Application of genomics to selective breeding of Atlantic salmon

Ross Houston

Introduction

- Salmonid aquaculture = large & expanding industry
- Atlantic salmon major farmed species

Global salmonid production (tonnes)

Source: FAO

- Infectious diseases are a serious problem for salmon farming worldwide
  - Viral diseases:
    - E.g. Infectious Salmon Anaemia (ISA), Infectious Pancreatic Necrosis (IPN), Pancreas Disease (PD)
  - Bacterial diseases:
    - E.g. Furunculosis, Salmon Rickettsial Syndrome (SRS)
  - Ectoparasites:
    - E.g. Sea lice (L. salmonis & C. rogercresseyi), Amoebic Gill Disease (AGD)

Control strategies can include vaccination, drugs, management, & selective breeding

- Infectious diseases are a serious problem for salmon farming worldwide
  - Viral diseases:
    - E.g. Infectious Salmon Anaemia (ISA), Infectious Pancreatic Necrosis (IPN), Pancreas Disease (PD)
  - Bacterial diseases:
    - E.g. Furunculosis, Salmon Rickettsial Syndrome (SRS)
  - Ectoparasites:
    - E.g. Sea lice (L. salmonis & C. rogercresseyi), Amoebic Gill Disease (AGD)

Is host resistance to these diseases heritable???
Control strategies can include vaccination, drugs, management, & selective breeding.

Viral diseases:
- E.g. Infectious Salmon Anaemia (ISA), Infectious Pancreatic Necrosis (IPN), Pancreas Disease (PD)

Bacterial diseases:
- E.g. Furunculosis, Salmon Rickettsial Syndrome (SRS)

Ectoparasites:
- E.g. Sea lice (L. salmonis & C. rogercresseyi), Amoebic Gill Disease (AGD)

Published evidence for host genetic variation in resistance:


Selective Breeding for Resistance

• Selection for resistance in salmon is achievable...
  • High fecundity and external fertilisation helps
  • Breeding value of candidates typically estimated from siblings

• ...but there are limitations:
  • Only half of the genetic variation is between-family variation
  • Disease challenge experiments each year are undesirable
  • Genomics tools can help:

Parents

Offspring selection candidates

Calculate breeding values

Offspring Disease Challenge

Subsequent generation

Take sample of DNA

Genotype for genetic markers

Calculate (genomic) breeding values

Genomics tools can help:

Selective Breeding for Resistance
Selective Breeding for Resistance

- Farmed salmon are close to wild ancestors
  - ~10 generations of selective breeding (fewer for disease resistance)
- New disease pressures in farmed environment
- Major effect QTL more common than in terrestrial species?

Introduction

- Genetic architecture of resistance important for application of genomics in breeding programmes
- Three case studies of genetic resistance in salmon:
  1. Infectious Pancreatic Necrosis (IPN) virus
  2. Pancreas Disease (PD) virus
  3. Sea lice (L. salmonis)

IPN Resistance

- Collected mortalities and survivors from ‘natural’ IPN virus outbreaks in seawater
  - Markers to assign to family and establish pedigree
    - heritability ~ 0.4
  - Genome scan for resistance loci in salmon genome:
    - Stage 1: QTL detection
      - Sparse markers, sire segregation
    - Stage 2: QTL confirmation & positioning
      - Denser markers, dam segregation

IPN Resistance

- Single locus explains almost all genetic variation in resistance in freshwater and seawater

...
IPN Resistance

• Large effect of QTL on mortality level of fry

<table>
<thead>
<tr>
<th>Dam haplotype</th>
<th>Sire haplotype</th>
<th>IPN mortality levels in fish carrying alternative QTL alleles (n = 341)</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>R</td>
<td>0%</td>
</tr>
<tr>
<td>S</td>
<td>S</td>
<td>1%</td>
</tr>
<tr>
<td>R</td>
<td>S</td>
<td>63%</td>
</tr>
</tbody>
</table>

Houston et al. (2010) Heredity

Next stages:

i. SNP markers to accurately predict QTL genotype in commercial salmon stocks

ii. Understanding of underlying biological mechanisms

• DNA sequencing technology is transforming animal breeding

- Reference genome sequences for A. salmon and rainbow trout
- Tools for high density SNP genotypes:
  - Genotyping by sequencing
  - SNP arrays


