Asian Seabass Genome Project: A status report

László Orbán, on behalf of the
The Asian Seabass Genome Consortium

Temasek Life Sciences Laboratory, Singapore; Dobzhansky Center, St. Petersburg, Russia; Pacific Biosystems, USA; SANBI, Capetown, South Africa; Max Planck Institute for Mol. Gen., Berlin, Germany; Chinese University of HK, Hong Kong; Centre for Comparative Genomics, Murdoch University, Australia; IGIB, New Delhi, India; Georgikon Faculty, University of Pannonnia, Keszthely, Hungary;
More people, less fish

Population 7 billion

Depleted oceans

Source: National Geographic
endoftheline.com
Outline

• Introducing the Asian seabass and the selection program;

• Status report on the Genome (and Transcriptome) Project;

• Early applications: phylogeography and sex change;

• Summary.
Asian seabass
(*Lates calcarifer*)

- Barramundi (AUS);
- Euryhaline, catadromous predator;
- Protandrous (male-first) hermaphrodite;
- Distributed over a wide geographical area;
- Cultured mostly by smaller farms, a few selection programs;
Asian seabass  
(*Lates calcarifer*)

- Barramundi (AUS);
- Euryhaline, catadromous predator;
- Protandrous (male-first) hermaphrodite;
- Distributed over a wide geographical area;
- Cultured mostly by smaller farms, few selection programs;
- TLL-AVA collaboration: Selection program in its 9th year – first mature F2s are ready for crosses;
- Increased growth rate.
Asian seabass selection program (timeline)

- 2004-2011: marker-assisted selection program with the Yue group
- Increased growth rate – F2 grows >20% faster than unselected;
- 2011: received S$10M grant from NRF for 5 yrs;
- Genomic selection – polygenic traits;
- Nutrigenomics and disease resistance;
- Start parallel program for salt tolerant tilapia.
Outline

• Introducing the Asian seabass and the selection program;

• **Status report on the Genome (and Transcriptome) Project**;

• Early applications: phylogeography and sex change;

• Summary.
Asian seabass genome – what was known

Genome size: 700 Mb
2n = 48
Genetic linkage map (high density)
Physical map
Expected gene count: 26,000-27,000
Repeat inventory

Visit Poster 88
on the
Sequencing and assembly of a
10 Mb region of the Asian
seabass genome containing
growth-associated QTLs

Asian seabass
Genome Assembly Plan

- **Illumina HiSeq 2500**
  - 2X100 nt PE
  - 80X Coverage

- **PacBio SMRT**
  - 80 Smart Cells
  - 30X Coverage

- **BAC end sequencing**
  - 1X Coverage

- **Repeat Inventory**
- **Transcriptome**
- **QTL on LG2**
- **Mate-pair libraries (?)**
- **Pooled BAC sequencing**
- **Gap filling (?)**

**Improved High Quality Draft Genome**

Shubha Vij & team
Lok Lab, CUHK, Hong Kong
Sivasubbu Lab, IGIB, New Delhi, India
Asian seabass - Genome Sequence Datasets

**Illumina HiSeq**
- 2X100 nt PE
- 500&750 bp
- 80X coverage

**PacBio SMRT**
- ~4 kb avg read length
- ~10 Kb library
- 30X coverage

**PacBio SMRT**
- ~8 kb avg read length
- ~20 Kb library
- 60X coverage

**Sanger seq**
- Two BAC libraries
- ~120 kb insert
- 11.5K seq
- 1X coverage

A partially inbred Asian seabass individual
Genome assembly - the first try

PacBio RS
Long reads

90X de novo assembly

# contigs 3,807
N50 1.2 Mb
Maximum 18.9 Mb

Siddarth Singh, Pacific Biosystems
Mike Schatz, CSH Labs
Final genome assembly

**PacBio RS**
Long reads

- **90X** de novo assembly
- # contigs: 3,807
- N50: 1.2 Mb
- Maximum: 18.9 Mb

**Illumina HiSeq**
Short reads

- **80X** denovo assembly
- # contigs: 11,969,803
- N50: 1,001 bp
- Maximum: 26,868 bp

