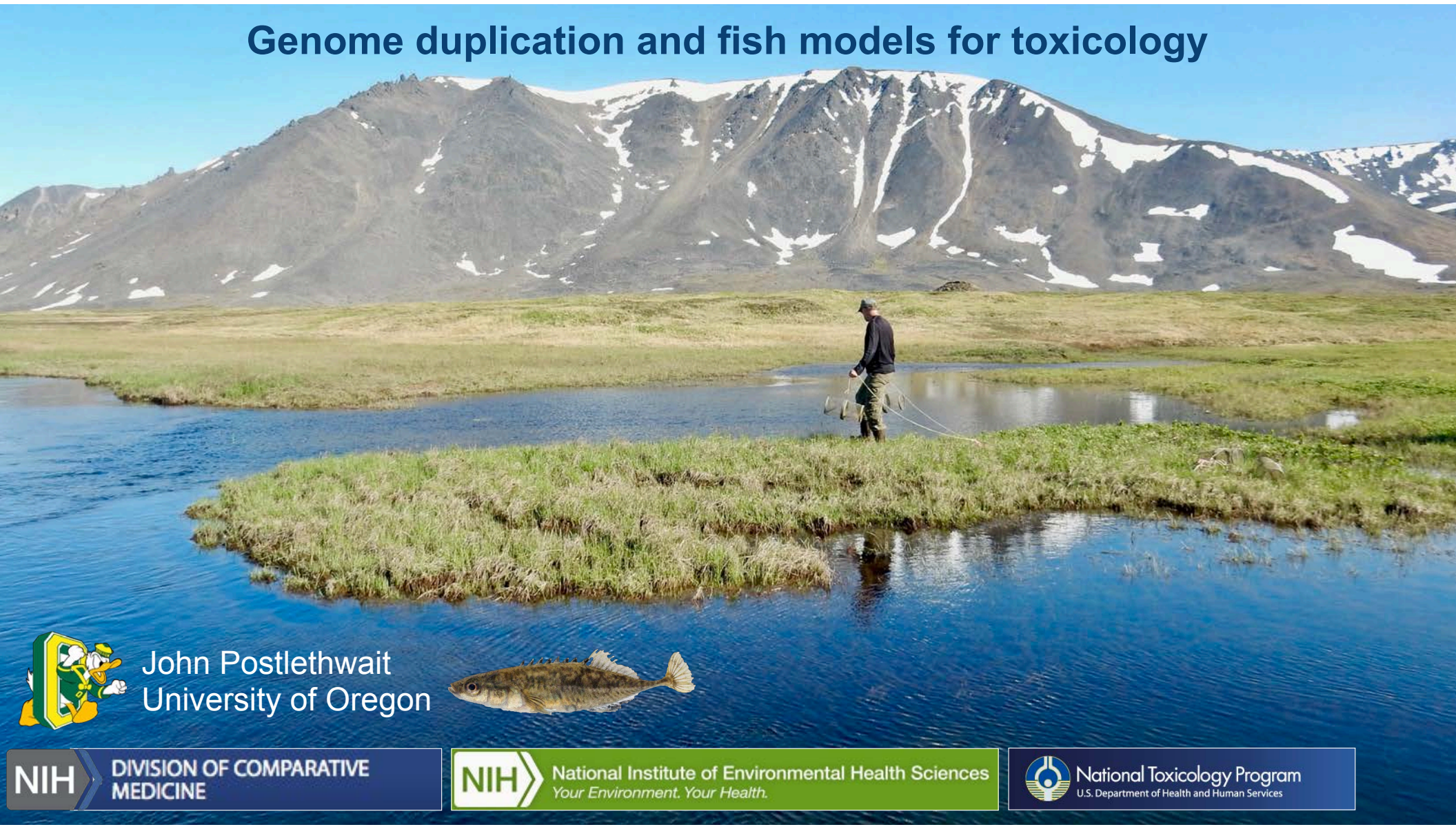
Genome duplication and fish models for toxicology



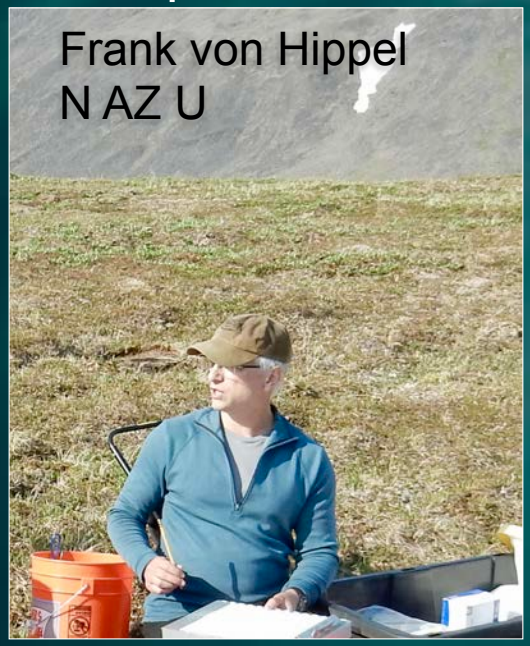
John Postlethwait
University of Oregon

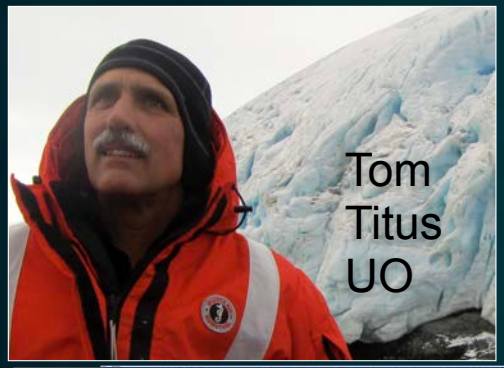
NIH DIVISION OF COMPARATIVE MEDICINE

NIH National Institute of Environmental Health Sciences
Your Environment. Your Health.

National Toxicology Program
U.S. Department of Health and Human Services

Genome duplication and fish models for toxicology



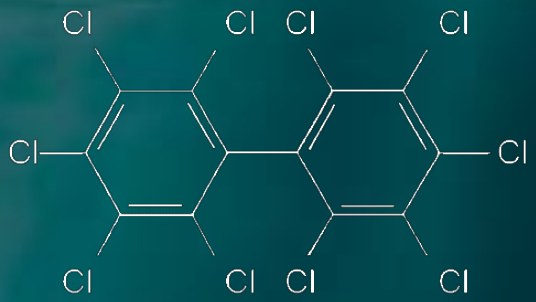


NIH DIVISION OF COMPARATIVE MEDICINE

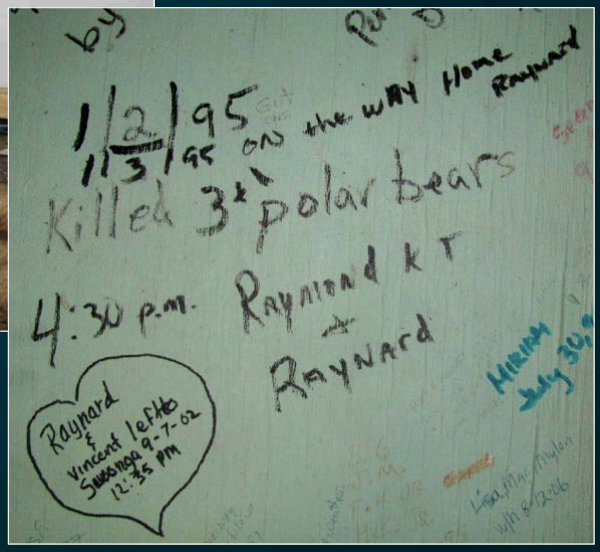
NIH National Institute of Environmental Health Sciences
Your Environment. Your Health.

The problem:

People living a subsistence lifestyle in the Arctic are highly exposed to persistent organic pollutants.

