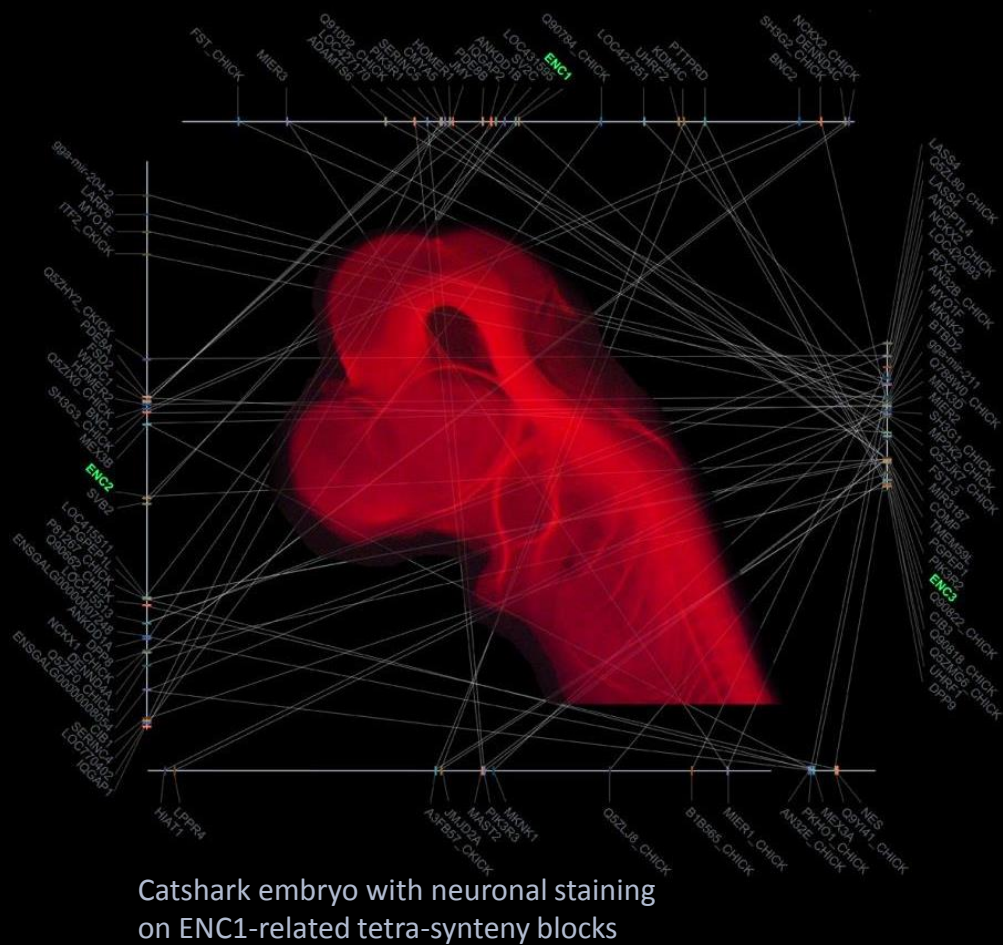


Assessing vertebrate homolog space ... from Evo-Devo viewpoint



Shigehiro Kuraku

Unit Leader

Phyloinformatics Unit

RIKEN CLST (Twitter: @clst_gras)

Adjunct Associate Prof.

Kwansei Gakuin University

Adjunct Associate Prof.

Kobe University

NGS & Phyloinfo in Kobe

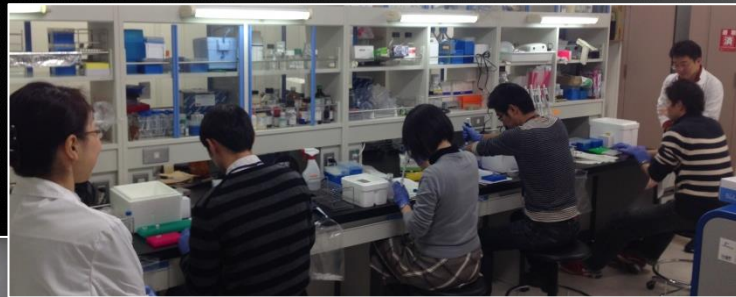
More than just a facility

- Strength in ChIP-seq/ChIP-exo & non-model vertebrates
- Experienced with limited samples
- Cost-saving protocols (e.g. iMate Protocol, Tatsumi *et al.*, 2015)
- Flexible project consultation for biologists

Illumina HiSeq 1500



Orbitrap Velos Pro for LC-MS/MS



NGS Library Prep Hands-on tutorial
January 20/21, 2015



Data analysis tutorial 'Sequence Informatics Afternoon'
April 23, 2014

- Hands-on tutorials (ChIP, data analysis, ...)
- Organizing meetings etc.
- Facility management consultation

Biological scope

Taxon: Vertebrata

Genes: regulatory genes (TFs, signaling ligands)

What in genome allowed characters unique to vertebrates?

e.g., organized brain, visceral organs, fins/limbs,

What molecular change triggered morphological evolution
in vertebrates?

Turtle shell



Lamprey head



Hagfish embryo



*Ota, Kuraku, & Kuratani
2007. Nature*

<https://whyevolutionistrue.wordpress.com>

<http://gallery.nanfa.org/>

Orthology: thoughts and concerns

Experience through collaborations with non-experts

‘Conservation of toolkit’ overtrusted
(Sean Carroll, 2005. etc.)

Is ‘Ortholog’ a jargon?

Tree-based methods underappreciated (cf. synteny)

Why?

Inaccessibility to experts’ instructions

No established standard method for tree inference

Sequence information scattered in many DBs

Tolerant peer review system for phylogenetics

Orthology: expectations!!



Phylogeny to phenotypic evolution

⋮

Inter-family co-evolution

e.g., massive loss of embryonic axis formation genes in mammals

Kuraku & Kuratani., 2011. GBE

Application of phylogeny-based approaches

Phylogeny-based genome scaffolding

e.g., ESPRIT *Dessimoz et al., 2011. Brief. Bioinfo.*

Agenda

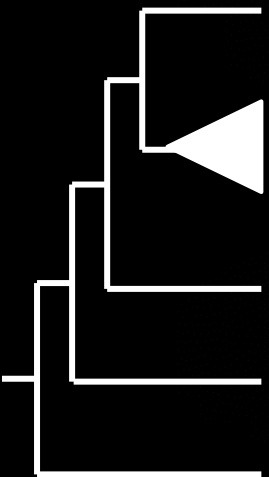
- 1) Revealing non-conservation of developmental 'toolkits'
'Cryptic pan-vertebrate genes'
=> markers for quality assessment of orthology predictions?
- 2) 'Two-round' whole genome duplications
Warning in using lamprey genomes
'lamprey dialect' misleading phylogeny inference?
access to less incomplete predicted gene set

Cryptic pan-vertebrate genes analyzed by my team

Hox14 – *Feiner et al, 2011. J Exp Zool B & Kuraku et al, 2008. PNAS*

Bmp16 – *Feiner et al, 2008. BMC Evol Biol*

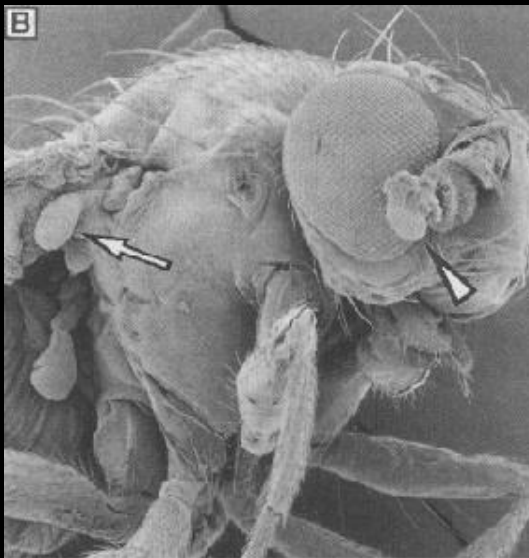
Pax4 & Pax10 – *Feiner et al, 2014. GBE & Manousaki et al, 2011. Evol Dev*



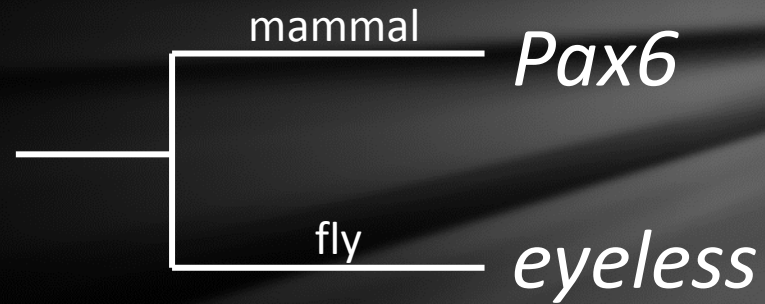
Taxon	<i>Hox14</i>	<i>Bmp16</i>	<i>Pax4</i>	<i>Pax10</i>
Mammals	–	–	+	–
Birds	–	–	–	–
Reptiles	–	+	+	+
Amphibians	–	–	–	+
Actinopterygian fish	–	+	+	+
Cartilaginous fish	+	+	?	+