Siddarth Singh, Pacific Biosystems
Mike Schatz, CSH Labs

Lok Lab, CUHK, Hong Kong
Sivasubbu Lab, IGIB, New Delhi, India
Current status and improvement efforts

- Validation by 80X Illumina-reads mapping
  - 96% of reads mapped as proper pairs

- Gene annotation underway

- Generated optical map data - *de novo* assembled and used to place PacBio genome contigs
  - 79.8% similarity between the optical map and sequence assembly

<table>
<thead>
<tr>
<th></th>
<th>PacBio</th>
<th>Optical Map</th>
</tr>
</thead>
<tbody>
<tr>
<td># contigs</td>
<td>3,807</td>
<td>3,333</td>
</tr>
<tr>
<td>Contig N50</td>
<td>1.2 Mb</td>
<td>6.2 Mb</td>
</tr>
<tr>
<td>Maximum</td>
<td>18.9 Mb</td>
<td>20.6 Mb</td>
</tr>
</tbody>
</table>

- Alternative Falcon assembly underway
Transcriptome sequencing and assembly

- >1 billion reads from various organs of multiple individuals and 3 NGS platforms were assembled in a step-wise manner

<table>
<thead>
<tr>
<th>Measurement</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total length (bp)</td>
<td>262,023,963</td>
</tr>
<tr>
<td>Number of contigs</td>
<td>267,616</td>
</tr>
<tr>
<td>Number of contigs ≥1 kb</td>
<td>70,588</td>
</tr>
<tr>
<td>Max length (bp)</td>
<td>31,251</td>
</tr>
<tr>
<td>Average length (bp)</td>
<td>979</td>
</tr>
</tbody>
</table>

- >80% of the expected protein-coding loci obtained, 58% of these represented by a predicted FL-cDNA sequence

Outline

- Introducing the Asian seabass and the selection program;
- Status report on the Genome (and Transcriptome) Project;
- Early applications: phylogeography and sex change;
- Summary.
Potential benefits of sequenced genomes

- Info on coding regions;
- Info on regulatory regions & pathways;
- Comparative/evolutionary genomics;
- New platforms (chips, RNAseq, GBS, methylome, etc.);
- Re-sequencing of variants;
- Rapid identification of mutations.
Molecular evidence points to the existence of two Asian seabass species

Ward et al., J. Fish Biol. (2008); Pethiyagoda & Gill, Zootaxa (2012)
Vij et al., Frontiers Marine Science (2014)
Resequencing 65 genomes reveals clear signs of admixture in the SEA region

<table>
<thead>
<tr>
<th>Indian region</th>
<th>S-E Asia/PH</th>
<th>AU/PG</th>
</tr>
</thead>
<tbody>
<tr>
<td>India-Eastern Coast</td>
<td>KH</td>
<td>AU</td>
</tr>
<tr>
<td>India-Western Coast</td>
<td>TH-EC</td>
<td>PG</td>
</tr>
<tr>
<td>VN</td>
<td>SG</td>
<td>ID</td>
</tr>
<tr>
<td>PH</td>
<td>ID-SJ</td>
<td>ID-SU</td>
</tr>
<tr>
<td>ID-K</td>
<td>AU-D</td>
<td>ID</td>
</tr>
<tr>
<td>VN</td>
<td>SG</td>
<td>ID</td>
</tr>
<tr>
<td>PH</td>
<td>ID-SJ</td>
<td>ID-SU</td>
</tr>
<tr>
<td>ID-K</td>
<td>AU-D</td>
<td>ID</td>
</tr>
<tr>
<td>VN</td>
<td>SG</td>
<td>ID</td>
</tr>
<tr>
<td>PH</td>
<td>ID-SJ</td>
<td>ID-SU</td>
</tr>
<tr>
<td>ID-K</td>
<td>AU-D</td>
<td>ID</td>
</tr>
<tr>
<td>VN</td>
<td>SG</td>
<td>ID</td>
</tr>
<tr>
<td>PH</td>
<td>ID-SJ</td>
<td>ID-SU</td>
</tr>
<tr>
<td>ID-K</td>
<td>AU-D</td>
<td>ID</td>
</tr>
<tr>
<td>VN</td>
<td>SG</td>
<td>ID</td>
</tr>
<tr>
<td>PH</td>
<td>ID-SJ</td>
<td>ID-SU</td>
</tr>
<tr>
<td>ID-K</td>
<td>AU-D</td>
<td>ID</td>
</tr>
<tr>
<td>VN</td>
<td>SG</td>
<td>ID</td>
</tr>
<tr>
<td>PH</td>
<td>ID-SJ</td>
<td>ID-SU</td>
</tr>
<tr>
<td>ID-K</td>
<td>AU-D</td>
<td>ID</td>
</tr>
<tr>
<td>VN</td>
<td>SG</td>
<td>ID</td>
</tr>
<tr>
<td>PH</td>
<td>ID-SJ</td>
<td>ID-SU</td>
</tr>
<tr>
<td>ID-K</td>
<td>AU-D</td>
<td>ID</td>
</tr>
<tr>
<td>VN</td>
<td>SG</td>
<td>ID</td>
</tr>
<tr>
<td>PH</td>
<td>ID-SJ</td>
<td>ID-SU</td>
</tr>
<tr>
<td>ID-K</td>
<td>AU-D</td>
<td>ID</td>
</tr>
</tbody>
</table>