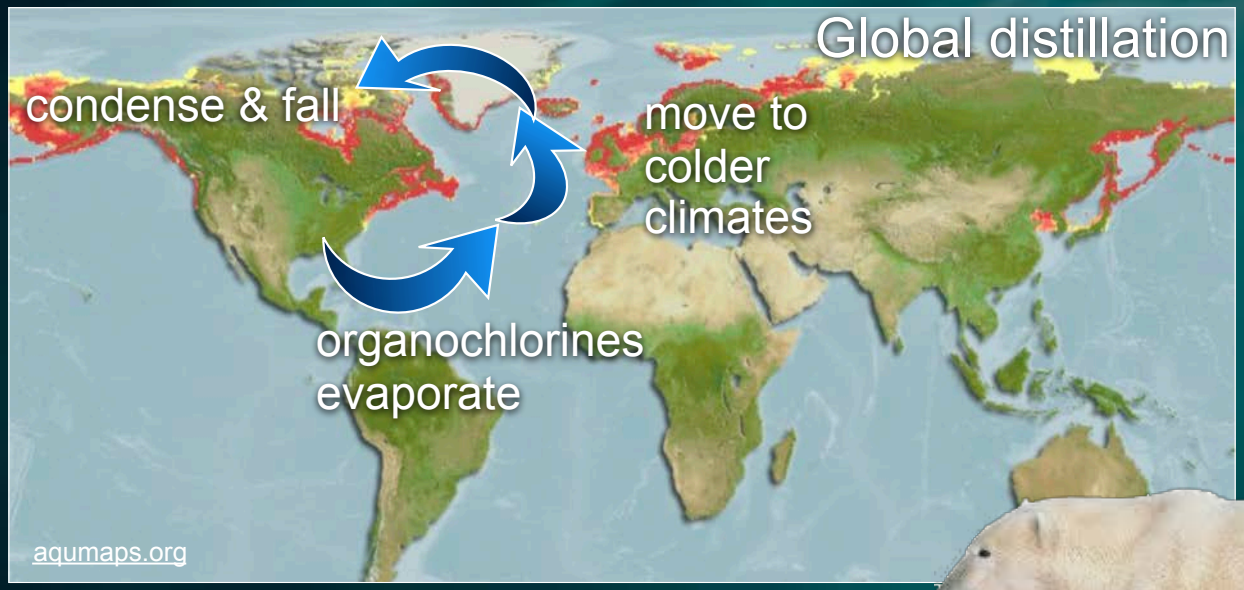


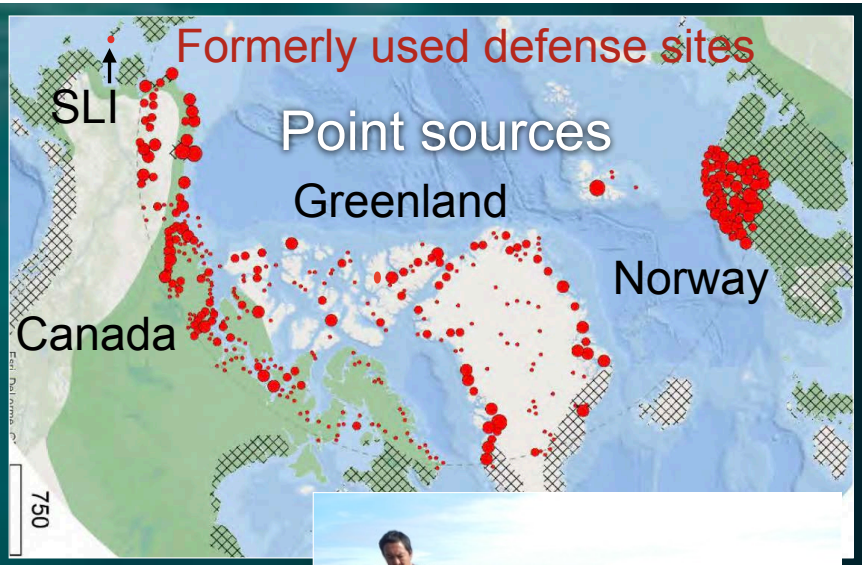
Annie Alowa, Sivuqaq Islander



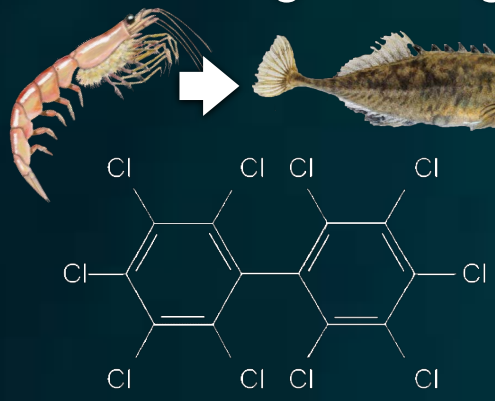
The problem:

Sources:





Biological magnification



The problem:

Point sources

Formerly used defense sites

SLI

Greenland

Norway

Canada

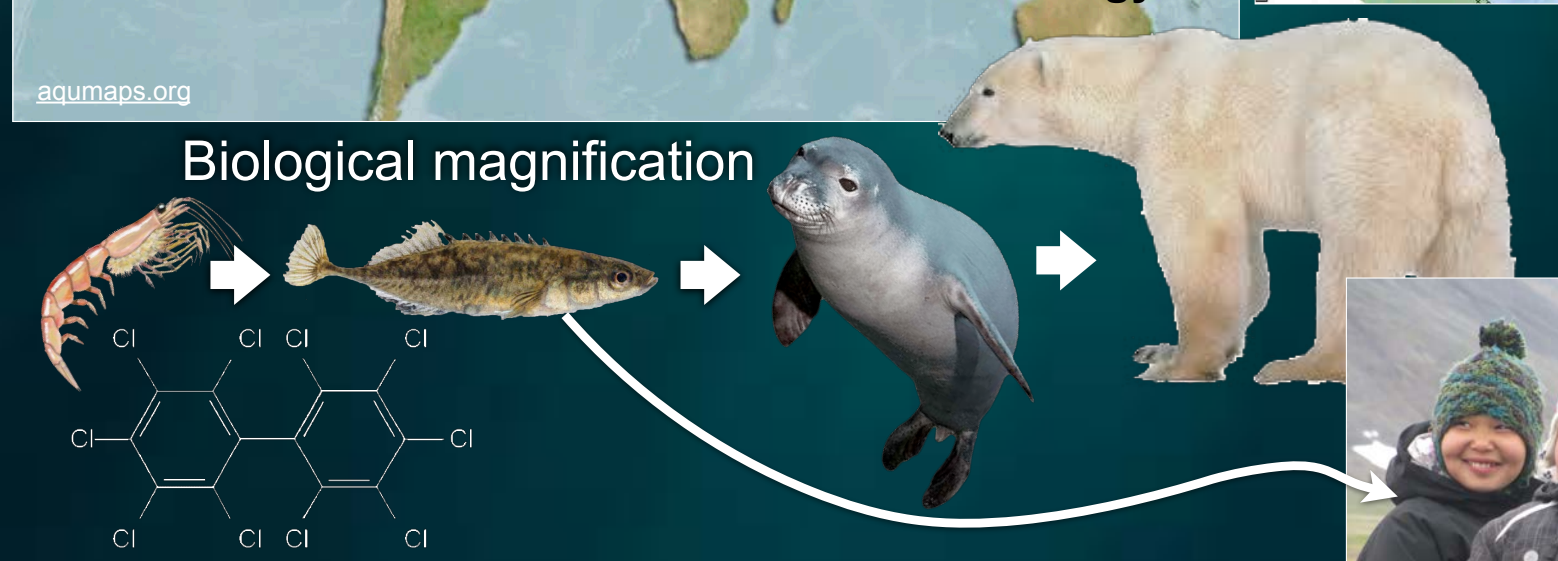
750

Global distillation

9-spine stickleback

Teleost fish are convenient models for toxicology.

aqumaps.org



Biological magnification

How to connect teleost genomes to human biology?

How to connect teleost genomes to human biology?

- Genome duplication in vertebrates
- Genome duplication complicates connectivity
- Ohnologs gone missing and lineage-specific evolution
- Application: St. Lawrence Island

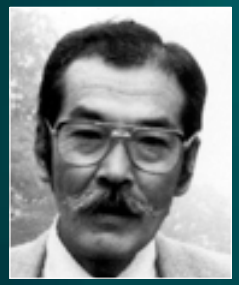
How to connect teleost genomes to human biology?

