

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 Pannonia, Keszthely, Hungary;

More people, less fish

Population 7 billion



Depleted oceans



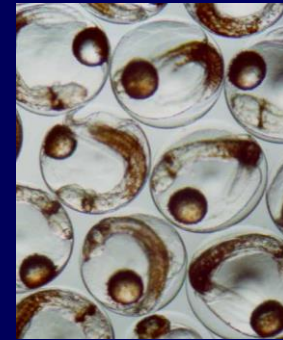
Source: National Geographic
endoftheline.com

Outline

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

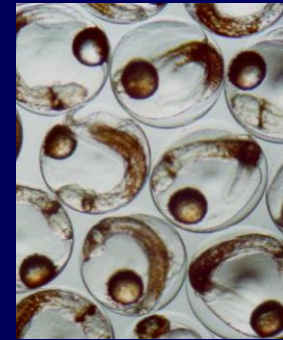
Asian seabass (*Lates calcarifer*)

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



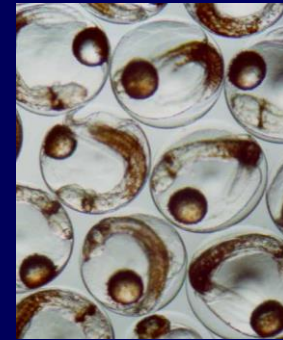
Asian seabass (*Lates calcarifer*)

- 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

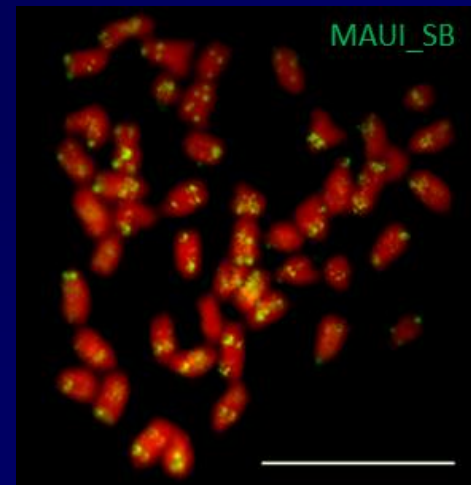
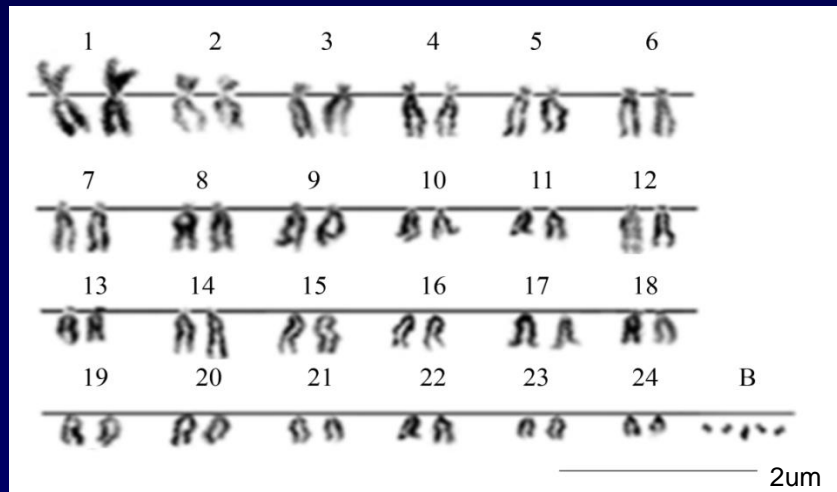
$2n = 48$

Genetic linkage map (high density)

