The genomics of fish migration

Christiaan Henkel

Naturalis
Biodiversity
Center

BASECLEAR

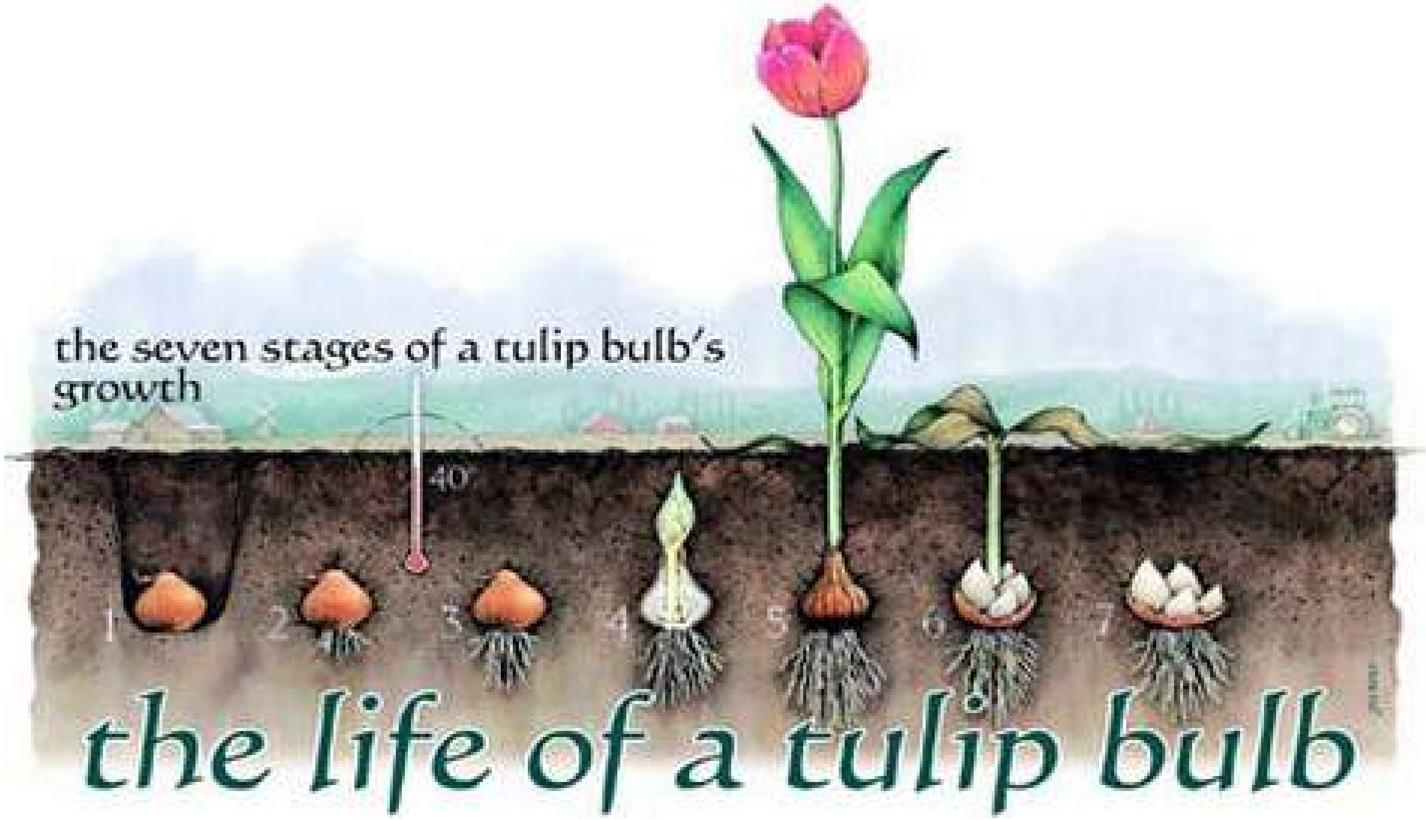
Universiteit
Leiden

hogeschool
Leiden

LUMC

Generade

Applied Genomics for Life


the seven stages of a tulip bulb's growth
40°
1
2
3
4
5
6
7
the life of a tulip bulb

illumina

PACIFIC BIOSCIENCES®

Oxford NANOPORE Technologies

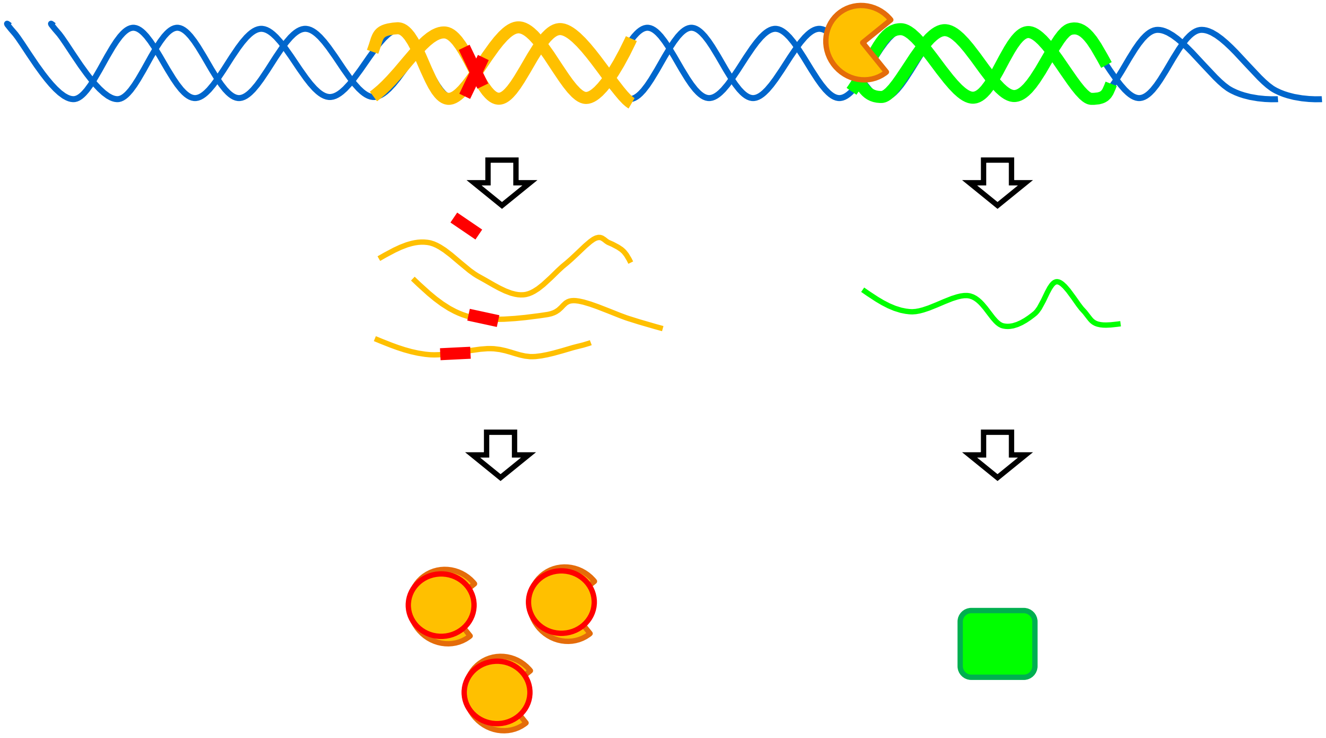
Genomics technology

Genomics applications

Genomics in aquaculture

The genomics of eel migration

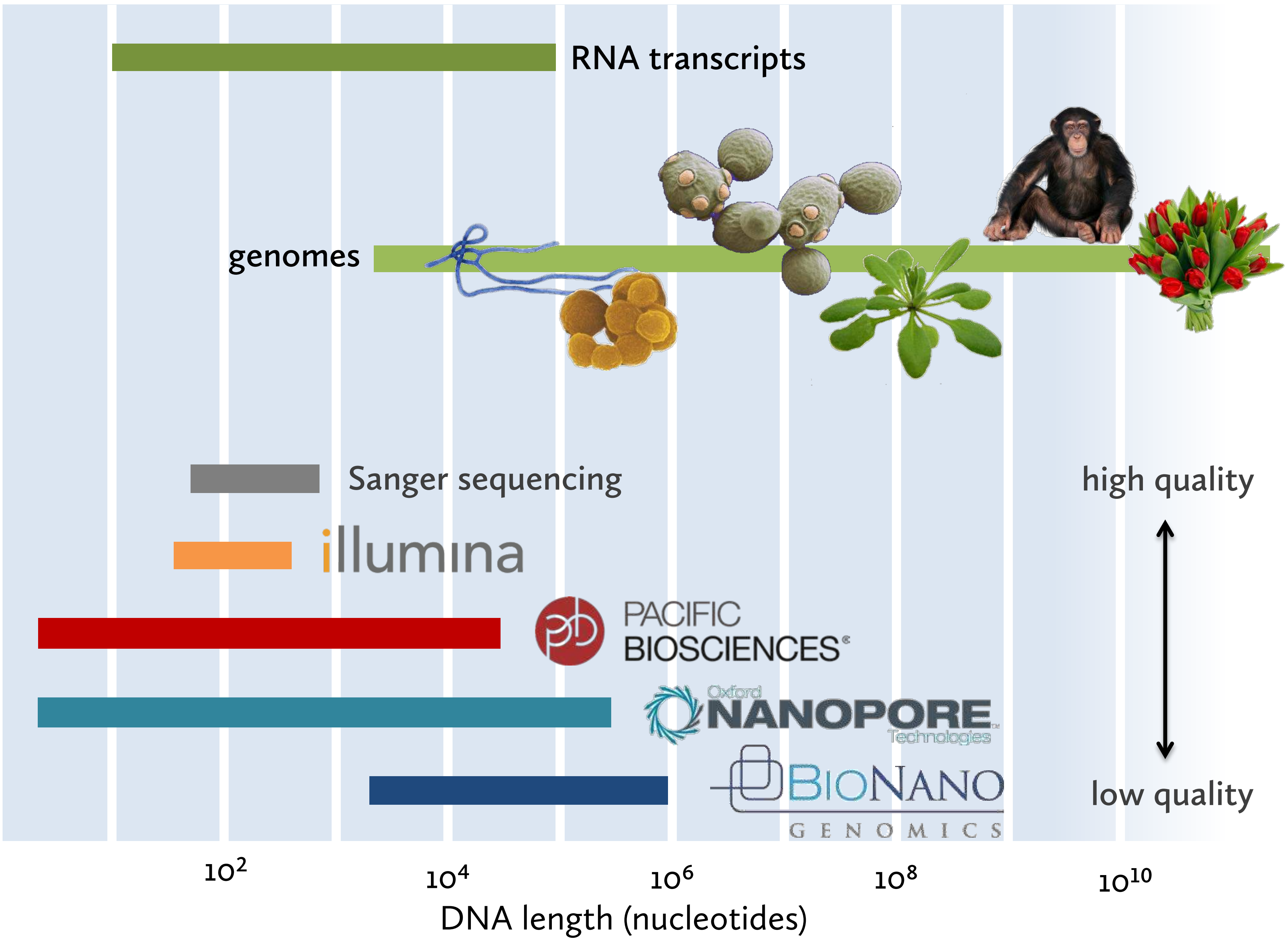
DNA-based information processing



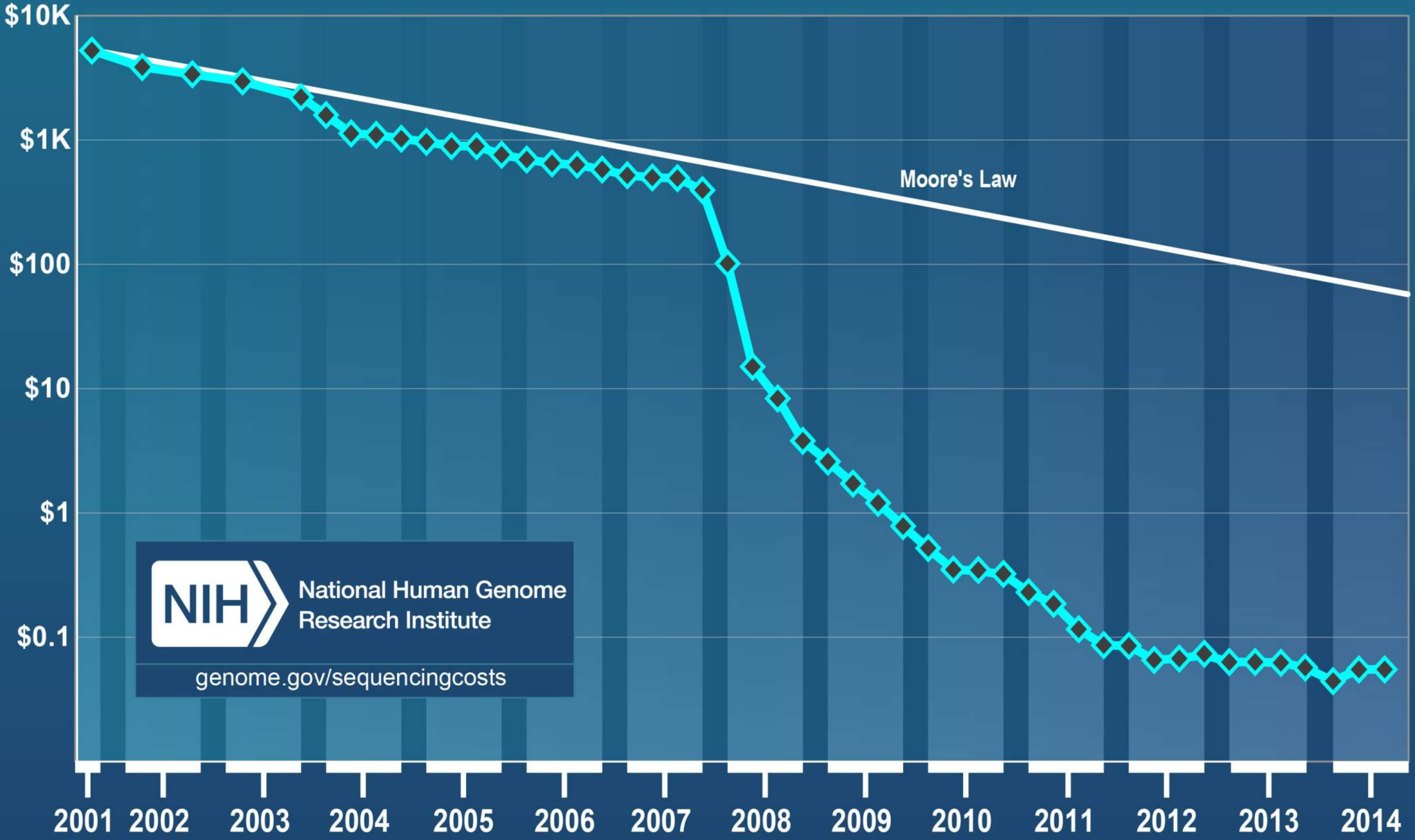
genome

mRNA

protein

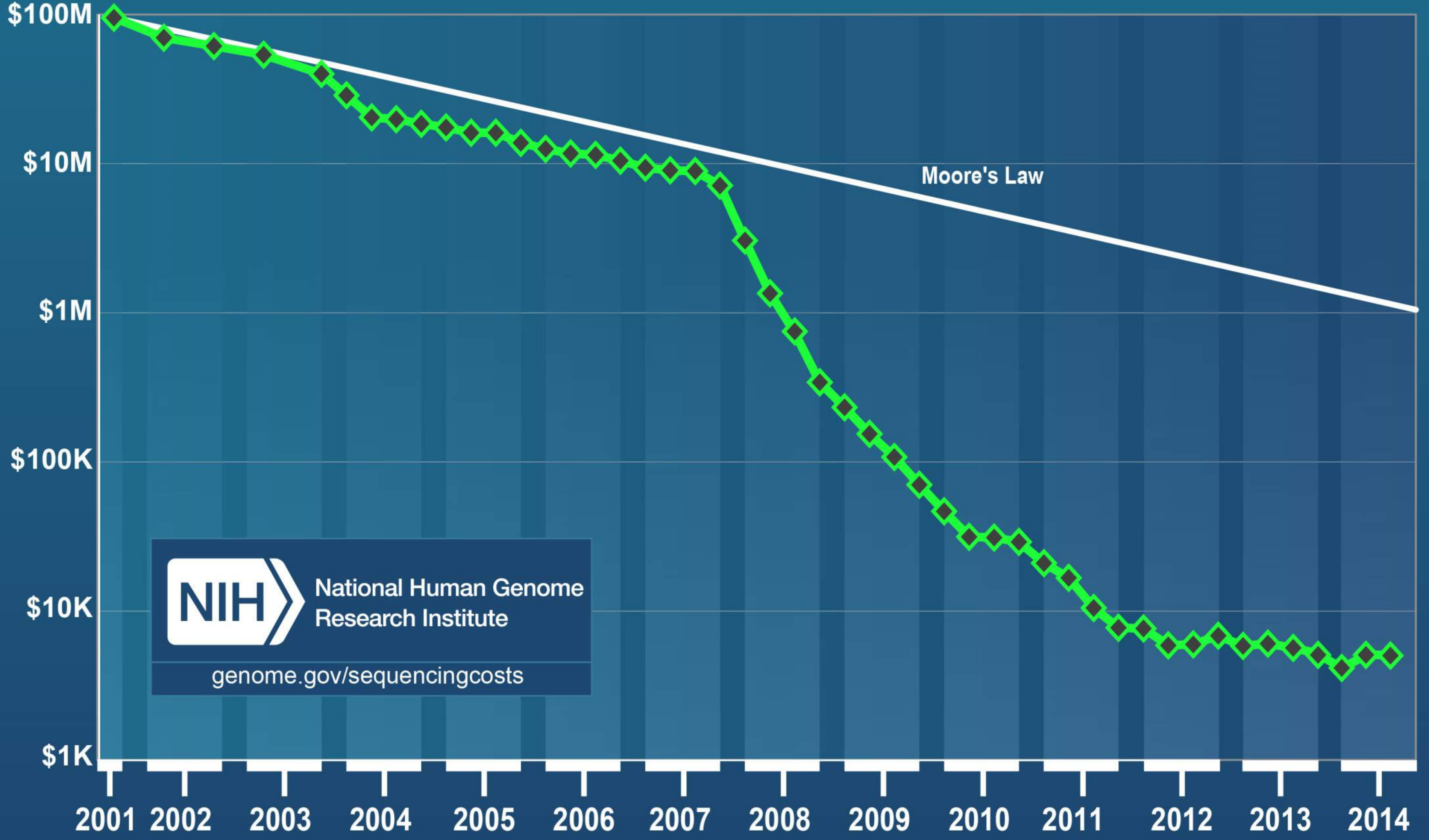

RNA transcripts
genomes
Sanger sequencing
illumina
PACIFIC BIOSCIENCES
Oxford NANOPORE Technologies
BIONANO GENOMICS
high quality
low quality
10^2
10^4
10^6
10^8
10^{10}
DNA length (nucleotides)

Cost per Raw Megabase of DNA Sequence



genome.gov/sequencingcosts

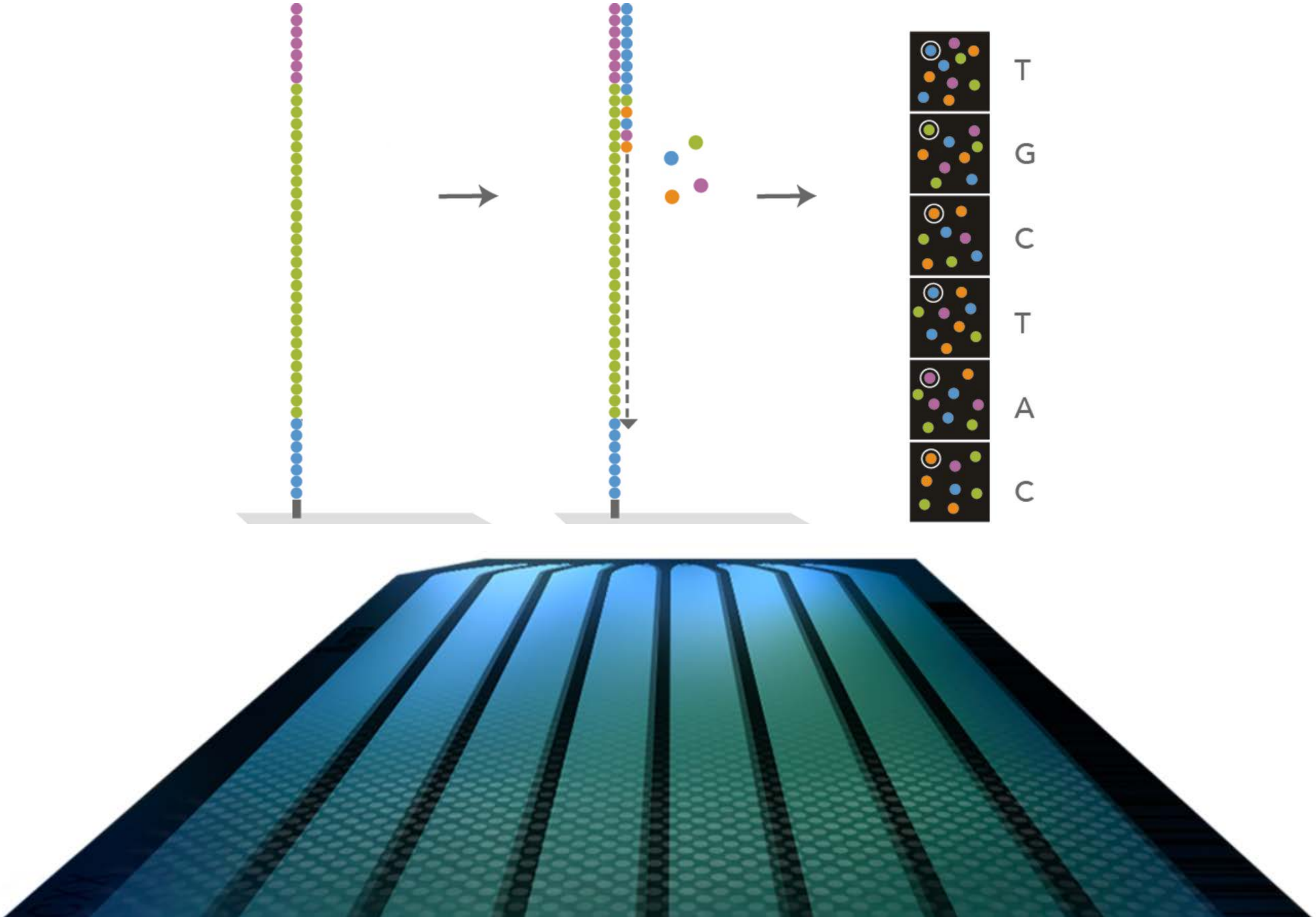
Cost per Genome



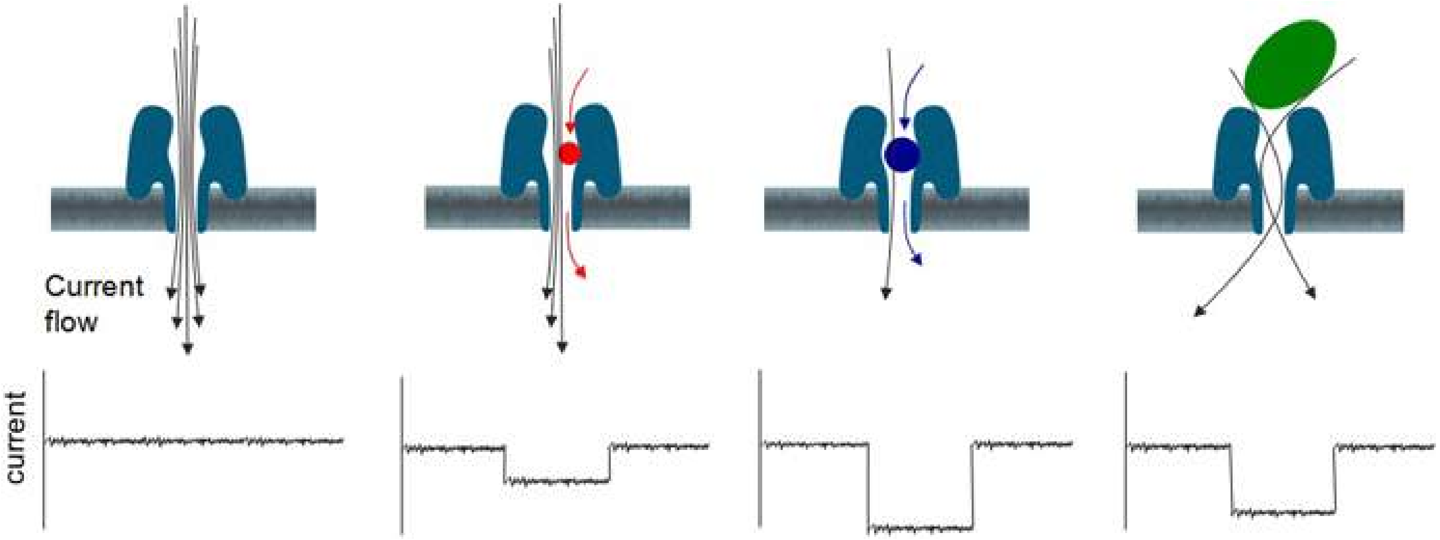
National Human Genome Research Institute