Pax6-*eyeless* orthology

Ectopic eyes in fly



Halder *et al.*, 1995. *Science*



Limitation of conventional methods

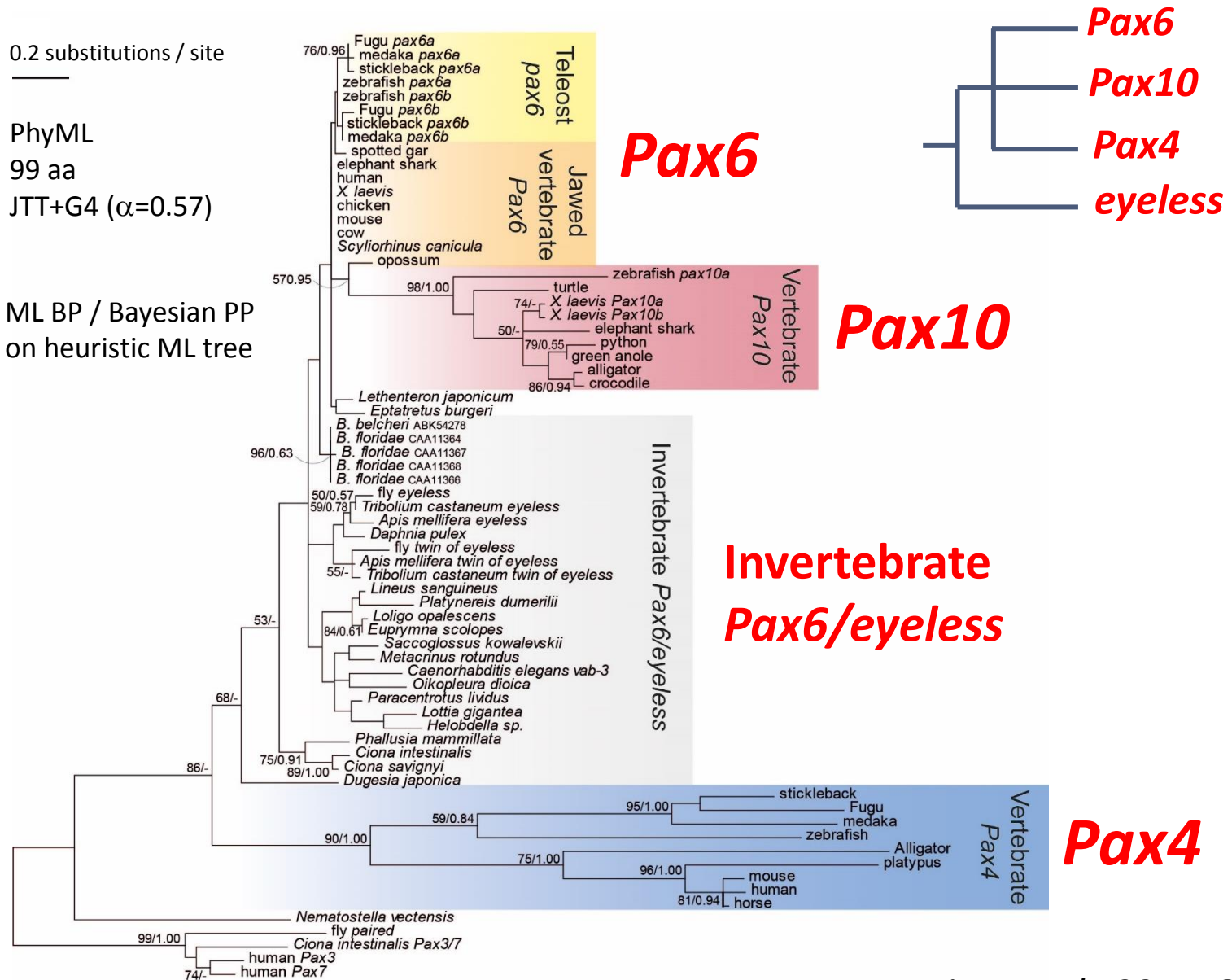
0.2 substitutions / site

PhyML

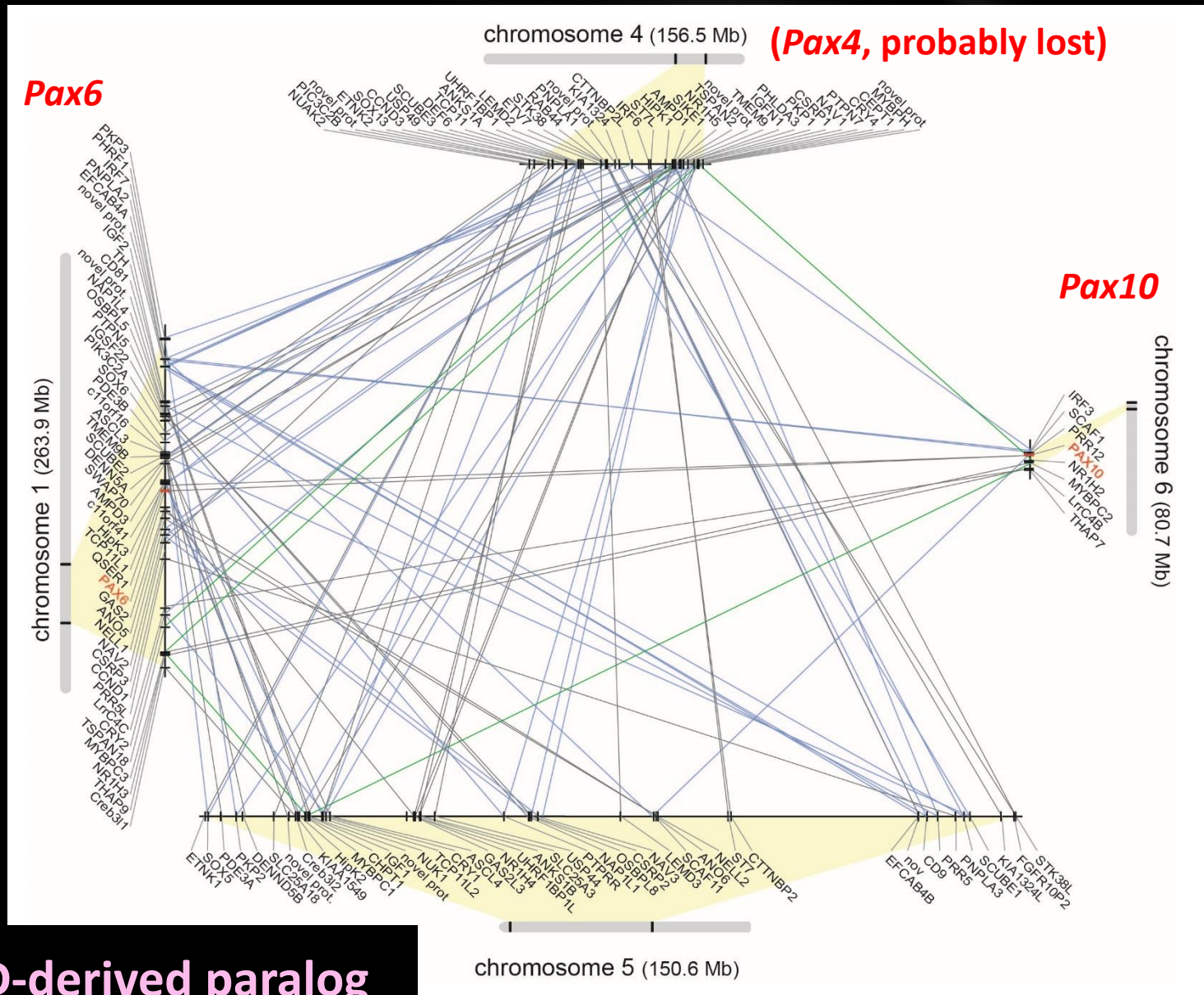
99 aa

JTT+G4 ($\alpha=0.57$)

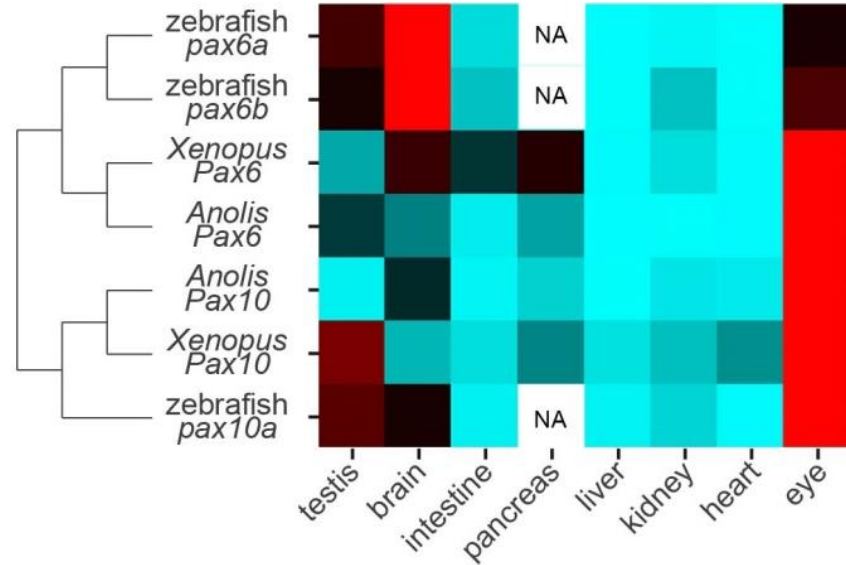
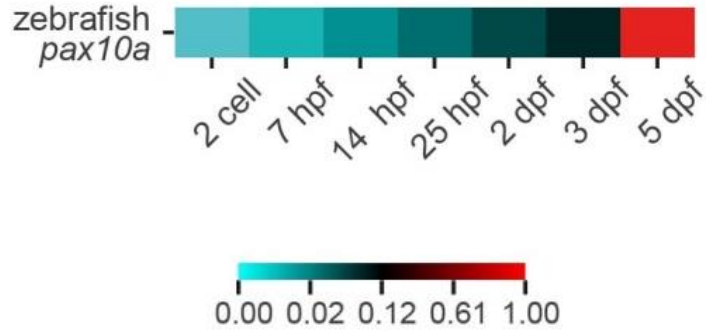
ML BP / Bayesian PP
on heuristic ML tree



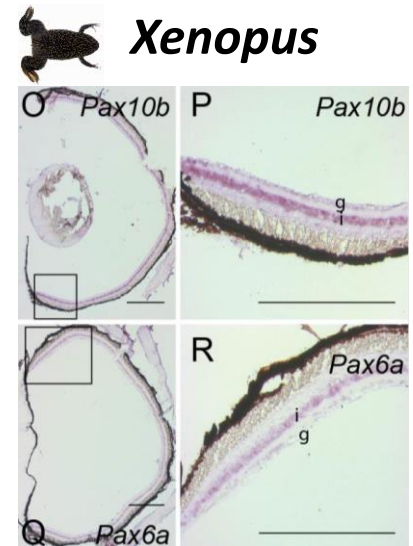
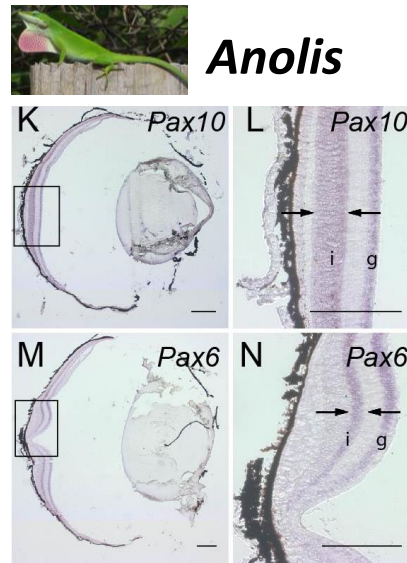
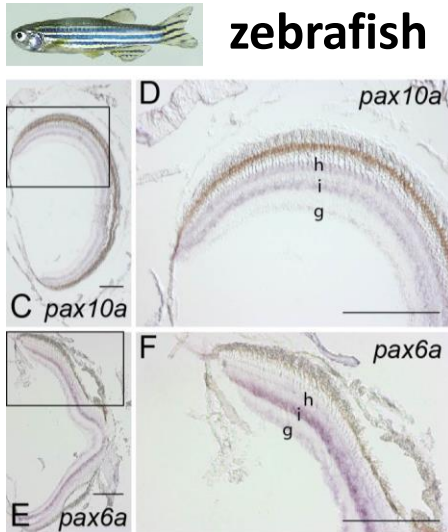
Conserved synteny in anole lizard genome



Pax10, not a useless gene



Pax10
Pax6

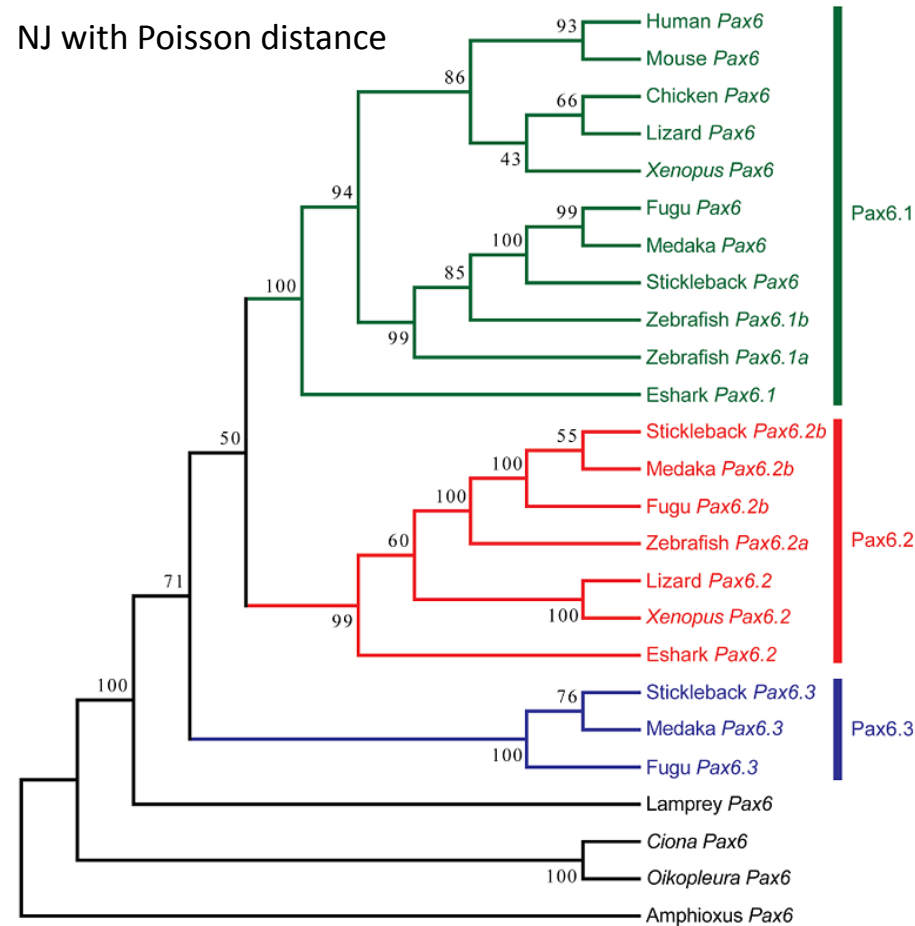


Sequencing of *Pax6* Loci from the Elephant Shark Reveals a Family of *Pax6* Genes in Vertebrate Genomes, Forged by Ancient Duplications and Divergences

Vydianathan Ravi^{1,2}, Shipra Bhatia^{2,3}, Philippe Gautier², Felix Loosli³, Boon-Hui Tay¹, Alice Tay¹, Emma Murdoch², Pedro Coutinho², Veronica van Heyningen², Sydney Brenner¹, Byrappa Venkatesh^{1,*}, Dirk A. Kleinjan^{2,*}

¹Institute of Molecular and Cell Biology, Agency for Science Technology and Research (A*STAR), Biopolis, Singapore, Singapore, ²MRC Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, United Kingdom, ³Institute of Toxicology and Genetics, Karlsruhe Institute of Technology, Eggenstein-Leopoldshafen, Germany

NJ with Poisson distance



Evolution of the Vertebrate Pax4/6 Class of Genes with Focus on Its Novel Member, the Pax10 Gene