Physical map

Expected gene count: 26,000-27,000

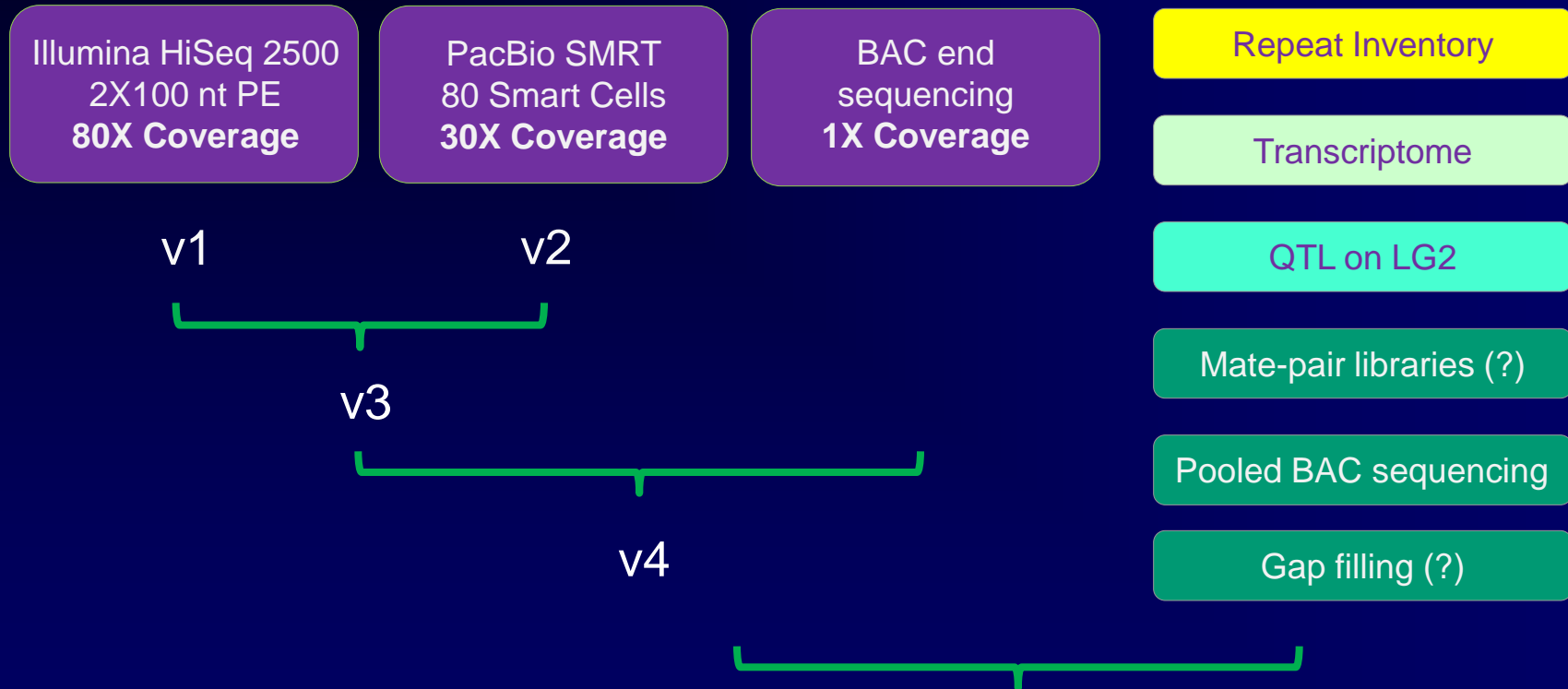
Repeat inventory



Asian seabass Genome Assembly Plan

Visit Poster 88

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



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



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

Siddarth Singh, Pacific Biosystems
Mike Schatz, CSH Labs

Illumina HiSeq Short reads

80X

denovo
assembly

# contigs	11,969,803
N50	1,001 bp
Maximum	26,868 bp

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