• Genome duplication in vertebrates

zebrafish

Ohnologs: duplicates from genome duplication

bony fish



Susumo Ohno
Evolution by Gene Duplication (1970)

Two rounds of genome duplication

jawed vertebrates

1 *Hox* cluster

e shark

cartilaginous fish

4 *Hox* clusters

How to connect teleost genomes to human biology?

• Genome duplication in vertebrates

zebrafish

rayfin fish

bony fish

Two rounds of genome duplication

jawed vertebrates

coelacanth

xenopus

chicken

mouse

human

lobefin fish

1 *Hox* cluster

e shark

cartilaginous fish

4 *Hox* clusters

How to connect teleost genomes to human biology?

• Genome duplication in vertebrates

zebrafish

medaka

cavefish

goldeye

eel

gar

teleost fish

rayfin fish

bony fish

Two rounds of genome duplication

coelacanth

xenopus

chicken

mouse

human

lobefin fish

jawed vertebrates

1 *Hox* cluster

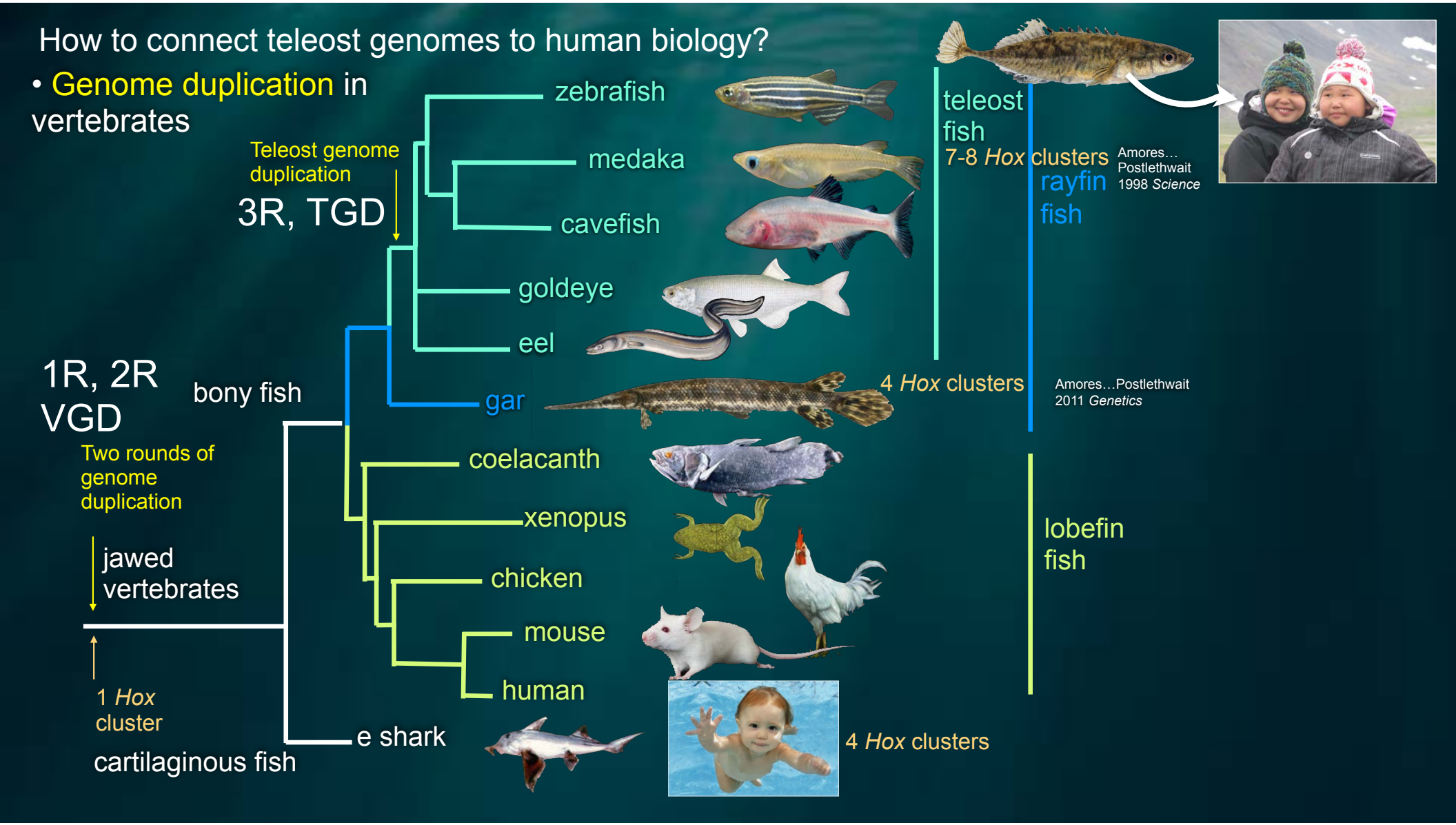
e shark

cartilaginous fish

4 *Hox* clusters

How to connect teleost genomes to human biology?

• Genome duplication in vertebrates



How to connect teleost genomes to human biology?

• Genome duplication in vertebrates

Teleost genome duplication

3R, TGD

1R, 2R VGD

Two rounds of genome duplication

jawed vertebrates

1 *Hox* cluster

bony fish

cartilaginous fish

zebrafish

medaka

cavefish

goldeye

eel

gar

coelacanth

xenopus

chicken

mouse

human

e shark

teleost fish
7-8 *Hox* clusters

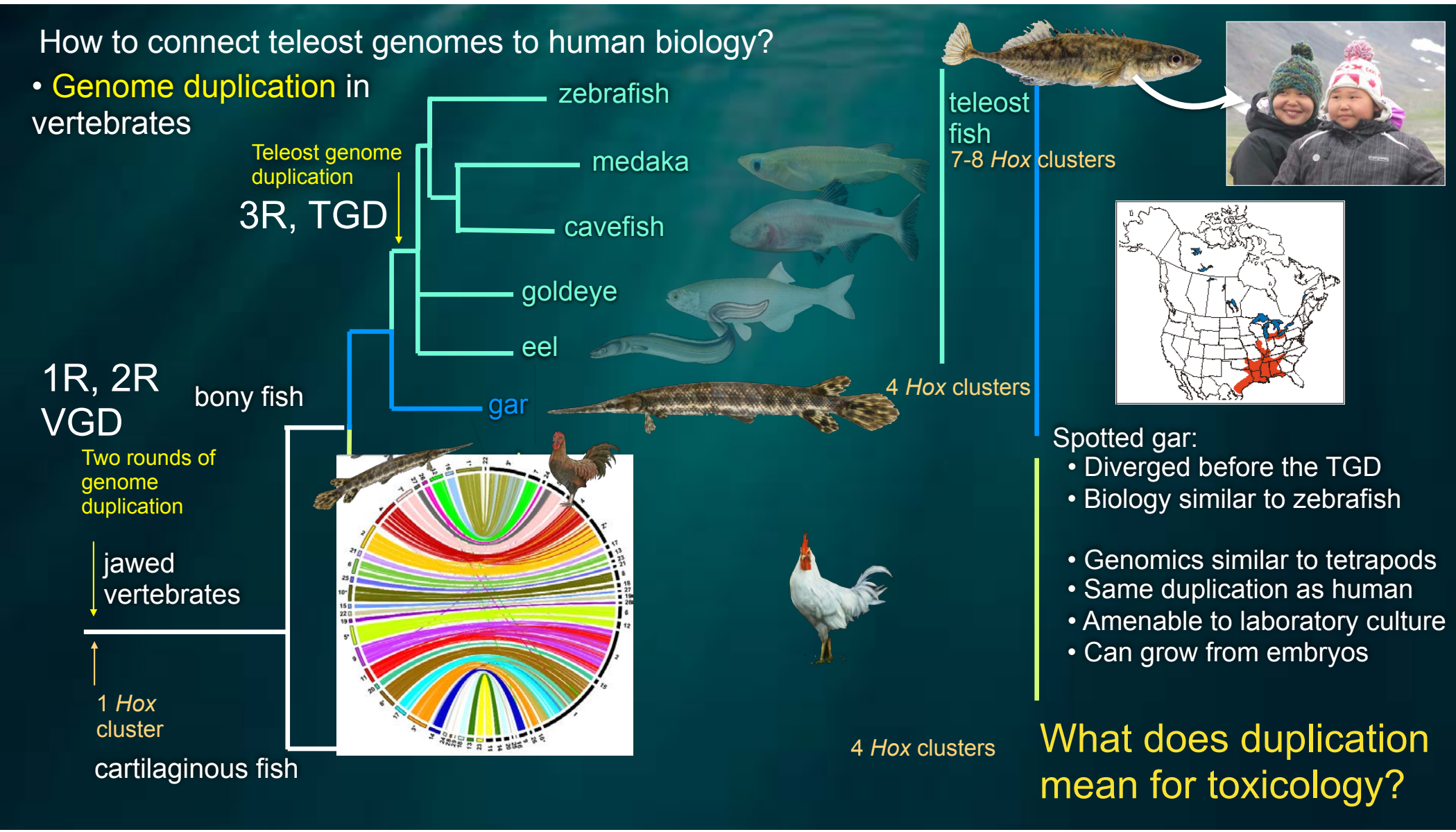
4 *Hox* clusters

4 *Hox* clusters

Spotted gar:
- Diverged before the TGD
- Biology similar to zebrafish

How to connect teleost genomes to human biology?

