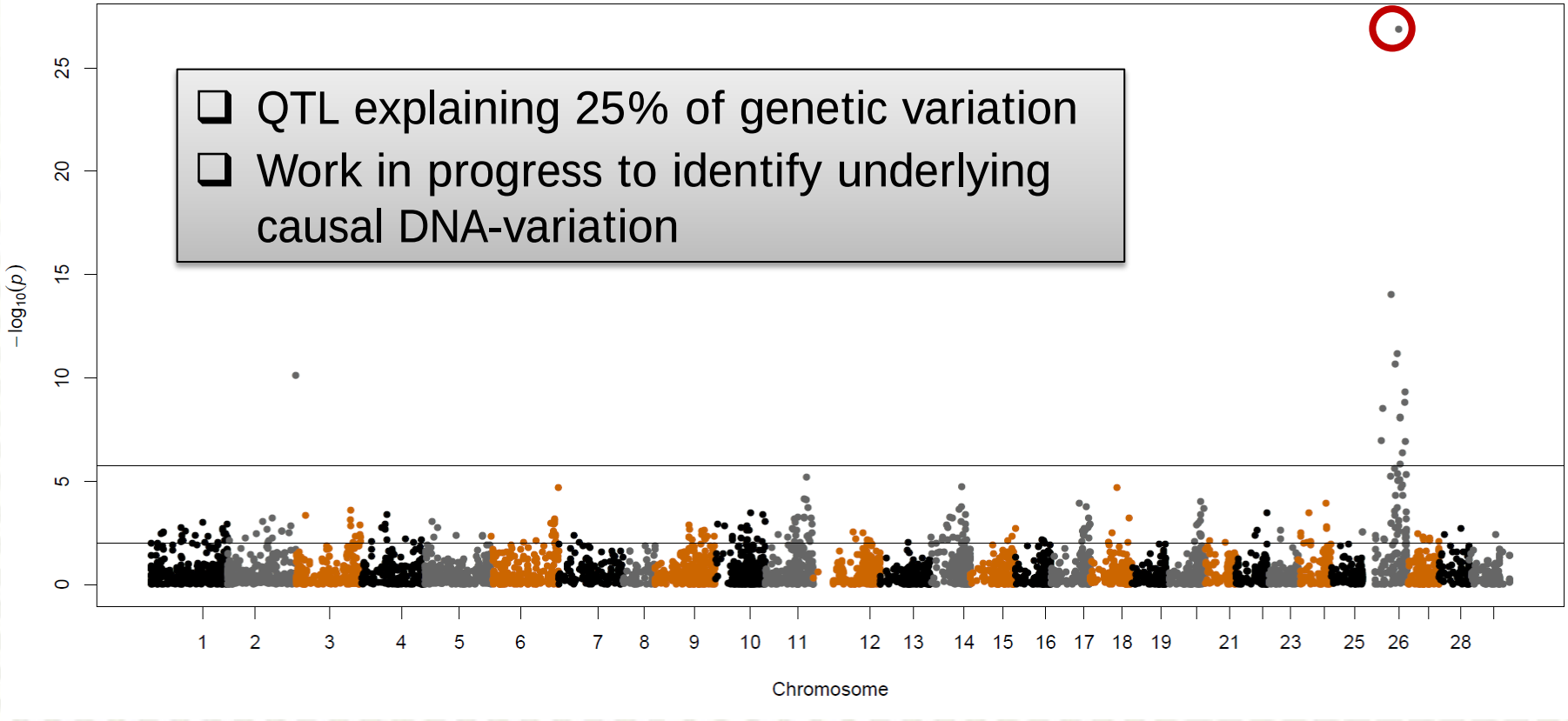
The Atlantic salmon karyotype



(Phillips, 2009)

Filet colour

- QTL explaining 25% of genetic variation
- Work in progress to identify underlying causal DNA-variation



Dramatic change in environment

11 generations of selective breeding not enough to draw large QTLs to fixation?