genome.gov/sequencingcosts

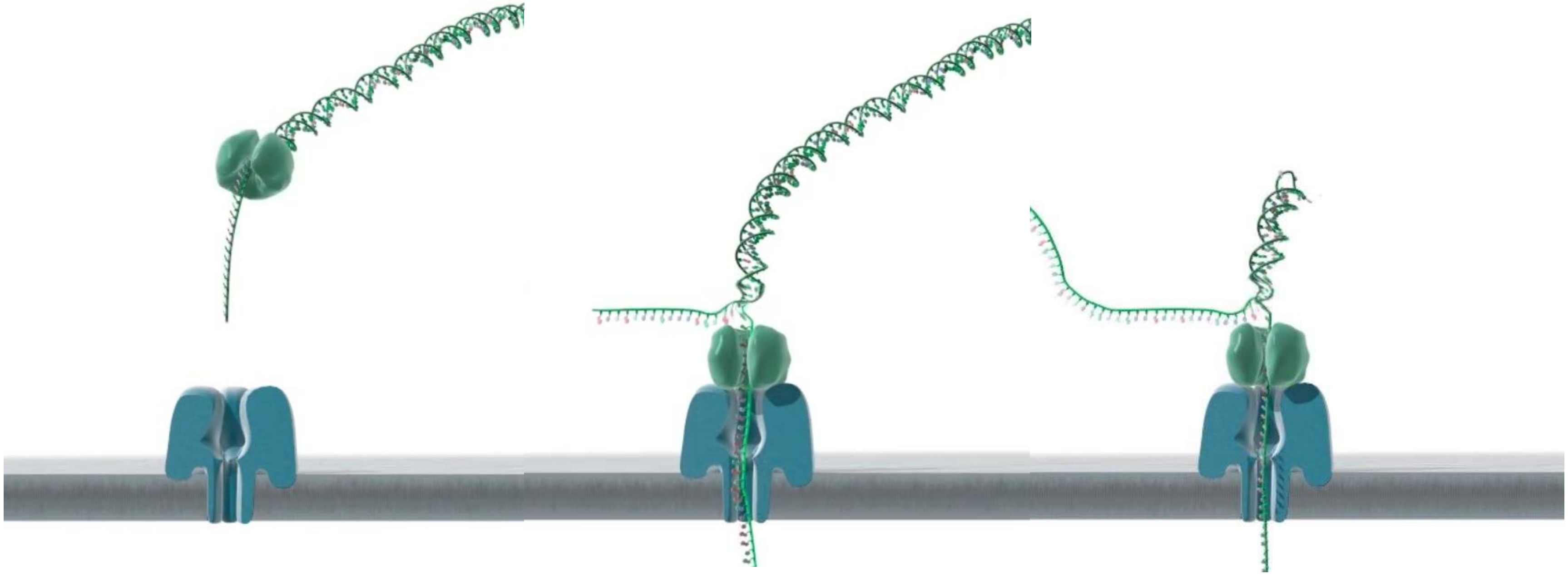
Illumina DNA sequencing



Nanopore molecule sensing



Nanopore DNA sequencing

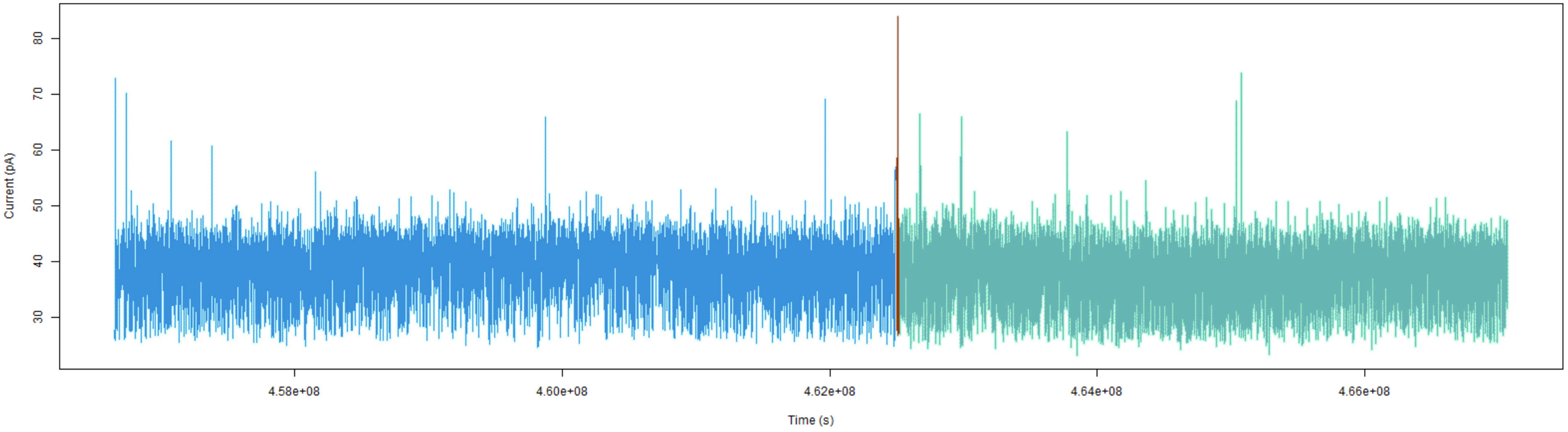


motor protein

brake protein

hairpin

Nanopore DNA sequencing

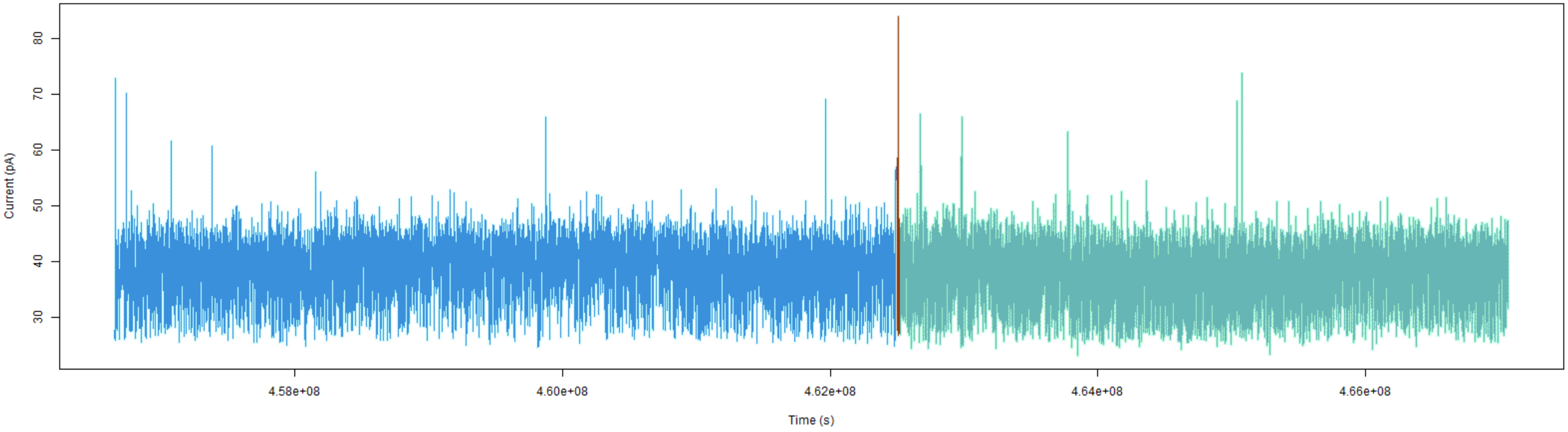


motor protein

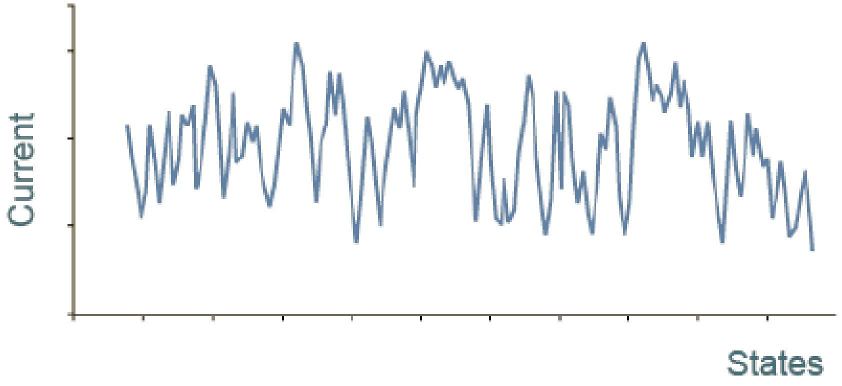
brake protein

hairpin

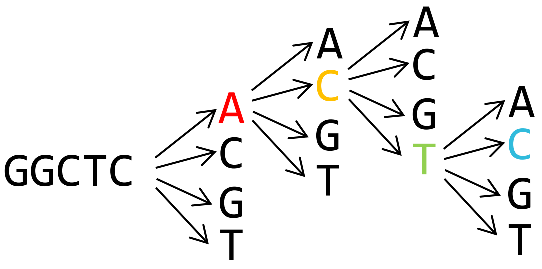
Nanopore DNA sequencing



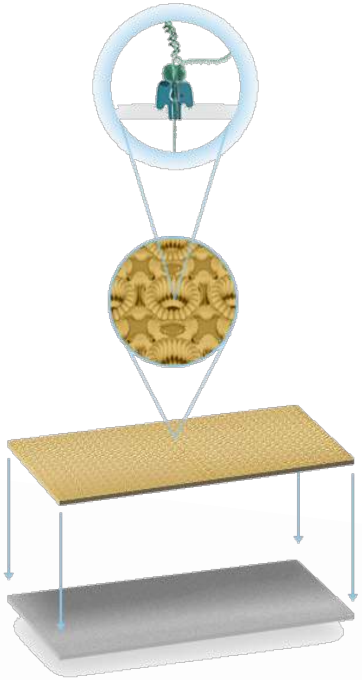
Squiggle plot



Base calling

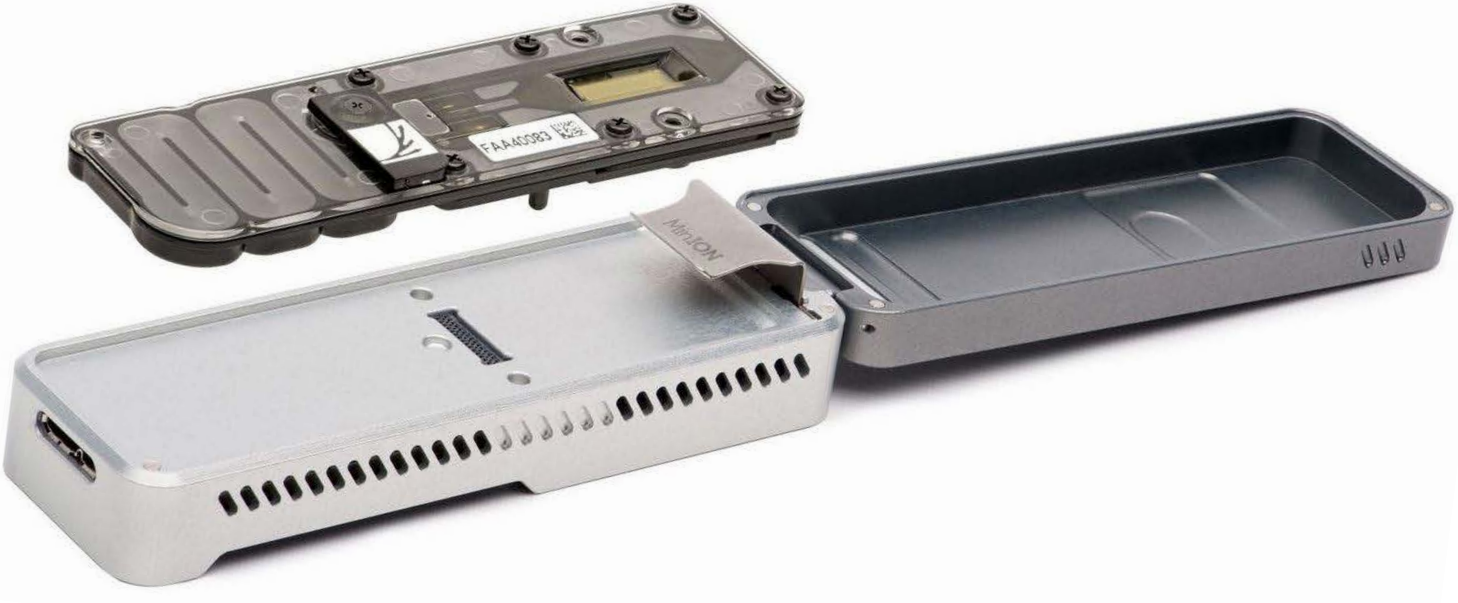


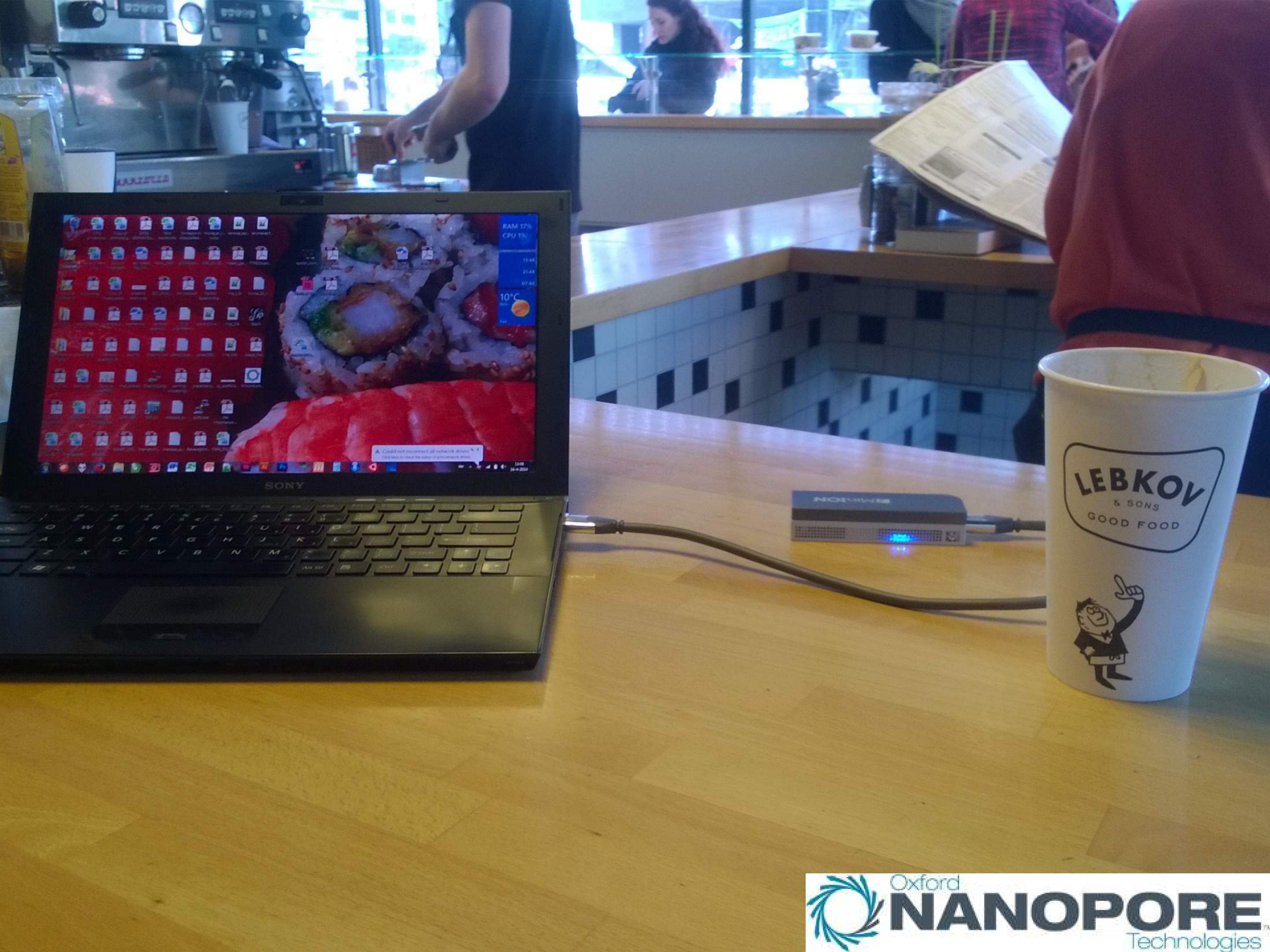
GGCTCACTCCCATAAGC



512 channels

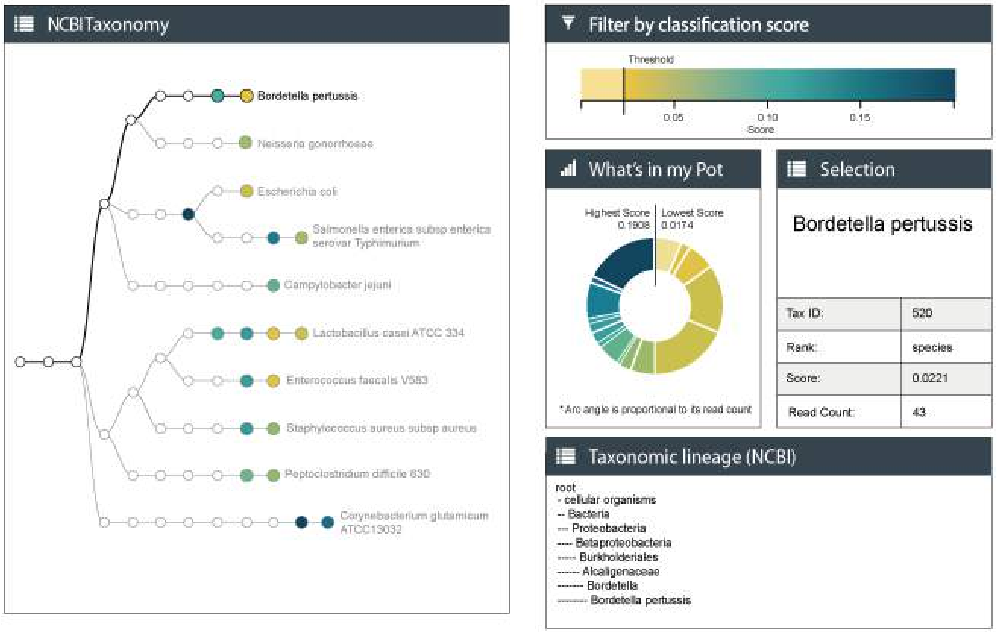
ASIC





Oxford NANOPORE Technologies

- Actual applications:
  - Tracking Ebola virus evolution
  - Rapid identification of bacteria/antibiotic resistances



Selection

Bordetella pertussis

Tax ID:	520
Rank:	species
Score:	0.0221
Read Count:	43

Taxonomic lineage (NCBI)

root
- cellular organisms
-- Bacteria
--- Proteobacteria
---- Betaproteobacteria
----- Burkholderiales
------ Alcaligenaceae
------- Bordetella
-------- Bordetella pertussis

- Astrobiology (MinIONs on the International Space Station)

Genomics technology

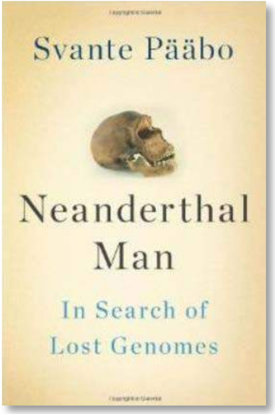
Genomics applications

Genomics in aquaculture

The genomics of eel migration

Genomes

Gene expression (RNA-seq) etc.



Ancient DNA

Variation

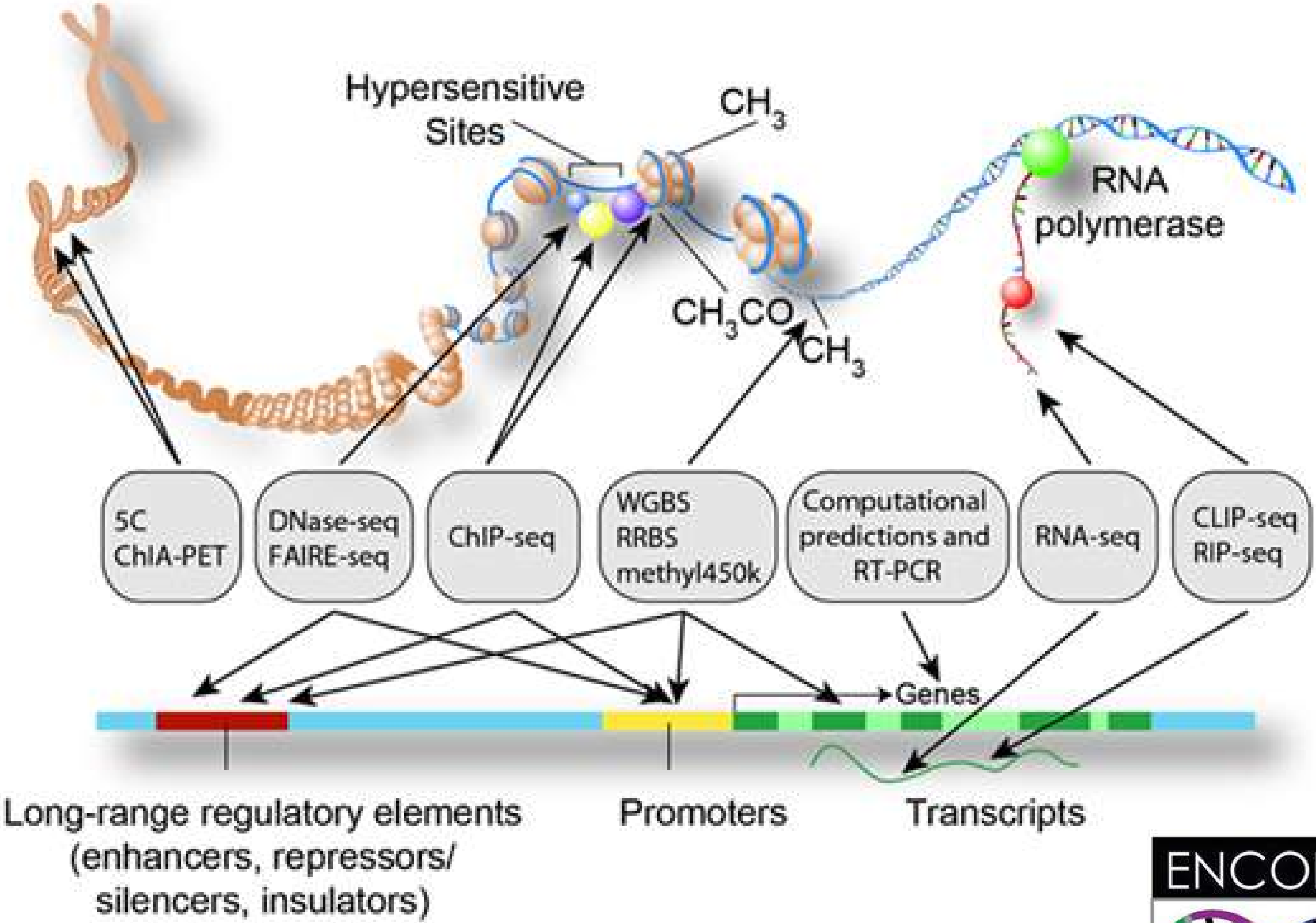
Basic NGS experiments

- RNA-seq
  Sequencing of cDNA

- ChIP-seq
  Sequencing of DNA associated with binding proteins
  (**Ch**romatin **I**mmuno**P**recipitation)

- Genome resequencing and genotyping
  Discovery of variation (SNPs and indels)

- Genome sequencing
  De novo assembly



Hypersensitive Sites

CH_3

CH_3CO

CH_3

RNA polymerase

5C
ChIA-PET

DNase-seq
FAIRE-seq

ChIP-seq

WGBS
RRBS
methyl450k

Computational predictions and RT-PCR

RNA-seq

CLIP-seq
RIP-seq

Genes

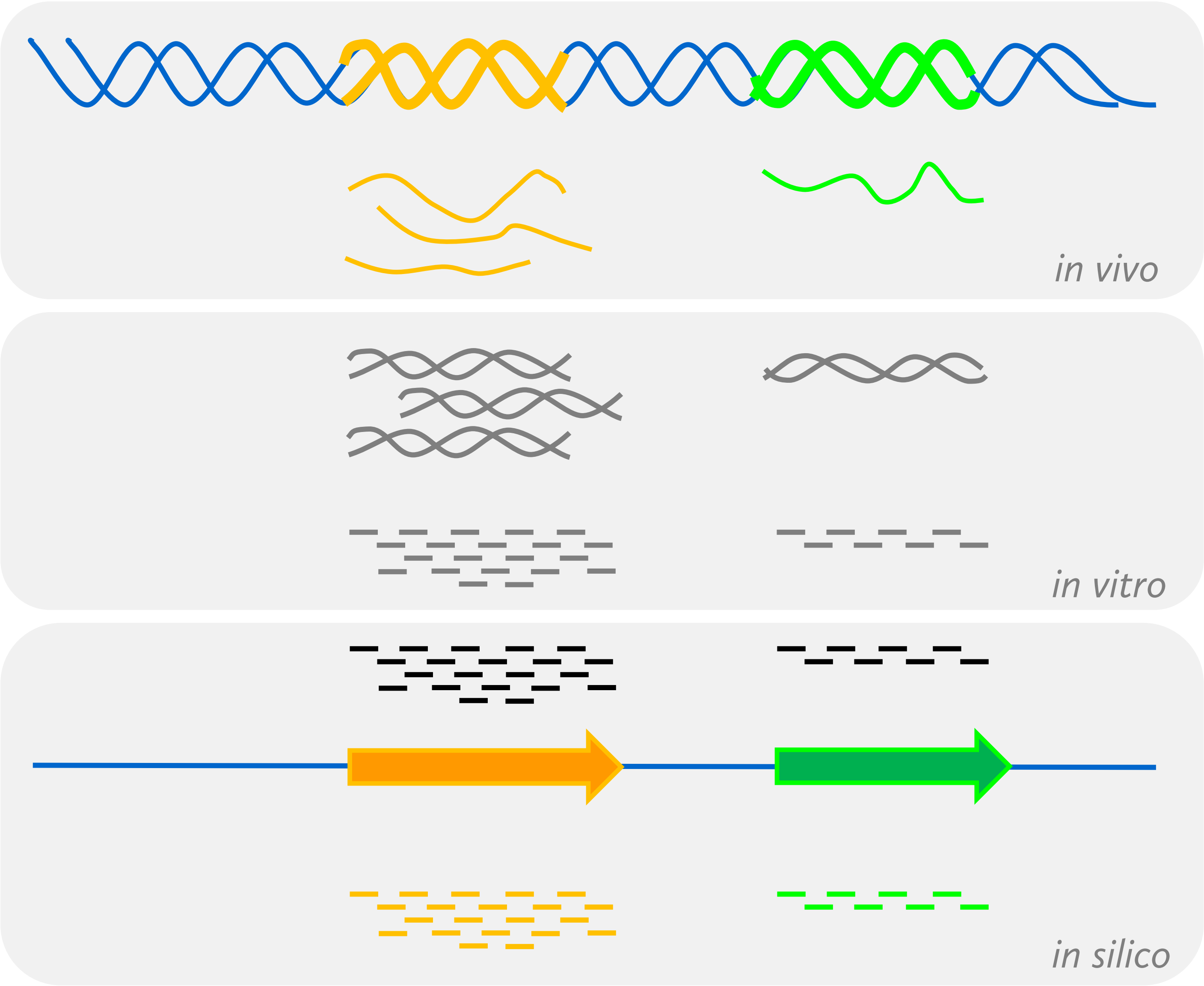
Long-range regulatory elements (enhancers, repressors/ silencers, insulators)