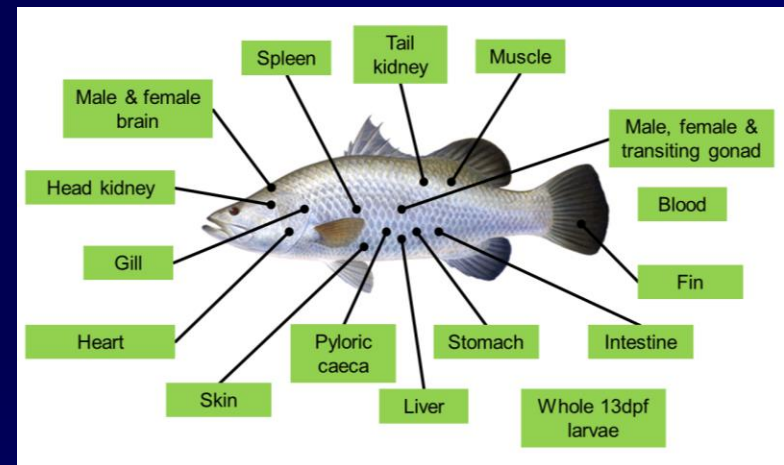
	PacBio	Optical Map
# contigs	3,807	3,333
Contig N50	1.2 Mb	6.2 Mb
Maximum	18.9 Mb	20.6 Mb

- 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

Total length (bp)	262,023,963
Number of contigs	267,616
Number of contigs ≥ 1 kb	70,588
Max length (bp)	31,251
Average length (bp)	979



- >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.

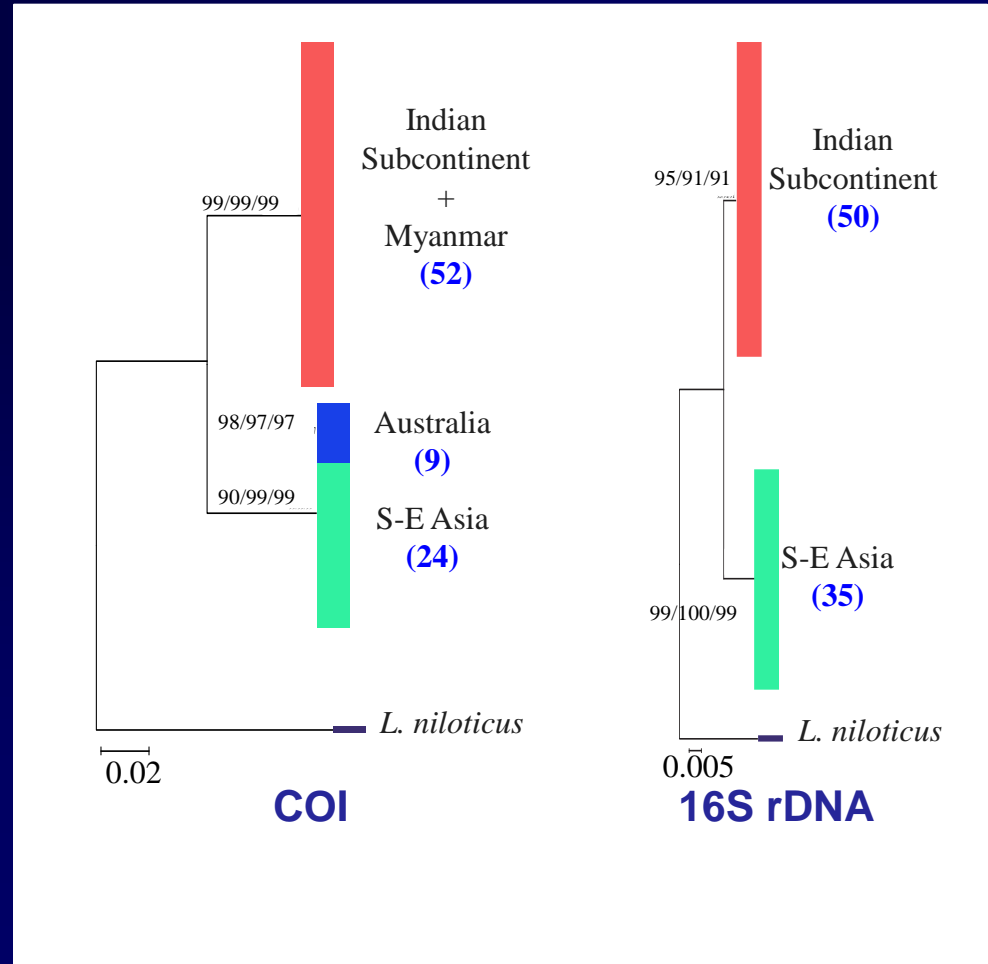
Potential benefits of sequenced genomes

- 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.