Davidson *et al. Genome Biology* 2010, **11**:403
http://genomebiology.com/2010/11/9/403

Genome Biology

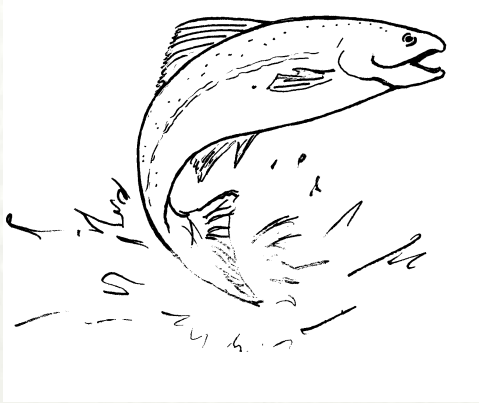
OPEN LETTER

Sequencing the genome of the Atlantic salmon (*Salmo salar*)

William S Davidson[1]*, Ben F Koop[2], Steven JM Jones[3], Patricia Iturra[4], Rodrigo Vidal[5], Alejandro Maass[6], Inge Jonassen[7], Sigbjorn Lien[8] and Stig W Omholt[8]

Abstract

The International Collaboration to Sequence the Atlantic Salmon Genome (ICSASG) will produce a genome sequence that identifies and physically maps all genes in the Atlantic salmon genome and acts as a reference sequence for other salmonids.



Funding:

FHF
FISKERI- OG HAVBRUKSNÆRINGENS FORSKNINGSFOND

InnovaChile CORFO

Genome BritishColumbia
Leading > Investing > Connecting

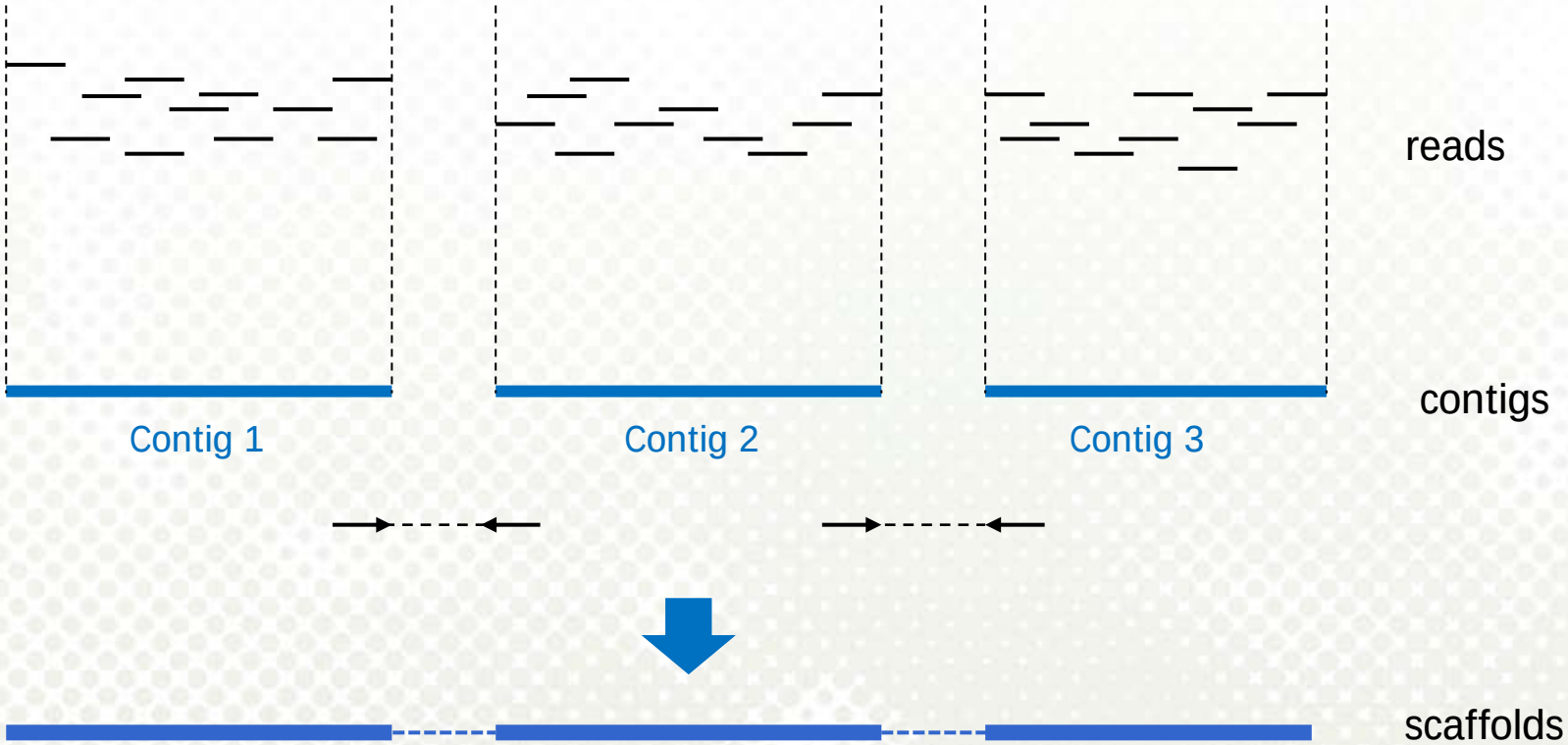
The Research Council of Norway

Workshop on Sequencing Salmonid Genomes
UMB, Ås, October 2005

Canada, Chile and Norway

formed the
International Collaboration to Sequence
the Atlantic Salmon Genome (ICSASG)
in April 2009 in Santiago

Main objective:
High quality reference genome sequence



- Contig N50 > 50 kb
- Scaffold N50 > 1.0 Mb

The Atlantic salmon genome

- ***Genome size:*** The salmon genome is quite similar to those of mammals with respect to size and overall base composition.
- ***Genome duplication:*** The common ancestor of salmonids underwent a whole genome duplication event between 20 and 120 million years ago. Thus, the extant salmonid species are considered pseudo-tetraploids whose genomes are in the process of reverting to a stable diploid state.
- ***Repetitive elements:*** The Atlantic salmon genome is highly repetitive, and at least 14 different DNA transposon families whose members are ~1.5 kb have been described.

Frequent repeat families in the Atlantic salmon genome

New Designation	Class	Subclass	Superfamily	Percent of Genome
CR1-UVic-16_Ssa	I	LINE	CR1	1.46
CR1-UVic-17_Ssa	I	LINE	CR1	2.86
CR1-UVic-19_Ssa	I	LINE	CR1	3.62
Crack-Uvic-18_Ssa	I	?	Crack	2.01
PGBD-UVic-10_Ssa	II	TIR	piggyBac	3.78
PGBD-UVic-11_Ssa	II	TIR	piggyBac	3.95
PGBD-UVic-12_Ssa	II	TIR	piggyBac	3.71
PGBD-UVic-13_Ssa	II	TIR	piggyBac	4.12
PGBD-UVic-14_Ssa	II	TIR	piggyBac	3.97
PGBD-UVic-15_Ssa	II	TIR	piggyBac	3.67
PGBD-UVic-9_Ssa	II	TIR	piggyBac	3.81
UVic-1_Ssa	II	TIR	Tc1/Mariner	3.50
UVic-2_Ssa	II	TIR	Tc1/Mariner	1.36
UVic-20_Ssa	I	LINE	?	0.87
UVic-21_Ssa	*	*	*	0.97
UVic-22_Ssa	*	*	*	3.76
UVic-3_Ssa	II	TIR	Tc1/Mariner	0.70
UVic-4_Ssa	II	TIR	Tc1/Mariner	1.41
UVic-5_Ssa	II	TIR	Tc1/Mariner	1.05
UVic-6_Ssa	II	TIR	Tc1/Mariner	2.88
UVic-7_Ssa	II	TIR	Tc1/Mariner	1.70
UVic-8_Ssa	II	TIR	Tc1/Mariner	2.90
				58.05