Promoters

Transcripts

ENCODE

RNA-seq



genome

mRNA

cDNA

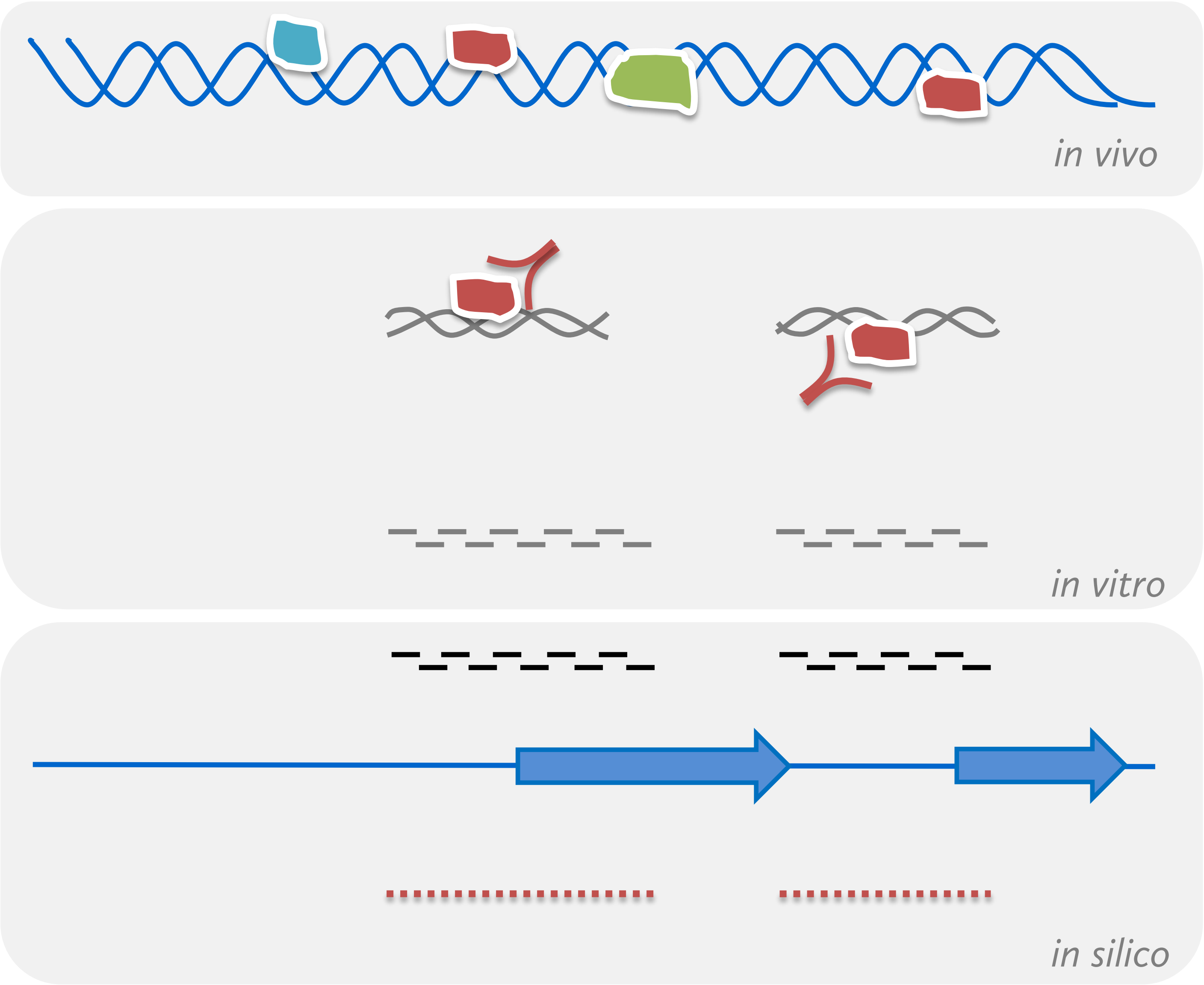
fragments

reads

reference

quantified
expression

ChIP-seq



genome

precipitate

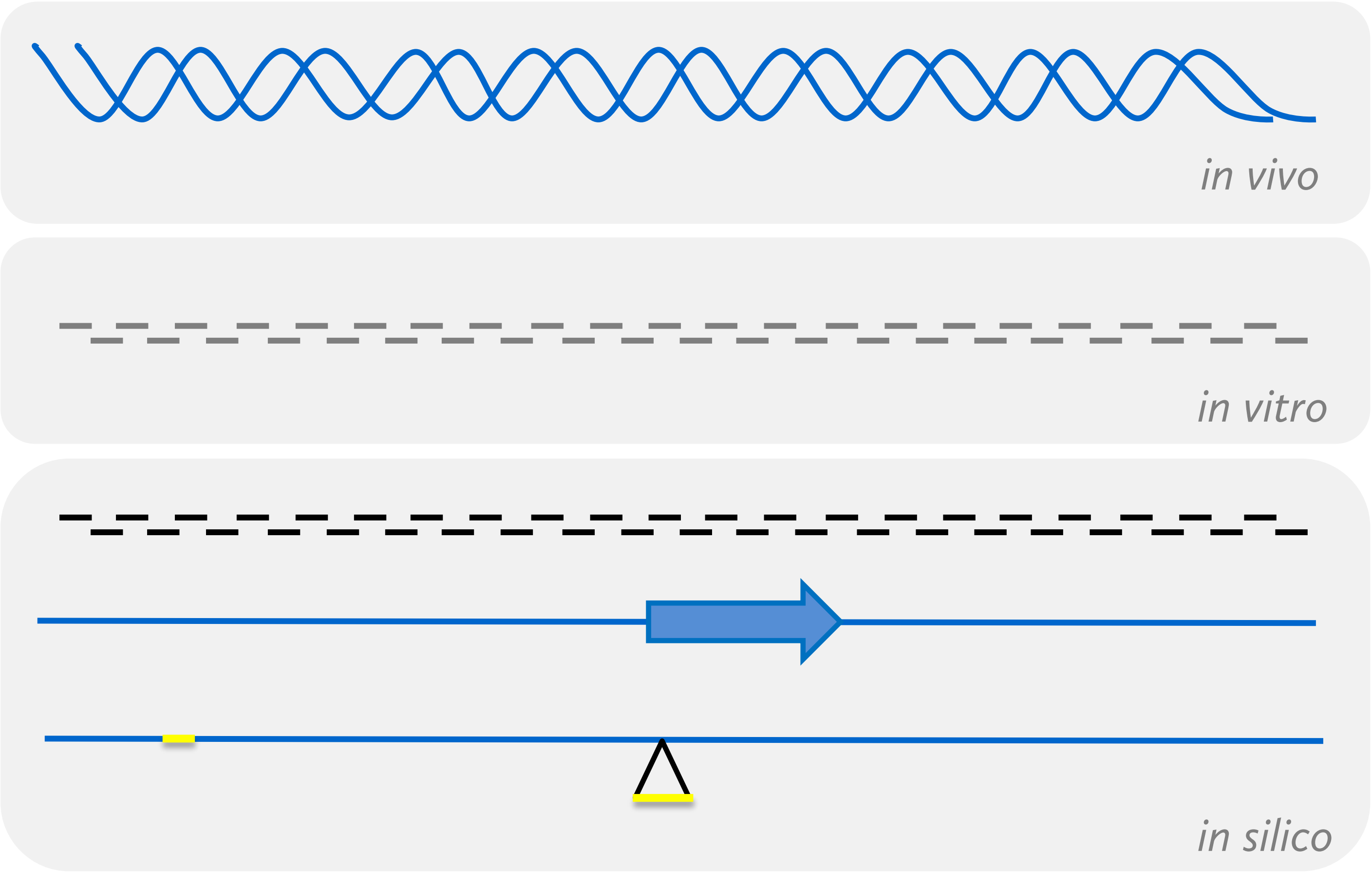
fragments

reads

reference

binding
sites

Genome resequencing



genome

fragments

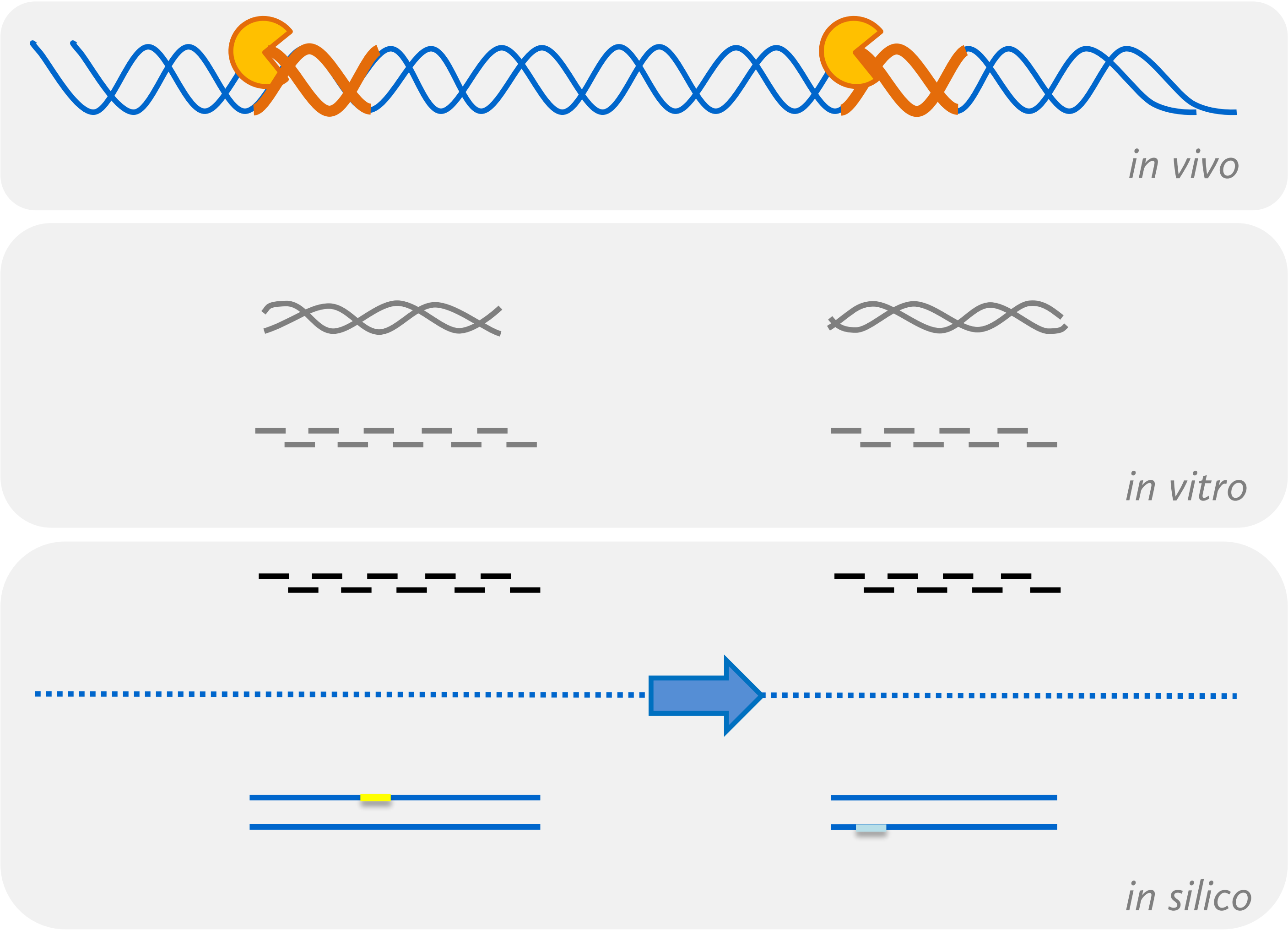
reads

reference

variation

single nucleotide
polymorphism (SNP)

RAD-tag genotyping



genome

Restriction-
Associated
DNA

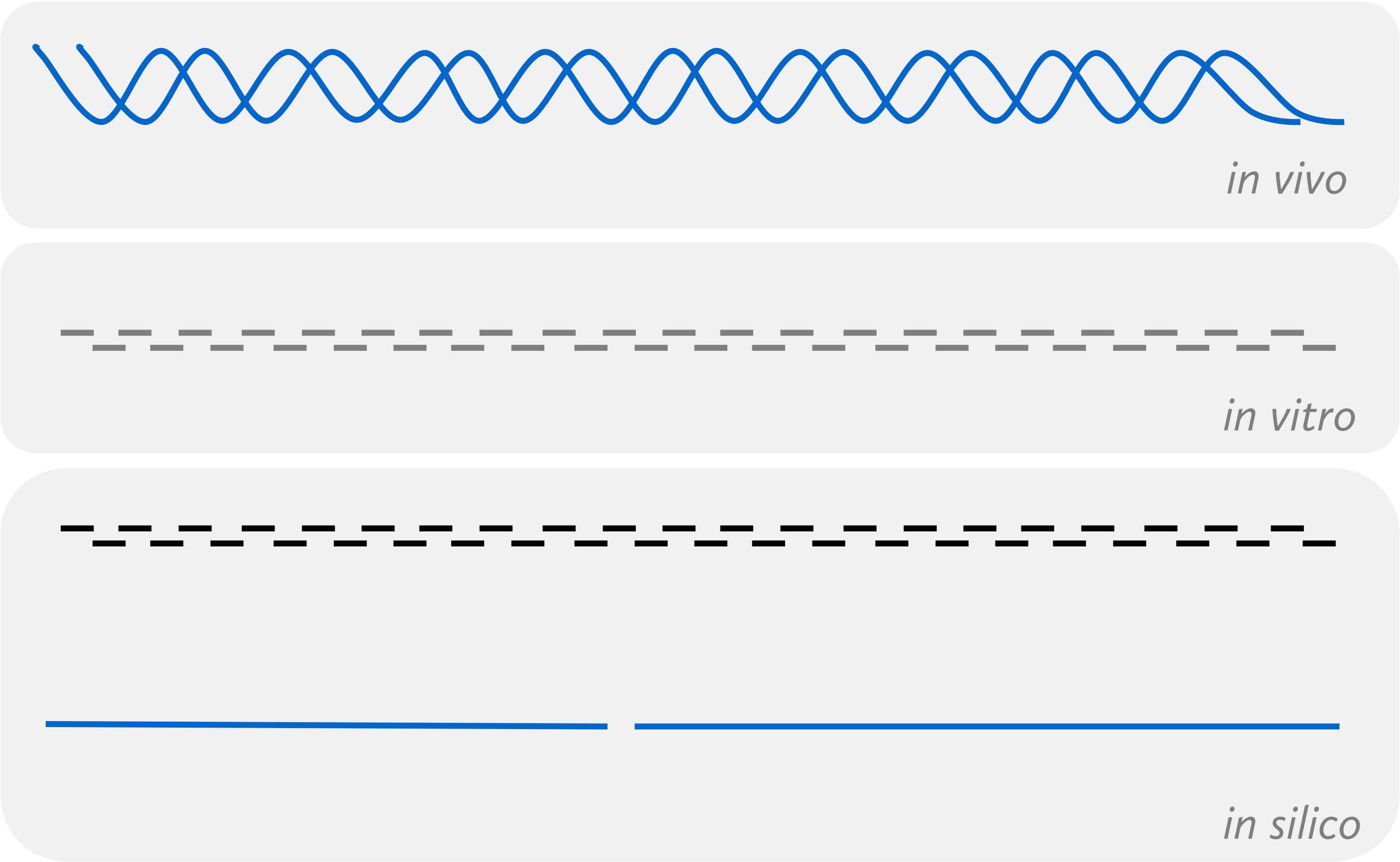
fragments

reads

reference

variation (SNPs)

Genome sequencing



genome

in vivo

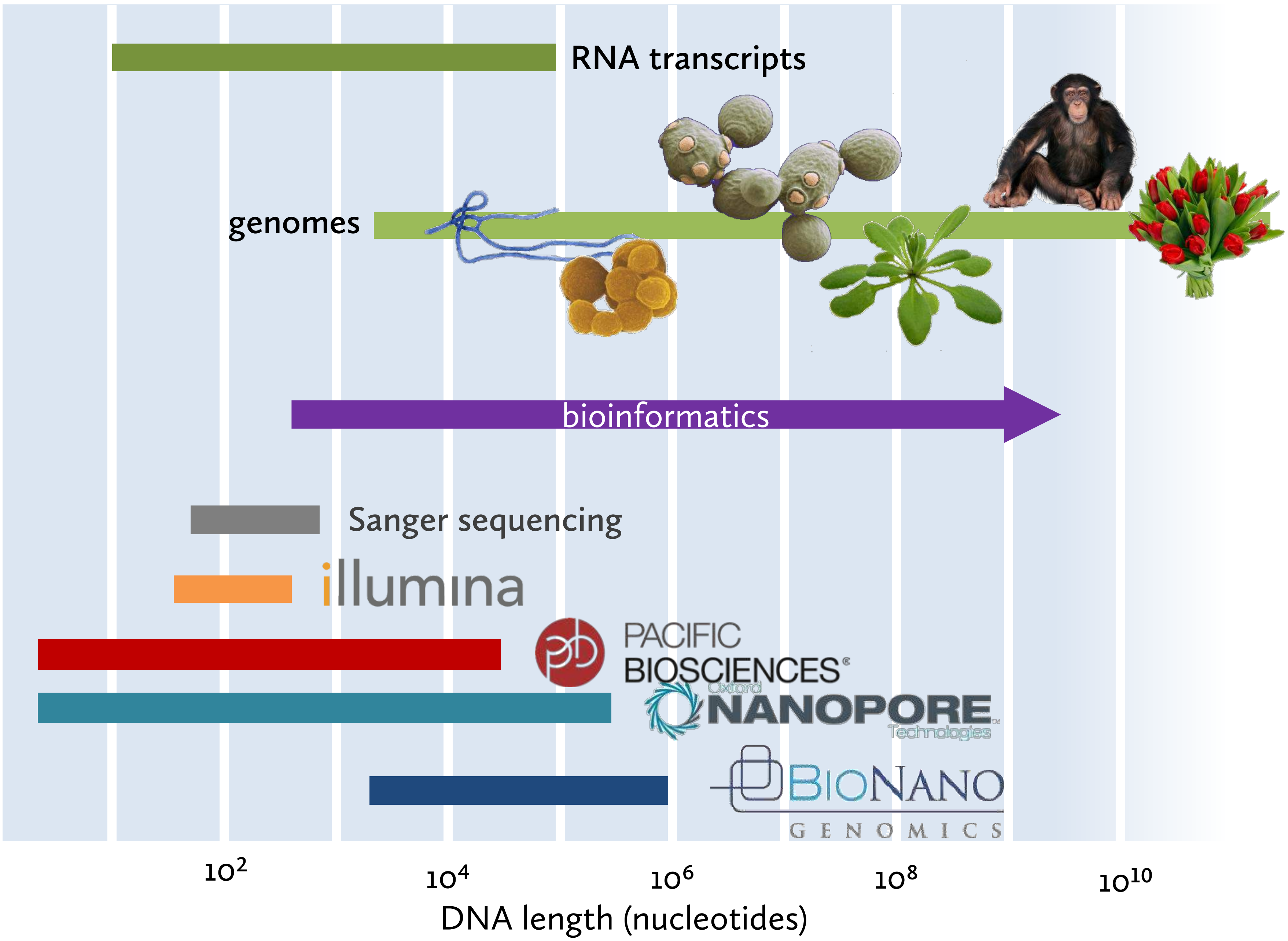
fragments

in vitro

reads

de novo
assembled
genome

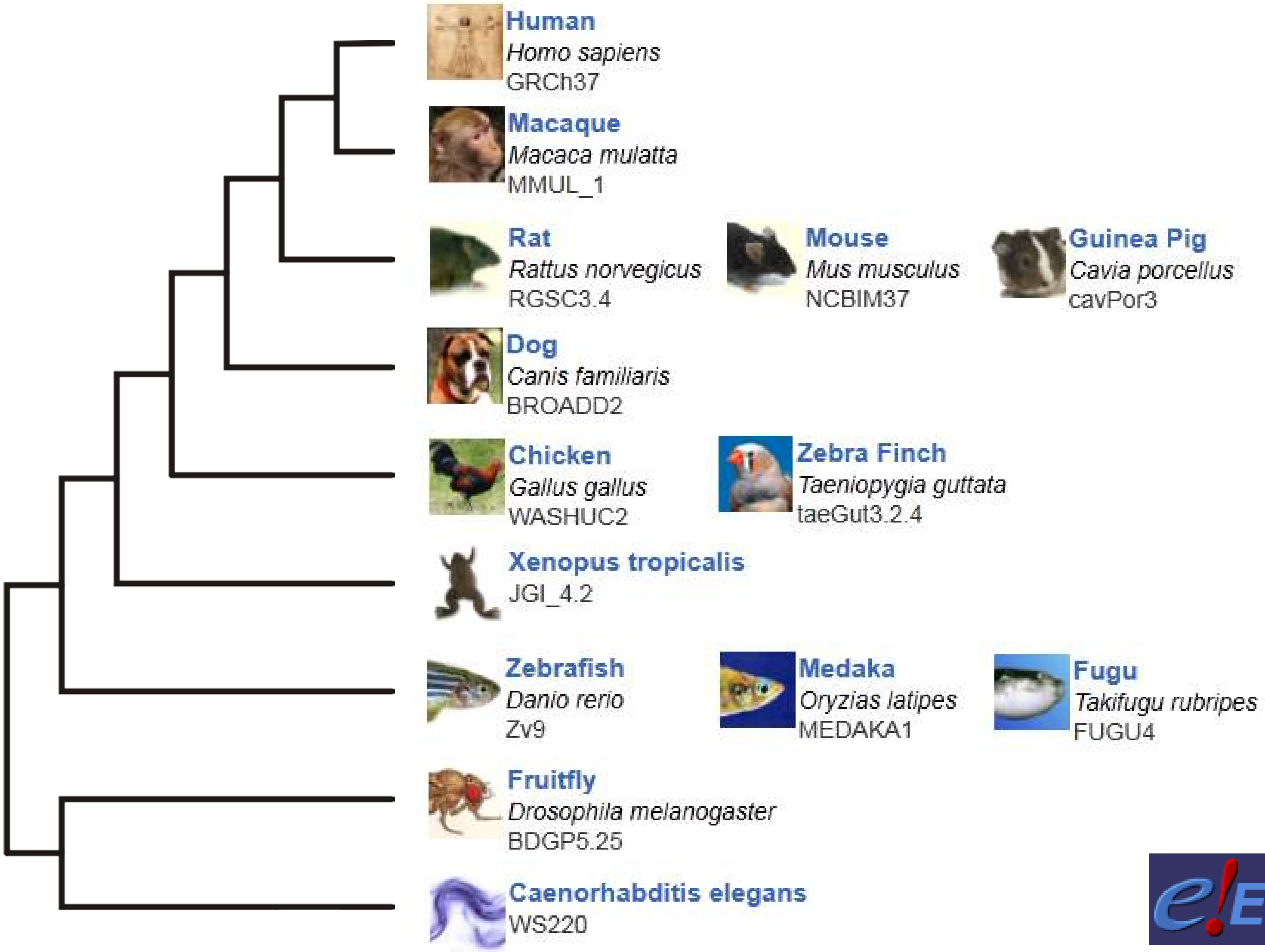
in silico


RNA transcripts
genomes
bioinformatics
Sanger sequencing
illumina
PACIFIC BIOSCIENCES®
Oxford NANOPORE Technologies
BIONANO GENOMICS
10^2
10^4
10^6
10^8
10^{10}
DNA length (nucleotides)

From genome to contigs to scaffolds

- Genomics technology
- Genomics applications
- Genomics in aquaculture
  - Fish genome sequencing
  - Carp genomes and transcriptomes
  - Rainbow trout genome
- The genomics of eel migration