Andrey Yurchenko, St. Petersburg, Russia; unpubl.
Asian seabass (*Lates calcarifer*)

- Problem: Protandrous (male-first) hermaphrodite;
- Zebrafish can be used as a model to understand sex change.

Natural sex reversal in Asian seabass

Unusual individuals: ‘reluctant’ males and primary females

Array-based transcriptomics: The transforming gonad returns to a near-undifferentiated stage before initiating feminization

Summary – Asian seabass Genome

- Genome is sequenced (170X);
- Assembly is based on PacBio data, Illumina validates;
- Optical mapping showed improvements;
- International consortium for annotation;
- Multiple benefits;
- (Seq and assembly of Mozambique tilapia genome is in progress.)
## Genome of the Mozambique tilapia

C-value (pg): 0.81-1.0 (Animal Genome Size Database)
Diploid Chrom number (n): 44
Genome Size: ca. 1 Gb

- Different approach from the ASB
  - 6 Short-insert PE (SIPE) – 60X
  - 3 Long-insert MP (1, 2, 3 kb) – 30X
  - Lucigen Long-insert MP (8kb) – 3X
  - Lucigen Fosmid MP (40kb) – 3X
  - Pacbio sequence data pending

- De novo assembly is underway
  - V1: SIPE-only assembly by MaSuRCA

<table>
<thead>
<tr>
<th>Assembly size</th>
<th>964 Mb</th>
</tr>
</thead>
<tbody>
<tr>
<td># of contigs</td>
<td>173,146</td>
</tr>
<tr>
<td>Max length</td>
<td>257.3 kb</td>
</tr>
<tr>
<td>N50 length</td>
<td>15.3 kb</td>
</tr>
</tbody>
</table>

Liew Woei Chang, Shen Xueyan & team
RGG:
Shubha Vij
Inna Kuznetcova
Woei Chang Liew
Xueyan Shen
Natascha May
Prakki Sri Datta
Jolly M. Saju
Purushothaman K.
Shawn Ngoh
Pranjali Bhandare

Farm teams
Genhua Yue & team
Huan Sein Lim & team

TLL Facilities

RGG Alumni:
Preethi Ravi
Rajini Sreenivasan
Laura Casas
Alex Chang
Mohd. Sorowar
Hossain
Xingang Wang
Richard Bartfai
Keh-Weei Tzung
Hsiao Yuen Kwan
Doreen Lau
Junhui Jiang

Collaborators:
Alan Christoffels (Capetown)
Si Lok (HK)
Matt Bellgard (Murdoch)
Mike Shatz (CSH)
PacBio
Aleks Komissarov & Andrey Yurchenko (St. Petersburg)
Heiner Kuhl (Berlin)
Dean Jerry (Queensland)
...

Funding:
National Research Foundation
AVA
EDB
MND
TEMASEK LIFESCIENCES LABORATORY
An outlook for the future

Past: MAS on seabass

Present: Genomic selection on A. seabass & tilapia

Near future: Integrated systems applied to several fish species

Photo: http://faithoncampus.com/