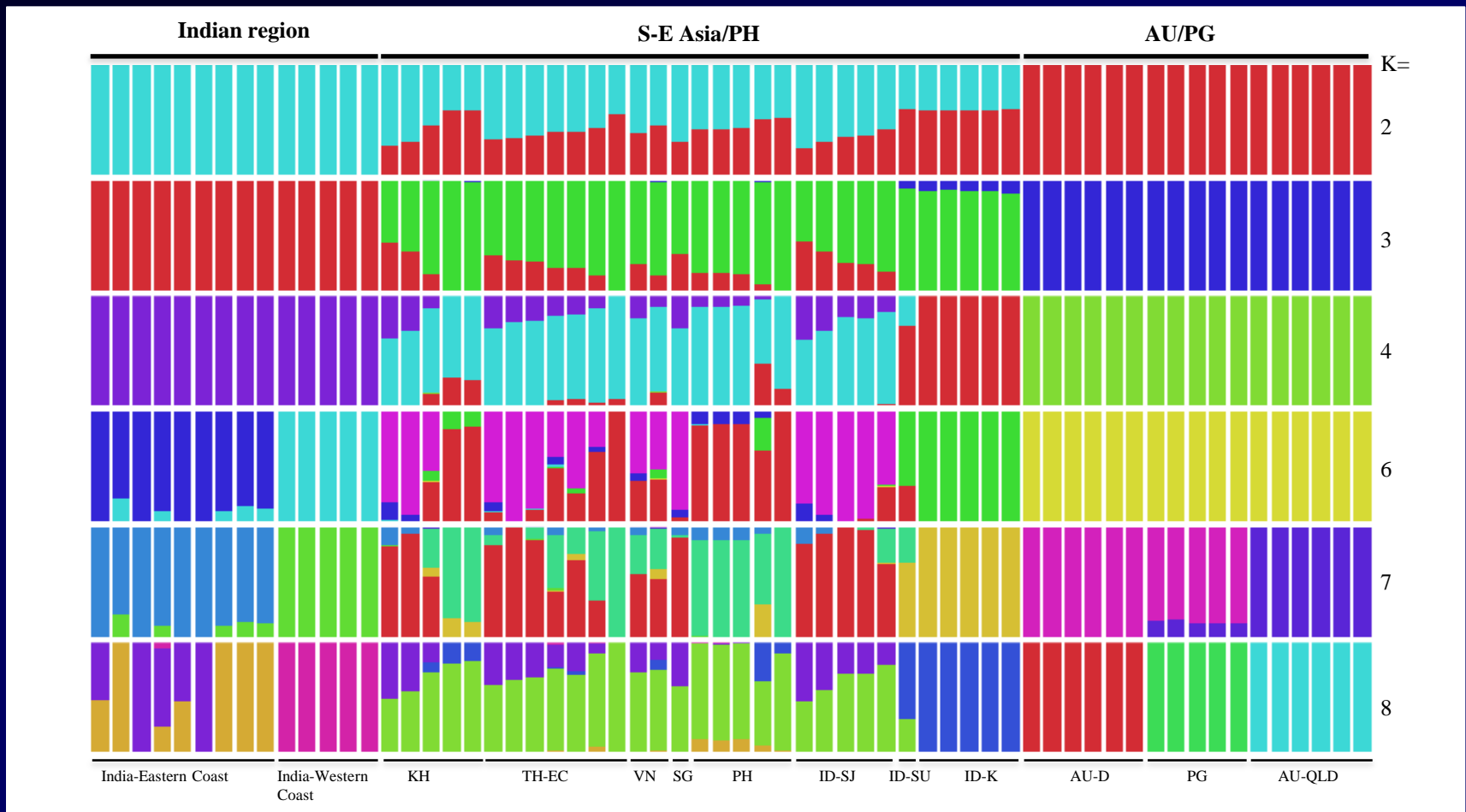
Molecular evidence points to the existence of two Asian seabass species