Model organisms



Human
Homo sapiens
GRCh37

Macaque
Macaca mulatta
MMUL_1

Rat
Rattus norvegicus
RGSC3.4

Mouse
Mus musculus
NCBIM37

Guinea Pig
Cavia porcellus
cavPor3

Dog
Canis familiaris
BROADD2

Chicken
Gallus gallus
WASHUC2

Zebra Finch
Taeniopygia guttata
taeGut3.2.4

Xenopus tropicalis
JGI_4.2

Zebrafish
Danio rerio
Zv9

Medaka
Oryzias latipes
MEDAKA1

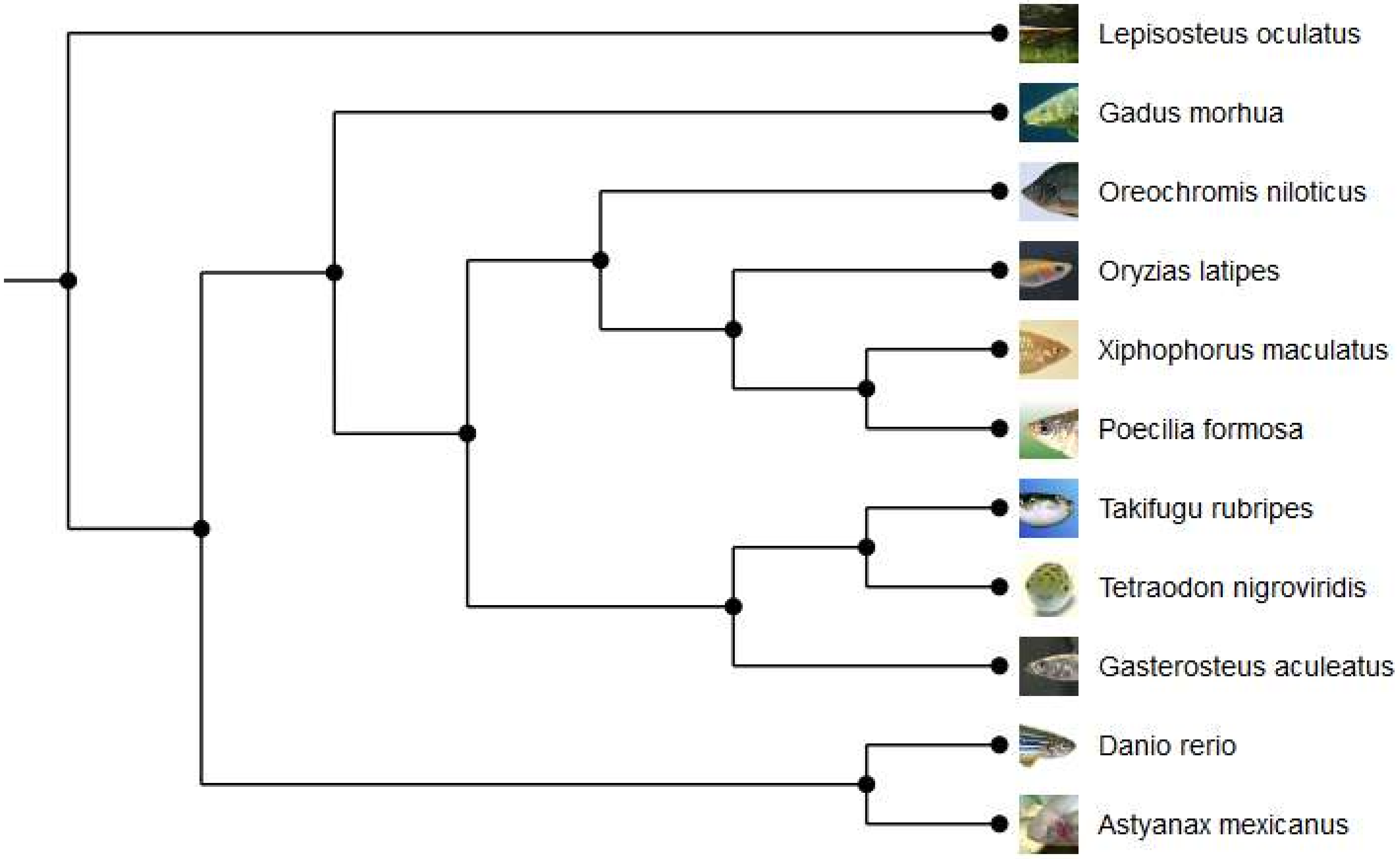
Fugu
Takifugu rubripes
FUGU4

Fruitfly
Drosophila melanogaster
BDGP5.25

Caenorhabditis elegans
WS220

e!Ensembl

Fish genomes



Lepisosteus oculatus

Gadus morhua

Oreochromis niloticus

Oryzias latipes

Xiphophorus maculatus

Poecilia formosa

Takifugu rubripes

Tetraodon nigroviridis

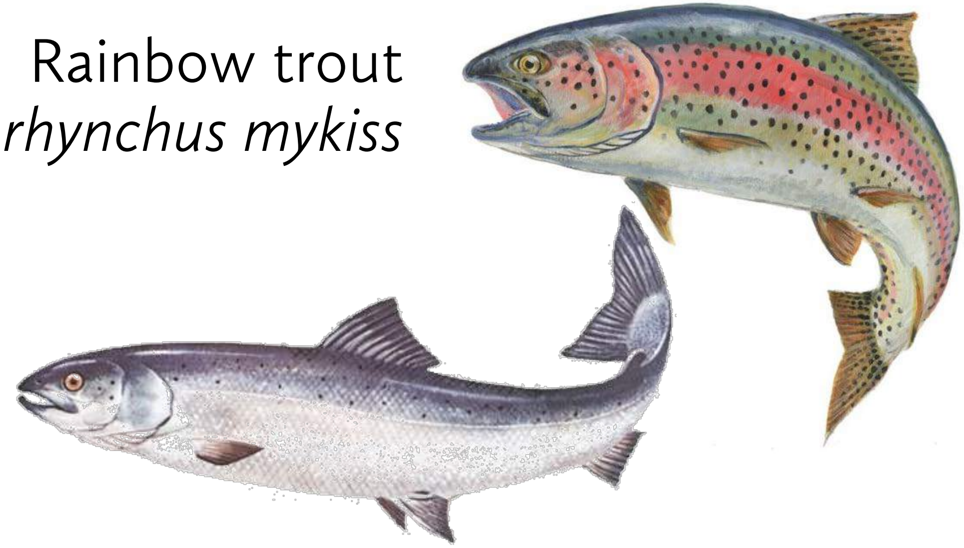
Gasterosteus aculeatus

Danio rerio

Astyanax mexicanus

e!Ensembl

More fish genomes



Rainbow trout
Oncorhynchus mykiss

Atlantic salmon
Salmo salar

Common carp
Cyprinus carpio

European eel
Anguilla anguilla

Common carp

Herman Spaink (Leiden)
Geert Wiegertjes (Wageningen)
Ron Dirks (ZF-screens)

© Wikipedia

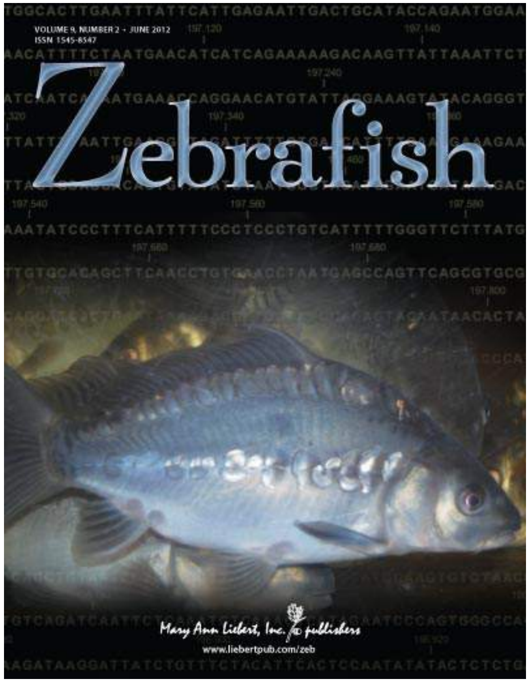
The common carp genome

- Carp is a close relative of zebrafish
- Important commercial species (aquaculture)
- Used for immunological studies
- Genome sequenced to support RNA-seq studies
- Carp has experienced a (full?) genome duplication (~12 Mya)

The common carp genome

- Homozygous, pseudotetraploid individual
- Genome size: ~1.5 Gb
- Scaffold N50: 17 Kb (not so good)

- Zebrafish genome size: 1.4–1.5 Gb
- Zebrafish repeat content: ~52%
- Carp repeat content: ~16%



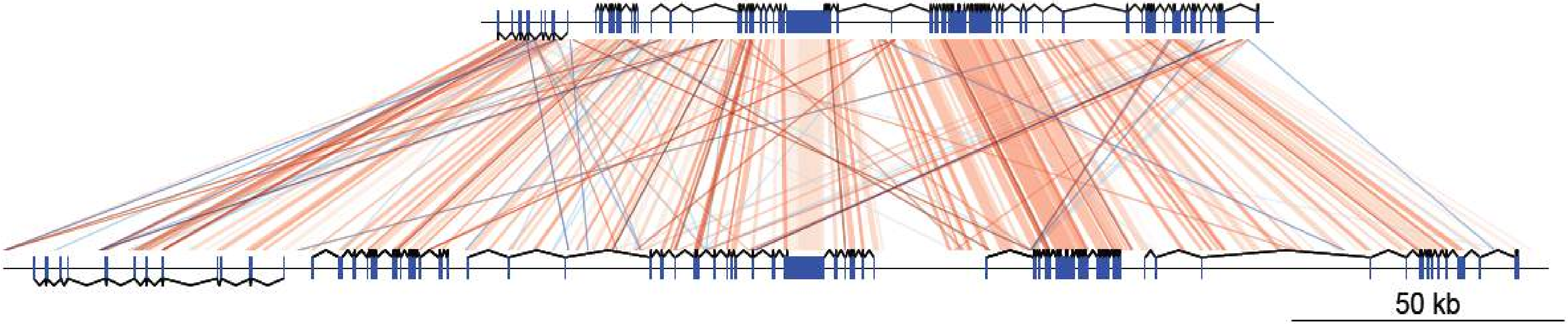
VOLUME 9, NUMBER 2 • JUNE 2012
ISSN 1545-8547

Zebrafish

Mary Ann Liebert, Inc. publishers
www.liebertpub.com/zeb

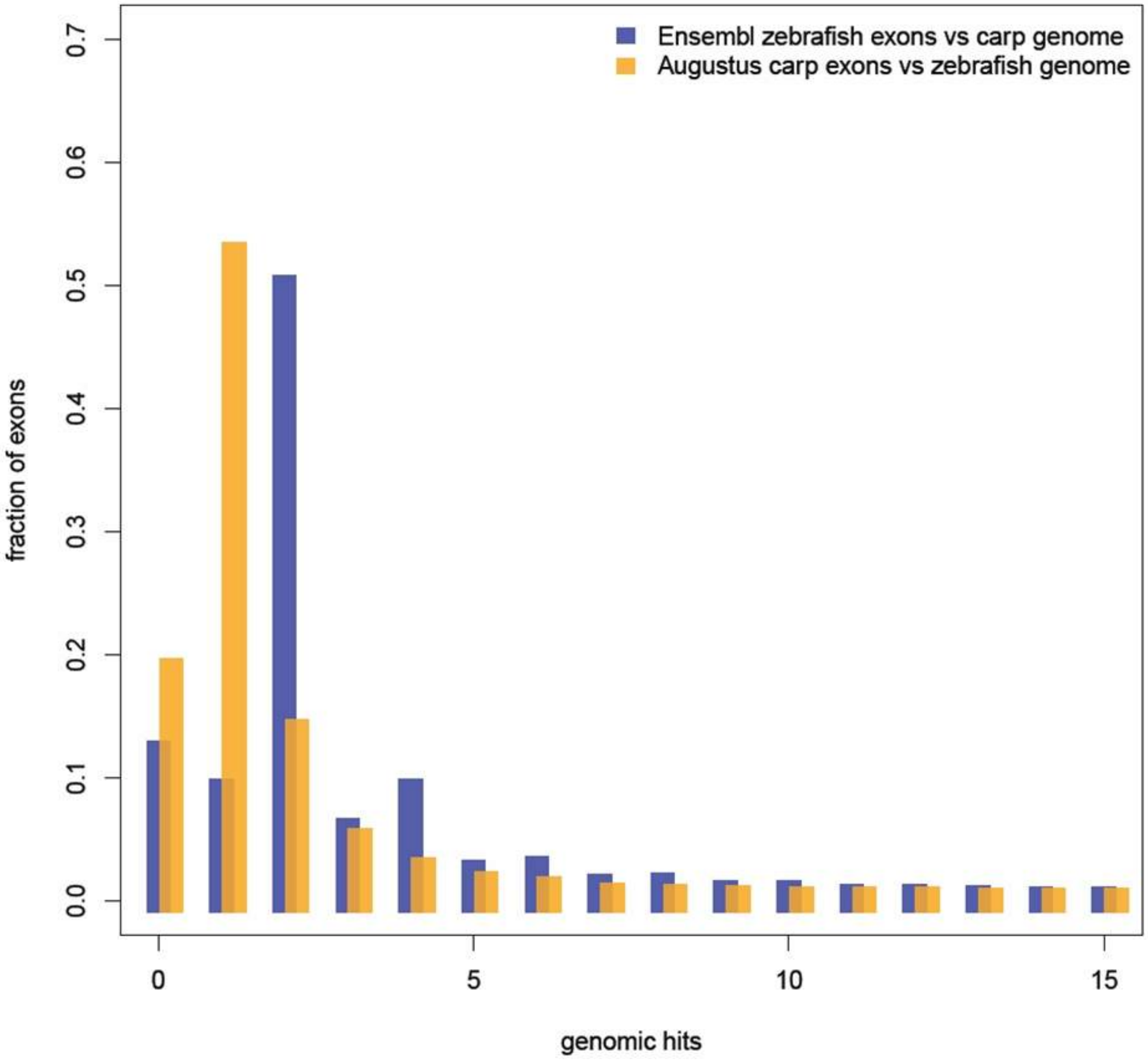
Comparing genomes

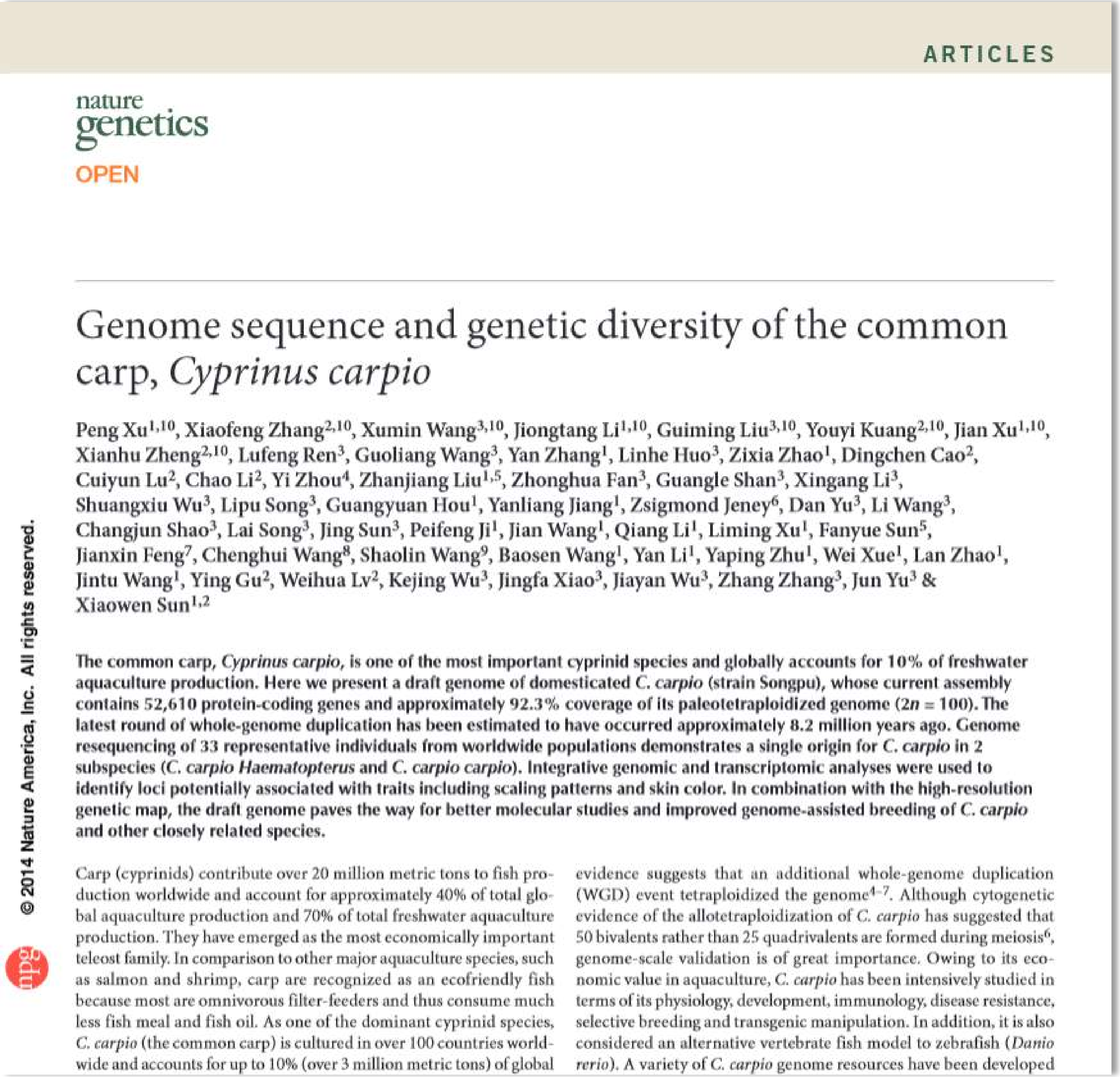
carp



zebrafish

Genome duplication





ARTICLES

nature genetics

OPEN

Genome sequence and genetic diversity of the common carp, *Cyprinus carpio*

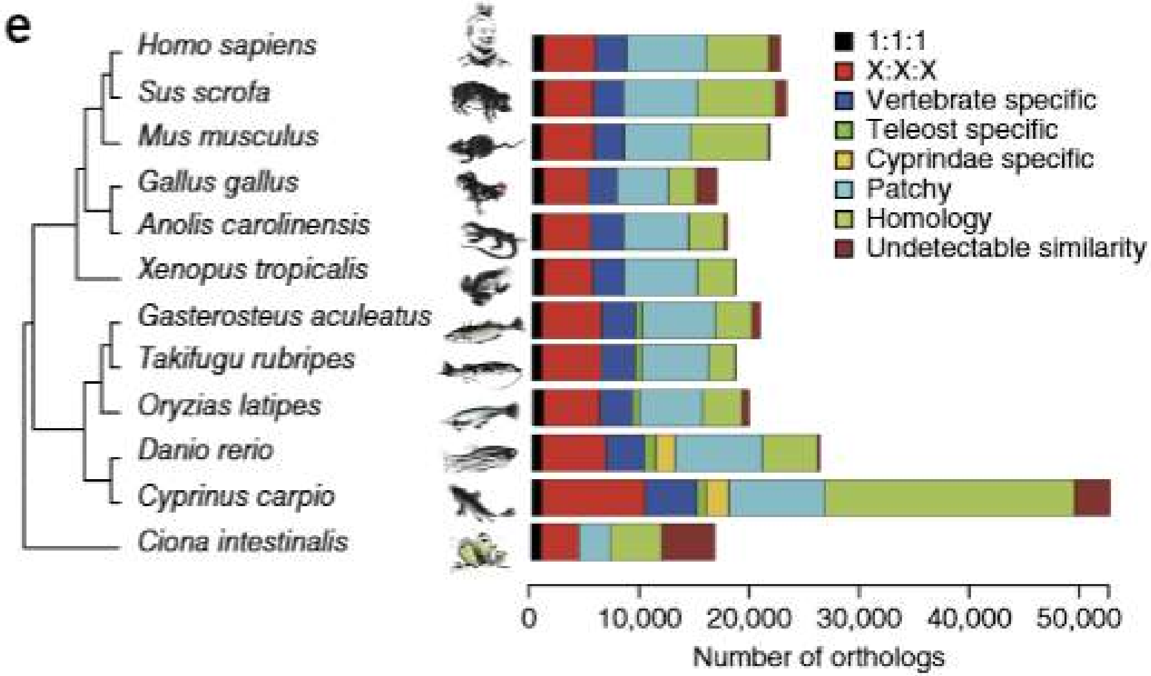
Peng Xu[1,10], Xiaofeng Zhang[2,10], Xumin Wang[3,10], Jiongtang Li[1,10], Guiming Liu[3,10], Youyi Kuang[2,10], Jian Xu[1,10], Xianhu Zheng[2,10], Lufeng Ren[3], Guoliang Wang[3], Yan Zhang[1], Linhe Huo[3], Zixia Zhao[1], Dingchen Cao[2], Cuiyun Lu[2], Chao Li[2], Yi Zhou[4], Zhanjiang Liu[1,5], Zhonghua Fan[3], Guangle Shan[3], Xingang Li[3], Shuangxiu Wu[3], Lipu Song[3], Guangyuan Hou[1], Yanliang Jiang[1], Zsigmond Jeney[6], Dan Yu[3], Li Wang[3], Changjun Shao[3], Lai Song[3], Jing Sun[3], Peifeng Ji[1], Jian Wang[1], Qiang Li[1], Liming Xu[1], Fanyue Sun[5], Jianxin Feng[7], Chenghui Wang[8], Shaolin Wang[9], Baosen Wang[1], Yan Li[1], Yaping Zhu[1], Wei Xue[1], Lan Zhao[1], Jintu Wang[1], Ying Gu[2], Weihua Lv[2], Kejing Wu[3], Jingfa Xiao[3], Jiayan Wu[3], Zhang Zhang[3], Jun Yu[3] & Xiaowen Sun[1,2]

The common carp, *Cyprinus carpio*, is one of the most important cyprinid species and globally accounts for 10% of freshwater aquaculture production. Here we present a draft genome of domesticated *C. carpio* (strain Songpu), whose current assembly contains 52,610 protein-coding genes and approximately 92.3% coverage of its paleotetraploidized genome ($2n = 100$). The latest round of whole-genome duplication has been estimated to have occurred approximately 8.2 million years ago. Genome resequencing of 33 representative individuals from worldwide populations demonstrates a single origin for *C. carpio* in 2 subspecies (*C. carpio Haematopterus* and *C. carpio carpio*). Integrative genomic and transcriptomic analyses were used to identify loci potentially associated with traits including scaling patterns and skin color. In combination with the high-resolution genetic map, the draft genome paves the way for better molecular studies and improved genome-assisted breeding of *C. carpio* and other closely related species.

Carp (cyprinids) contribute over 20 million metric tons to fish production worldwide and account for approximately 40% of total global aquaculture production and 70% of total freshwater aquaculture production. They have emerged as the most economically important teleost family. In comparison to other major aquaculture species, such as salmon and shrimp, carp are recognized as an ecofriendly fish because most are omnivorous filter-feeders and thus consume much less fish meal and fish oil. As one of the dominant cyprinid species, *C. carpio* (the common carp) is cultured in over 100 countries worldwide and accounts for up to 10% (over 3 million metric tons) of global

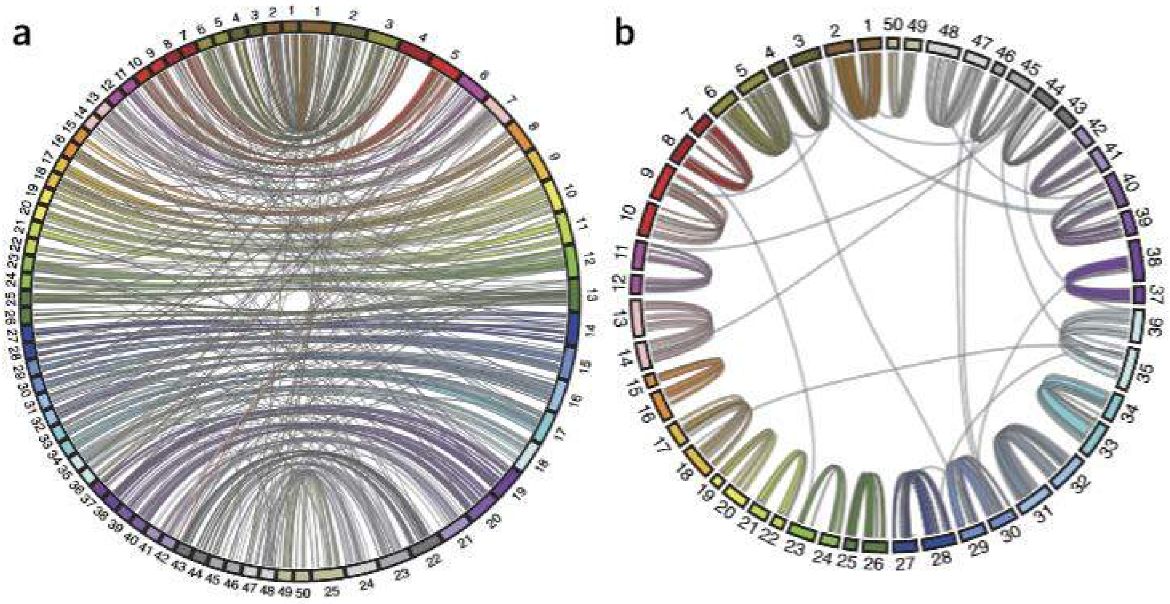
evidence suggests that an additional whole-genome duplication (WGD) event tetraploidized the genome[4–7]. Although cytogenetic evidence of the allotetraploidization of *C. carpio* has suggested that 50 bivalents rather than 25 quadrivalents are formed during meiosis[6], genome-scale validation is of great importance. Owing to its economic value in aquaculture, *C. carpio* has been intensively studied in terms of its physiology, development, immunology, disease resistance, selective breeding and transgenic manipulation. In addition, it is also considered an alternative vertebrate fish model to zebrafish (*Danio rerio*). A variety of *C. carpio* genome resources have been developed

Genome duplication



- Carp has twice as many genes as zebrafish
- Zebrafish have more genes than humans
- Vertebrates have more genes than non-vertebrate chordates