Some salmon repeats are very long (3-5kb) and highly similar (>97%)

Double haploid salmon 'Sally'

A double haploid salmon has been produced by mitotic androgenesis

The Atlantic Salmon Genome Sequencing Project (Phase I)

Libraries:

- 3-4 Kb plasmid library
- 40 Kb Fosmid library
- 130 kb BAC library

Sanger sequencing

'Reads':

- 200,000 BAC reads 600bp Q20 = 0.075x coverage
- 5,291,005 Fosmid reads 630bp Q20 = 1x coverage
- 12,173,913 plasmid reads 750bp Q20 = 2.8x coverage

Total: 3.875x coverage or ~17,664,918 reads

The Atlantic Salmon Genome Sequencing Project (Phase II)

J. Craig Venter INSTITUTE

Next generation sequencing technology:

illumina PE + MP

+

PACIFIC BIOSCIENCES

+

BAC by BAC sequencing

Contig N50= 20kb

The Atlantic Salmon Genome Sequencing Project

Other objectives:

- Integrate sequence with genetic and physical maps
- Anchor sequence to chromosomes
- Annotate the salmon genome sequence
- Make annotated genome available through genome browser(s)

MSRGF_jcf2339334026

SNP_ID	Chr	Order	Female	Male	Meioses	Scaffold	Length	Start	End
GCR_cBin4736_Ctg1_83	ssa01	**346**	133	49.9	2339	jcf2339334026	10 885 339	210 663	210 564
ESTNV_31404_229	ssa06	**31**	8.6	1.9	2998	jcf2339334026	10 885 339	747 152	747 251
ESTNV_34763_2153	ssa03	**274**	110.3	57.7	1591	jcf2339334026	10 885 339	800 939	800 840
ESTNV_34763_1056	ssa03	**277**	110.9	57.7	739	jcf2339334026	10 885 339	803 602	803 523
GCR_cBin40305_Ctg1_70	ssa06	**28**	8.3	1.8	2606	jcf2339334026	10 885 339	1 099 896	1 099 802
BASS15_B7_A06_540	ssa09	**250**	91.9	5.7	1776	jcf2339334026	10 885 339	2 232 385	2 232 286
ESTNV_17881_371	ssa09	**249**	91.9	5.7	1778	jcf2339334026	10 885 339	2 249 828	2 249 927
ESTNV_15783_204	ssa09	**252**	92	5.7	2080	jcf2339334026	10 885 339	2 308 085	2 307 996
GCR_cBin23719_Ctg1_361	ssa09	**248**	91.8	5.7	2224	jcf2339334026	10 885 339	2 723 458	2 723 360
GCR_cBin6611_Ctg1_379	ssa09	**251**	92	5.7	2644	jcf2339334026	10 885 339	2 955 615	2 955 714
GCR_cBin35853_Ctg1_91	ssa09	**253**	92.3	5.7	1484	jcf2339334026	10 885 339	3 493 694	3 493 793
GCR_cBin21225_Ctg1_149	ssa09	**254**	92.3	5.7	2971	jcf2339334026	10 885 339	3 852 921	3 852 822
ESTNV_32148_1877	ssa09	**257**	92.9	5.7	2569	jcf2339334026	10 885 339	5 031 019	5 031 118
GCR_cBin35196_Ctg1_476	ssa09	**256**	92.9	5.7	2658	jcf2339334026	10 885 339	5 066 264	5 066 165
ESTV_19379_358	ssa09	**209**	83.2	5.1	1863	jcf2339334026	10 885 339	6 109 912	6 109 813