• Genome duplication in vertebrates



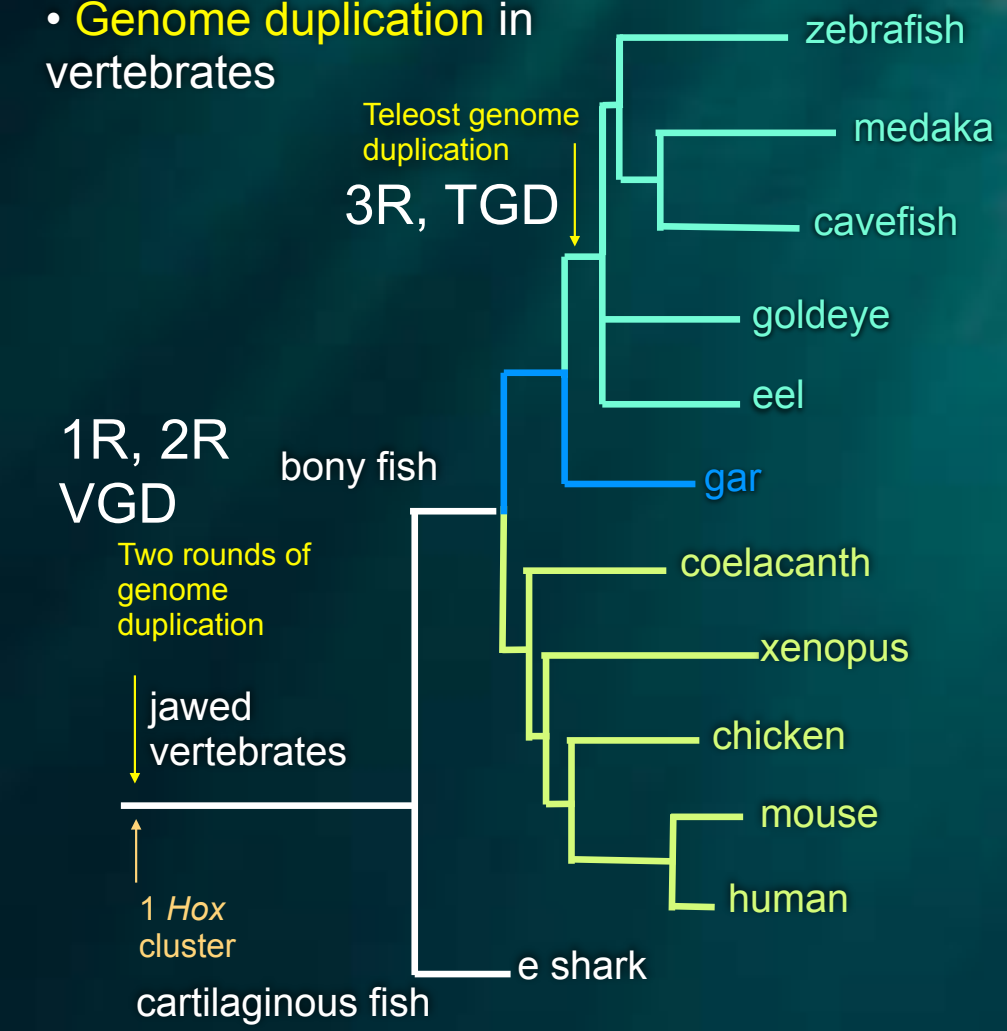
Spotted gar:
- Diverged before the TGD
- Biology similar to zebrafish

- Genomics similar to tetrapods
- Same duplication as human
- Amenable to laboratory culture
- Can grow from embryos

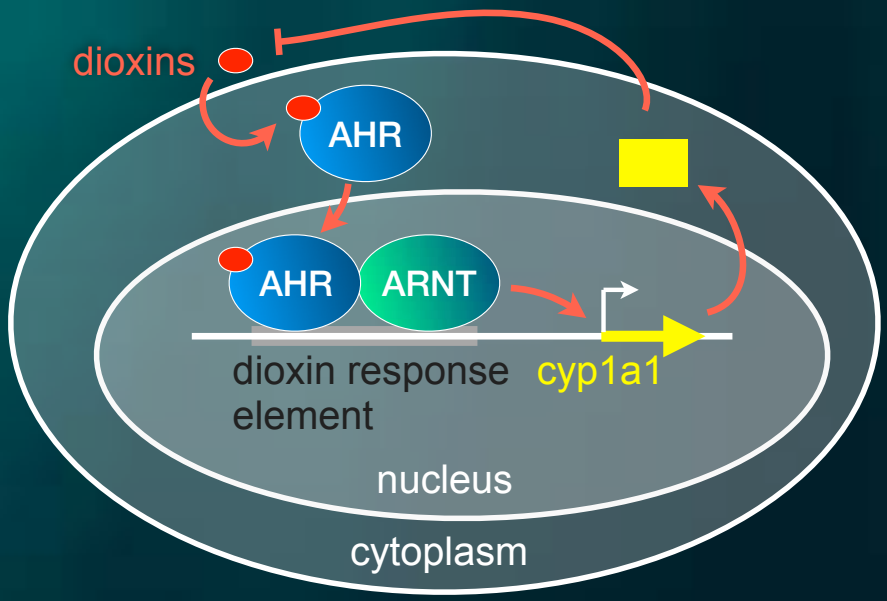
What does duplication mean for toxicology?

What does duplication mean for toxicology?