SNP Discovery and Genotyping by RAD Sequencing

- Illumina sequencing of pooled, barcoded samples
- Applied to RR and SS homozygotes from IPNV-challenged families
- Compared SNP genotypes for concordance with QTL

RAD Seq

Whole Genome Seq

Adapted from Etter et al (2009)

22K SNPs → 50 linked SNPs → 10 screened → 2 associated

Association between SNP and IPN mortality in test (n=4000) and validation (n=5000) populations of salmon fry

<table>
<thead>
<tr>
<th>Population</th>
<th>RAD01 SNP Genotype</th>
<th>Mortality Rate (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RR</td>
<td>0.10 (0.03)</td>
</tr>
<tr>
<td></td>
<td>RS</td>
<td>0.17 (0.01)</td>
</tr>
<tr>
<td></td>
<td>SS</td>
<td>0.60 (0.01)</td>
</tr>
<tr>
<td>TEST</td>
<td>0.11 (0.01)</td>
<td>0.25 (0.01)</td>
</tr>
<tr>
<td></td>
<td>0.63 (0.01)</td>
<td></td>
</tr>
<tr>
<td>VALIDATION</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Commercial application:

- SNP panel used as genetic test to select resistant broodstock

Houston et al. (2012) BRC Genomics

• Microarray & RNA-Seq comparison of RR vs SS host response to infection in resistant vs susceptible fish

1 day post-infection
7 days
20 days

Gene expression susceptible fry

- Interferons & other cytokines

Gene expression resistant fry

- Resistance mechanism early in host-pathogen interaction
- Mapping early DE genes → positional / functional candidates
**PD Resistance**

- Near monogenic host resistance to a virus is atypical
  - Resistance to other diseases likely to be controlled by a mix of QTL and a polygenic component

- Pancreas Disease (salmonid alphavirus)
  - Morbidity & mortality at post-smolt stage
  - Causative agent – Salmonid alphavirus
  - Six subtypes, geographic specificity

**PD Resistance**

- Fry challenge model
  - Naive fry (Marine Harvest) challenged with water from tank of IP-injected parr
  - Total n = 3,949
  - 150 full-sib families
  - Sire half-sib structure
  - Mortality ~ 60% overall

- Sampled survivors and mortalities from challenge
  - Genotyped for SNP markers, assigned to family
  - Estimated $h^2$
    - Model: $Y_{ij} = \mu + Sire + Dam + e_{ij}$
    - Observed, LOGIT, PROBIT scales

- QTL mapping using sparse SNP panel
  - Two-stage mapping strategy as with IPN
  - Chr 3 QTL confirmed in separate population

- Fig 3 PD resistance QTL mapped in population of post-smolts (Baranski et al)
  - Position of fry & post-smolt QTL on chromosome coincide

- Analysis Method
  - Mortality (logit) 0.34 (0.05)
  - Underlying liability scale 0.54 (0.07)
  - Logit link scale 0.46 (0.04)
Pancreas disease resistance is highly heritable ($h^2 \approx 0.5$). The genetic architecture is ‘intermediate’. A QTL on Chr 3 affects resistance in fry & post-smolts. SNPs associated with QTL applied in selective breeding.

Sea lice challenge
- Pedigreed population (Landcatch) challenged with L. salmonis larvae
  - 96 larvae per fish
  - 725 fish sampled 7 days post infection
  - Individual lice counts using stereo microscope
  - Weight & length measurements and fin clip sample

Genotyping and QC
- Samples genotyped using Axiom 132 K SNP array
- QC filtering of SNPs and individuals (Mendel error, MAF, etc.)
  - 111 K SNPs retained
  - 29 sires, 61 dams, 534 offspring

SNP array is informative in farmed Scottish, farmed Norwegian and wild European Atlantic salmon populations.