Nathalie Feiner^{1,2,3}, Axel Meyer^{1,2}, and Shigehiro Kuraku^{1,2,4,*}

¹Department of Biology, University of Konstanz, Germany

²International Max-Planck Research School (IMPRS) for Organismal Biology, University of Konstanz, Germany

³Present address: Department of Zoology, University of Oxford, United Kingdom

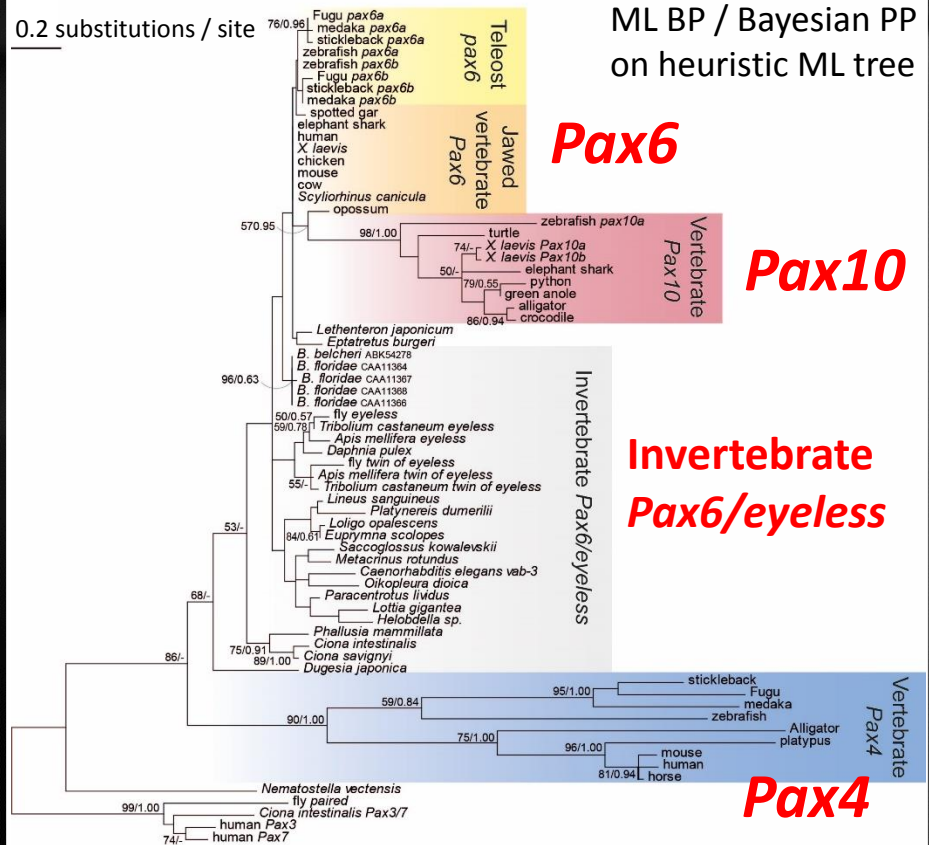
⁴Present address: Genome Resource and Analysis Unit, RIKEN Center for Developmental Biology, Chuo-ku, Kobe, Hyogo, Japan

*Corresponding author: E-mail: shigehiro-kuraku@cdb.riken.jp.

Accepted: June 14, 2014

Data deposition: This project has been deposited at EMBL under the accessions HF567444–HF567455.

0.2 substitutions / site



Sequencing of *Pax6* Loci from the Elephant Shark Reveals a Family of *Pax6* Genes in Vertebrate Genomes, Forged by Ancient Duplications and Divergences

Vydianathan Ravi^{1,3}, Shipra Bhatia^{2,3}, Philippe Gautier², Felix Loosli³, Boon-Hui Tay¹, Alice Tay¹, Emma Murdoch², Pedro Coutinho², Veronica van Heyningen², Sydney Brenner¹, Byrappa Venkatesh^{1*}, Dirk A. Kleinjan^{2*}

¹Institute of Molecular and Cell Biology, Agency for Science Technology and Research (A*STAR), Biopolis, Singapore, Singapore, ²MRC Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, United Kingdom, ³Institute of Toxicology and Genetics, Karlsruhe Institute of Technology, Eggenstein-Leopoldshafen, Germany

Evolution of the Vertebrate *Pax4/6* Class of Genes with Focus on Its Novel Member, the *Pax10* Gene

Nathalie Feiner^{1,2,3}, Axel Meyer^{1,2}, and Shigehiro Kuraku^{1,2,4,*}

¹Department of Biology, University of Konstanz, Germany

²International Max-Planck Research School (IMPRS) for Organismal Biology, University of Konstanz, Germany

³Present address: Department of Zoology, University of Oxford, United Kingdom

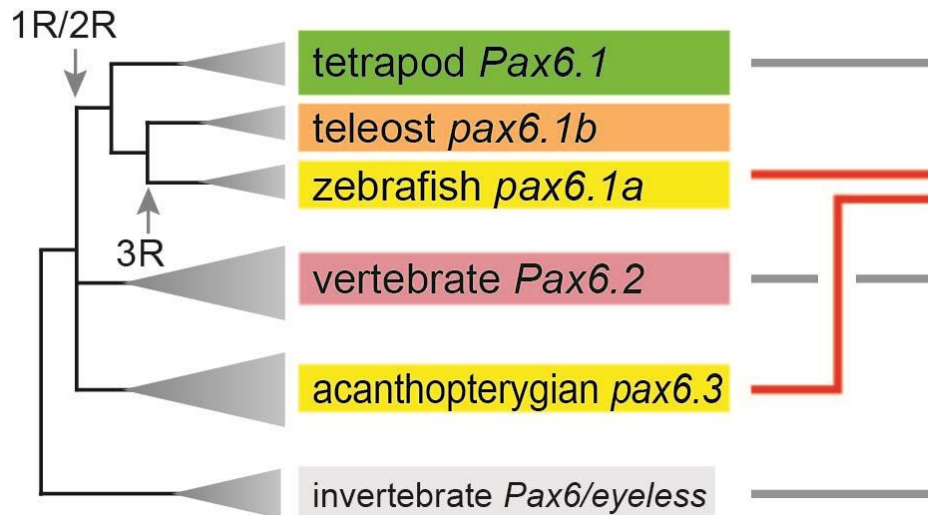
⁴Present address: Genome Resource and Analysis Unit, RIKEN Center for Developmental Biology, Chuo-ku, Kobe, Hyogo, Japan

*Corresponding author: E-mail: shigehiro-kuraku@cdb.riken.jp.

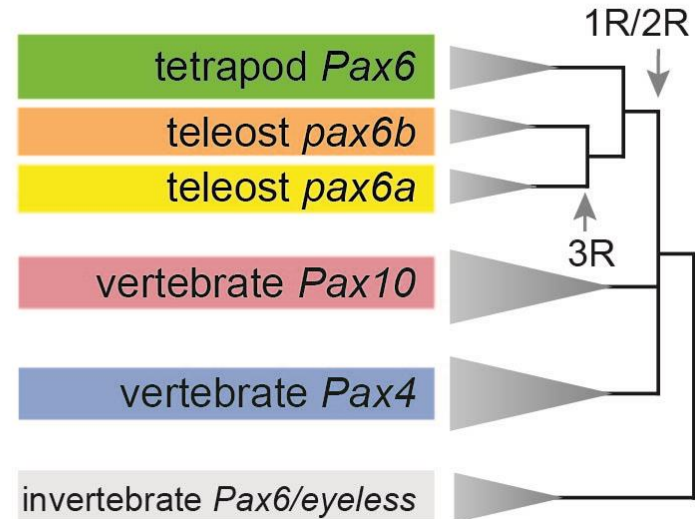
Accepted: June 14, 2014

Data deposition: This project has been deposited at EMBL under the accessions HF567444–HF567455.

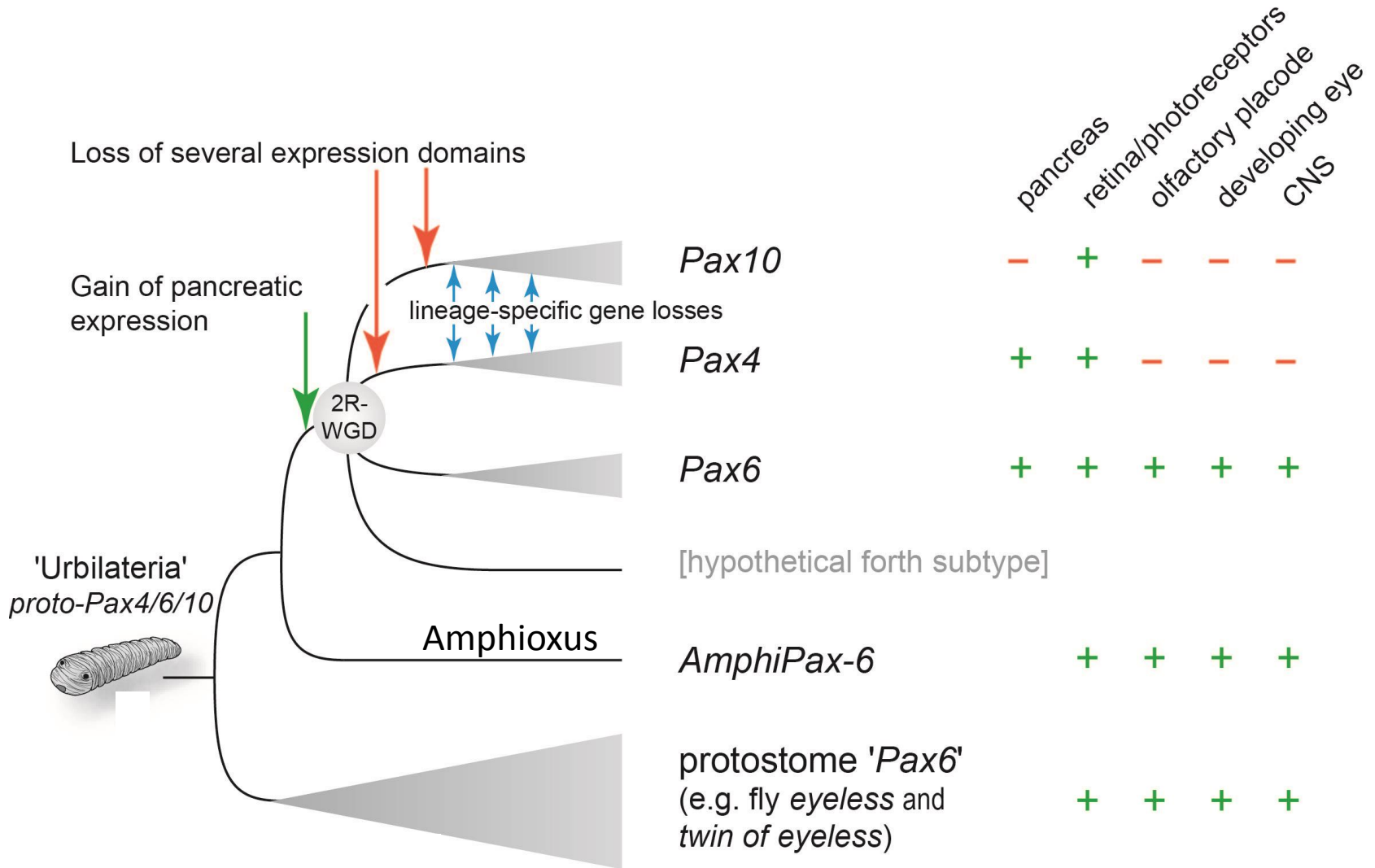
Hypothesis 1: (Ravi et al., 2013)



Hypothesis 2: (our study)