Ward et al., J. Fish Biol. (2008);
Pethiyagoda & Gill, Zootaxa (2012)
Vij et al., Frontiers Marine Science (2014)



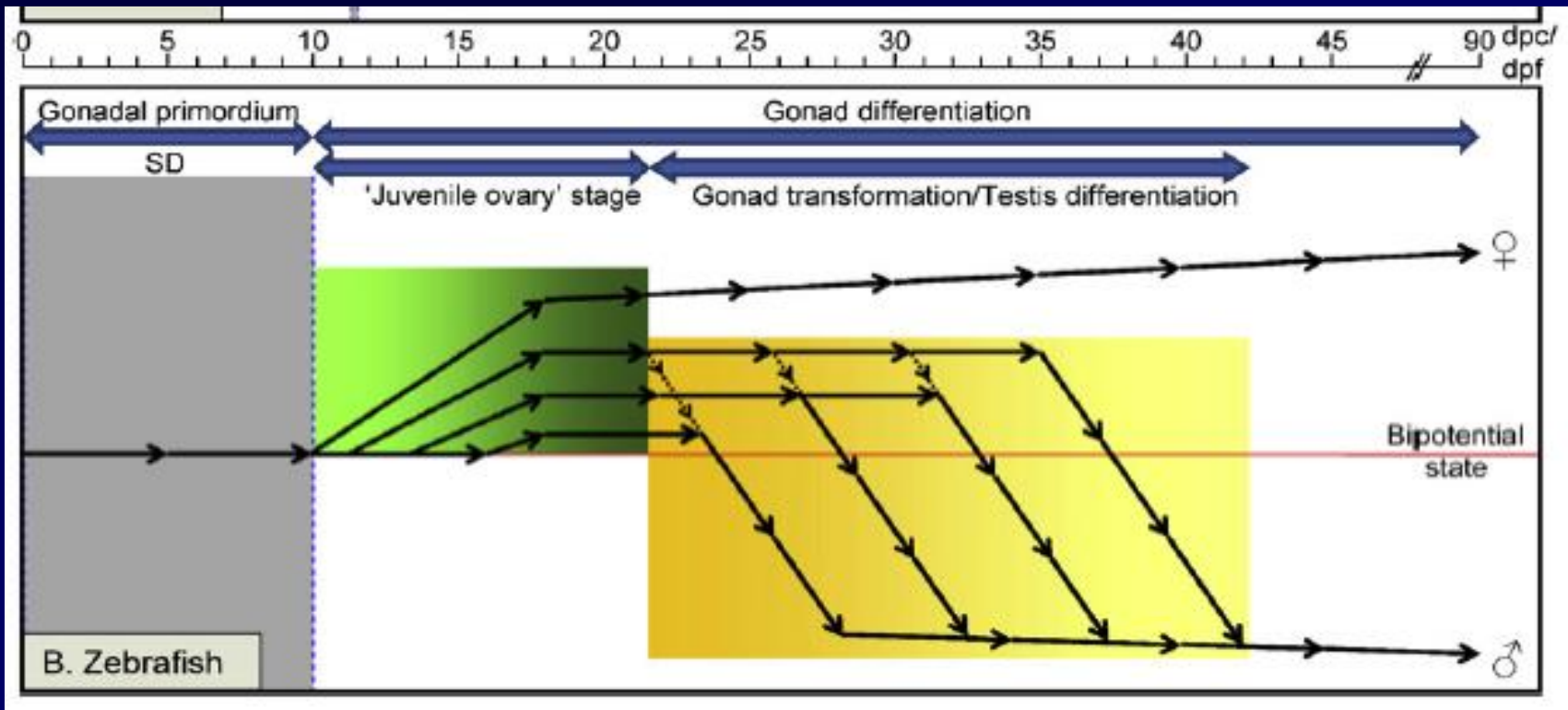
Resequencing 65 genomes reveals clear signs of admixture in the SEA region