Genome duplication



carp vs. zebrafish

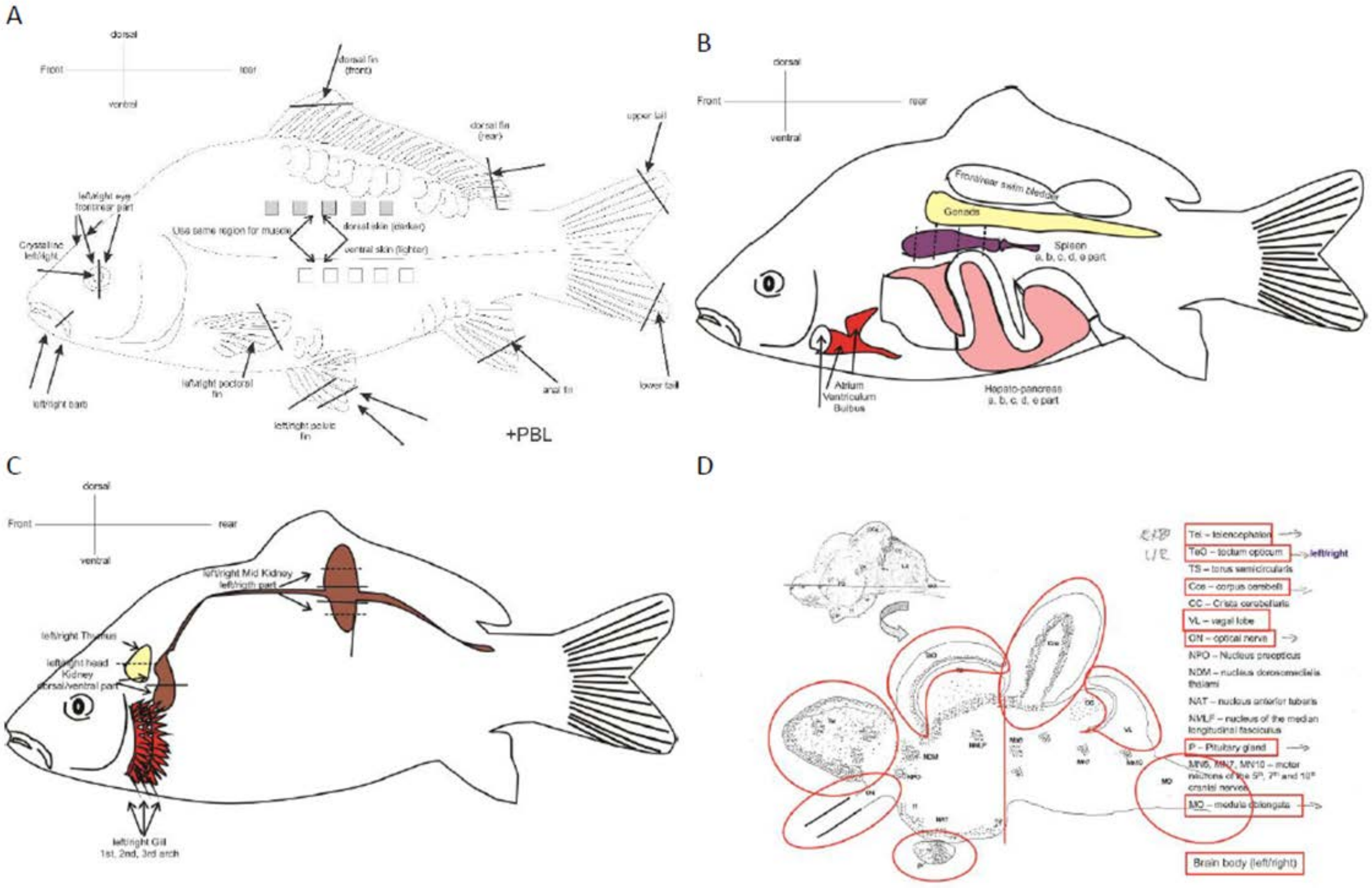
orthologues
homologous genes
separated by
speciation

carp vs. carp

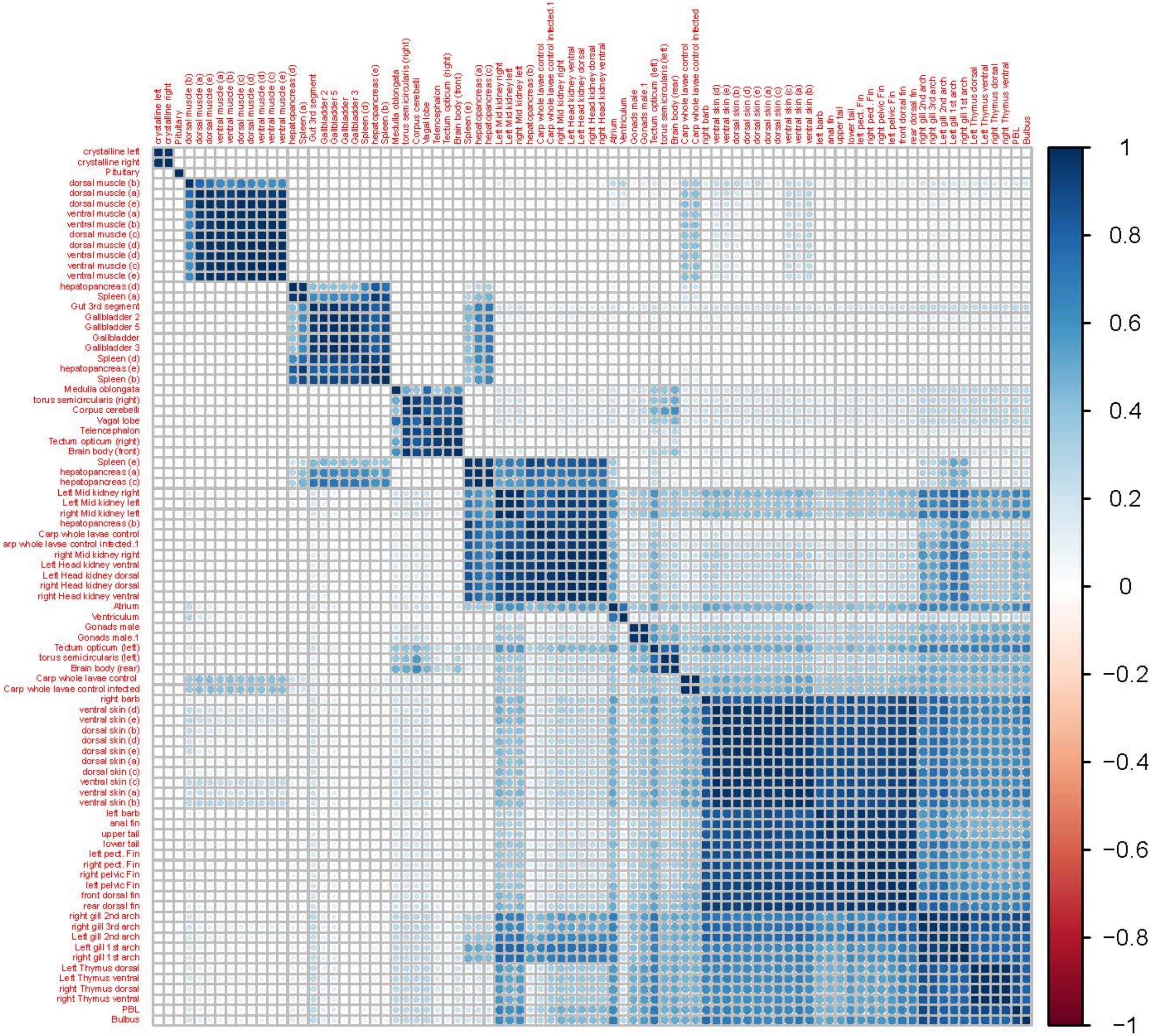
paralogues
homologous genes
separated by
gene duplication

Carp transcriptome atlas

- >100 tissue samples



Carp transcriptome atlas



nature
COMMUNICATIONS

ARTICLE

Received 9 Jan 2014 | Accepted 14 Mar 2014 | Published 22 Apr 2014

DOI: 10.1038/ncomms4657

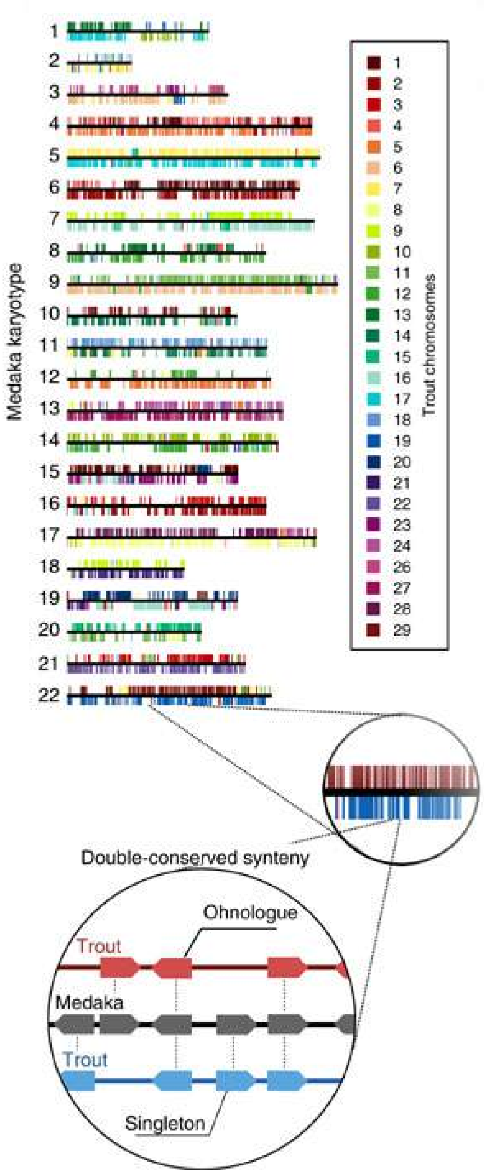
OPEN

The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates

Camille Berthelot[1,2,3,4,5], Frédéric Brunet[6,*], Domitille Chalopin[6,*], Amélie Juanchich[7,*], Maria Bernard[4,8], Benjamin Noël[4], Pascal Bento[4], Corinne Da Silva[4], Karine Labadie[4], Adriana Alberti[4], Jean-Marc Aury[4], Alexandra Louis[1,2,3], Patrice Dehais[9], Philippe Bardou[9], Jérôme Montfort[7], Christophe Klopp[9], Cédric Cabau[9], Christine Gaspin[10,11], Gary H. Thorgaard[12], Mekki Boussaha[8], Edwige Quillet[8], René Guyomard[8], Delphine Galiana[6], Julien Bobe[7], Jean-Nicolas Volff[6], Carine Genêt[8], Patrick Wincker[4,13,14], Olivier Jaillon[4,13,14], Hugues Roest Crollius[1,2,3] & Yann Guiguen[7]

Genome duplication in salmonids

- Salmonids share a whole-genome duplication (~100 Mya)
- This genome duplication is independent from the carp genome duplication



Rainbow trout transcriptomics

- RNA-seq on red and white muscle tissue
- Swimming and resting fish

OPEN ACCESS Freely available online

PLOS | ONE

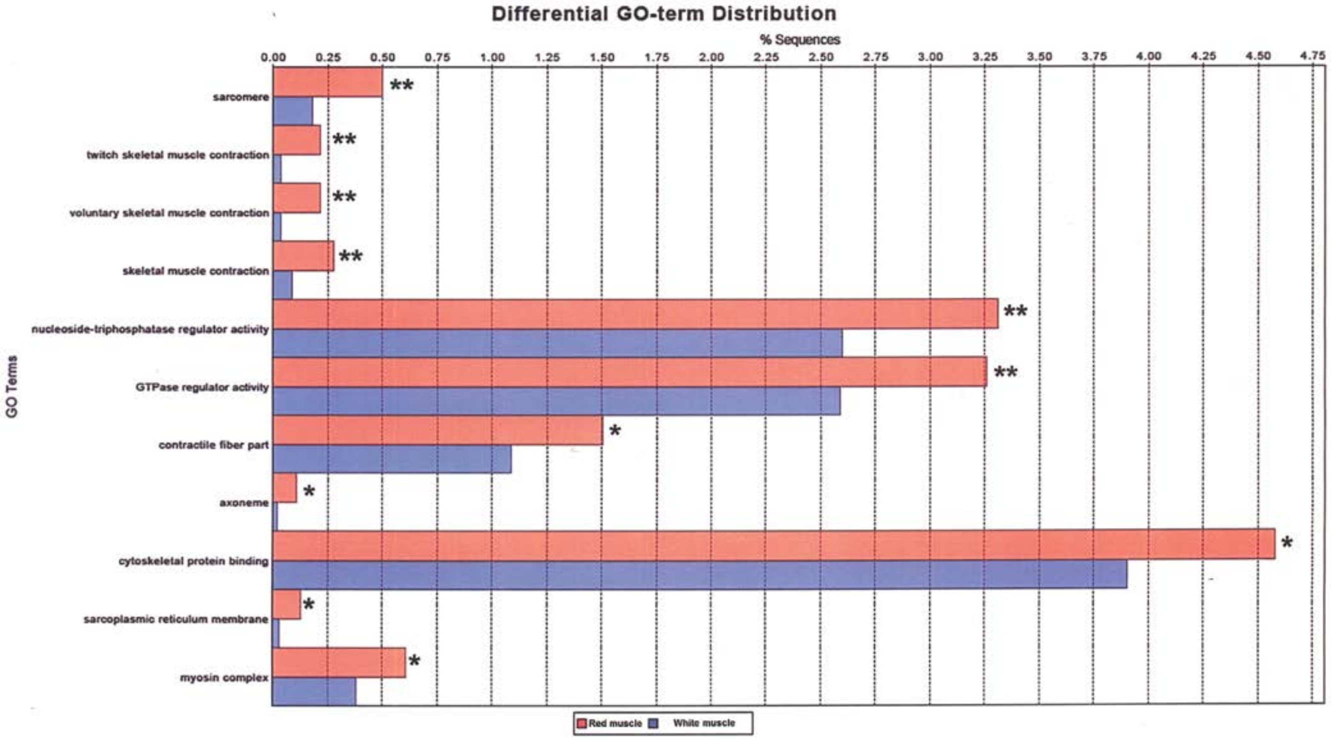
Deep RNA Sequencing of the Skeletal Muscle Transcriptome in Swimming Fish

Arjan P. Palstra[1*¤], Sergi Beltran[2], Erik Burgerhout[3,4], Sebastiaan A. Brittijn[3,4], Leonardo J. Magnoni[1], Christiaan V. Henkel[4], Hans J. Jansen[4], Guido E. E. J. M. van den Thillart[3,4], Herman P. Spaink[3,4], Josep V. Planas[1]

1 Departament de Fisiologia i Immunologia, Facultat de Biologia, Universitat de Barcelona and Institut de Biomedicina de la Universitat de Barcelona (IBUB), Barcelona, Spain, **2** Unitat de Bioinformàtica, Centres Científics i Tecnològics (CCiT-UB), Universitat de Barcelona, Barcelona, Spain, **3** Molecular Cell Biology, Institute of Biology, Leiden University (IBL), Sylvius Laboratory, Leiden, The Netherlands, **4** ZF-screens, Leiden, The Netherlands

Rainbow trout transcriptomics

- RNA-seq on red and white muscle tissue
- Swimmers and resters



Genomics technology

Genomics applications

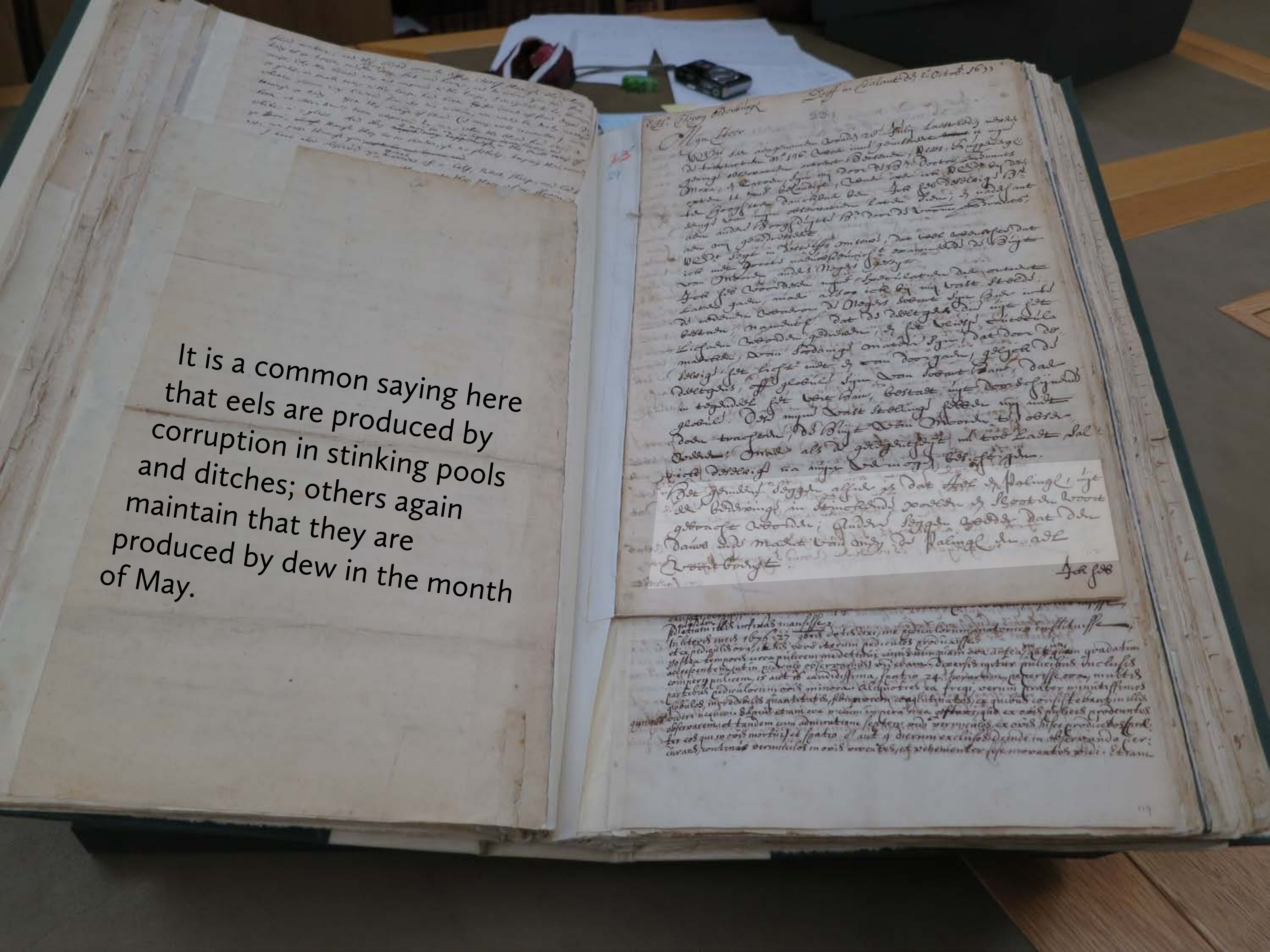
Genomics in aquaculture

The genomics of eel migration

- The eel life cycle
- Population structure (RAD-tags)
- Pituitary transcriptome (RNA-seq)
- Genome duplication

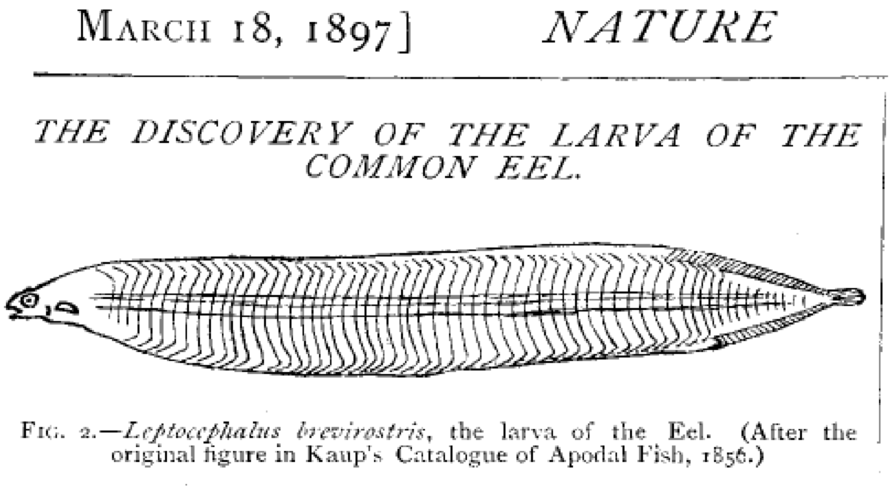
© Arne Kuilman/Flickr

Antoni van Leeuwenhoek (1632–1723)



It is a common saying here that eels are produced by corruption in stinking pools and ditches; others again maintain that they are produced by dew in the month of May.

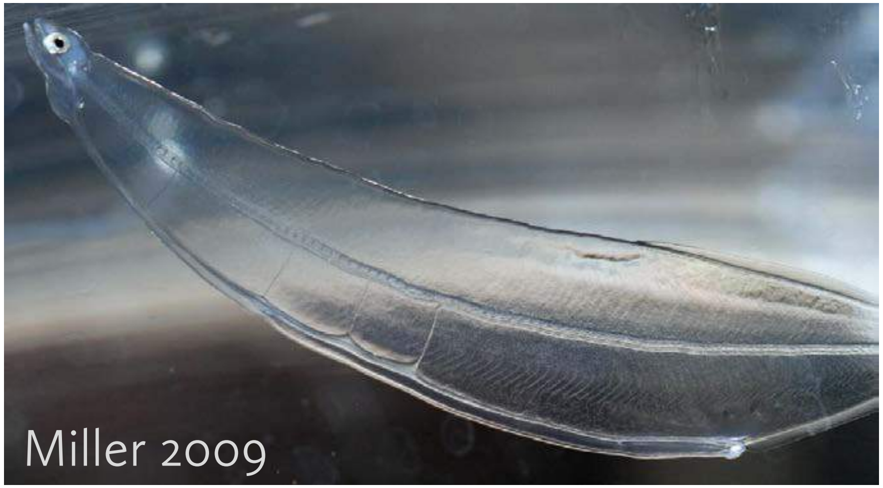
Eel larvae (leptocephali)



MARCH 18, 1897] NATURE

THE DISCOVERY OF THE LARVA OF THE COMMON EEL.

FIG. 2.—*Leptocephalus brevirostris*, the larva of the Eel. (After the original figure in Kaup's Catalogue of Apodal Fish, 1856.)



Miller 2009

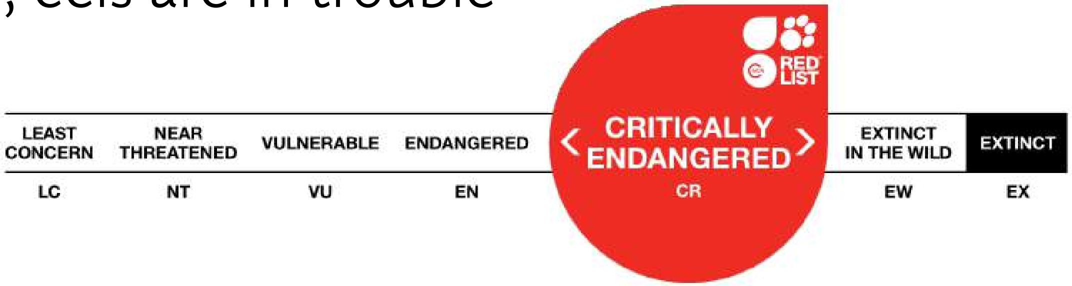
Eel biology

- The eel is a representative of the most primitive teleosts (superorder Elopomorpha, diverged quickly after the teleost-specific genome duplication)
- Elopomorpha exhibit indirect development, with a distinct larval stage (leptocephalus)
- Eel species have a long and complex life cycle
- Only one population (panmictic), but some interbreeding with American eels
- Eels are weird (elongated, toxic, fluorescent)

Why eel genomics?

- The European eel (*Anguilla anguilla*) is an economically important species

- Currently, eels are in trouble



Why eel genomics?

- Eels do not reproduce in captivity
- Genomics tools are being used to study
  - population structure (RAD-tag sequencing)
  - physiology (RNA-seq, microarrays)
  - development and evolution (genome sequencing)

Moeizame palingseks in Volendam

Viskweek De paling wordt met uitsterven bedreigd, en deze vissen zijn niet te kweken. Nederlandse biologen proberen het met zwemtraining en celimplantaten.

The eel life cycle

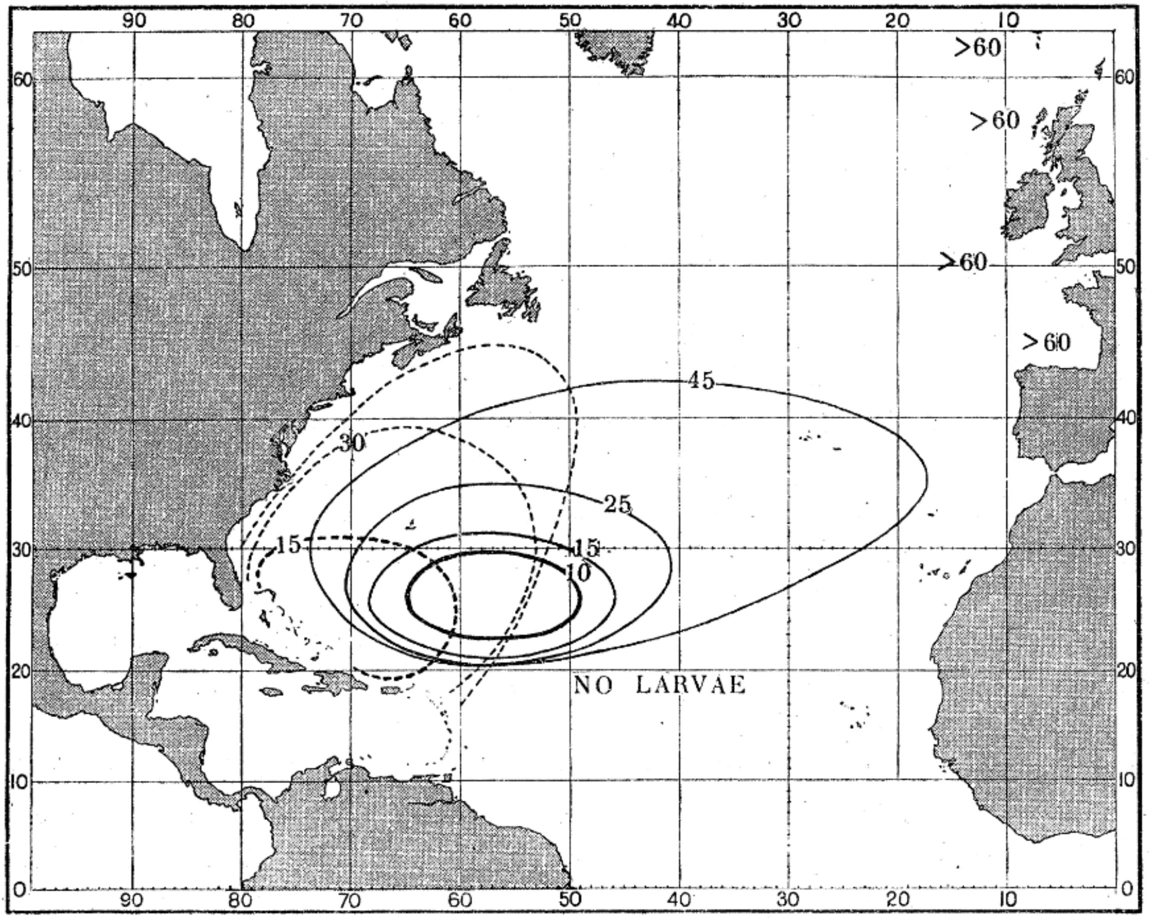
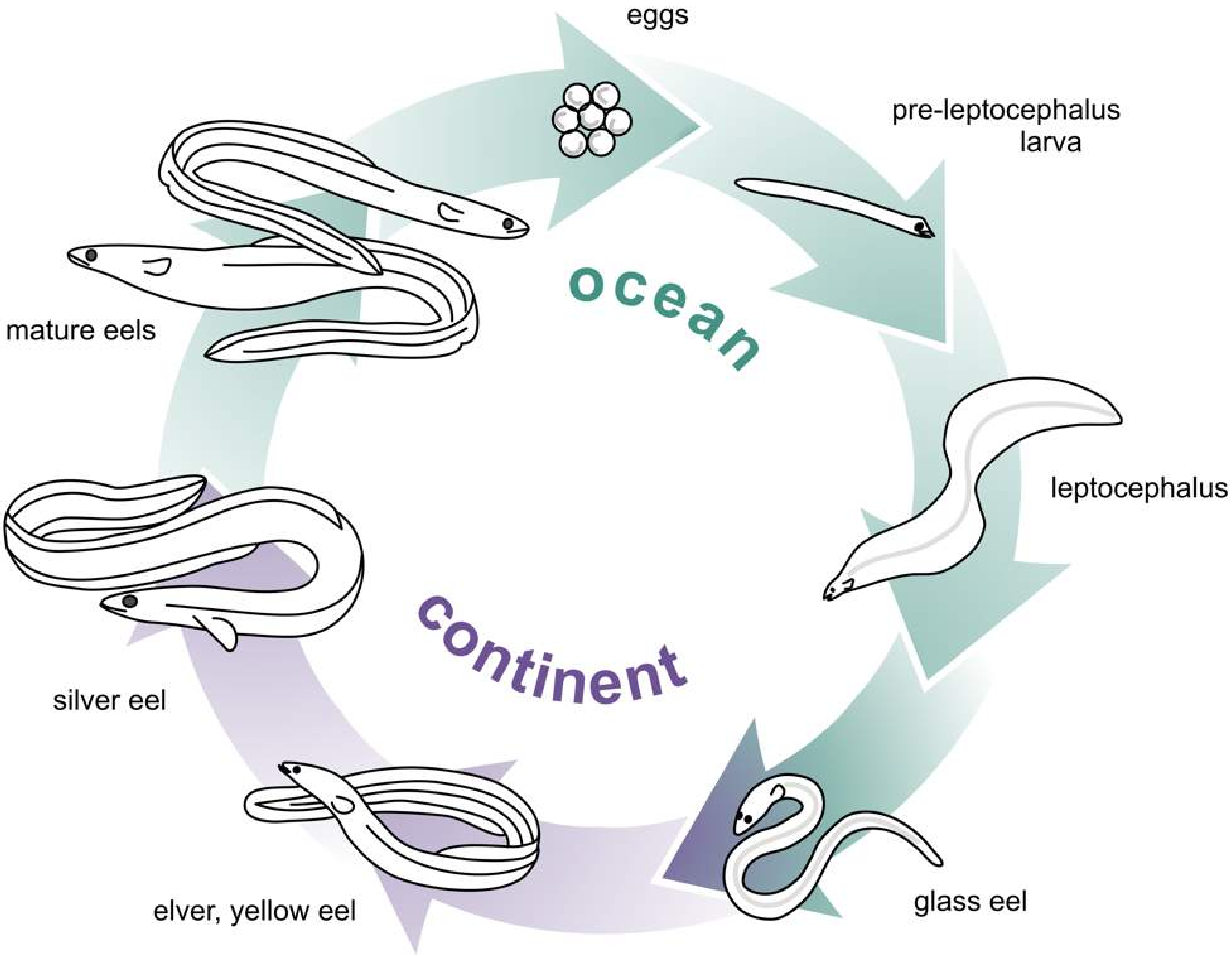


FIG. 4.—European Eel (*Anguilla vulgaris*) and American Eel (*Anguilla rostrata*); breeding areas and distribution of larvæ shown by curves (dotted for the American, continuous for the European species). The heavily-drawn innermost curves embrace the breeding areas of the two species. The curves show limits of occurrence, *i.e.*, specimens less than 25 mm. in length have only been found inside the 25 mm. curve, etc.