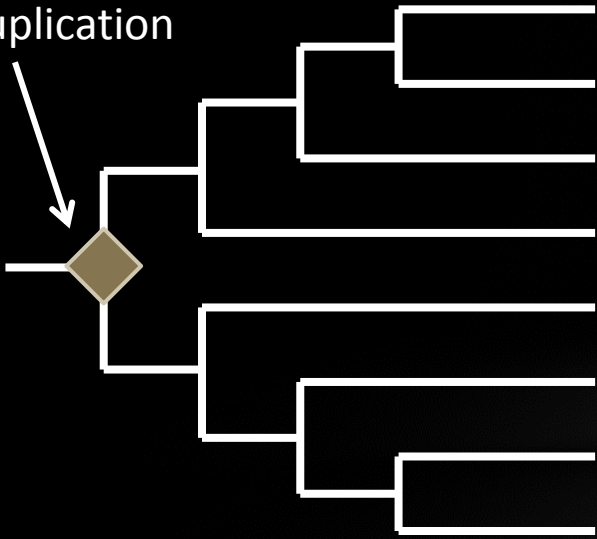


Revised *Pax6*-*eyeless* relationship

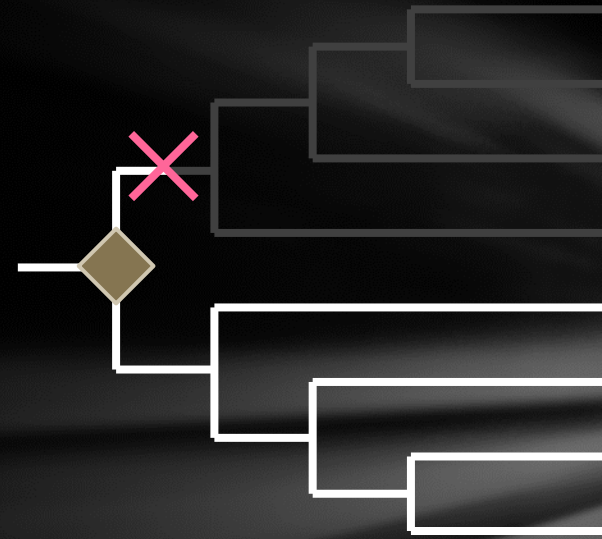


No loss

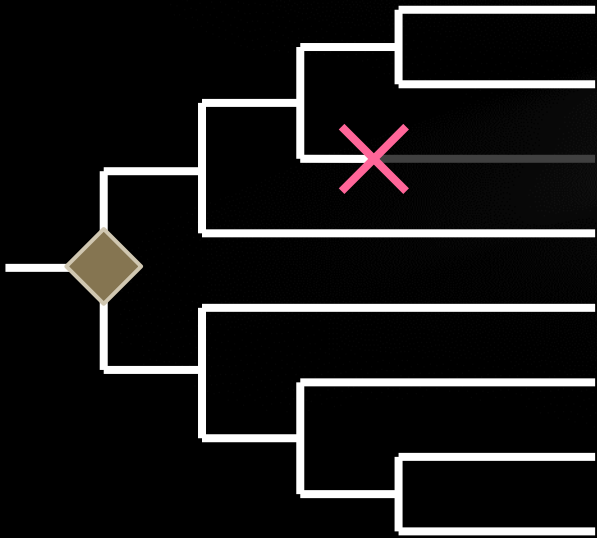
gene duplication



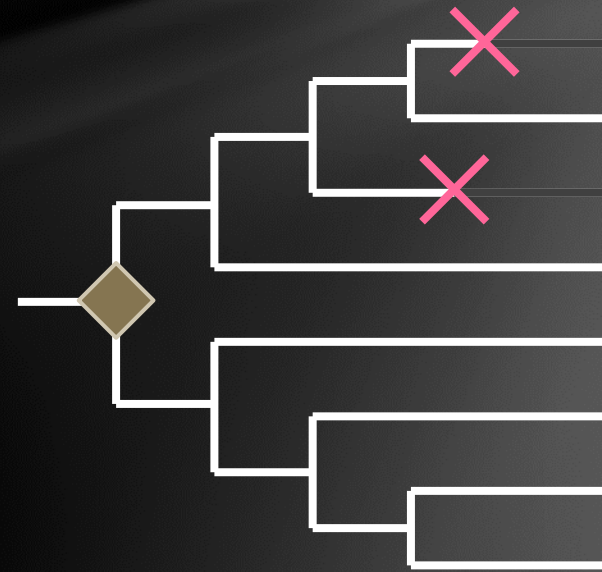
Loss just after duplication



Loss long after duplication

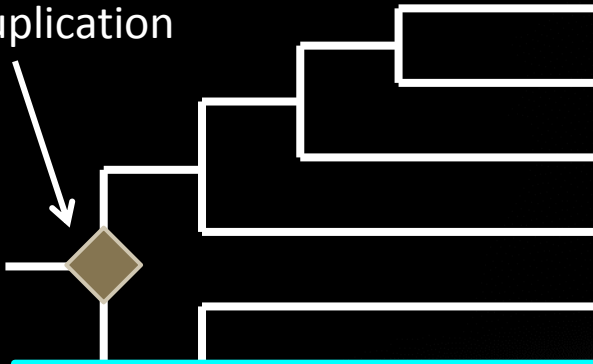


Multiple losses



No loss

gene duplication

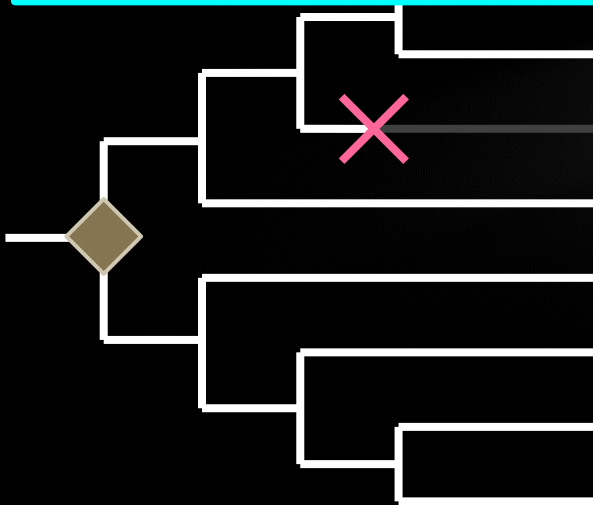


Loss of canonical functions/expressions

Hox14 – axial elements (somites, nt)

Bmp16 – DV-axis formation

Pax4 & *Pax10* – CNS



Loss just after duplication



Multiple losses

Agenda

1) Revealing non-conservation of developmental 'toolkits'
'Cryptic pan-vertebrate genes'

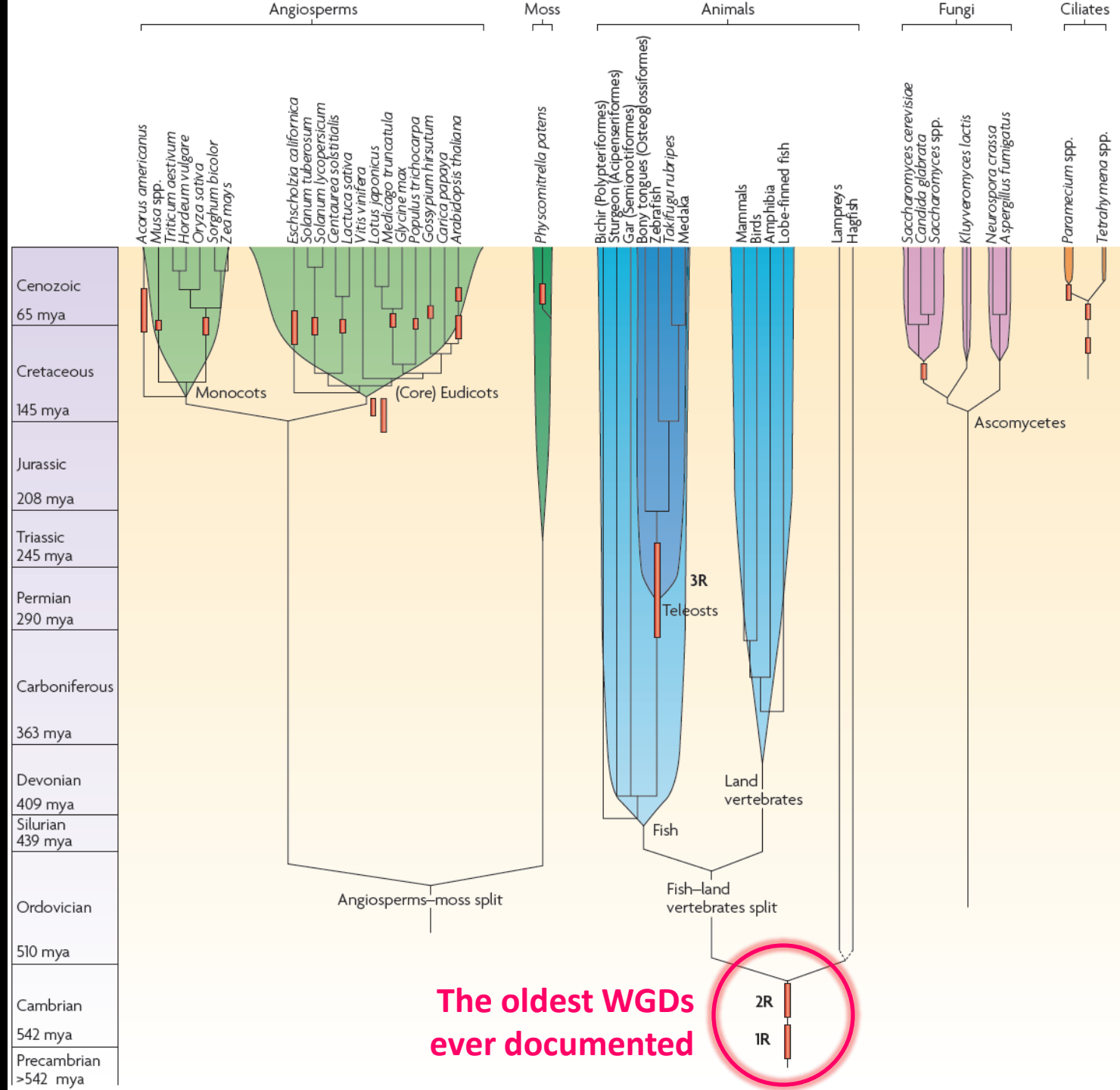
2) 'Two-round' whole genome duplications

Warning in using lamprey genomes

'lamprey dialect' misleading phylogeny inference?

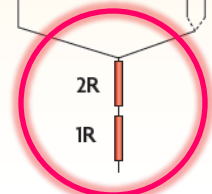
access to less incomplete predicted gene set

Whole genome duplications (WGDs)



Van de Peer, Maere,
 & Meyer. 2009.
 Nat Rev Genet

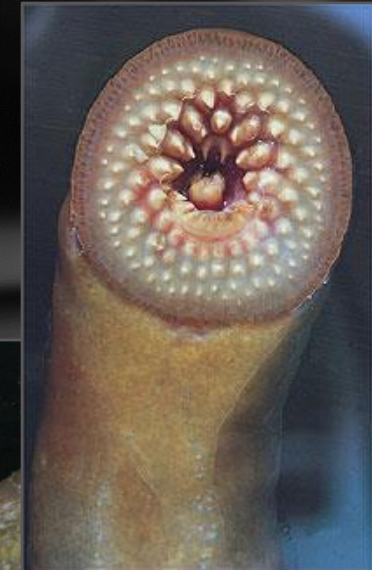
The oldest WGDs
 ever documented



Cyclostomata (hagfishes and lampreys)

