VERIFICATION OF ISOGENIC CLONAL LINES IN THE ATLANTIC SALMON (Salmo salar) THROUGH ddRADseq

Münevver Oral¹§, John B Taggart¹, Brendan J McAndrew¹, David J Penman¹, José Mota-Velasco², Per Gunnar Fjelldal³ and Tom Hansen³

§ Presenter
¹ Institute of Aquaculture, School of Natural Sciences, University of Stirling, Stirling FK9 4LA, Scotland, UK
² Landcatch, Natural Selection Ltd, Stirling FK9 4NF, Scotland, UK
³ Institute of Marine Research (IMR), Matre Research Station, NO-5984 Matredal, Norway

25th June 2015, Santiago de Compostela, Spain, ISGAXII
I. Background: Why does it matter to produce clonal lines?

Clonal Lines = Genetically Identical Individuals

- Homogenity (decreases variation in experiments)
- Standardisation of the research – refined experimental designs (3Rs)
- Speed of generation (2 consecutive production cycles via Gynogenesis or Androgenesis)
- Reveals genetic variation for many traits
- QTL identification and whole genome sequencing projects
I. Background: How to produce clonal lines?

Gynogenesis (G) All maternal inheritance
Androgenesis (A) All paternal inheritance

Spontaneous rise of:
- Haploids (due to failure in shock to deploy diploidy)
- Meiotic Gynogenetics (due to failure in the time of shock – produced by blocking 2nd polar body exclusion therefore enables to ‘capture’ the results of any crossover events – undesired heterozygosity
- Essential to verify the isogenic nature of clonal lines

FAO, Inbreeding and Broodstock Management
Chapter 6, Chromosome Set Manipulations.
I. Background: Teleost Specific-WGD

WGD results in paralogs loci.

Sequence variants found in duplicated genomes:
1. Paralogues Sequence Variants (PSVs) - fixed sites, no polymorphism
2. SNPs – allelic polymorphism
3. Multisite Variants (MSVs) – polymorphism found across paralogs

Aim of the study

• Verification of optimised genome irradiation protocol in Salmon
• Verification of successful production of isogenic clonal lines
II. Materials and Methods
Production of Clonal fish

- Sperm was diluted to $5 \times 10^8$ ml$^{-1}$ and irradiated at 170 $\mu$W.cm$^{-2}$ with 254nm UV light

- Pressure shocks used 4400-4800 min°C post-fertilization

See Online
AquaExcel_deliverables_optimsation of G1 fish production in salmon
Experimental Design
Putative Clonal Lines

Haploid Family
Parents + Progeny
(PSVs/MSVs)

Outbred Founders

G1 Family
G1: Homozygous Clone Founders

No progeny

G2 fish (putative clonal lines)

DH_Fams: DH1_Fam, DH2_Fam, DH3_Fam, DH4_Fam, DH5_Fam

DH_Fams: G2 fish (putative clonal lines)
II. Material and Methods
Double Digest RADseq (ddRADseq)

1. RE double digest (Sbfi & Sphl)
2. Add adaptors (P1+P2)
3. Size select and amplify library

Size Selection excludes very small (A) & very big fragments (B) away from the library.

Sequence & Analyse
III. Results: Sequencing & RAD tag summary

**Raw Reads:** All reads have been produced by sequencer

35,862,448 million raw reads (17,931,224 paired end)

**Filtered reads:** Reads with right barcodes & adapters combination

30,958,609 filtered reads

- G1: 1,457
- DH1: 1,238
- DH2: 1,174
- DH3: 1,174
- DH4: 1,199
- DH5: 1,158
- Haploid: 489

Reads used by Stacks to create individual paired-end markers

Total RAD markers identified in each FAM

Loci retrieved in 70% of the samples

*Stacks package (Catchen et al., 2011).*

mo19@stir.ac.uk
### III. Results

**Distribution of RAD alleles in G1 FAM**

<table>
<thead>
<tr>
<th>Map types available</th>
<th>RAD alleles (total loci)</th>
<th>Potential Paternal contributor loci</th>
<th>% of Potential Contributor Loci</th>
</tr>
</thead>
<tbody>
<tr>
<td>ab/aa</td>
<td>431</td>
<td>13</td>
<td>3.0</td>
</tr>
<tr>
<td>aa/bb</td>
<td>175</td>
<td>1</td>
<td>0.6</td>
</tr>
<tr>
<td>ab/UNK</td>
<td>18</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>ab/ab</td>
<td>314</td>
<td>93</td>
<td>29.6</td>
</tr>
<tr>
<td>ab/cc</td>
<td>7</td>
<td>1</td>
<td>14.3</td>
</tr>
<tr>
<td>aa/ab</td>
<td>445</td>
<td>11</td>
<td>2.5</td>
</tr>
<tr>
<td>ab/ac</td>
<td>40</td>
<td>8</td>
<td>20.0</td>
</tr>
<tr>
<td>cc/ab</td>
<td>5</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>UNK/ab</td>
<td>21</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>ab/cd</td>
<td>1</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>TOTAL</td>
<td>1457</td>
<td>127</td>
<td>8.7</td>
</tr>
</tbody>
</table>