Schmidt 1923/1925


eggs
pre-leptocephalus larva
leptocephalus
glass eel
elver, yellow eel
silver eel
mature eels
ocean
continent

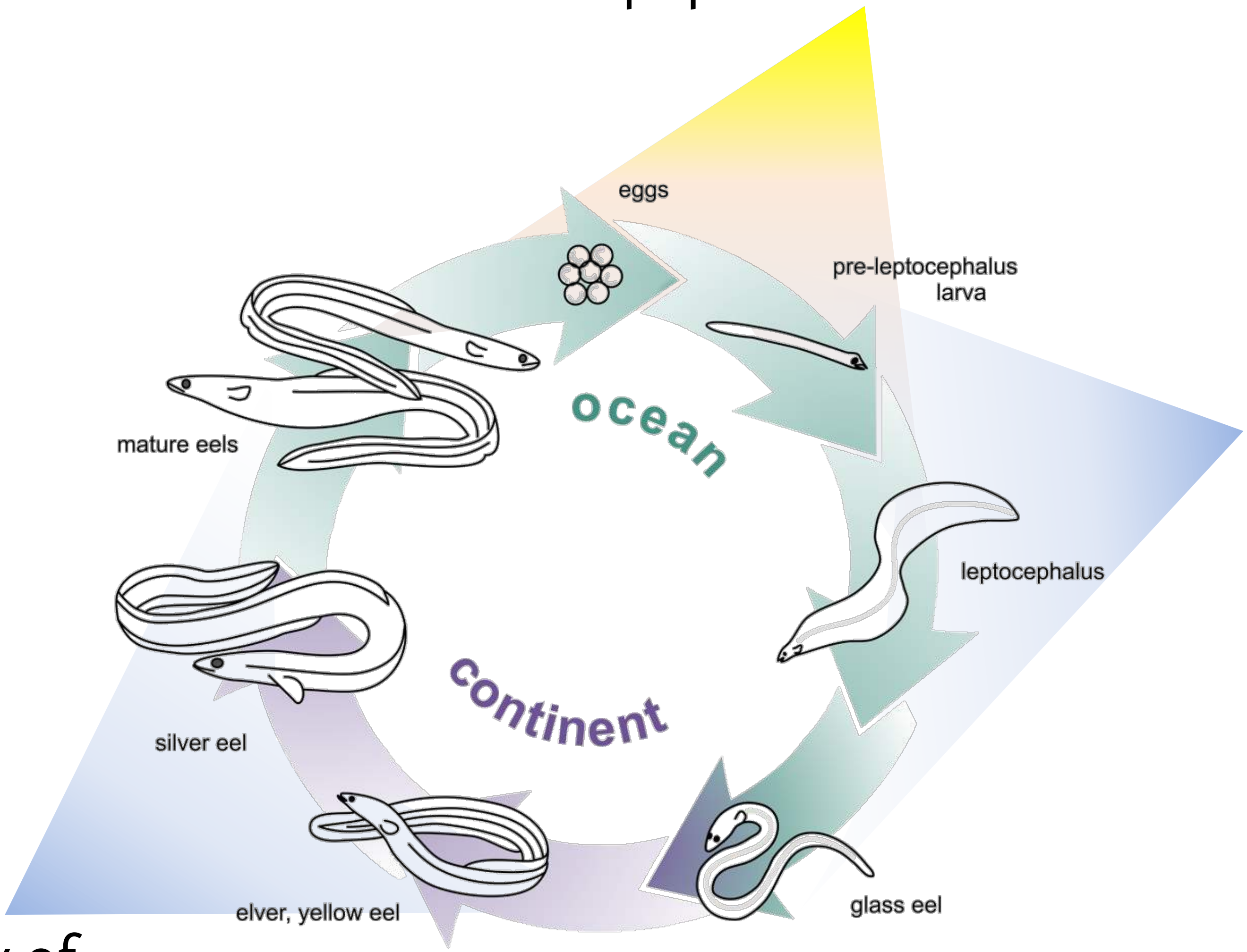
Eel larva,
8 days old

Leptocephalus larvae


Glass eels

© Arne Kuiman/Flickr

Silver eel



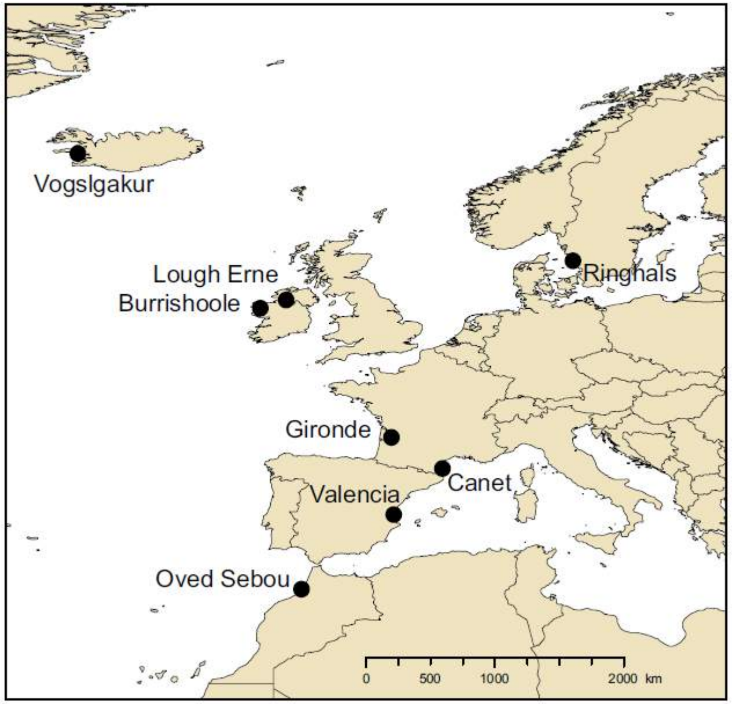
population structure

evo-devo

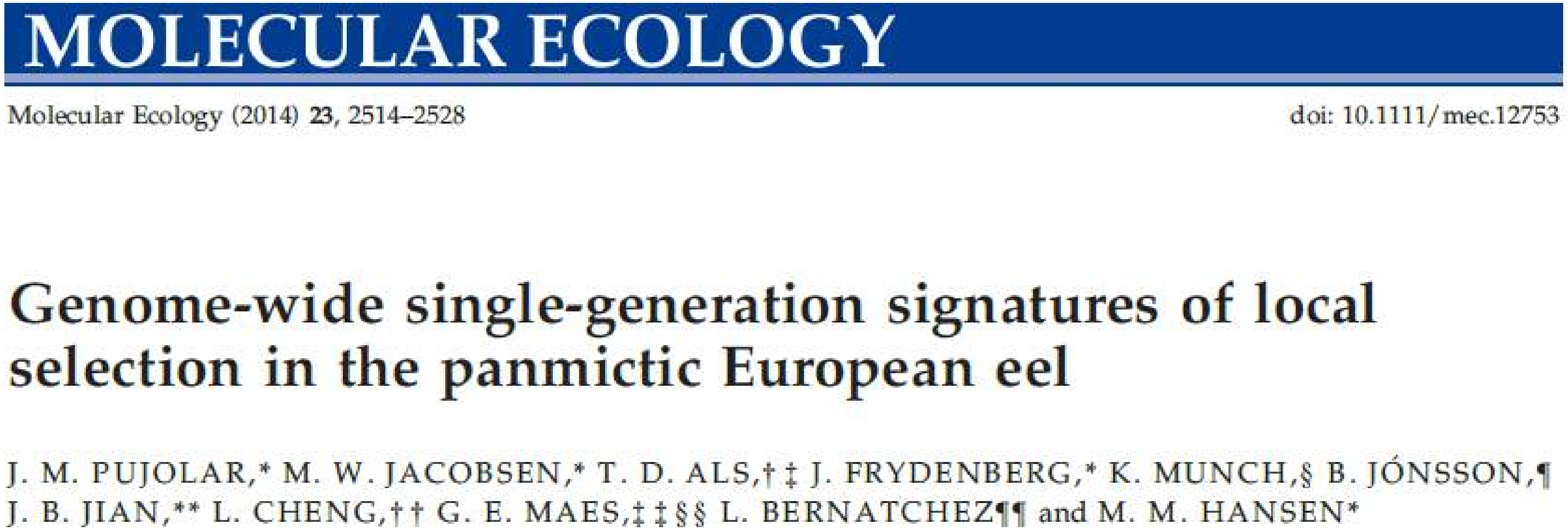
physiology of maturation

Population structure of European eels

- Panmictic or not?
  (i.e. one single interbreeding population, or with substructure?)
- Studied by RAD-tag SNP genotyping
- 259 glass eels from 8 locations
- 50354 SNPs indexed



Population structure of European eels



MOLECULAR ECOLOGY

Molecular Ecology (2014) **23**, 2514–2528 doi: 10.1111/mec.12753

Genome-wide single-generation signatures of local selection in the panmictic European eel

J. M. PUJOLAR,* M. W. JACOBSEN,* T. D. ALS,†‡ J. FRYDENBERG,* K. MUNCH,§ B. JÓNSSON,¶ J. B. JIAN,** L. CHENG,†† G. E. MAES,‡‡§§ L. BERNATCHEZ¶¶ and M. M. HANSEN*

- 5.9% hybrids with American eels (*A. rostrata*) in Iceland, introgression in mainland Europe
- Almost no population substructure, but local selection for calcium signalling, Wnt signalling, circadian rhythm, etc.

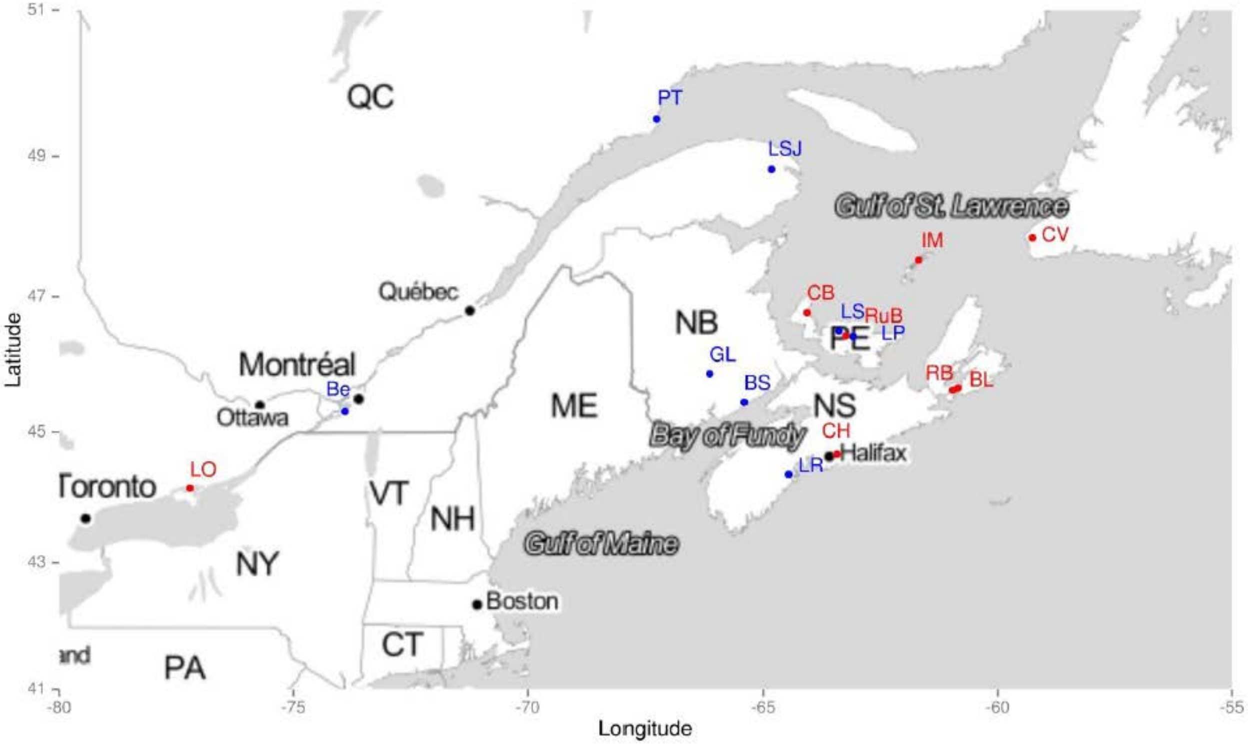
Local selection in American eels

- Unambiguously panmictic
- However, large differences in phenotypes between
  - Freshwater eels: long-lived, large, almost exclusively female
  - Brackish water eels: early migrants, smaller, some males

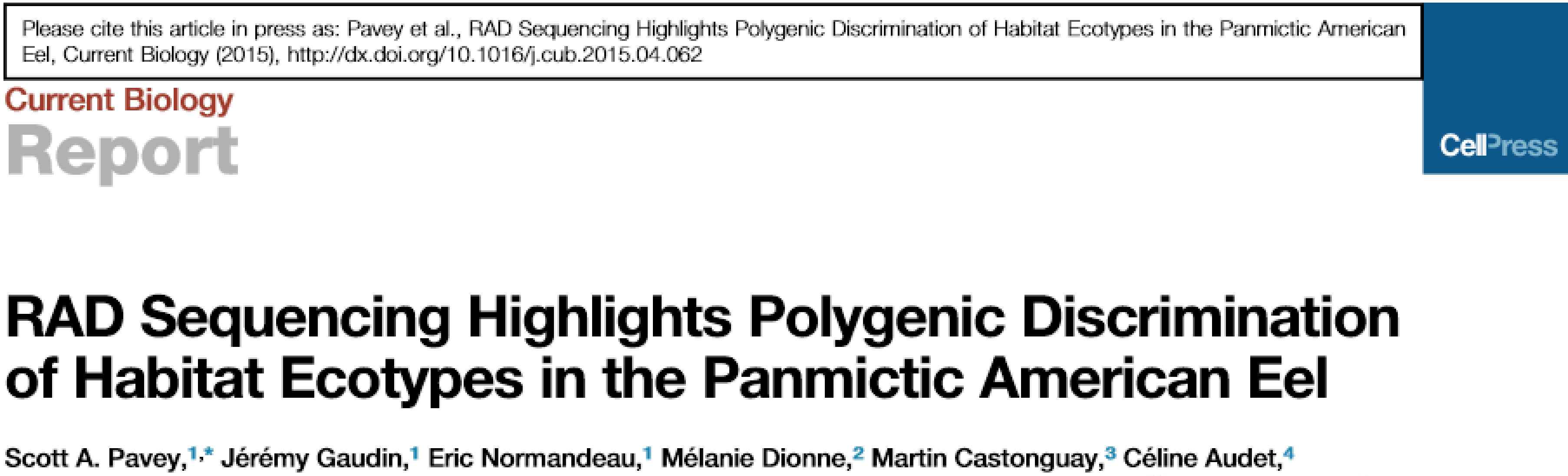
- Phenotypic plasticity or local selection?

Local selection in American eels

- Studied by RAD-tag SNP genotyping
- Eels from 16 locations
- 42424 SNPs indexed



Local selection in American eels



Please cite this article in press as: Pavey et al., RAD Sequencing Highlights Polygenic Discrimination of Habitat Ecotypes in the Panmictic American Eel, Current Biology (2015), http://dx.doi.org/10.1016/j.cub.2015.04.062

Current Biology

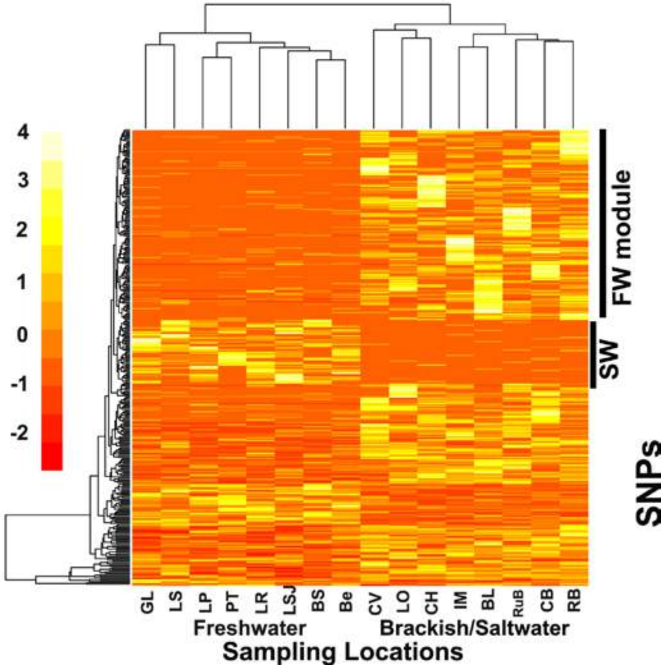
Report

CellPress

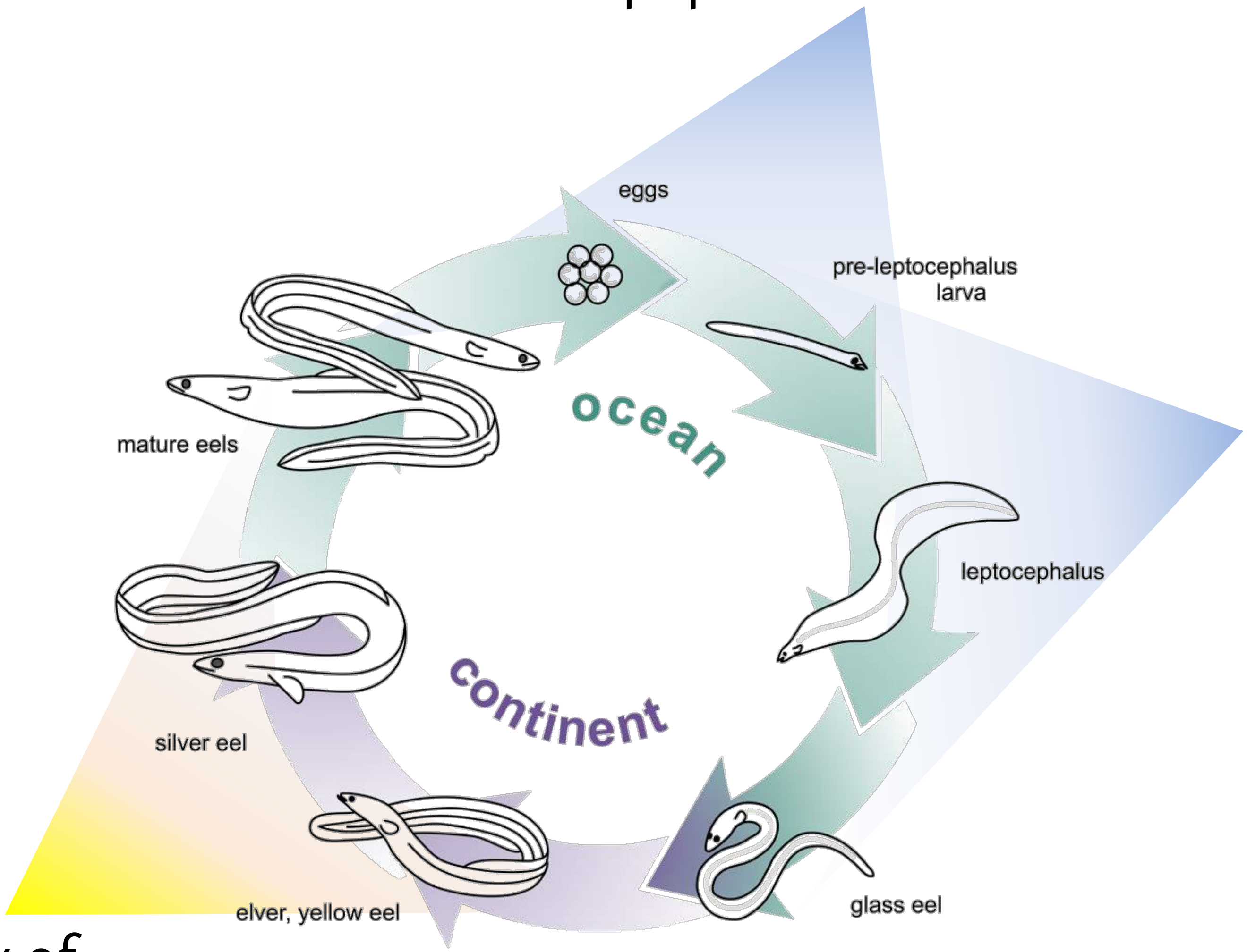
RAD Sequencing Highlights Polygenic Discrimination of Habitat Ecotypes in the Panmictic American Eel

Scott A. Pavey,[1,*] Jérémy Gaudin,[1] Eric Normandeau,[1] Mélanie Dionne,[2] Martin Castonguay,[3] Céline Audet,[4] and Louis Bernatchez[1]

- 331 SNP differences between freshwater/saltwater eels
- Associated genes: developmental, urinary transport, respoiratory system



population structure



evo-devo

physiology of maturation

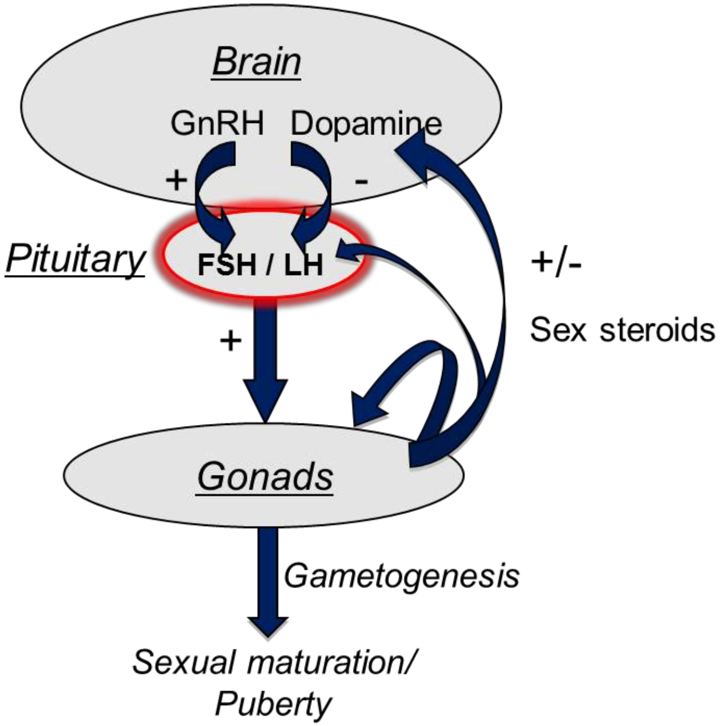
The European eel genome

- Sequenced and assembled in Leiden 2010–2011
- Funding from The Netherlands, Japan, Norway, France
- ~1.1 Gbp genome size
- Scaffold N50 = 78 Kbp
- Fragmented draft assembly because of short read sequencing (Illumina) and high levels of heterozygosity
- (We will try to improve this using nanopore sequencing)

The European eel genome

- Physiology of maturation
  (RNA-seq and microarray studies of gonads, fin clips, pituitary)
- Evolution and development
  (body plan, life cycle, genome evolution)

RNA-seq on the pituitary gland



Brain

GnRH Dopamine

\+

\-

Pituitary

FSH / LH

\+

+/-

Sex steroids

Gonads

Gametogenesis

Sexual maturation/
Puberty

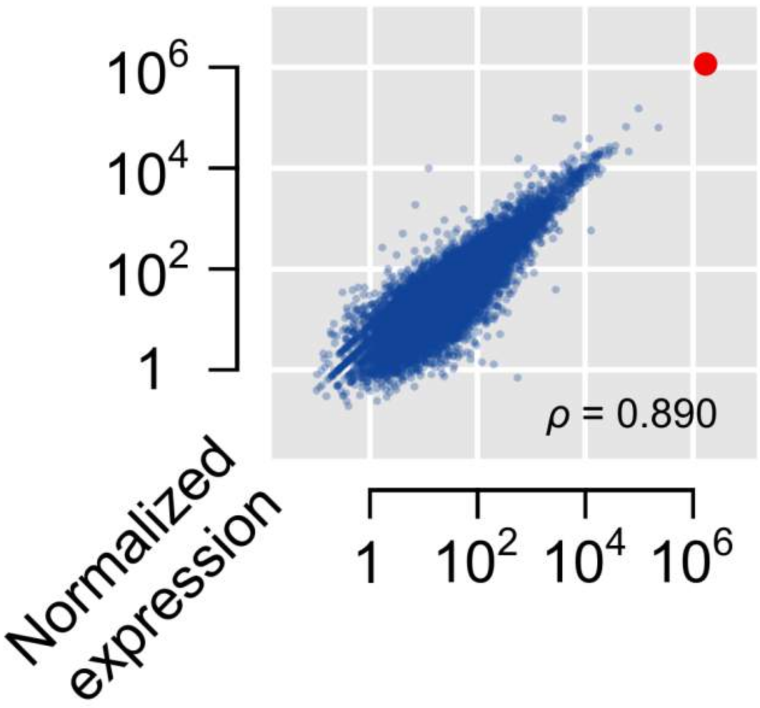
RNA-seq on the pituitary gland

OPEN ACCESS Freely available online

PLOS | ONE

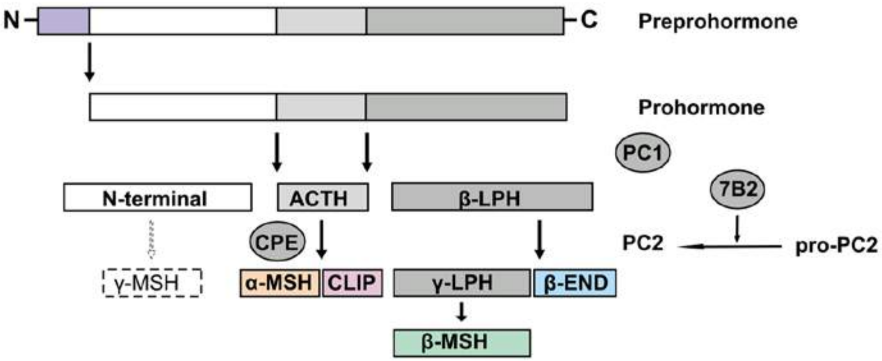
The Pituitary Gland of the European Eel Reveals Massive Expression of Genes Involved in the Melanocortin System