Japanese lamprey
Lethenteron japonicum

sea lamprey
Petromyzon marinus

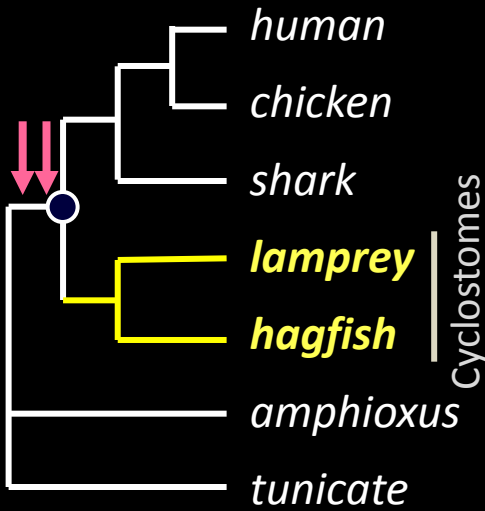


Inshore hagfish
Eptatretus burgeri

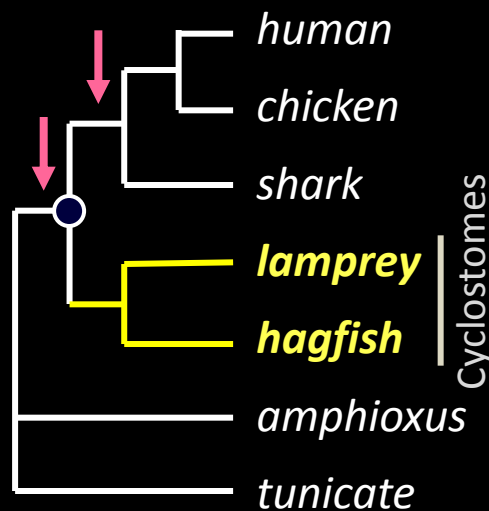
By K.G. Ota, S. Kuraku and S. Kuratani

Timing of two-round (2R) WGDs

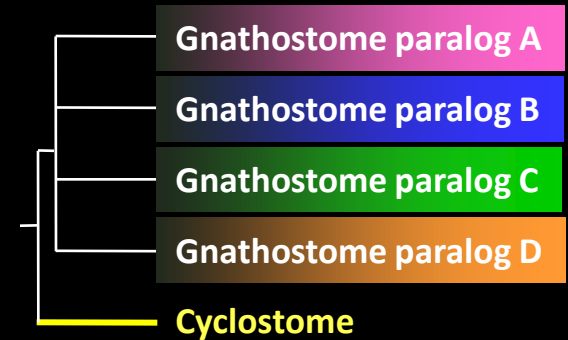
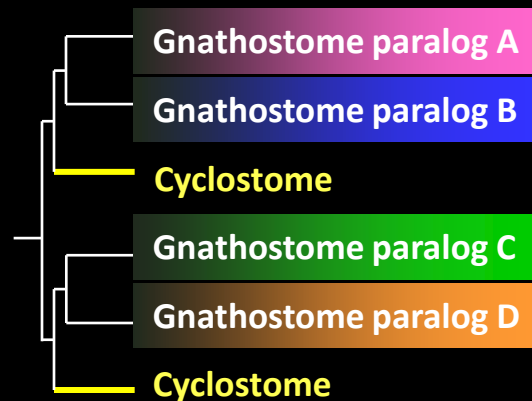
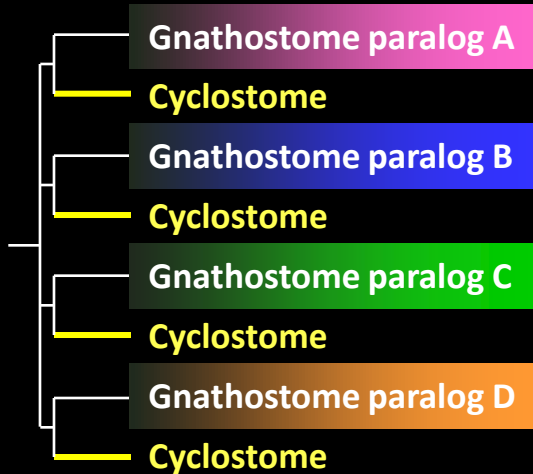
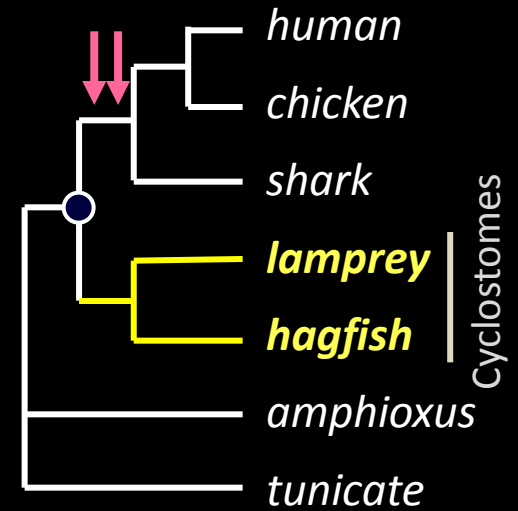
Hypothesis A



Hypothesis B



Hypothesis C



Timing of two-round (2R) WGDs

Firm evidence of '2R' from tetra-syteny blocks

Larhammar et al., 2002. Genome Res., etc.

Independent analyses under different methodologies

2009 Integrative analysis (55 gene families)

Kuraku et al., 2009. Mol. Biol. Evol.

2013 Two lamprey genomes

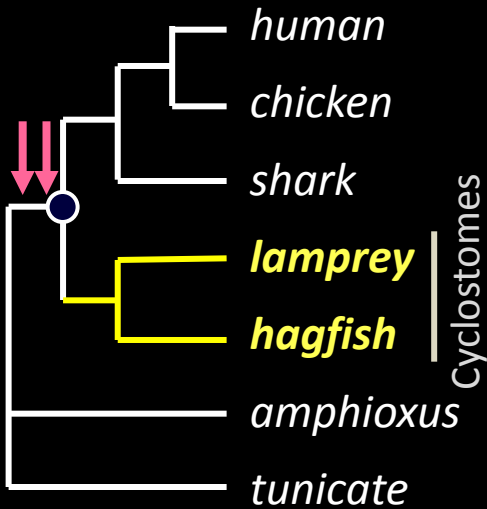
Petromyzon marinus (Smith, Kuraku, et al., 2013. Nature Genet)

Lethenteron japonicum (Mehta et al., 2013. PNAS)

↓ ? Phylome analysis with dense taxon-sampling ?

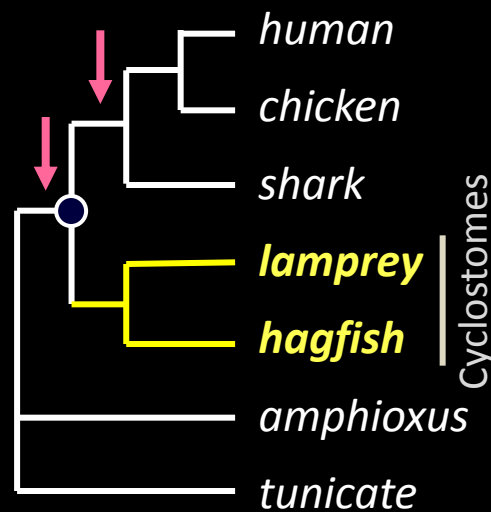
Timing of two-round (2R) WGDs

Hypothesis A



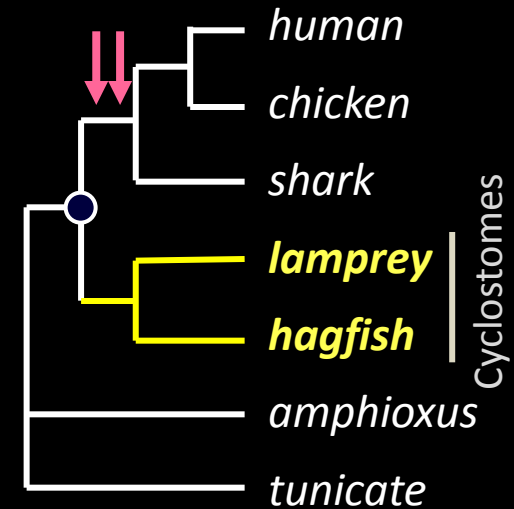
- Mol. phylogeny of 55 gene families
Kuraku et al., 2009. MBE
- Globin gene phylogeny
Hoffmann et al., 2010. PNAS
- Sea lamprey genome analysis
Smith, Kuraku et al., 2013. Nature Genetics

Hypothesis B



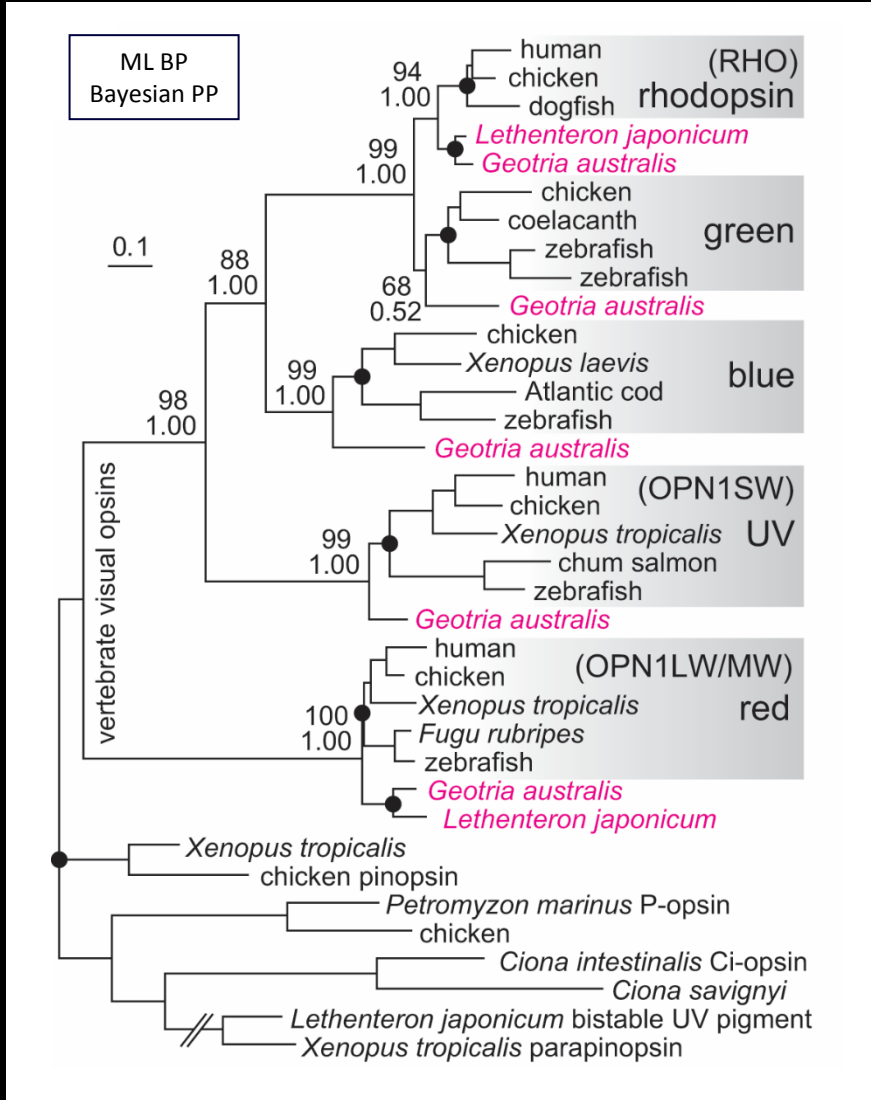
- Composition of Hox/Dlx clusters
Neidert et al., 2001. PNAS
Irvine et al., 2002. J Exp Zool B
Force et al., 2002. J Exp Zool B etc
- Mol. phylogeny of 33 gene families
Escriva et al., 2002. MBE
- Amphioxus genome
Putnam et al., 2008. Nature