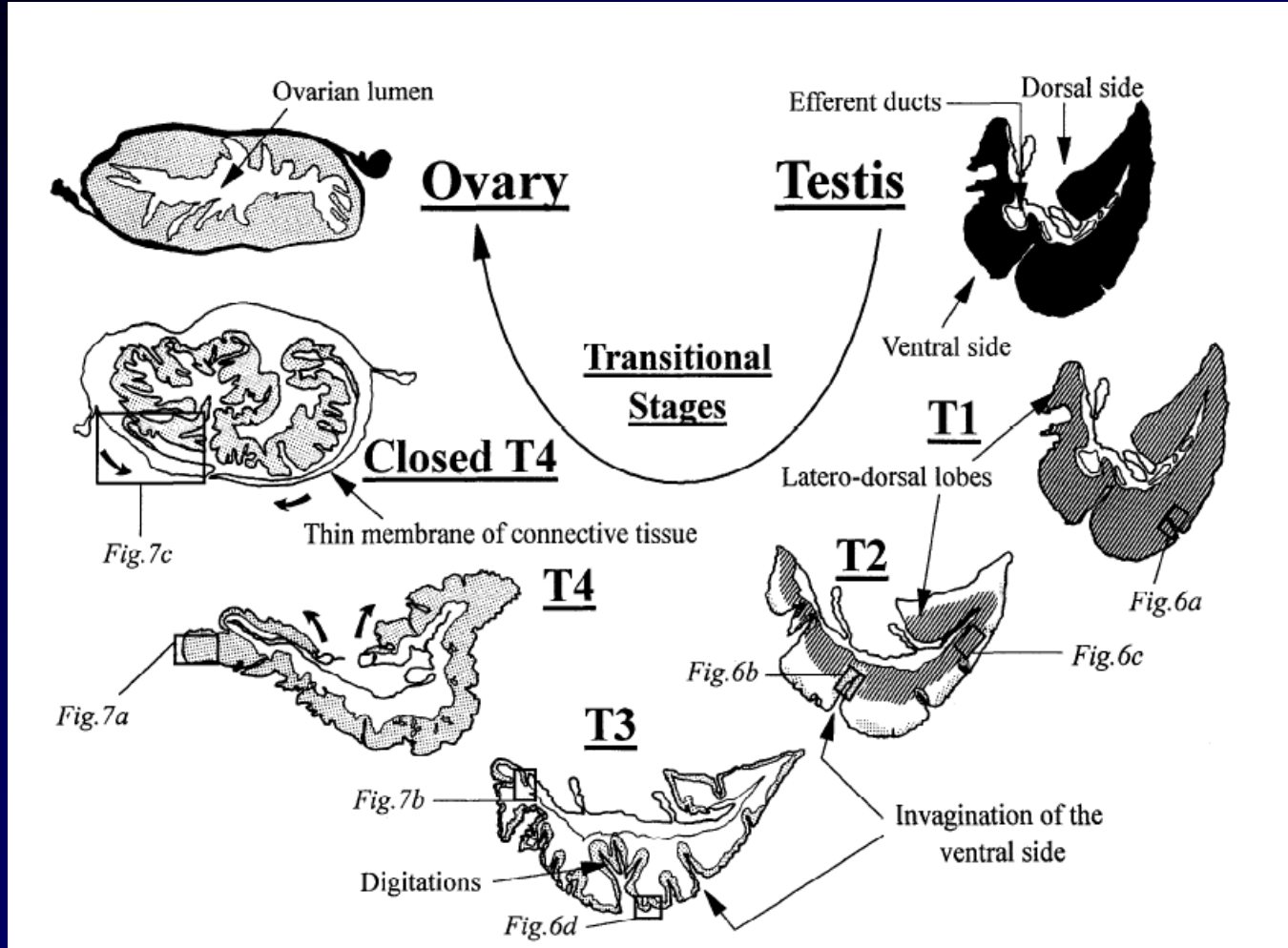
Andrey Yurchenko, St.
Petersburg, Russia; unpubl.