ESTNV_33105_1177	ssa09	**211**	84.1	5.4	2482	jcf2339334026	10 885 339	6 831 985	6 831 886
ESTV_16452_1030	ssa09	**213**	84.7	5.4	712	jcf2339334026	10 885 339	7 150 351	7 150 294
GCR_cBin46790_Ctg1_21	ssa09	**230**	87.4	5.6	238	jcf2339334026	10 885 339	7 391 419	7 391 489
ESTV_16287_822	ssa09	**212**	84.7	5.4	2410	jcf2339334026	10 885 339	7 426 525	7 426 591
ESTNV_34928_1678	ssa09	**219**	85.3	5.6	2039	jcf2339334026	10 885 339	7 636 085	7 636 180
ESTNV_29649_636	ssa09	**217**	85.3	5.6	2465	jcf2339334026	10 885 339	7 788 762	7 788 663
ESTNV_34379_298	ssa09	**214**	85.3	5.6	2294	jcf2339334026	10 885 339	7 790 079	7 790 147
ESTNV_21259_337	ssa09	**216**	85.3	5.6	2834	jcf2339334026	10 885 339	7 796 521	7 796 620
GCR_cBin42637_Ctg1_200	ssa09	**215**	85.3	5.6	1268	jcf2339334026	10 885 339	7 886 258	7 886 159
ESTNV_32419_1390	ssa19	**24**	20.9	3.1	285	jcf2339334026	10 885 339	7 973 791	7 973 692
GCR_cBin23341_Ctg1_419	ssa19	**30**	23.3	3.1	2252	jcf2339334026	10 885 339	7 984 596	7 984 497
GCR_cBin12500_Ctg1_64	ssa19	**27**	21.4	3.1	1623	jcf2339334026	10 885 339	8 479 137	8 479 225
GCR_cBin38814_Ctg1_115	ssa19	**28**	21.5	3.1	2772	jcf2339334026	10 885 339	8 495 008	8 494 909
GCR_cBin38814_Ctg1_79	ssa19	**26**	21.4	3.1	2795	jcf2339334026	10 885 339	8 495 044	8 494 945
GCR_cBin38814_Ctg1_29	ssa19	**29**	21.7	3.1	2778	jcf2339334026	10 885 339	8 495 073	8 494 995
ESTV_20150_714	ssa27	**117**	48	1.5	2350	jcf2339334026	10 885 339	9 430 952	9 431 051
GCR_cBin20876_Ctg1_29	ssa17	**4**	0.8	0.2	1666	jcf2339334026	10 885 339	9 670 357	9 670 279
ESTNV_35090_1063	ssa11	**123**	57.5	2	1464	jcf2339334026	10 885 339	10 843 955	10 843 856

Development of SalmonHD SNP-chips

SNP-detection

- *Whole genome resequencing;* 28 parents + 4 double haploids, 10-15x coverage with Illumina PE (2x100bp)
- *Reference;* CIGENE Celera assembly (contigs_2012-02-01)
- *Software;* Bowtie -> FreeBayes -> in house SNP-filtering pipeline

1. ***ssaXHD:*** 952 389 putative SNPs
   - 7K SNPs: 7 000
   - Exonic SNPs: 109 366
   - Likely nonsynonymous SNPs: 6 021
   - Genome distribution
   - SNPs in 'telomeric' contigs: 13 871
   - SNPs in highly homeologous regions: 16 579

Affymetrix® AXIOM

2. ***ssaHD:*** 200K SNPs selected based on;
   - Call rate & clustering (SNPs preferred)
   - Genome distribution
   - Predicted function
   - LD with other SNPs, haplotype tags