Hypothesis C



- ParaHox clusters
Furlong et al., 2007. MBE

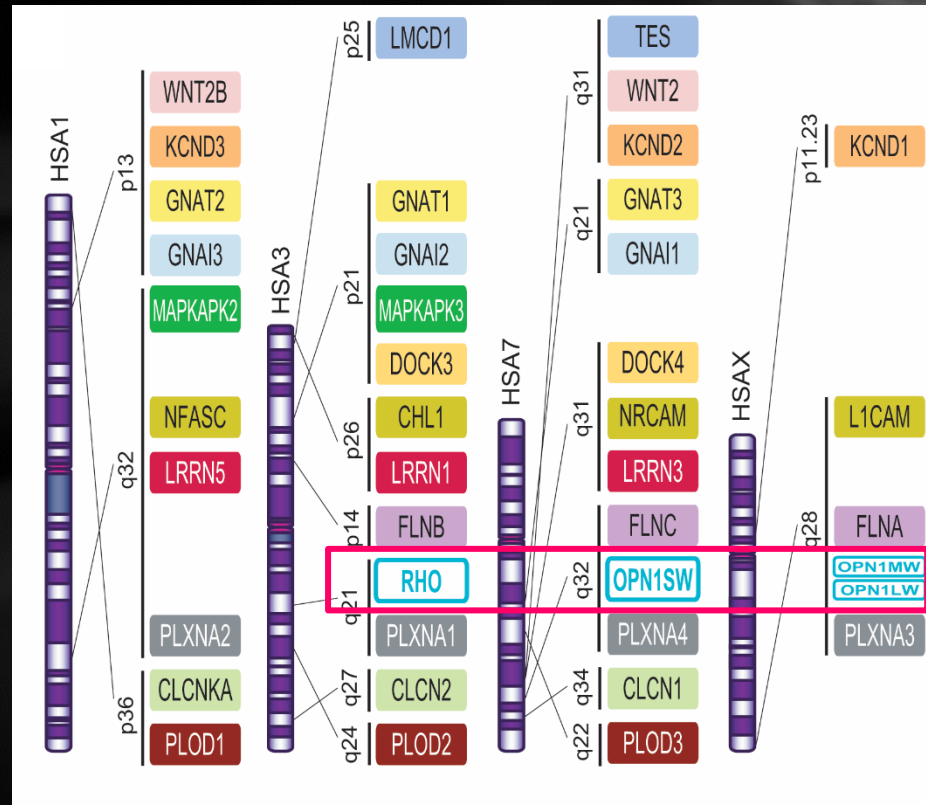
Visual opsins



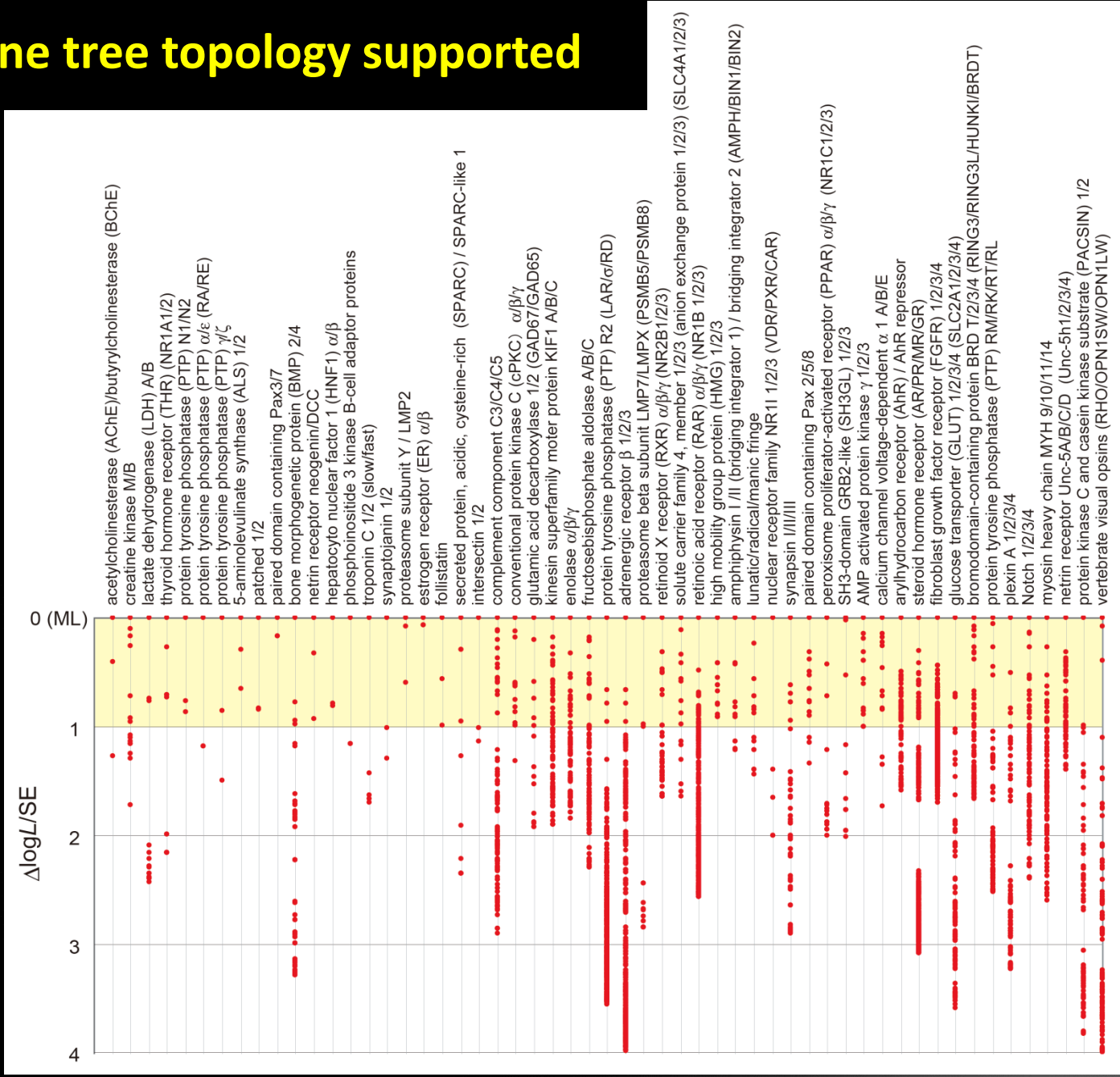
Exhaustive ML, 11 OTUs
JTT+ Γ_4 ($\alpha=0.92$), 280 sites

1:1 relationships between cyclostome and gnathostome genes

Kuraku et al., 2009. *Mol. Biol. Evol.* 26:47-



Not only one tree topology supported

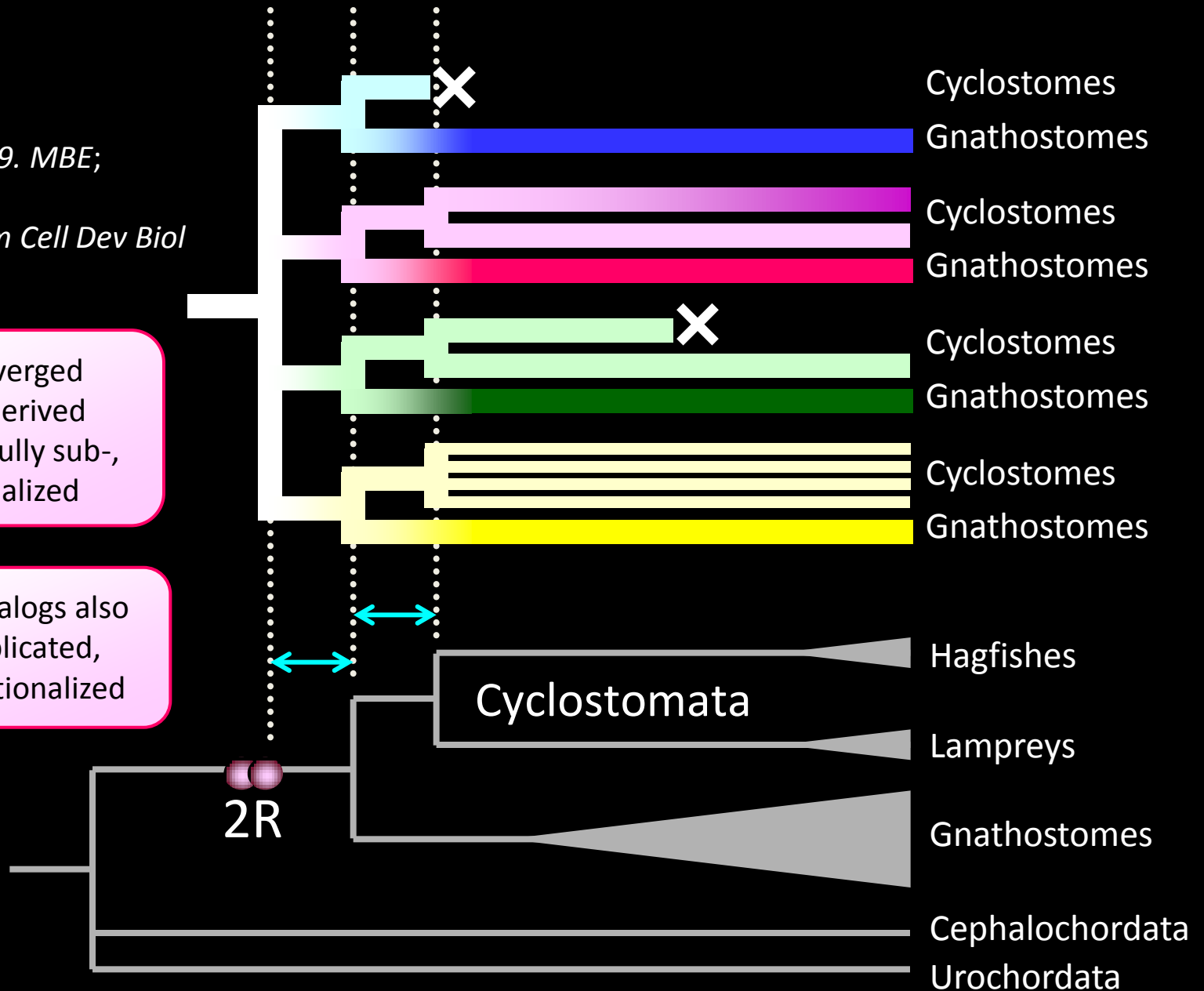


What Characterizes 'Cyclostomata' ?

Kuraku et al., 2009. MBE;
Reviewed in
Kuraku, 2013. Sem Cell Dev Biol

Cyclostomes diverged before the 2R-derived paralogs were fully sub-, or neo-functionalized

Cyclostome paralogs also secondarily duplicated, lost or neofunctionalized

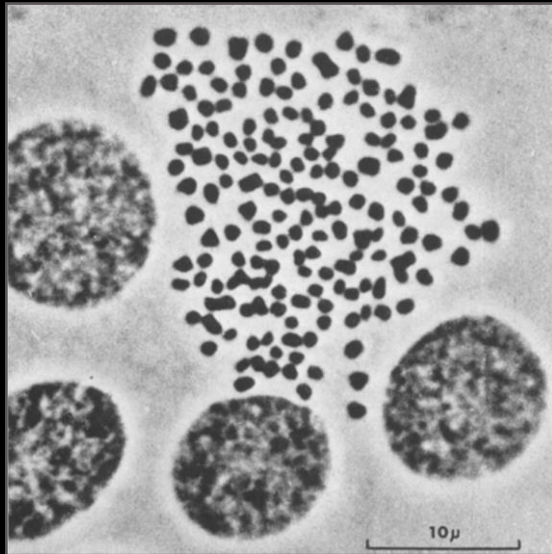




Sea lamprey genome project

Sequenced at Wash. Univ. Genome Institute

Petromyzon marinus (2n=168)



Potter and Rothwell, 1970

International consortium

Smith, Kuraku, et al. 2013
Nature Genetics

The screenshot shows the Ensembl genome browser interface for the sea lamprey (*Petromyzon marinus*) genome assembly Pmarinus_7.0. The page includes a search bar, a navigation menu (BLAST/BLAT, BioMart, Tools, Downloads, More), and a search bar with the text "Search Lamprey...". The main content area is divided into several sections:

- Genome assembly: Pmarinus_7.0**: Includes a search bar for "Search Lamprey...", a "Go" button, and example coordinates (e.g., O42160_PETMA or GL476598:172728-528162 or prion). It also has links for "More information and statistics" and "Download DNA sequence (FASTA)".
- Gene annotation**: Includes a search bar for "Search Lamprey...", a "Go" button, and example coordinates (e.g., O42160_PETMA or GL476598:172728-528162 or prion). It also has links for "More information and statistics" and "Download DNA sequence (FASTA)".
- Comparative genomics**: Includes a search bar for "Search Lamprey...", a "Go" button, and example coordinates (e.g., O42160_PETMA or GL476598:172728-528162 or prion). It also has links for "More information and statistics" and "Download DNA sequence (FASTA)".
- Variation**: Includes a search bar for "Search Lamprey...", a "Go" button, and example coordinates (e.g., O42160_PETMA or GL476598:172728-528162 or prion). It also has links for "More information and statistics" and "Download DNA sequence (FASTA)".

http://www.ensembl.org/Petromyzon_marinus/Info/Index

Coding genes: 10,415

Incomplete genome assembly: Pax6 missing

Incomplete gene annotation: Fgf8/17-A missing

=> Access to the consortium gene set??

Timing of two-round (2R) WGDs

Firm evidence of '2R' from tetra-syteny blocks

Larhammar et al., 2002. Genome Res., etc.

Independent analyses with different methodologies

2009 Integrative analysis (55 gene families)

Kuraku et al., 2009. Mol. Biol. Evol.

2013 Two lamprey genomes

Petromyzon marinus (Smith, Kuraku, et al., 2013. Nature Genet)

Lethenteron japonicum (Mehta et al., 2013. PNAS)

Analyzing systematic bias, etc.

↓ ? Phylome analysis with dense taxon-sampling ?



Lamprey genes/genomes

Protein-coding genes

High GC, long UTR, frequent simple repeats ('QQQQ...', 'HHHH...', etc.)