Asian seabass (*Lates calcarifer*)

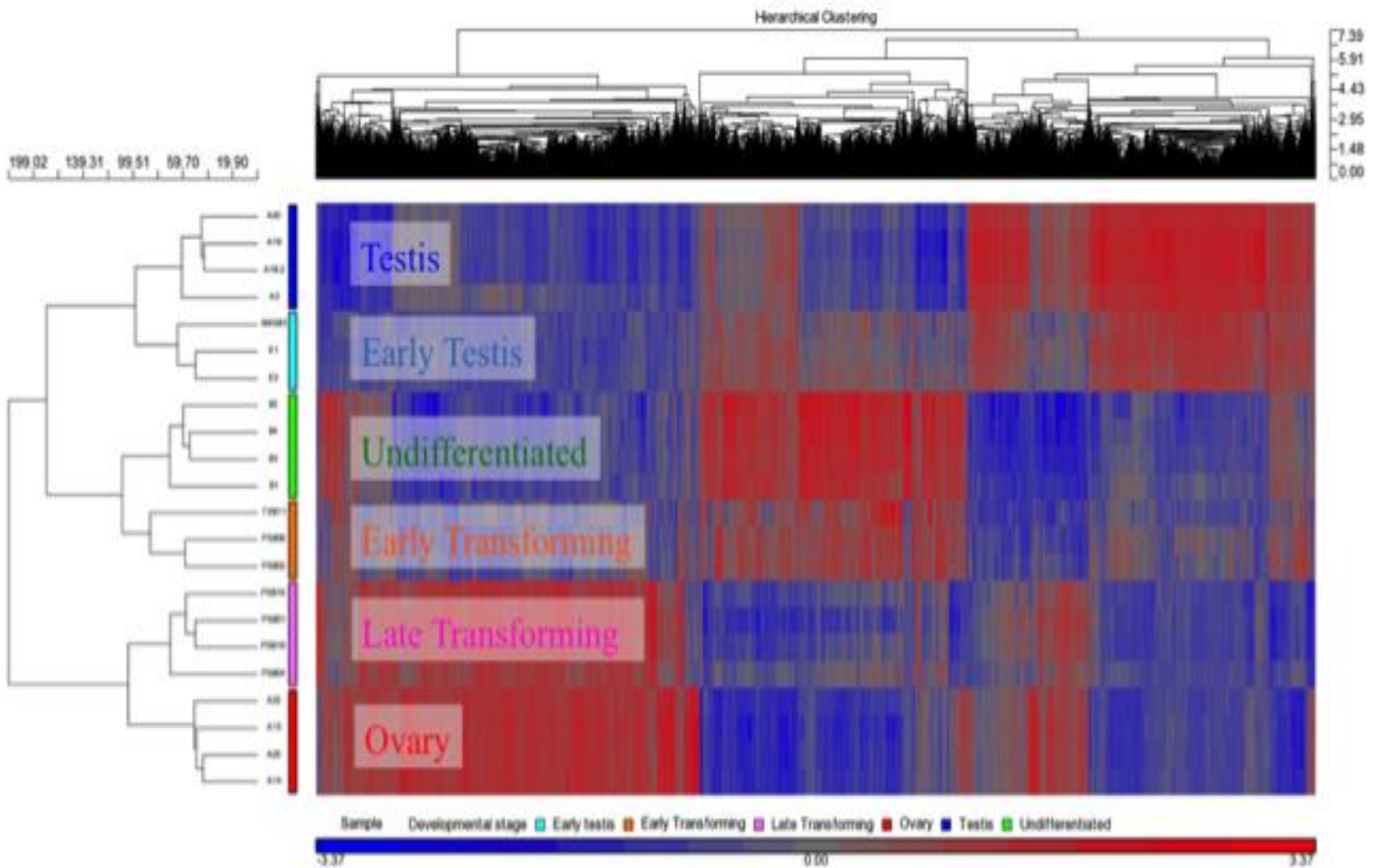
- Problem: Protandrous (male-first) hermaphrodite;
- Zebrafish can be used as a model to understand sex change.