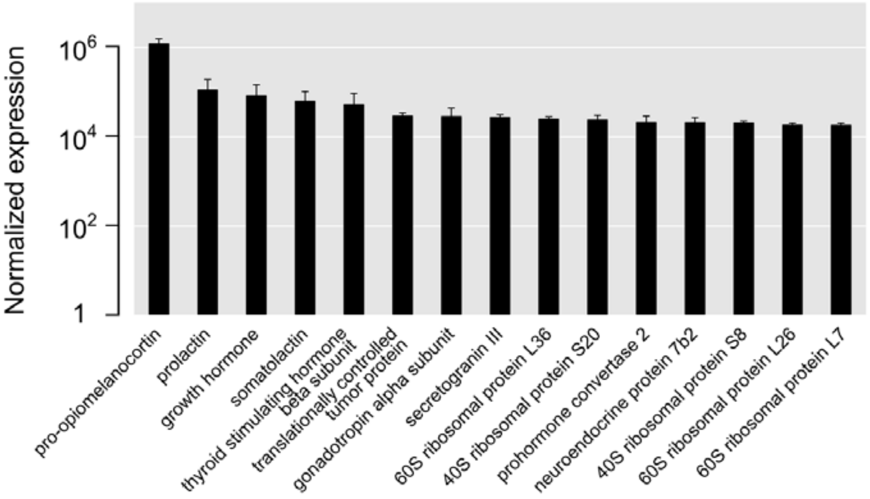
Eirill Ager-Wick[1], Ron P. Dirks[2], Erik Burgerhout[2,3], Rasoul Nourizadeh-Lillabadi[1], Daniëlle L. de Wijze[2,3], Herman P. Spaink[3], Guido E. E. J. M. van den Thillart[2,3], Katsumi Tsukamoto[4], Sylvie Dufour[5], Finn-Arne Weltzien[1], Christiaan V. Henkel[1,2,3]*

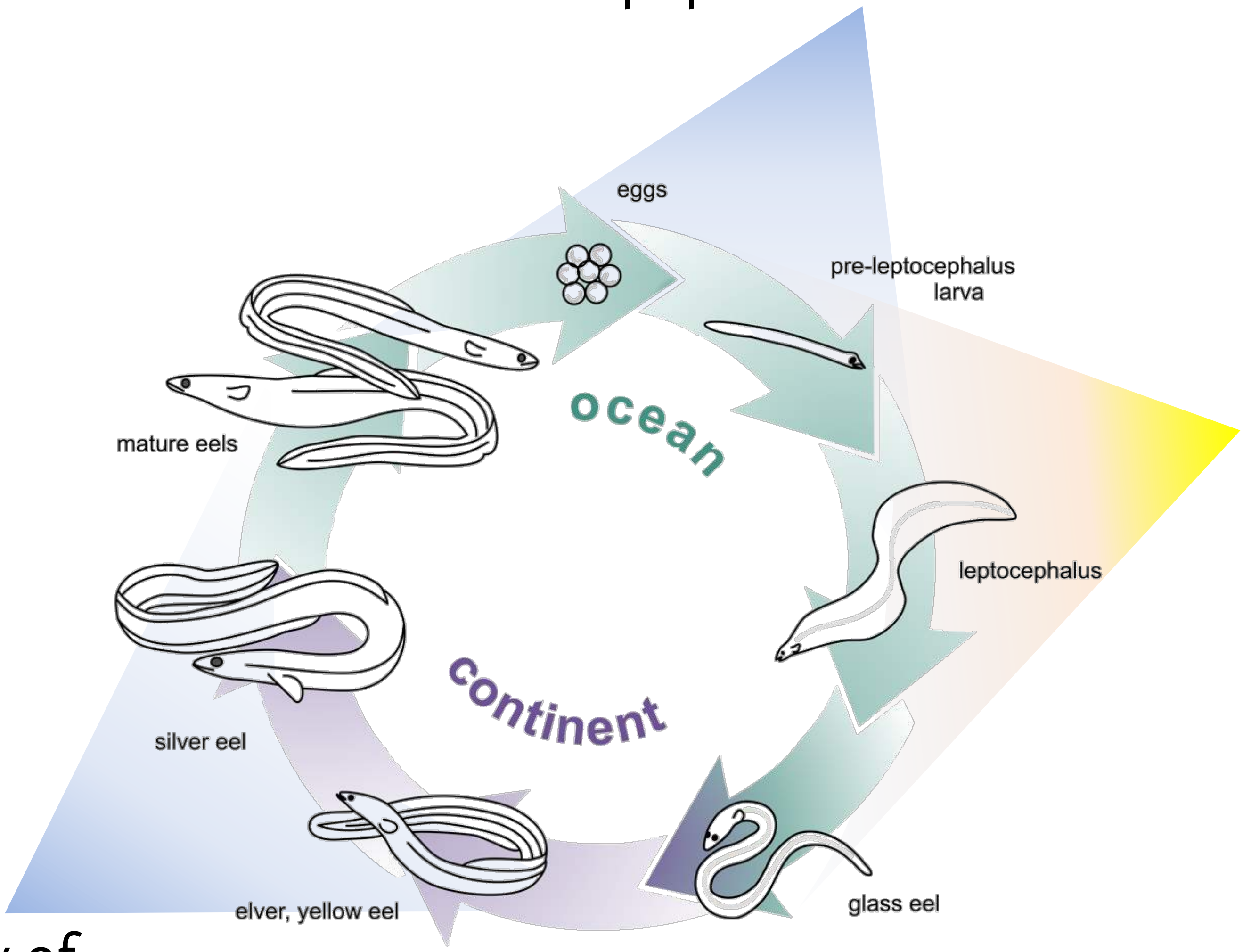


RNA-seq on the pituitary gland

- Silver eel pituitary:
  - >30% of mRNA codes for pro-opiomelanocortin (POMC)
  - POMC affects stress levels, skin colour, energy metabolism
  - Very high expression of POMC-processing enzymes







population structure

evo-devo

physiology of maturation

OPEN ACCESS Freely available online

PLoS one

Primitive Duplicate Hox Clusters in the European Eel's Genome

Christiaan V. Henkel[1]*, Erik Burgerhout[1], Daniëlle L. de Wijze[2], Ron P. Dirks[1], Yuki Minegishi[2], Hans J. Jansen[1], Herman P. Spaink[2], Sylvie Dufour[3], Finn-Arne Weltzien[4], Katsumi Tsukamoto[5], Guido E. E. J. M. van den Thillart[2]

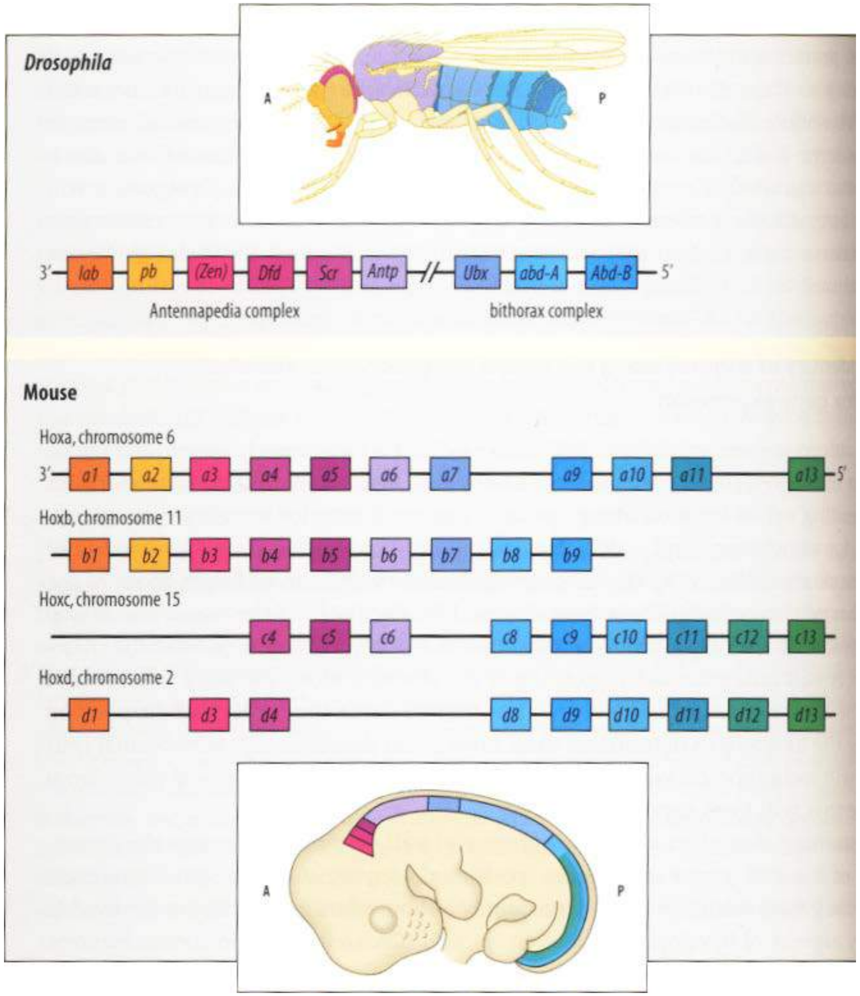
1 ZF-screens B.V., Leiden, The Netherlands, 2 Institute of Biology, Leiden University, Leiden, The Netherlands, 3 UMR BOREA, CNRS 7208, Muséum National d'Histoire Naturelle, Paris, France, 4 Norwegian School of Veterinary Science, Oslo, Norway, 5 Atmosphere and Ocean Research Institute, The University of Tokyo, Kashiwa, Chiba, Tokyo, Japan

Hox genes and clusters

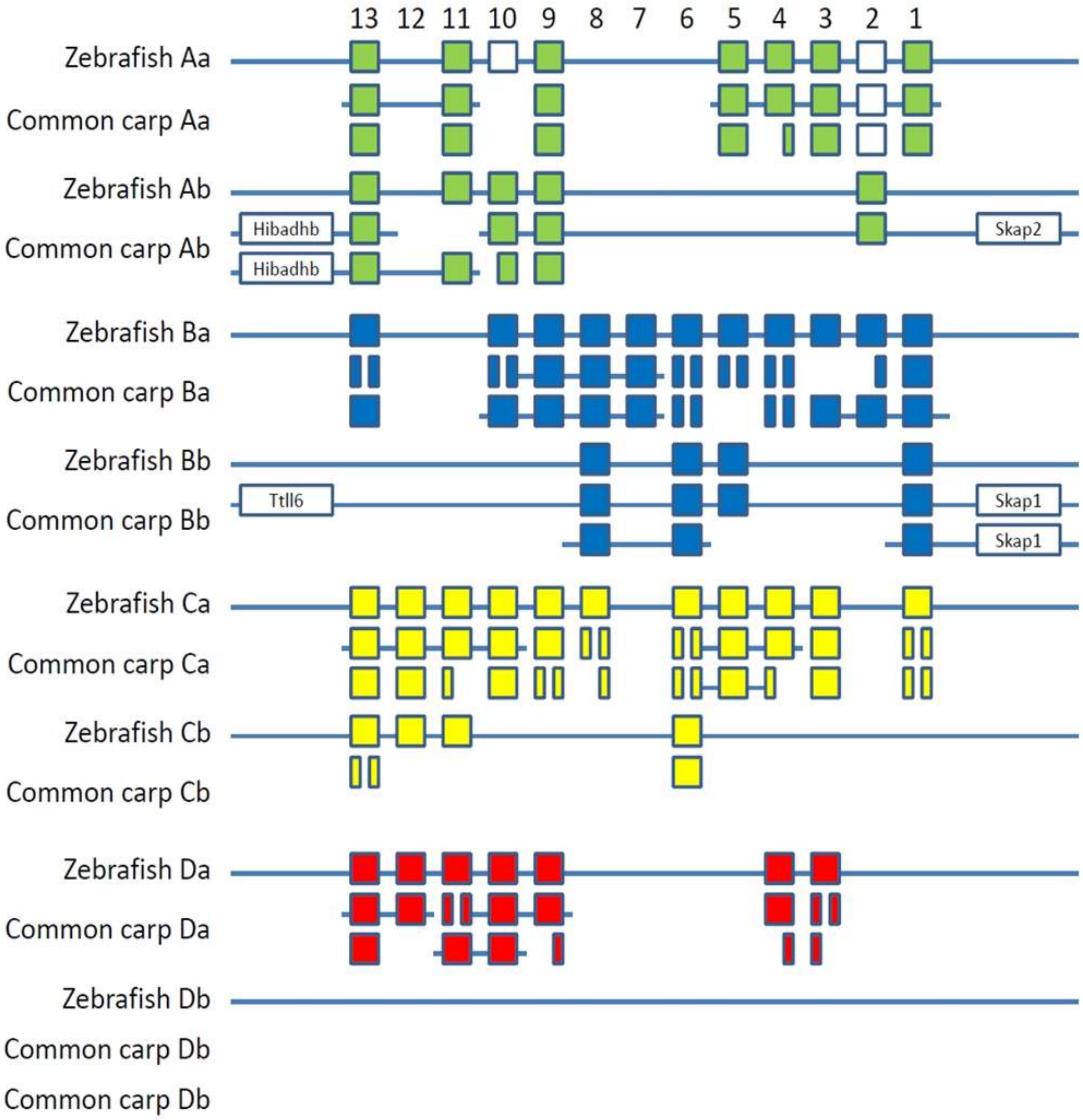
- Hox genes are transcription factors involved in the determination of cell fate during development
- Hox gene count ~ developmental complexity
- Hox genes proliferate through duplication
  (tandem gene duplication and full genome duplication)
- Usually, duplicate genes disappear, unless functional

Hox-clusters

- Hox genes encode transcription factors involved in patterning the embryo (body plan)
- Occur in tightly organized clusters in most animals
- Cluster counts increase through genome duplication
- Tetrapods have four clusters

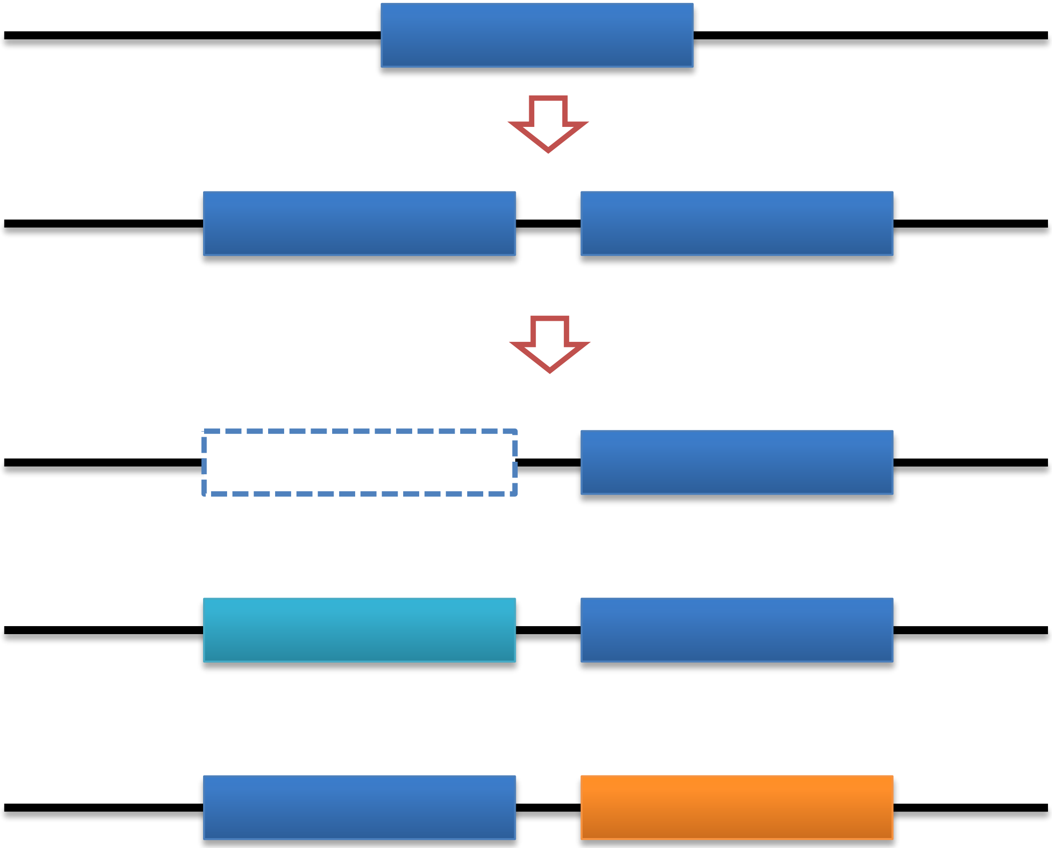


Hox cluster duplications in carp



- Zebrafish:
  51 genes,
  7 clusters
- Carp:
  88 genes
- 40% gene loss
  after duplication?

The fate of duplicated genes



Original gene

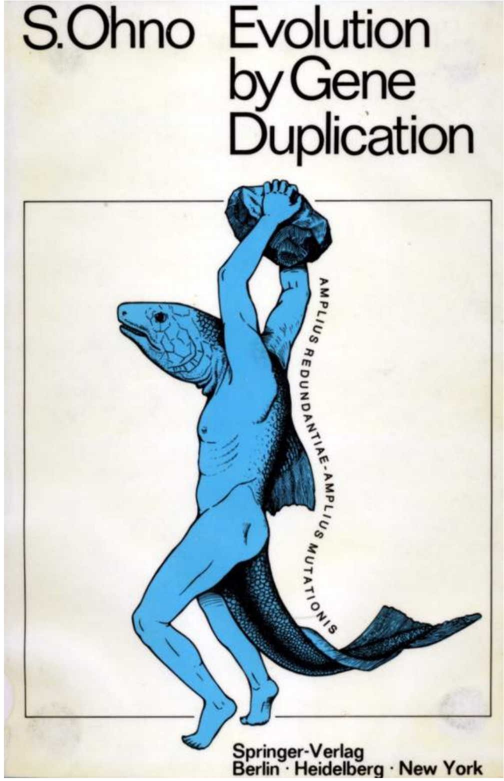
Two copies

One copy disappears

Subfunctionalization

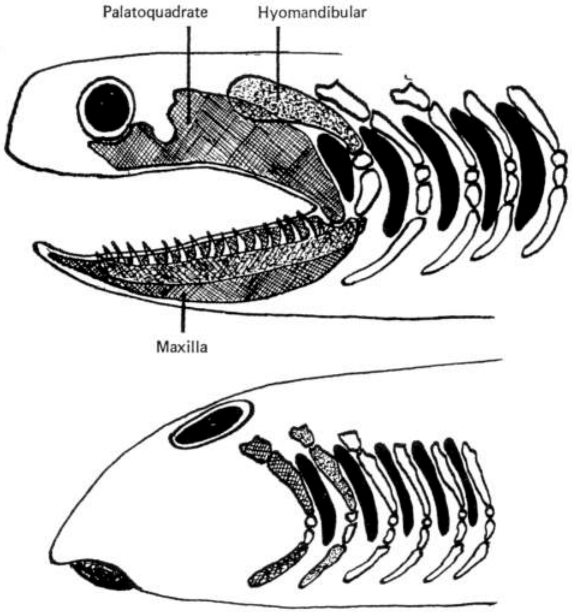
Neofunctionalization

Gene duplication



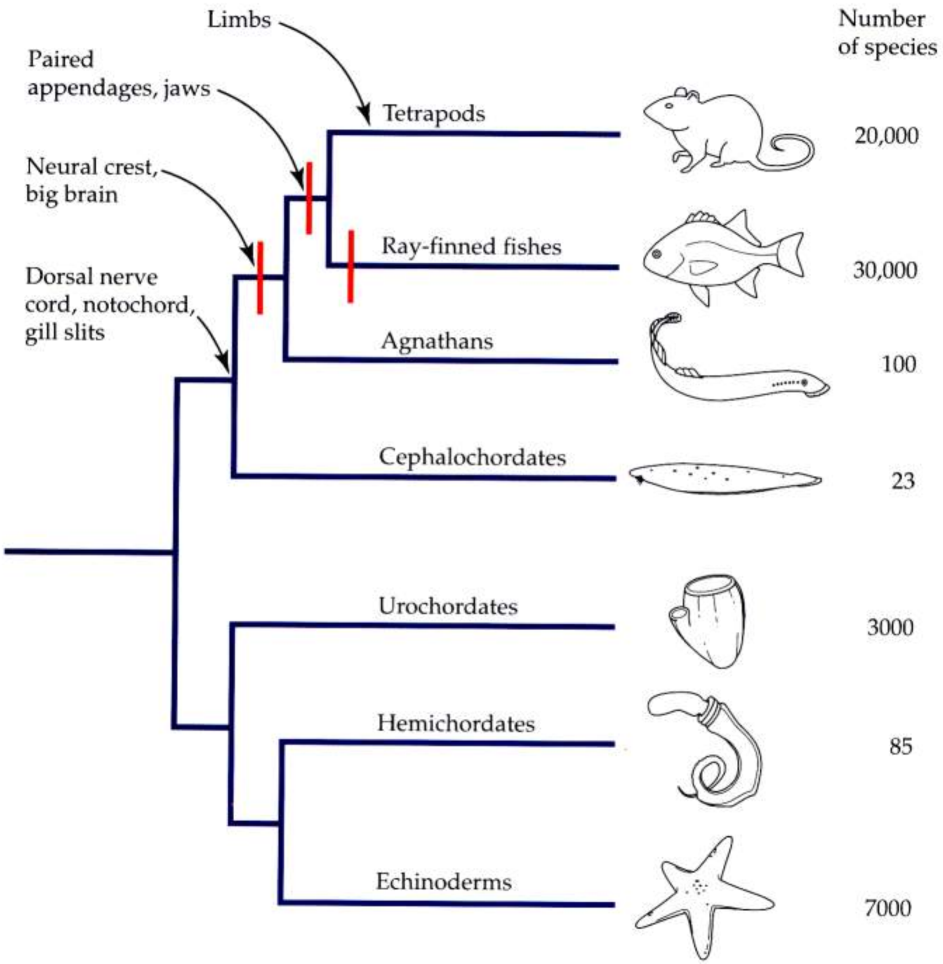
4. Morphological Changes Due to Functional Diversification of a Duplicated Regulatory Gene

A great morphological change in evolution has also been accomplished by the modification of a redundant copy of the previous structure. The first major anatomical improvement that occurred to early vertebrates was the development of jaws.