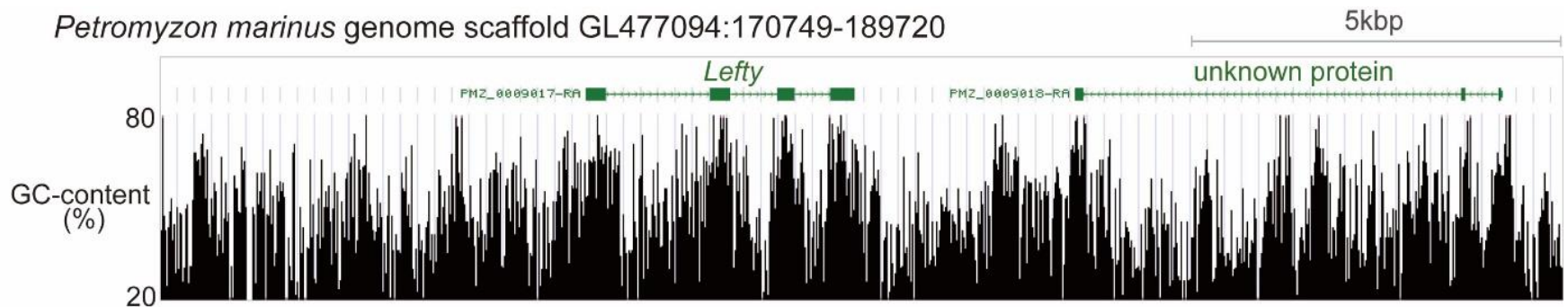
Ambiguous orthology to gnathostome genes

cf. Bmp2/4-B, DlxA, Hh-A, HoxW10a,

Genomic landscape

Abundant repetitive elements including foreign Tc1 *Kuraku et al., 2012*

High GC in exons

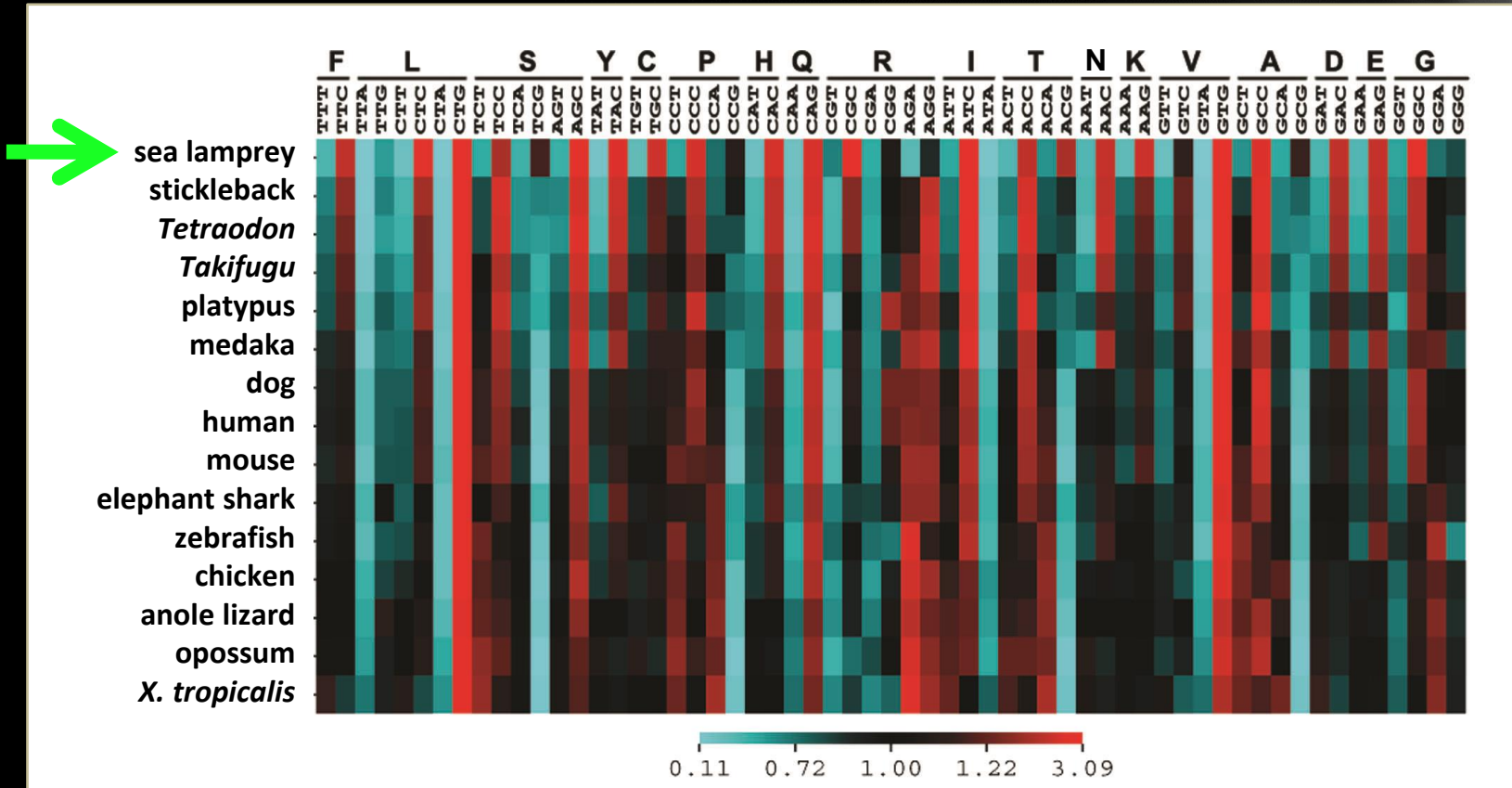




Peculiar nature of lamprey genes

Codon usage bias

Methods: RSCU (Sharp et al., 1986) and ENc (Wright, 1990)



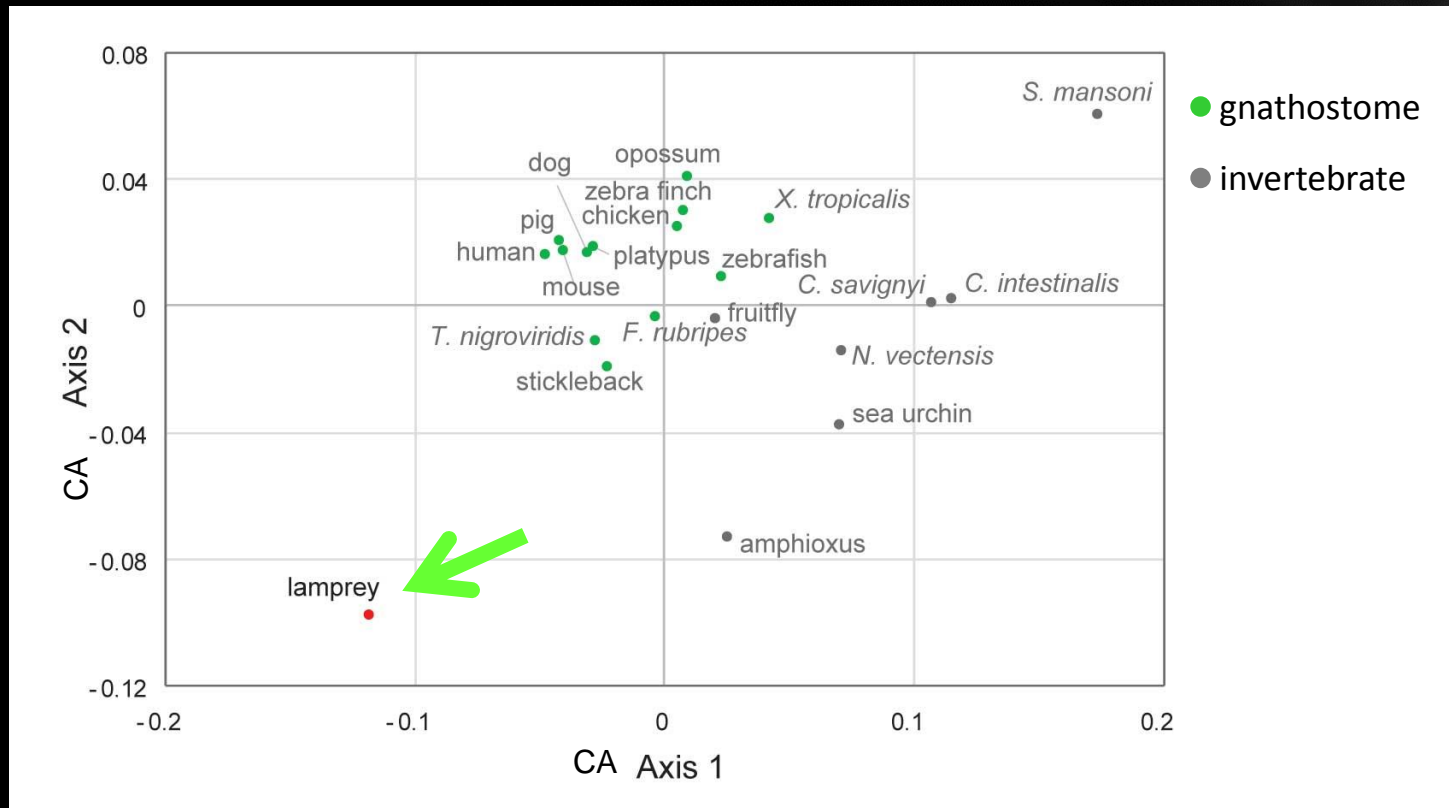
Heavy use of GC-rich codons in translation



Peculiar nature of lamprey genes

Amino acid composition

Methods: Correspondence analysis for frequencies of 20 amino acids

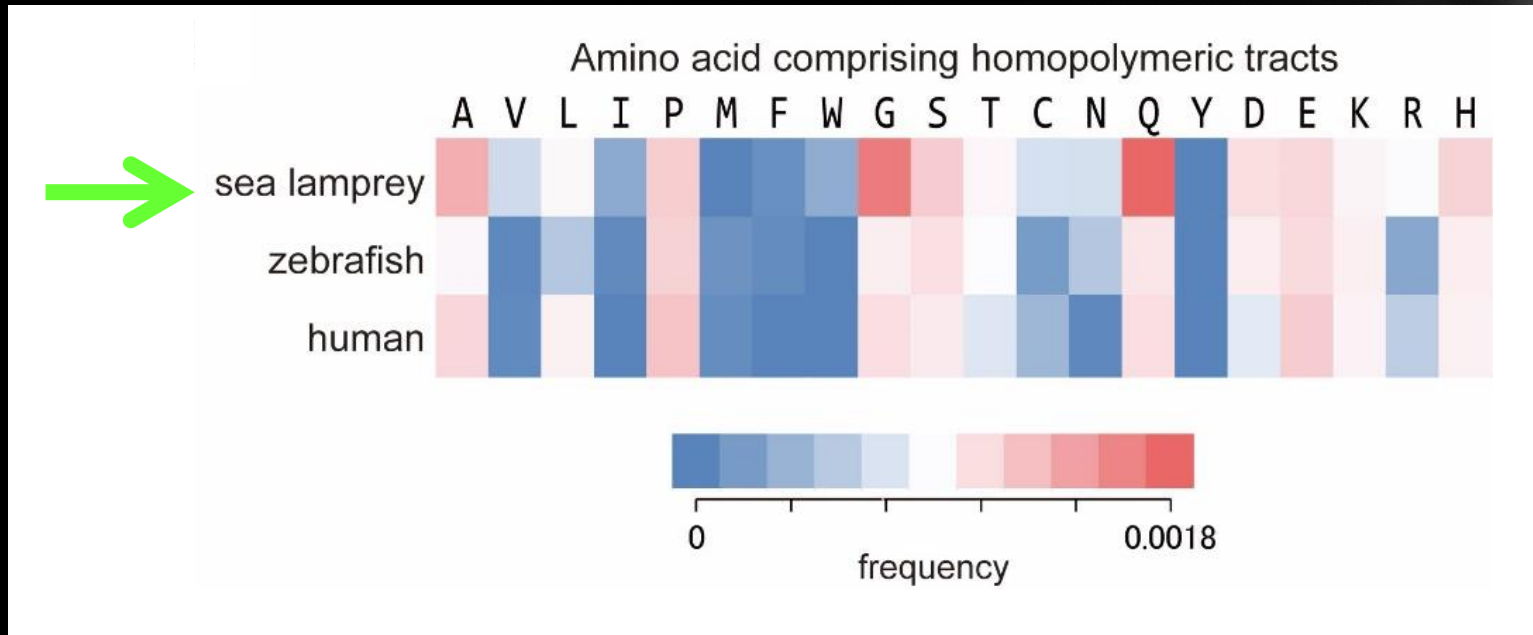


Genome-wide deviation of 'gene model' in lamprey



Peculiar nature of lamprey genes

Unique pattern of homopolymeric amino acid (HPAA) tracts



Independent HPAA insertions can mislead phylogeny

cf. Tank EM, ..., Langeland JA. *Evolution & Development* 2009, 11:343-353.