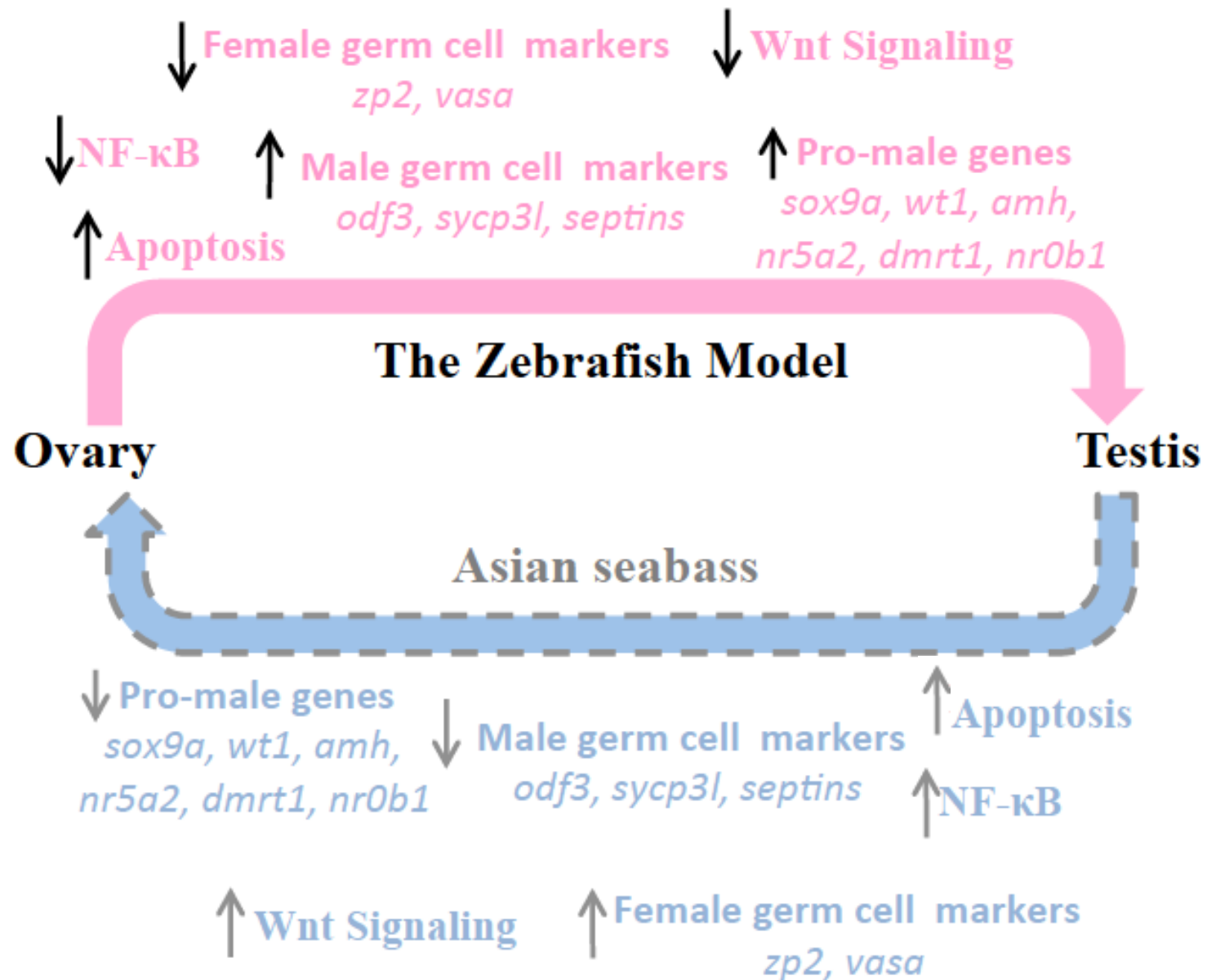
Natural sex reversal in Asian seabass



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

Assembly size	964 Mb
# of contigs	173,146
Max length	257.3 kb
N50 length	15.3 kb

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:



Near future:
Integrated systems applied to
several fish species

Present:
Genomic selection on
A. seabass & tilapia

Past: MAS on
seabass