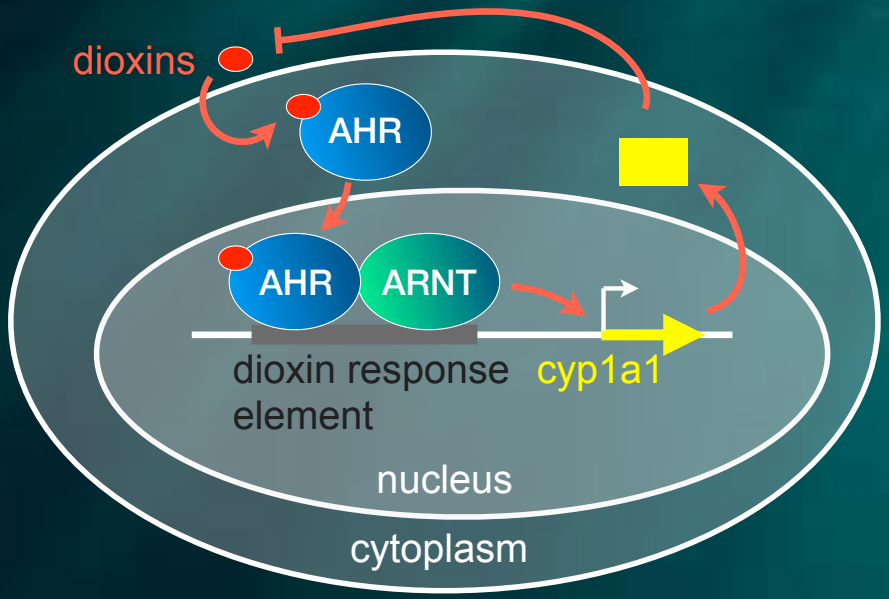
- Genome duplication in vertebrates



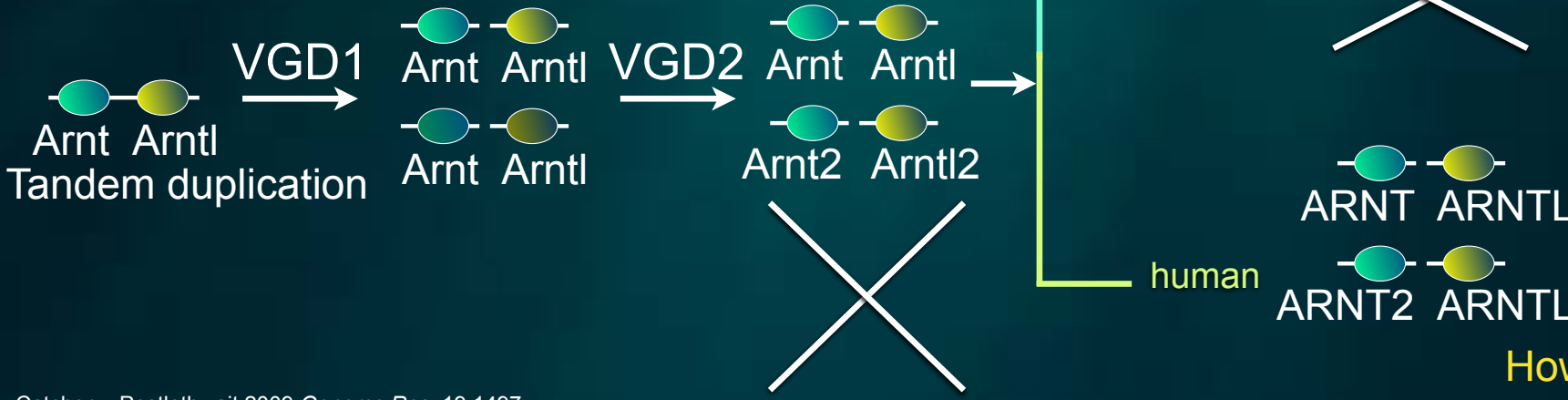
How to connect teleost genomes to human biology?



What does duplication mean for toxicology?



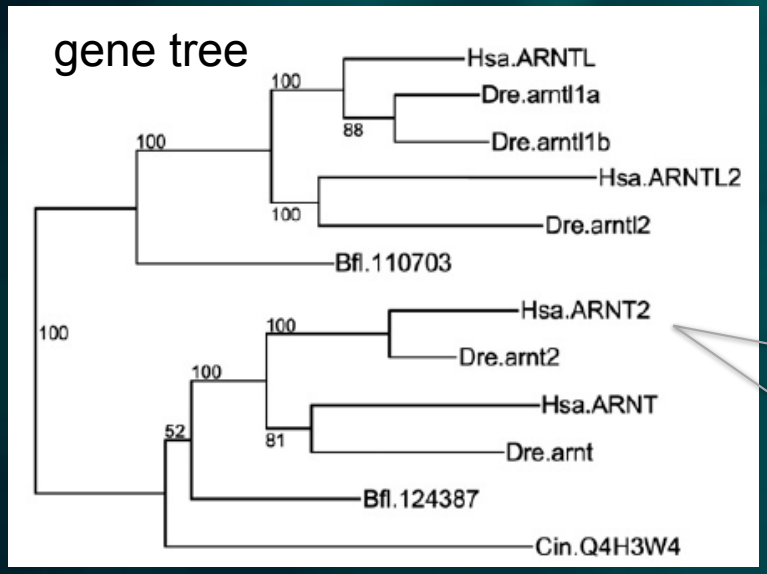
How to connect teleost genomes to human biology?



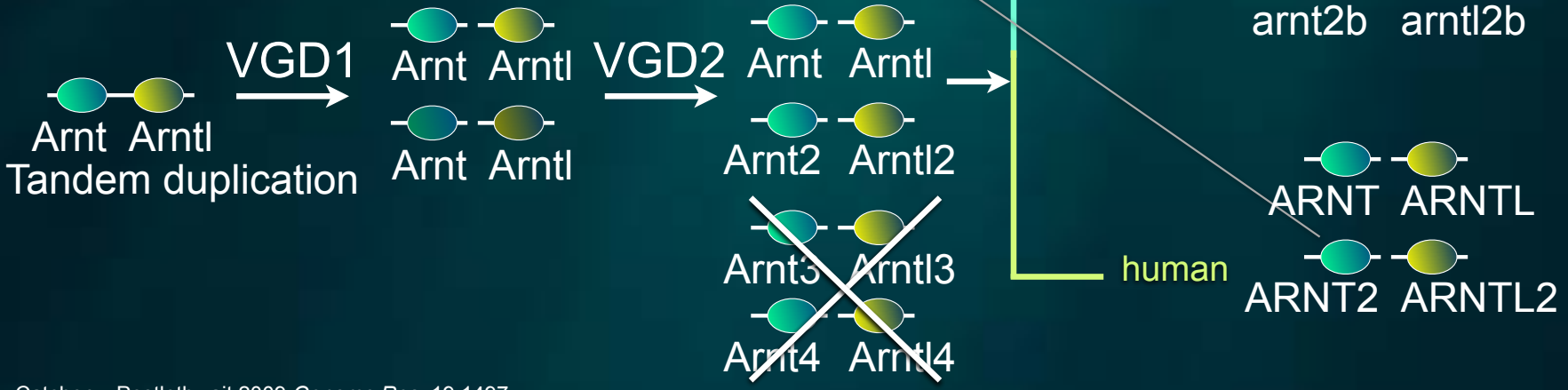
How do we know?

Catchen…Postlethwait 2009 *Genome Res.* 19:1497

How do we know?

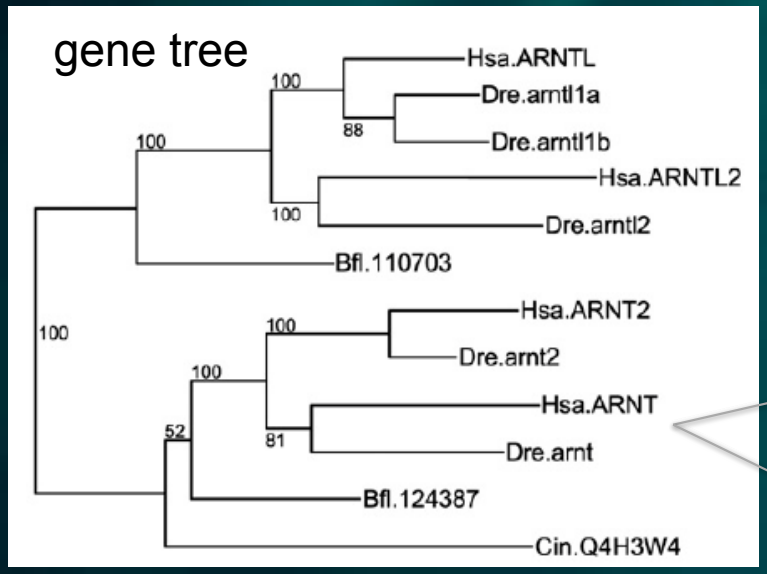


How to connect teleost genomes to human biology?



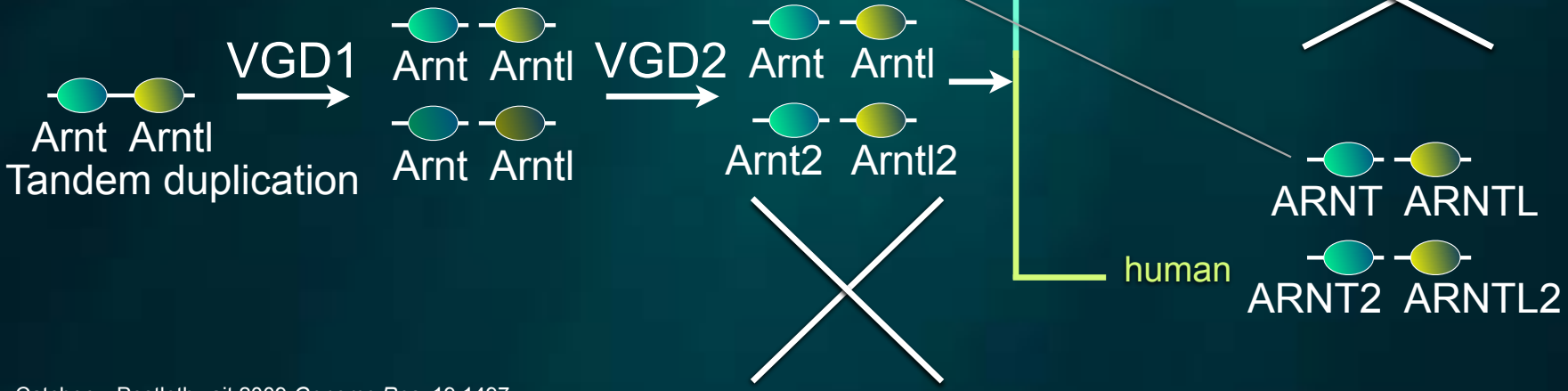
Catchen…Postlethwait 2009 *Genome Res*. 19:1497

How do we know?