↑
HPAA
insertion

Petromyzon marinus (sea lamprey) EmxA

MFQPPTKRCFTIESLVAKDGGDSVSAATL **QQQQQQQQQQQQQQQQQQQQQQ** PAYPIPGGANPL
TCGAQPPHPF **AAAAAAAAAAAA** SRAGMPELFFHDAHHLQPLLAVPAMHPSGHHHHHLSH
PPPSLPLFGGPQGRDAISLYPWL LHRPRYLGHRYPGADGNAESLLHSPFARK **PKRIRTA**
SPSQLLRLEHAF EKNHYVGAERKQLASSLSLSETQVKVWFQNRRTKYKRQKLEEE GPESP
QKKKSSHINRWRQATNQSGDEIDVTSDD ← homeodomain

Eptatretus burgeri (inshore hagfish) EmxA

MFQPVAKRCFTIESLVAKESDPSPNLLHVSYGSGSHSGIGSVHSGPHYGPGRPGYGPHQGDL
FFPMEGVHPGPVVPVPSMASHPMTGQSLHPAPSIPLLGAPHSRDHFSLYPWL FHRPPYLGH
RFRGADGGPETLL LHGPFARK **PKRIRTA** **SPSQLLRLEHAF EKNHYVGAERKQLAGSLT**
TETQVKVWFQNRRTKYKRQKMEEE GPESPHKKKGSHHINRWRLATNQSSGDEIDVTSDD

↑
HPAA
insertion

Petromyzon marinus (sea lamprey) EmxB

MFQPATTKRCFTIESLVAKDCPAPTSRSE **QQQQQQQQQQQQ** HEAPLRPAALS FAGTHNHQHHQ
HQQHHHSSASVS **AAAAAAAAAAAA** FVPAFPQPTARALYPHPAELLYSDPGHHPGSAGG
PLQVPALPPHHHLQSHPLFGPPQRDPMTFYPWLLNRHRYLGHRYPGPETGHEGLLFPGPL
ARK **PKRIRTA** **SPSQLLRLEHAF EKNHYVGSERKQLASSLSLSETQVKVWFQNRRTKHKR**
QKLEEE GPDEQQKKKGTHHVNRWRMATKQPSSSEDIDVTSDD

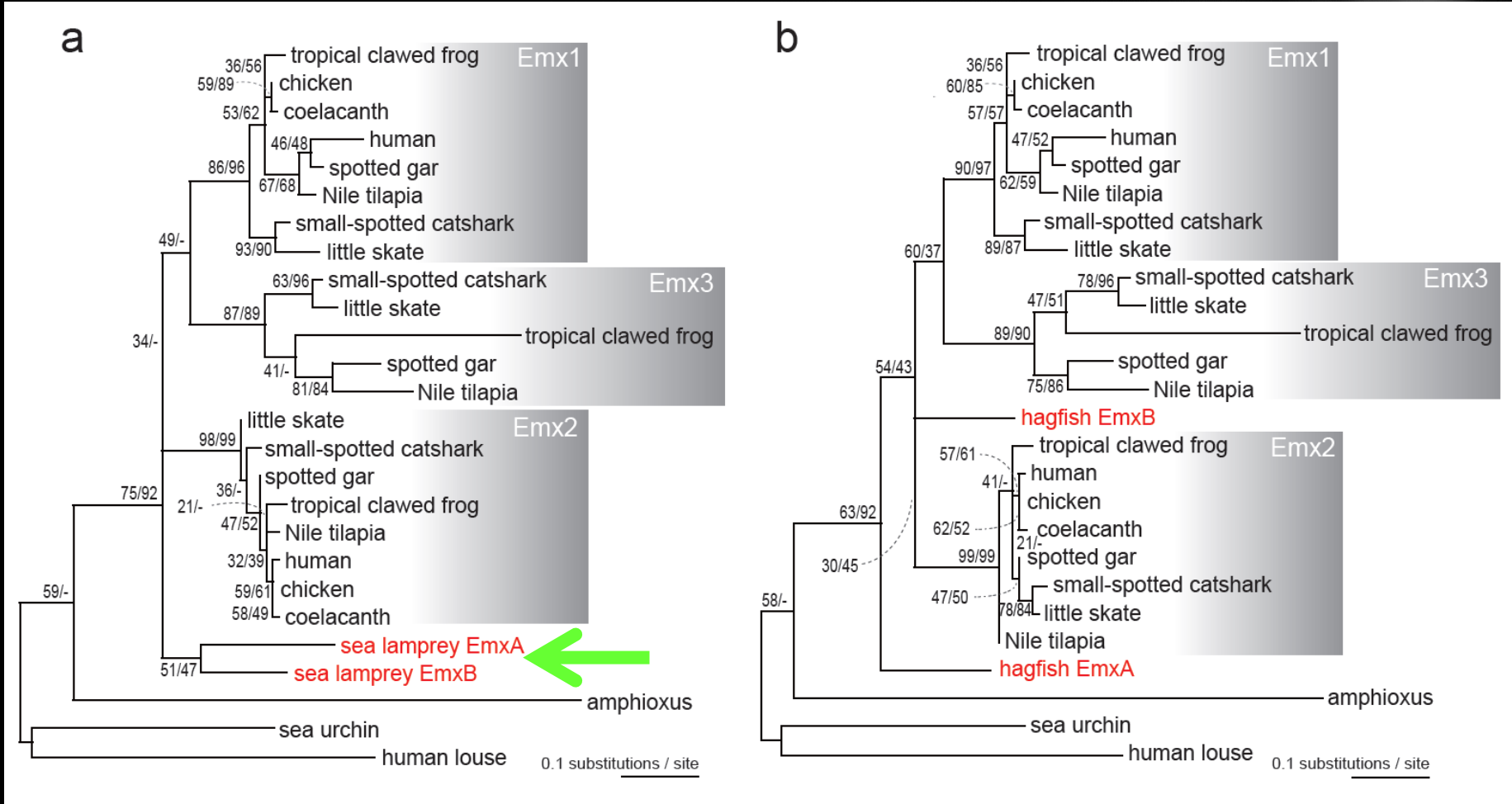
Eptatretus burgeri (inshore hagfish) EmxB

MFQPAAKRCFTIESLVAKDCPAPGGASNRDPEPLRPAALS FAGAPVHSQPGTAFMPGFAQP
AGRTLPHADLMYPEPGPHHAASGPLIPALPVGAPLQPSHHPLFGPPQRDPMTFYPWLL
NRHRYLSHRYPGPDSCAESLLFPGP FARK **PKRIRTA** **SPSQLLRLEHAF EKNHYVGSERK**
QLASSLSL TETQVKVWFQNRRTKHKRQKLEEE GPDEHQKKKGTHHINRWRMATKQTSSEDI
DVTSSEN

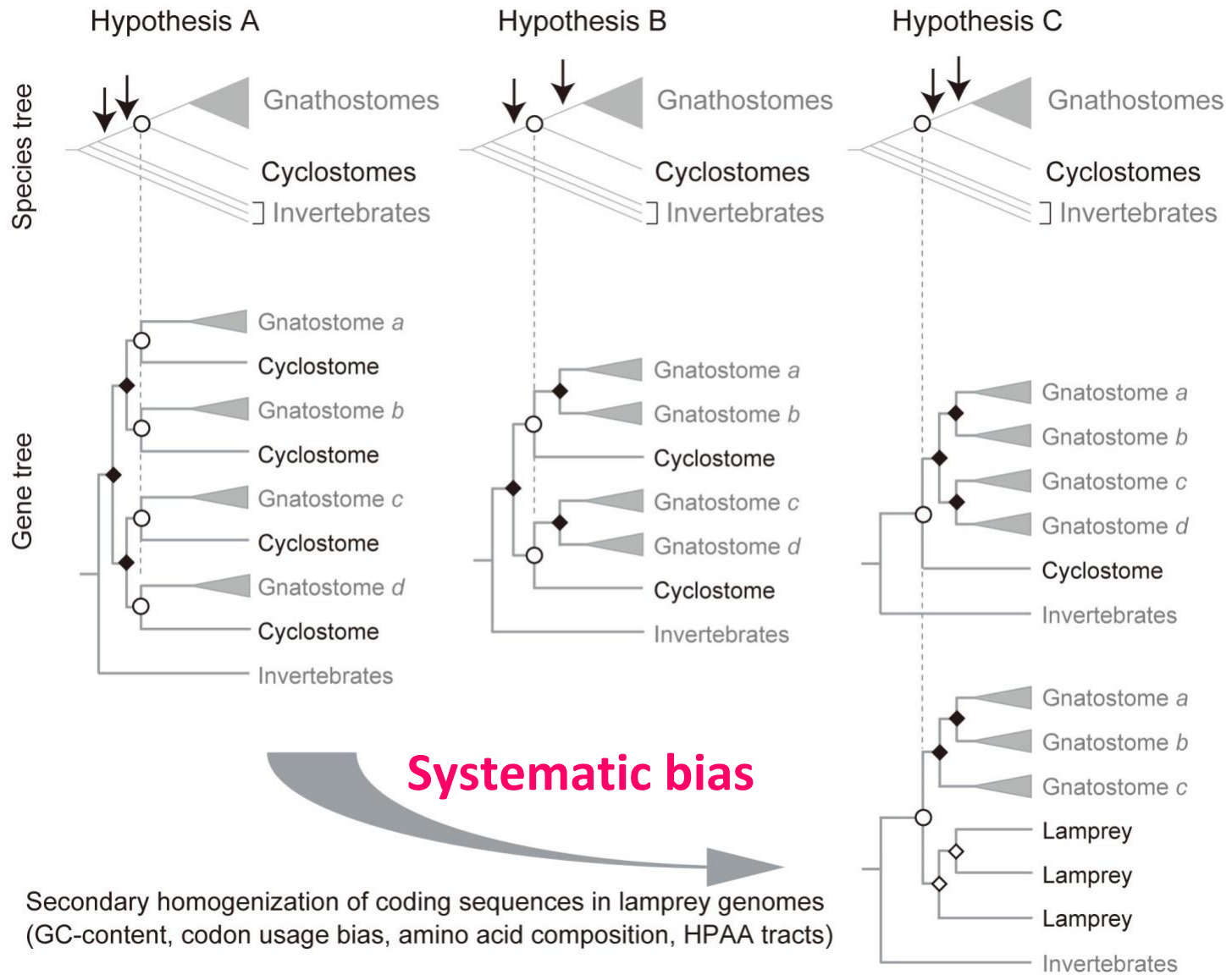
EmxA-B proximity supported only for lamprey

Tree inference with hagfish and/or lamprey

Various alignment and tree inference programs tested



Why highly incongruent ?



Arising questions and concerns

Why wasn't QfO initiated earlier?

Any clever way to cope with 'lamprey dialect'?

'Non-homogeneous' methods?

More taxon sampling?

Taxon sampling



Genome size: 6.65 Gb

Species in yellow - Currently sequenced in-house

Arising questions and concerns

Why wasn't QfO initiated earlier?

Any clever way to cope with 'lamprey dialect'?

'Non-homogeneous' methods?

More taxon sampling?

Basically all cyclostome (lamprey and hagfish) genes have ambiguity in orthology

This is also true for genes of eel, arowana, etc. in relation to '3R' (teleost fish-specific) WGD

reviewed in Kuraku, 2013. Sem Cell Dev Biol

How do we support and monitor 'quest by non-experts'?

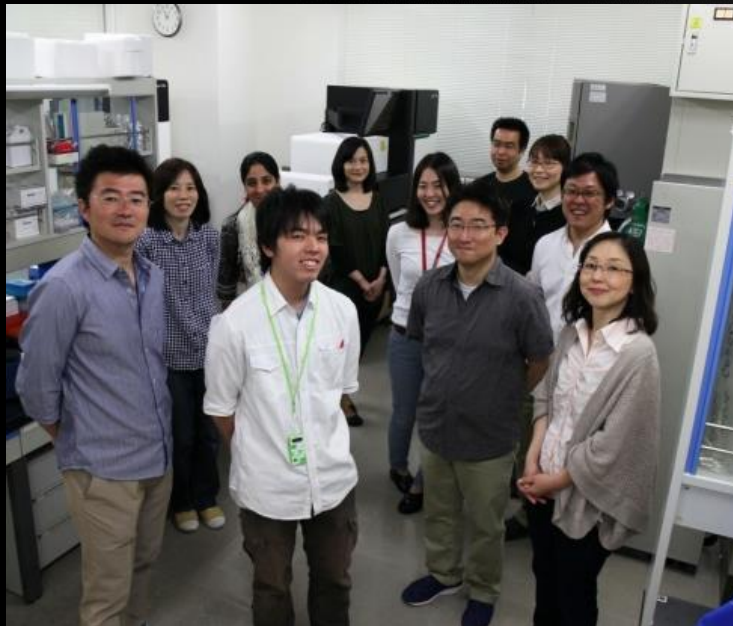
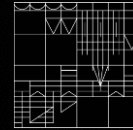
Acknowledgements

Sea Lamprey Genome Consortium

Meyer Lab in Univ. Konstanz

Kuratani Lab in RIKEN Kobe

Lab Members in RIKEN Kobe (since March '12)



Osamu Nishimura

Chiharu Tanegashima

Kaori Tatsumi

Munazah Andrabi

Yuichiro Hara

Mitsutaka Kadota

Kaori Tanaka

Sayo Saito

Fumio Motone

Yuka Kageyama

Alumni

Kazu Itomi

Tetsutaro Hayashi

Miyuki Noro

Kana Miyagishi

Kenichiro Uno

Open to consultation about NGS
Possibility for Ph.D. and postdoc