“Natural selection merely modified, while redundancy created” (1970)

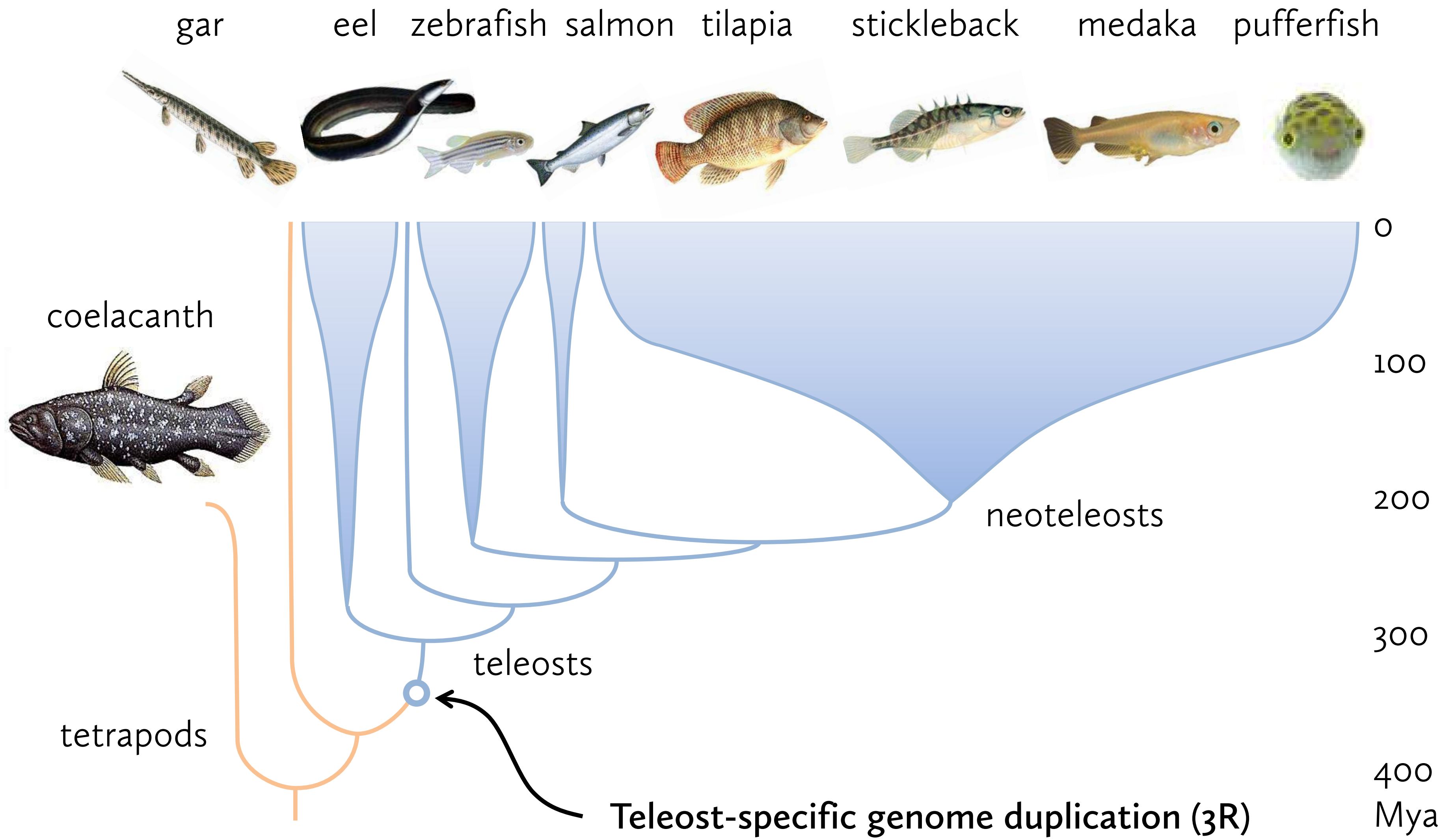
Genome duplication and innovation



Lynch 2007

- Three rounds of genome duplication in vertebrate evolution
- Duplicate Hox genes/clusters are freed from functional constraints
- This leads to evolutionary novelties and increases in complexity
- Or morphological flexibility: teleosts are not more complex than tetrapods

Phylogenetic position



gar eel zebrafish salmon tilapia stickleback medaka pufferfish

coelacanth

0

100

neoteleosts

200

teleosts

300

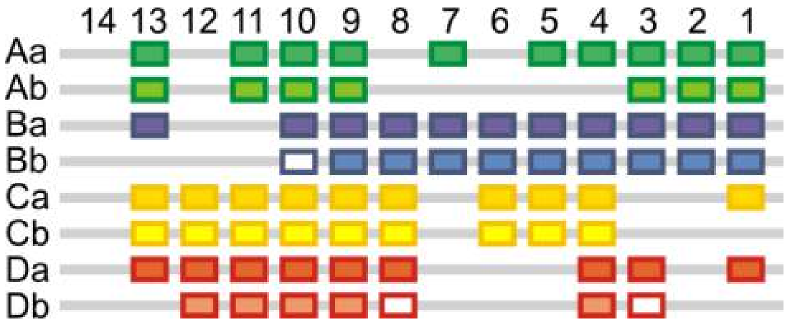
tetrapods

400
Mya

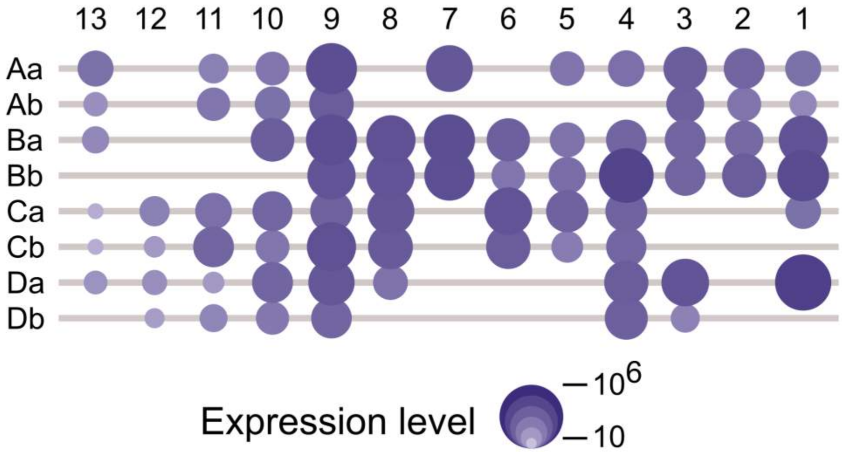
Teleost-specific genome duplication (3R)

Near et al., PNAS 2012

Hox gene clusters of *A. anguilla*

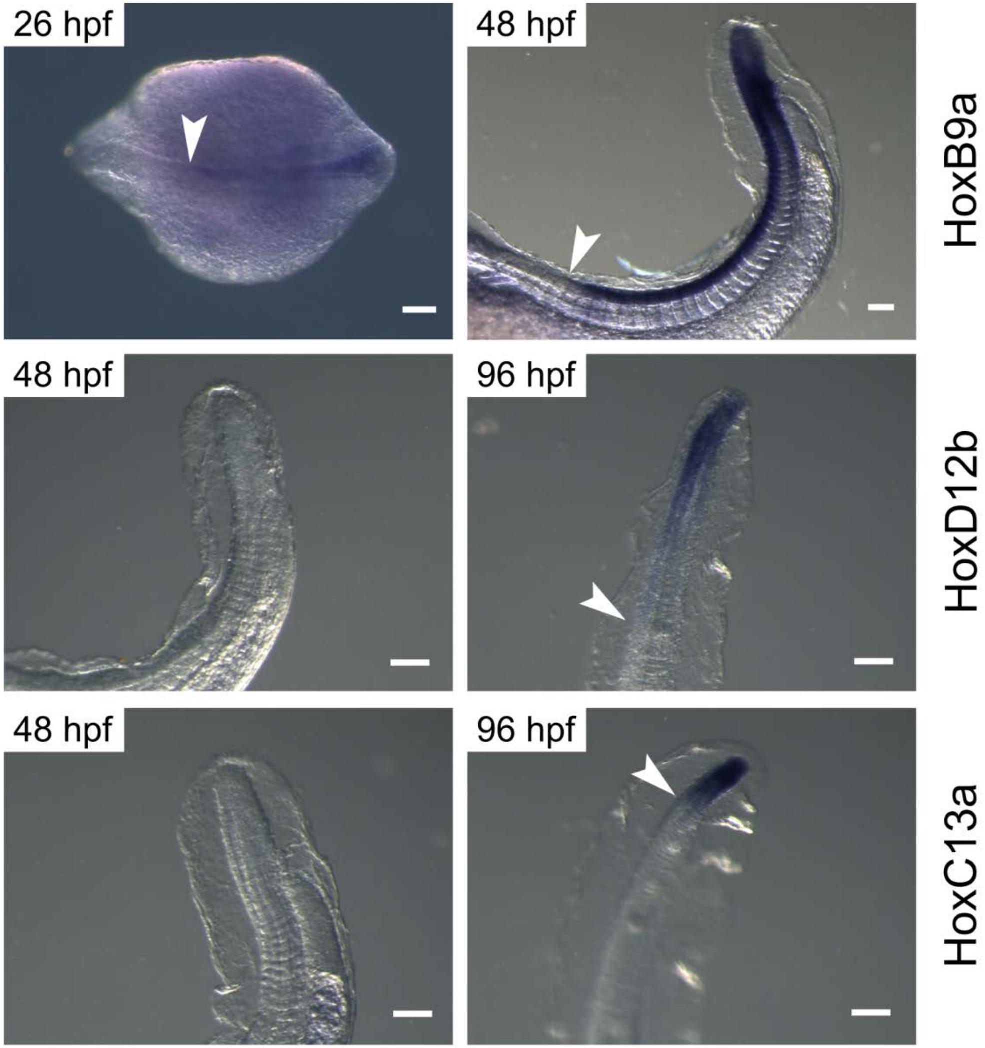


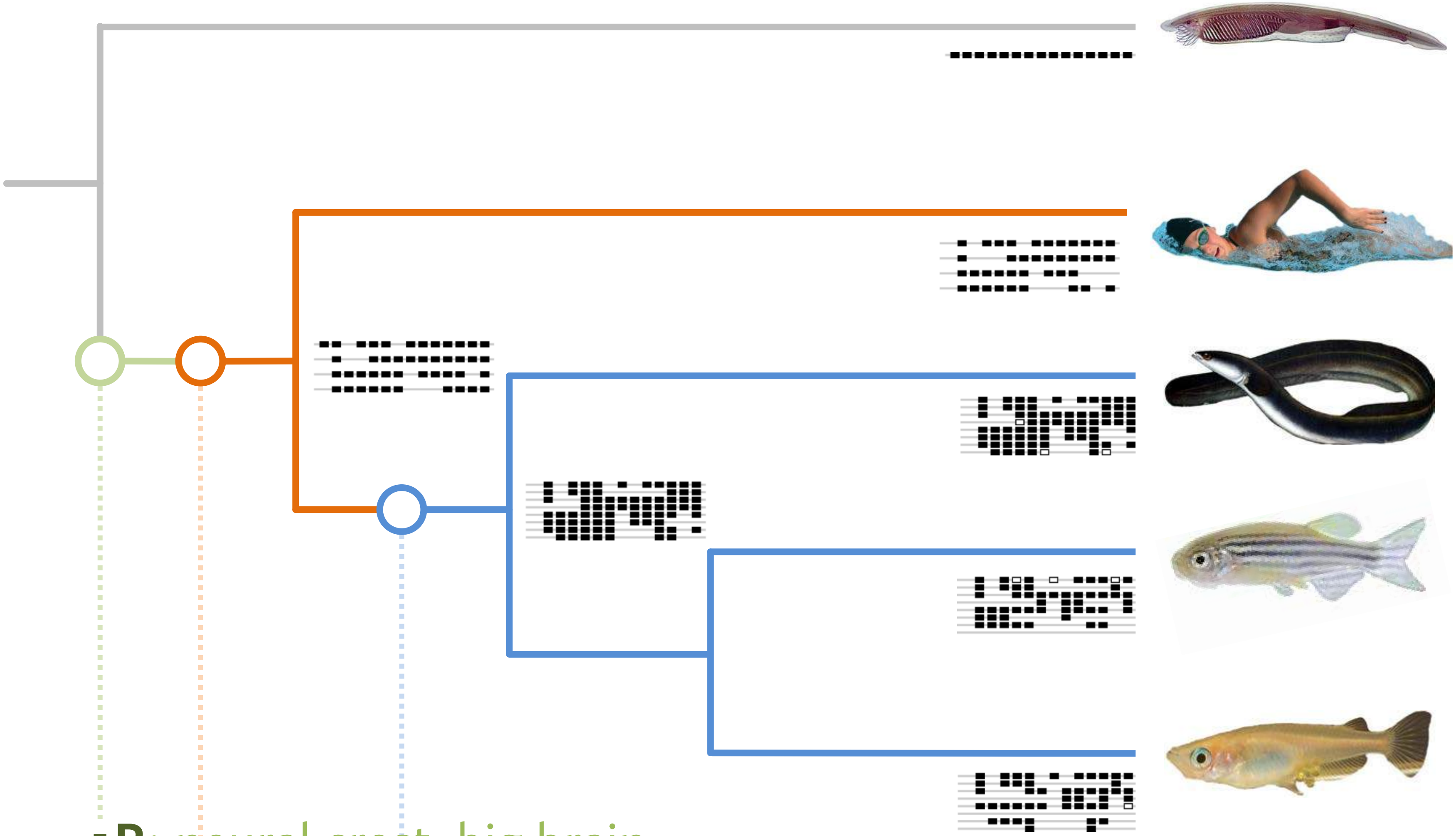
	14	13	12	11	10	9	8	7	6	5	4	3	2	1
Aa		■		■	■	■		■		■	■	■	■	■
Ab		■		■	■	■						■	■	■
Ba		■			■	■	■	■	■	■	■	■	■	■
Bb					□	■	■	■	■	■	■	■	■	■
Ca		■	■	■	■	■	■		■	■	■			■
Cb		■	■	■	■	■	■		■	■	■			
Da		■	■	■	■	■	■				■	■		■
Db			■	■	■	■	□				■	□		



- 73 genes in 8 clusters
- All are expressed at 26 hours post-fertilization

Hox expression





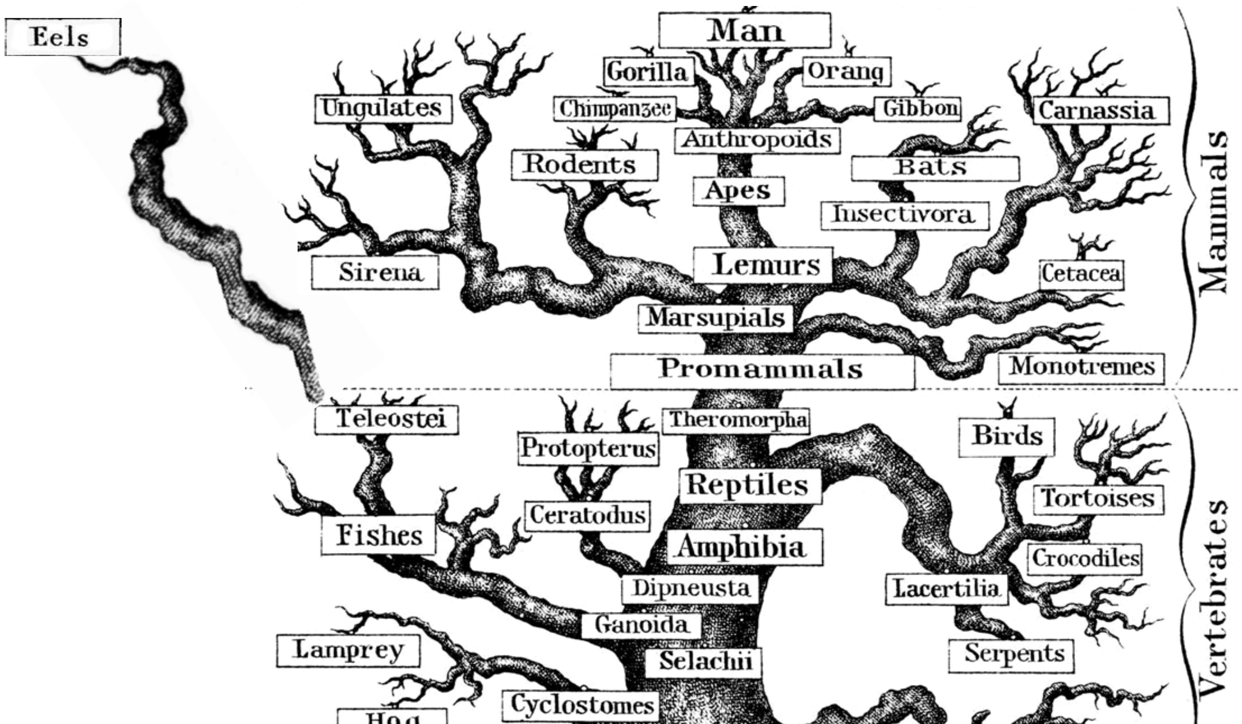
1R: neural crest, big brain

2R: jaws, paired appendages

3R: increased morphological plasticity?

Eel Hox paradoxes

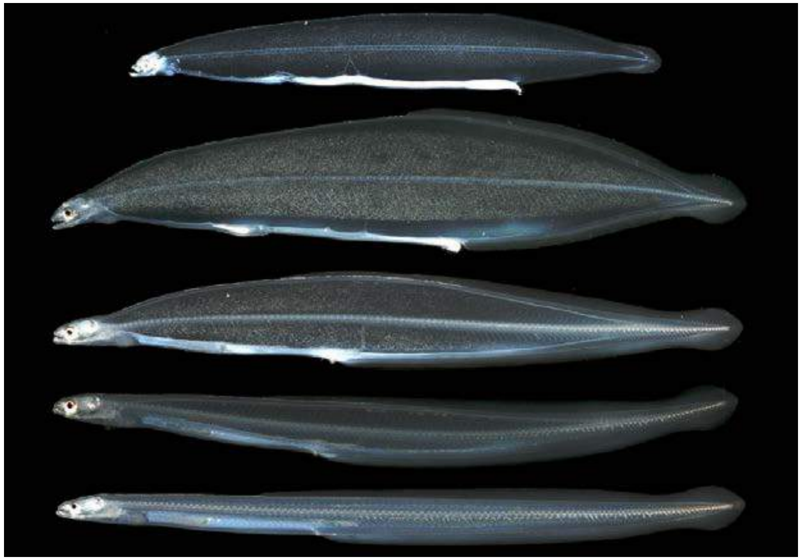
- The eel’s Hox clusters have lost very few genes since the teleost-specific genome duplication (~300 Mya)
- More (stable) Hox genes = higher developmental complexity ?

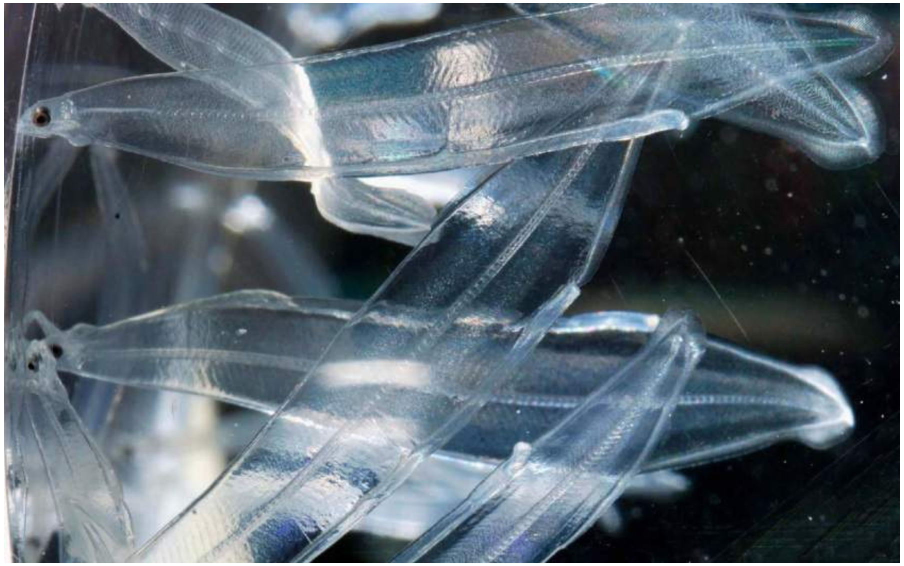


Eel Hox hypotheses

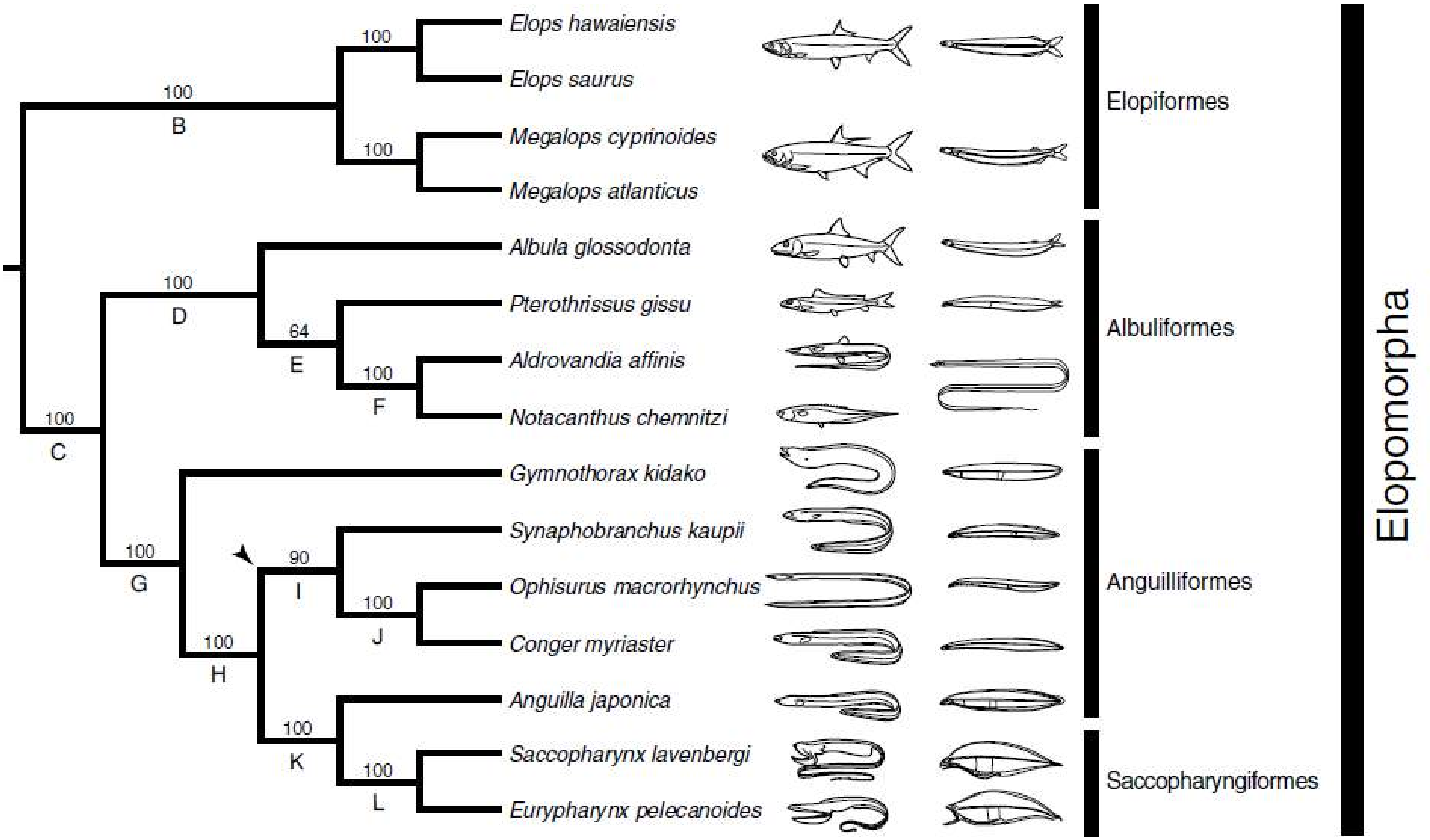
- Eels have a strange, elongated from
- However, many of their Elopomorpha relatives do not
- Amongst eels, a lot of diversity in mechanisms

- The leptocephalus larval form emerged quickly after the genome duplication





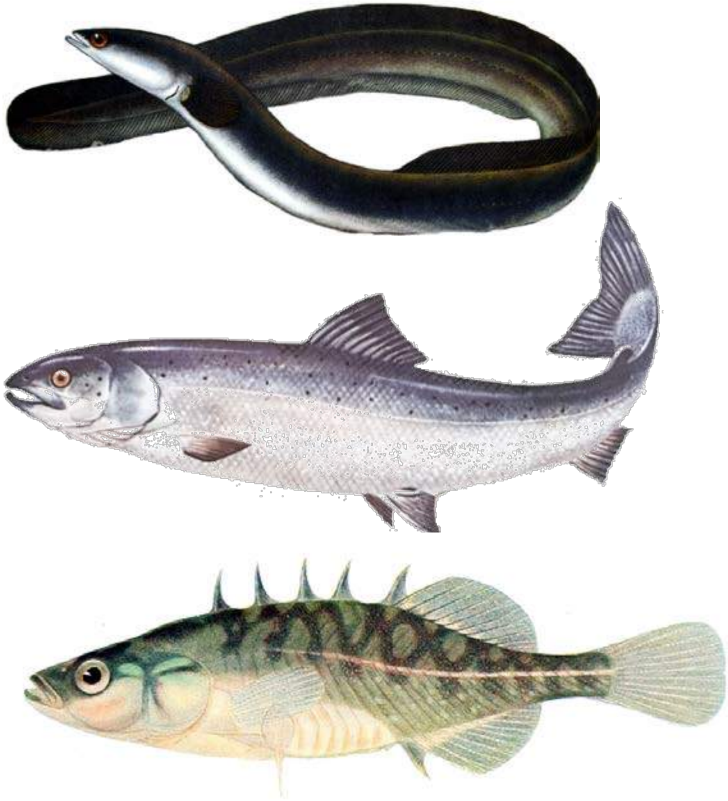
Leptocephalus larvae



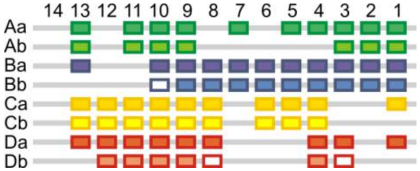
Inoue 2004

Hox clusters and complex life cylces?

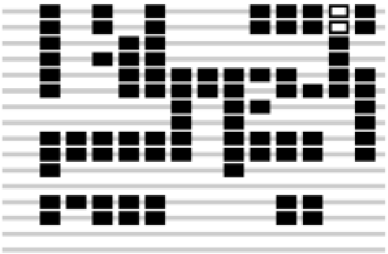
- The eel life cycle is tied to increased developmental complexity
- Is this a general phenomenon for migratory fish species?



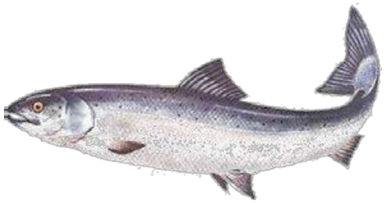
Hox clusters and complex life cylces?

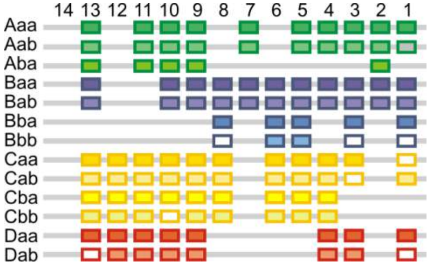


3 genome duplications (ancient)
73 Hox genes
stable and functional
extra larval stage



4 genome duplications (recent)
88 Hox genes
unstable, functional?
no extra larval stage

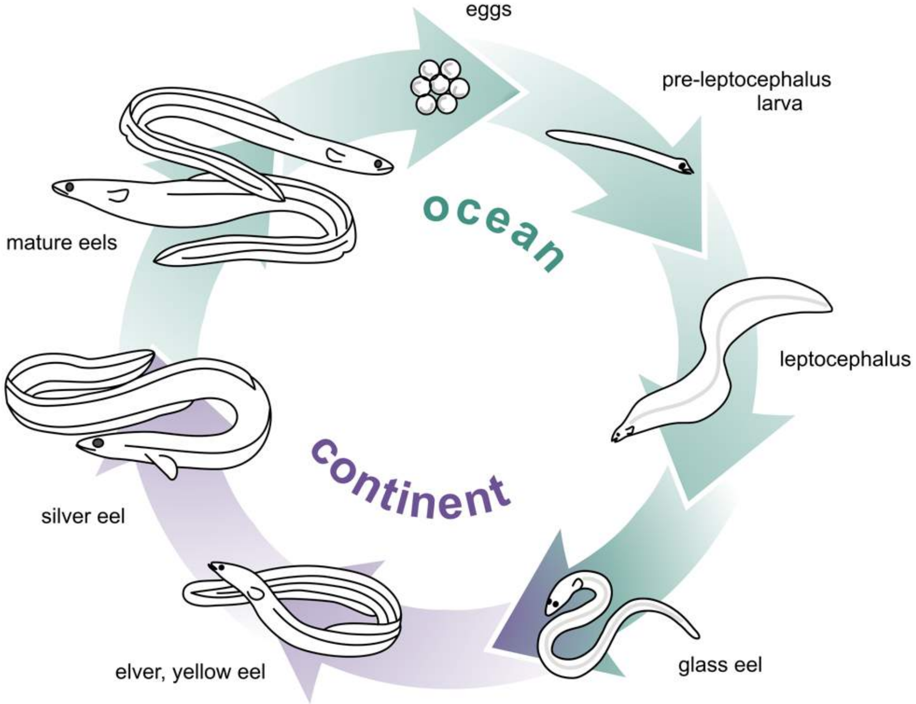




4 genome duplications (ancient)
118 Hox genes
stable, functional?
extra larval stages

Genomics and the eel life cycle

- SNP genotyping uncovers the eel population structure and selection pressures
- Transcriptomics illuminate eel physiology at distinct life stages
- Genomics explain the origin of the eel body plans



Kumagai 2013