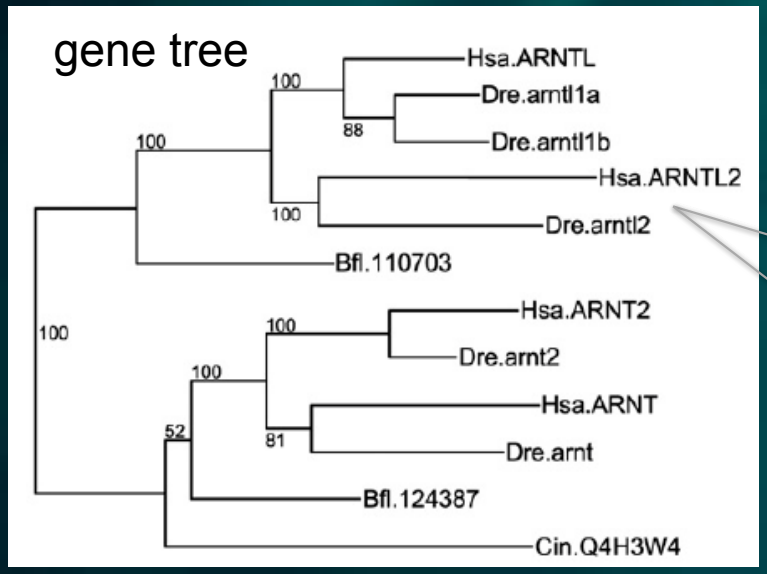
How to connect teleost genomes to human biology?

orthologs



Catchen…Postlethwait 2009 *Genome Res*. 19:1497

How do we know?



How to connect teleost genomes to human biology?

orthologs

arnt arntl1a

arntl1b

TGD

zebrafish

arnt2 arntl2

Arnt Arntl

Tandem duplication

VGD1

Arnt Arntl

Arnt Arntl

VGD2

Arnt Arntl

Arnt2 Arntl2

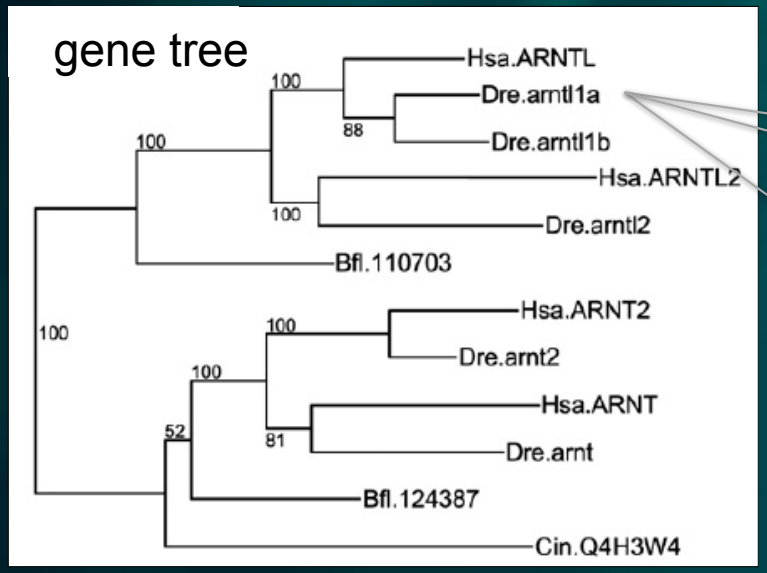
ARNT ARNTL

human

ARNT2 ARNTL2

Catchen…Postlethwait 2009 *Genome Res*. 19:1497

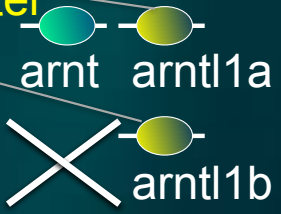
How do we know?

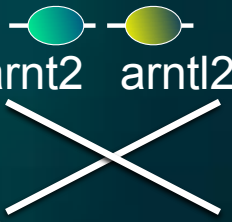


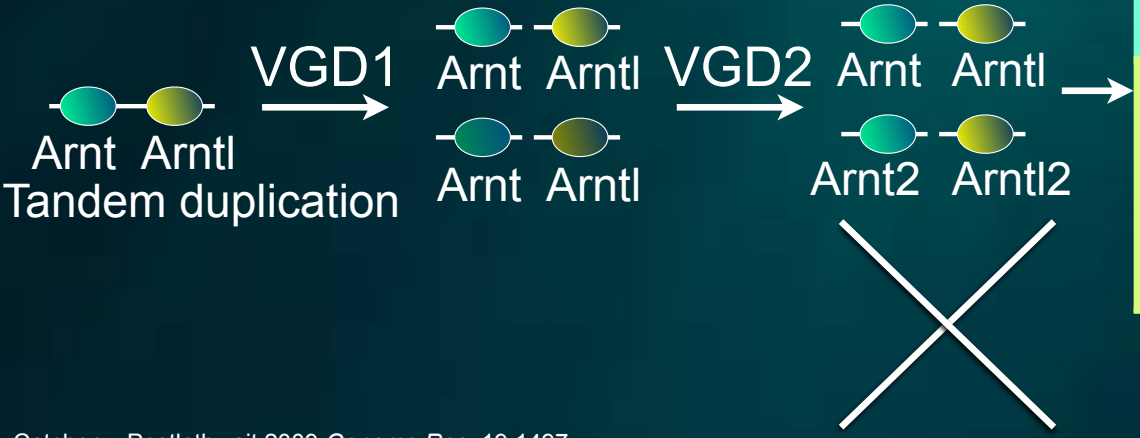
co-orthologs

How to connect teleost genomes to human biology?

Tree shows duplication after human-fish divergence.

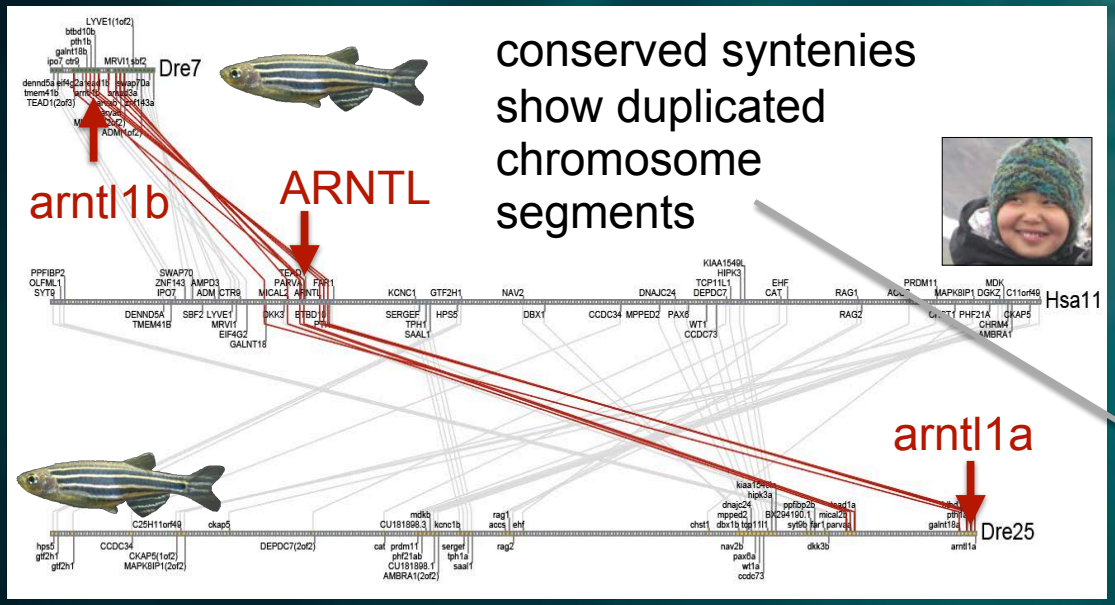






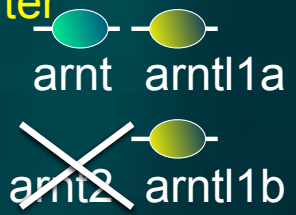
Catchen…Postlethwait 2009 *Genome Res.* 19:1497

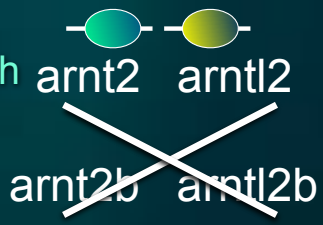
How do we know?