AquaGen™

CIGENE
Centre for Integrative Genetics

Use of «*SalmonHD SNP-chip*» to improve assemblies

Main objectives

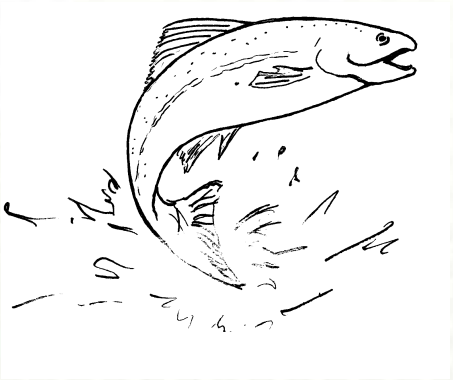
- Large scale error detection in scaffolds/contigs
- Accurate spitting and editing of scaffolds
- Separate highly homeologous sequences
- Anchor majority of sequences to chromosomes

In combination with 'homeologous mapping'

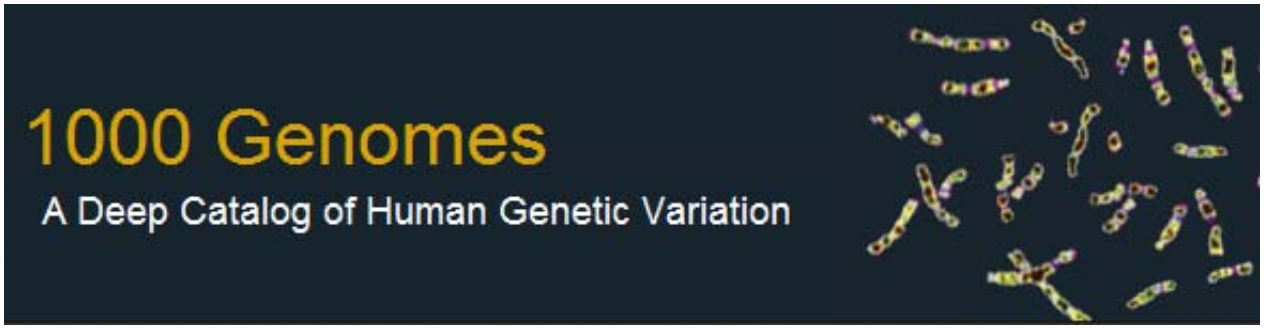
- Resolve duplicated regions in the salmon
- Detect larger differences in duplicated regions (deletions, insertions, inversions)
- Build the foundation for analyzing levels and mechanisms of diploidization and duplicate gene evolution in Atlantic salmon

Most important traits in breeding goal

- **Growth and product quality**:
  - Weight
  - Fillet-weight and thickness
  - Fillet- colour and lipid content
  - Fillet texture



- **Diseases:**
  - Infectious pancreatic necrosis (IPN)
  - Infectious salmon anemia (ISA)
  - Pancreas Disease (PD)
  - Salmon Rickettsial Syndrome (SRS)
- **Sea lice resistance** (*Lepeophtheirus salmonis*)



ARTICLE

doi:10.1038/nature11632

An integrated map of genetic variation from 1,092 human genomes

The 1000 Genomes Project Consortium*

The Aqua Genome Project

Genome sequences for Atlantic salmon and Atlantic cod

FUGE Genotyping Platform / The Norwegian Sequencing Centre

WP1 – 1000 genomes salmon and cod projects

WP2 – Dissection of the genomic architecture underlying economically important traits

WP3 – Selectional forces and landscape genomics

WP4 – Phenotypic plasticity and epigenetics of teleost development

WP5 – Functional genomics: combining genomic sequence data with transcriptome profiling and functional testing to dissect the molecular mechanism of disease resistance and other important traits

Create a sustainable basis for aquaculture and fisheries management of Atlantic salmon and cod by studying the genetic basis of phenotypic variation and revealing the genomic effects of selection and adaptation to changing environmental conditions.

CEES
Centre for Ecological and Evolutionary Synthesis

Nofima

AquaGen™

CIGENE
Centre for Integrative Genetics


Thank you for your attention!
CIGENE
Centre for Integrative Genetics