- **8.7% potential paternal contributor loci, but WGD.??!**

**G1 Family with 6DHs Progeny**

- DH1 Family
- DH2 Family
- DH3 Family
- DH4 Family
- DH5 Family

**Outbred Founders**

**Putative Clonal Lines**

- G1: Homozygous Clone Founders
III. Results:
Distribution of RAD alleles in DH Fams

<table>
<thead>
<tr>
<th>Family</th>
<th>79%</th>
<th>78%</th>
<th>83%</th>
<th>84%</th>
<th>86%</th>
</tr>
</thead>
<tbody>
<tr>
<td>DH1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DH2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DH3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DH4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DH5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

♂ / ♀

[Legend for the pie charts]

ab/aa  aa/bb  ab/ab  aa/ab  ab/cc
UNK/ab  ab/ac  ab/UNK  ab/cd  cc/ab
III. Results:
Investigation of putative sire contribution

Potential Putative Paternal Contributor Loci ??

- BLAST_NCBI_WGS of Salmon

- NO convincing sign of any paternal contribution to offspring
  - Repetitive elements (transposons)
  - PSVs / MSVs
  - Noise of salmon genome

- BLAST_NCBI_RefSEQ

- It was used to prove the existence of repetitive elements
### III. Results: Control test to identify true SNPs

<table>
<thead>
<tr>
<th></th>
<th>G1_FAM</th>
<th>DH1_FAM</th>
<th>DH2_FAM</th>
<th>DH3_FAM</th>
<th>DH4_FAM</th>
<th>DH5_FAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total RAD loci</td>
<td>1457</td>
<td>1238</td>
<td>1174</td>
<td>1174</td>
<td>1199</td>
<td>1158</td>
</tr>
<tr>
<td>Potential sire cont loci</td>
<td>127</td>
<td>325</td>
<td>270</td>
<td>320</td>
<td>336</td>
<td>262</td>
</tr>
<tr>
<td>All female cont loci</td>
<td>1330</td>
<td>913</td>
<td>904</td>
<td>854</td>
<td>863</td>
<td>896</td>
</tr>
<tr>
<td>Further investigated</td>
<td>30</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>

#### G1_FAM Frequencies
- 57% True SNPs
- 20% Paralogous loci
- 23% Repetitive elements

#### DH_FAM Mean Frequencies
- 34% True SNPs
- 44% Paralogous loci
- 22% Repetitive elements

Estimated true SNP markers in G1_Fam: **758**

Estimated true SNP markers in each DH_Fams: **301**

Frequency of haploid derived heterozygous putative SNPs were 30%
IV. Conclusion

✓ Verification of optimised genome irradiation procedure for the Atlantic salmon

✓ Verification of isogenic nature of 5 clonal lines in the Atlantic salmon

• ddRADseq is a cost-effective and quick method, generating hundreds of diagnostic markers
Thanks, any questions??

mo19@stir.ac.uk
The financial support of the Turkish Government, Ministry of Education for my PhD scholarship (MEB-YLSY).

The research leading to these results has received funding from the European Union’s Seventh Framework Programme (FP7/2007-2013) under grant agreement no 262336. This publication reflects the views only of the author, and the European Union cannot be held responsible for any use which may be made of the information contained therein.
II. Material Methods:
Why PCR duplicates cannot be removed from ddRADseq paired end reads?
II. Material Methods: Difference between meiotic and mitotic gynogenesis

- **Oogonia**
- **Genome duplication and replication**
- **Meiosis I**
  - 1st p.b.
  - **Ovulation**
- **Meiosis II**
  - "Early" shock
  - **Fertilisation**
- **Mitosis I**
- **Outcome**
  - "Meiotic" gynogenetic (2n)
  - Haploid gynogenetic (n)
  - No shock
  - "Late" shock
  - "Mitotic" gynogenetic (2n)

**UV irradiation of sperm**