Sea Lice Resistance
- Create High Density SNP chip
  - Illumina sequencing of diverse populations
    - Combination of RAD-Seq, RR-Seq and RNA-Seq
    - Millions of candidates to ~132K verified, polymorphic assays
    - Sequencing haploid fish to remove paralogous variation

SNP array development & application
- Validation and utility of Affymetrix SNP chip
  - 1. Can be used to detect population structure based on genomic similarity
  - 2. Y-specific probes accurately predict phenotypic sex of juveniles

Houston et al. (2014), BMC Genomics

Quantitative Genetic parameters
- Animal model in ASRemi used to estimate $h^2$ of lice count
  - Sex: fixed effect, animal (pedigree): random effect
  - Body weight as a covariate
- Genomic (IBS) relationship matrix calculated using Genabel
  - Equivalent model with G matrix replacing A matrix

<table>
<thead>
<tr>
<th>Mean lice count</th>
<th>Standard Deviation</th>
<th>$h^2$ (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>25.6</td>
<td>12.4</td>
<td>0.24 (0.08)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.20 (0.07)</td>
</tr>
</tbody>
</table>

Consistent $h^2$ estimates using pedigree and genomic matrix.
Sea Lice Resistance

- Genome-wide association analysis
  - Genabel mmscore & ASReml

- The most significant SNP
  - Located on distal end of chromosome 3, $P < 10^{-5}$
  - Derived from RNA-Seq
  - Occurs in a gene for which the protein product is known to be involved in the host response of the epidermis to sea lice infection

<table>
<thead>
<tr>
<th>SNP Genotype</th>
<th>AA (n=15)</th>
<th>AB (n=94)</th>
<th>BB (n=537)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lice Count (SE)</td>
<td>36.4 (3.4)</td>
<td>29.8 (1.3)</td>
<td>24.3 (0.6)</td>
</tr>
</tbody>
</table>

- Resistance allele is common in this population

Sea Lice Resistance

- Summary:
  - We created a publicly-available high-density SNP array
  - Can detect population structure & predict phenotypic sex
  - $h^2$ estimated at 0.24 (pedigree) and 0.20 (genomic)
  - The genetic architecture is polygenic
  - Genomic prediction & selection for resistance underway

Overall Summary

- Host resistance to infectious diseases has a genetic component
  - PD and IPN viruses ~ 50% variation
  - Sea lice ~ 25% variation
  - But the genetic architecture of resistance varies....

Genetic Architecture of Resistance

- Polygenic: No evidence for 'major' QTL
- Intermediate: ~30% genetic variation explained by 3 QTL
- Nearly Monogenic: ~80% genetic variation explained by single QTL

Overall Summary

- Host resistance to infectious diseases has a genetic component
  - PD and IPN viruses ~ 50% observed variation genetic
  - Sea lice ~ 25% observed variation genetic
  - But the genetic architecture of resistance varies....

- Genomic tools have a major role for salmon disease research and application
  - Past success for IPN virus & significant progress for other diseases
  - Working with aquaculture industry is key
  - Application of genomics is tailored to genetic architecture of resistance
Acknowledgements

Main collaborators:

- Steve Bishop
- Chris Haley
- Alan Archibald
- Serap Gonen
- Natalie Lowe
- John Taggart
- Michael Bekart
- James Bren
- Brendan McAndrew
- Derrick Guy
- Alastair Hamilton
- Alan Tinch
- Jose Mota-Velasco
- Ashie Norris
- Fiona Brew

Karim Gharti
Richard Talbot
Tim Cauard
John Davey

Genotyping Strategies

- Rough comparison of SNP chip & genotyping by sequencing

<table>
<thead>
<tr>
<th></th>
<th>De Novo SNP Chip</th>
<th>RAD-Seq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time to set up</td>
<td>1 – 2 years</td>
<td>1 – 6 months</td>
</tr>
<tr>
<td>Up front cost</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>Cost per sample</td>
<td>$50 - 100</td>
<td>$10 - 30</td>
</tr>
<tr>
<td>SNP density</td>
<td>30 – 1000 K</td>
<td>1 – 20 K</td>
</tr>
<tr>
<td>Bioinformatics requirement</td>
<td>Medium</td>
<td>High</td>
</tr>
<tr>
<td>Population-specificity</td>
<td>Variable</td>
<td>High</td>
</tr>
<tr>
<td>Accuracy of data</td>
<td>Very High</td>
<td>High</td>
</tr>
<tr>
<td>Suitable sample size</td>
<td>Large</td>
<td>Small - Medium</td>
</tr>
</tbody>
</table>