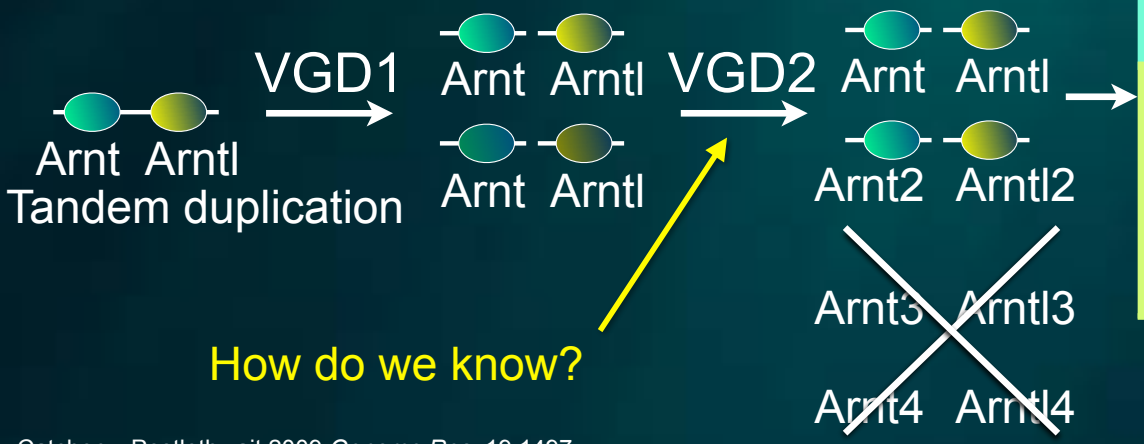


How to connect teleost genomes to human biology?

Tree shows duplication after human-fish divergence.

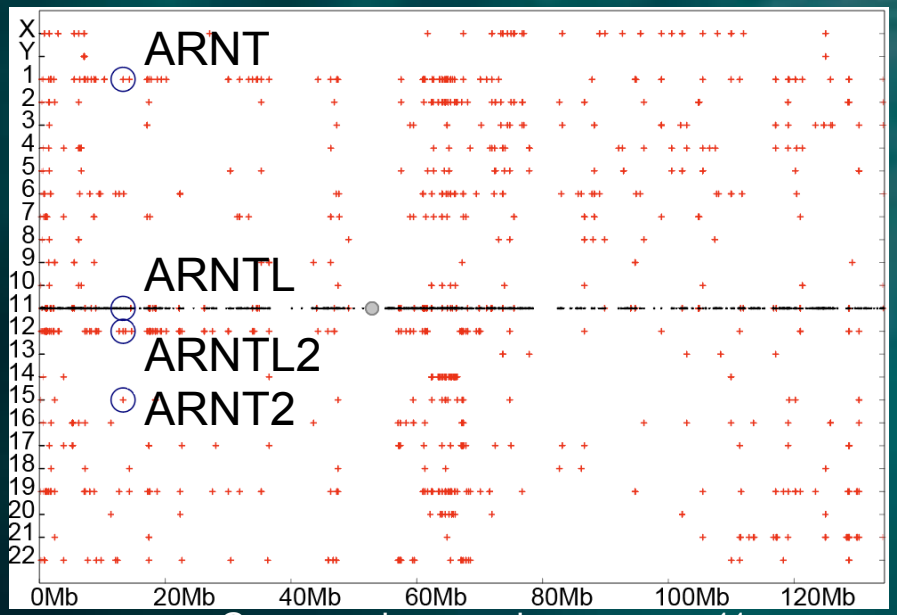






Catchen…Postlethwait 2009 *Genome Res.* 19:1497

How do we know?

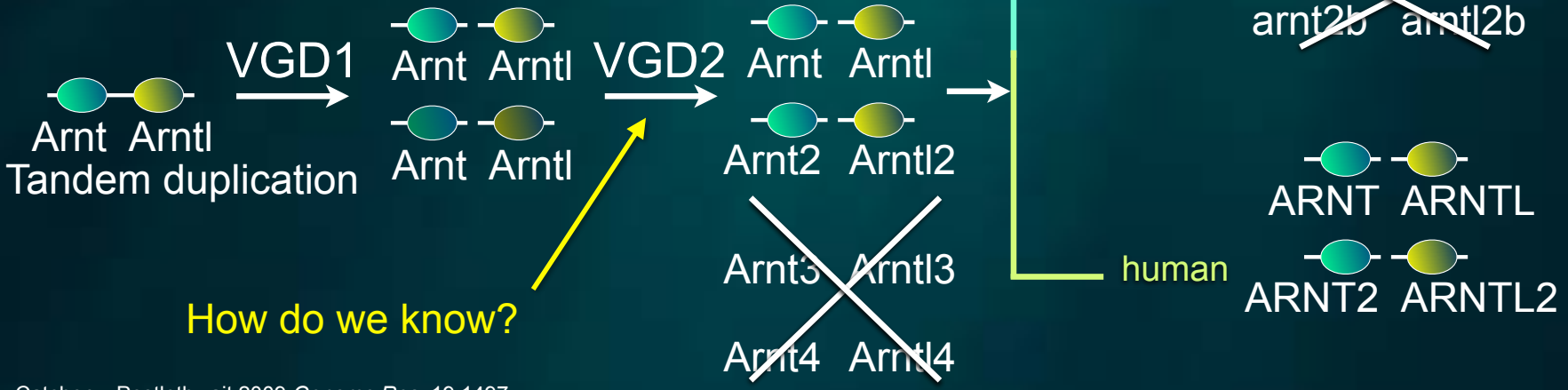


Human chromosomes

Genes on human chromosome 11

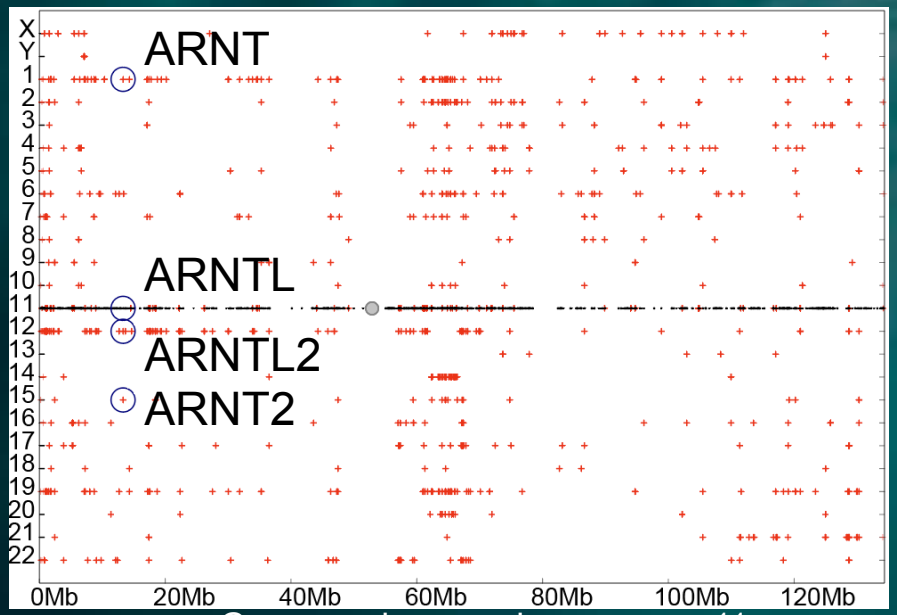
How to connect teleost genomes to human biology?

Duplicated chromosome segments in human genome



Catchen…Postlethwait 2009 *Genome Res.* 19:1497

How do we know?



Human chromosomes

Genes on human chromosome 11

How to connect teleost genomes to human biology?

Duplicated chromosome segments in human genome

Arnt Arntl
Tandem duplication

VGD1

Arnt Arntl
Arnt Arntl

VGD2

Arnt Arntl
Arnt2 Arntl2
Arnt3 Arntl3
Arnt4 Arntl4

How do we know?

TGD

zebrafish

human

arnt arntl1a
arnt2 arntl1b
arnt2 arntl2
arnt2b arntl2b

ARNT ARNTL
ARNT2 ARNTL2

How do fish genes relate to human genes?

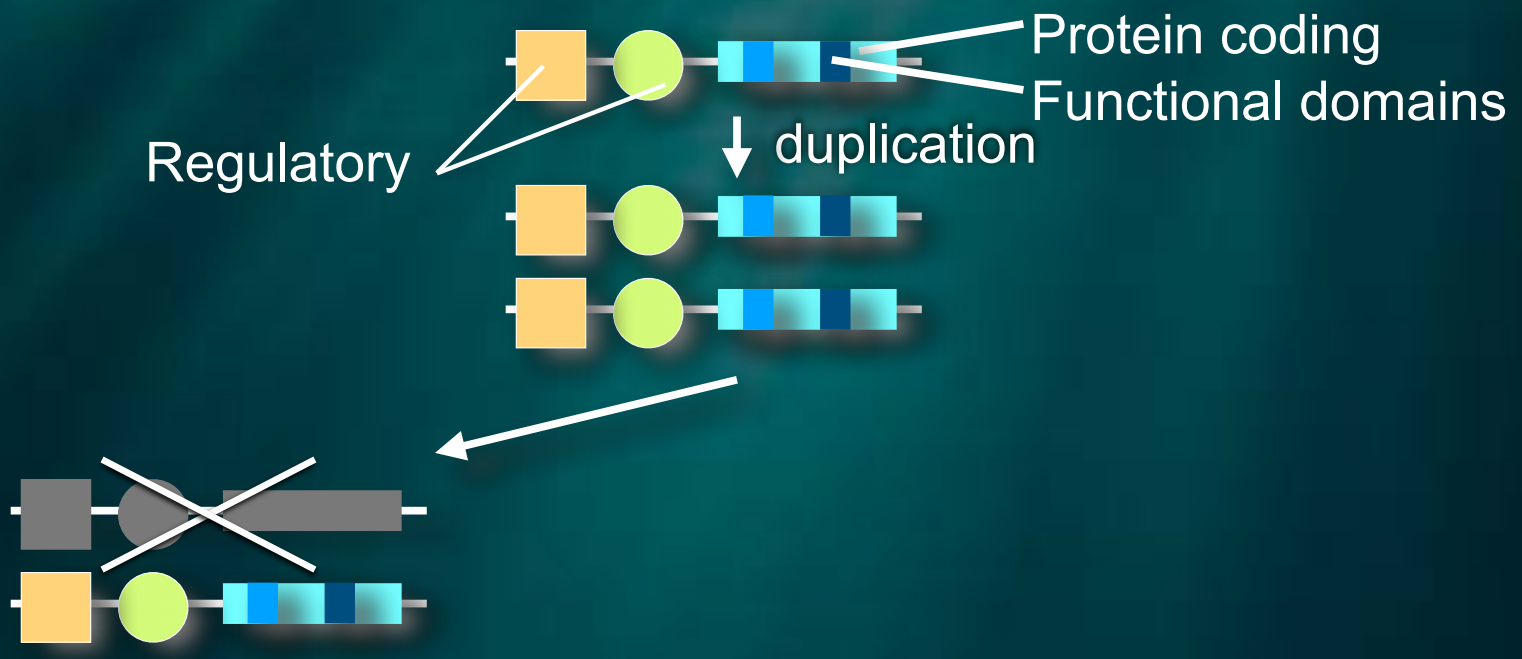
• Genome duplication complicates connectivity

Catchen…Postlethwait 2009 *Genome Res.* 19:1497

How to connect teleost genomes to human biology?

- Genome duplication in vertebrates
- Genome duplication complicates connectivity
- Ohnologs gone missing and lineage-specific evolution
- Application: St. Lawrence Island

- Genome duplication complicates connectivity



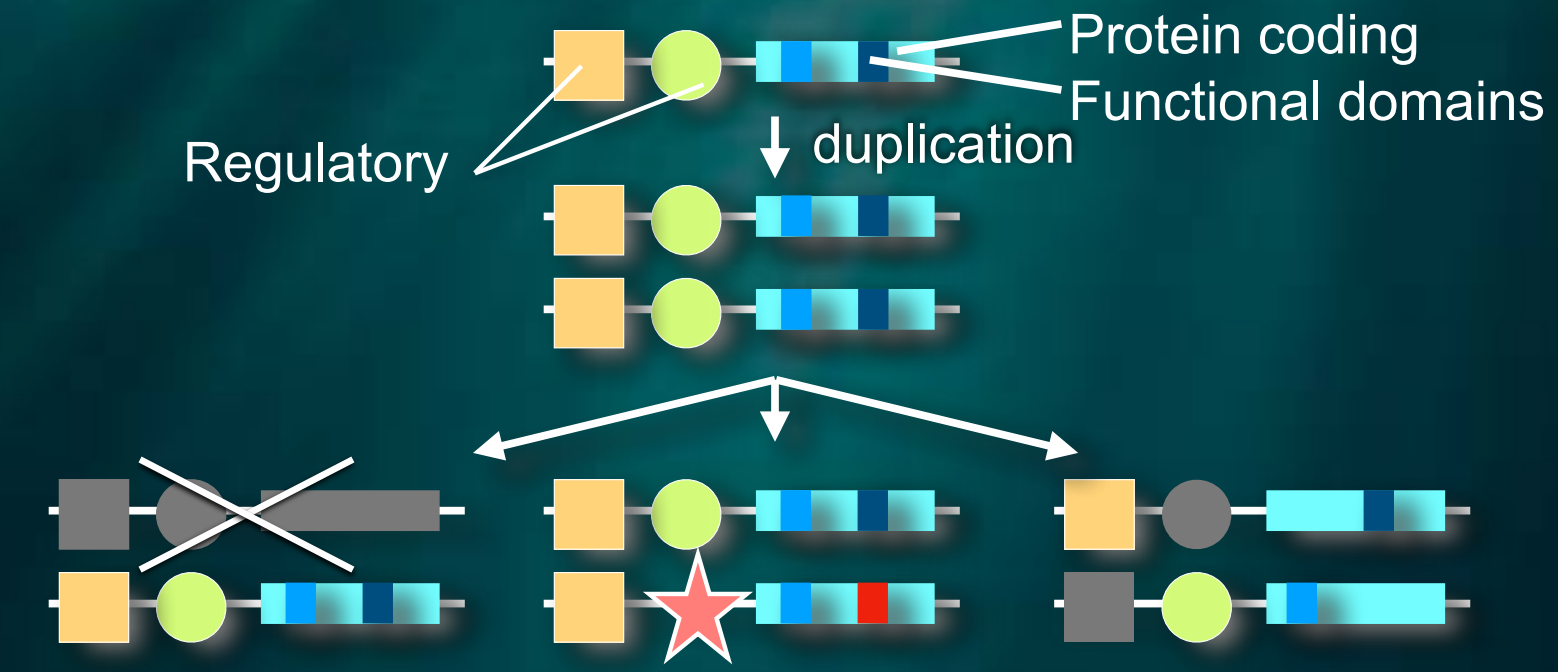
nonfunctionalization

Gene loss: most frequent fate

Alan Force

Force…Postlethwait 1999 *Genetics*

• Genome duplication complicates connectivity



nonfunctionalization

Gene loss: most frequent fate

neofunctionalization

New functions evolve from ancestral genes

subfunctionalization

Reciprocal loss of ancestral functions

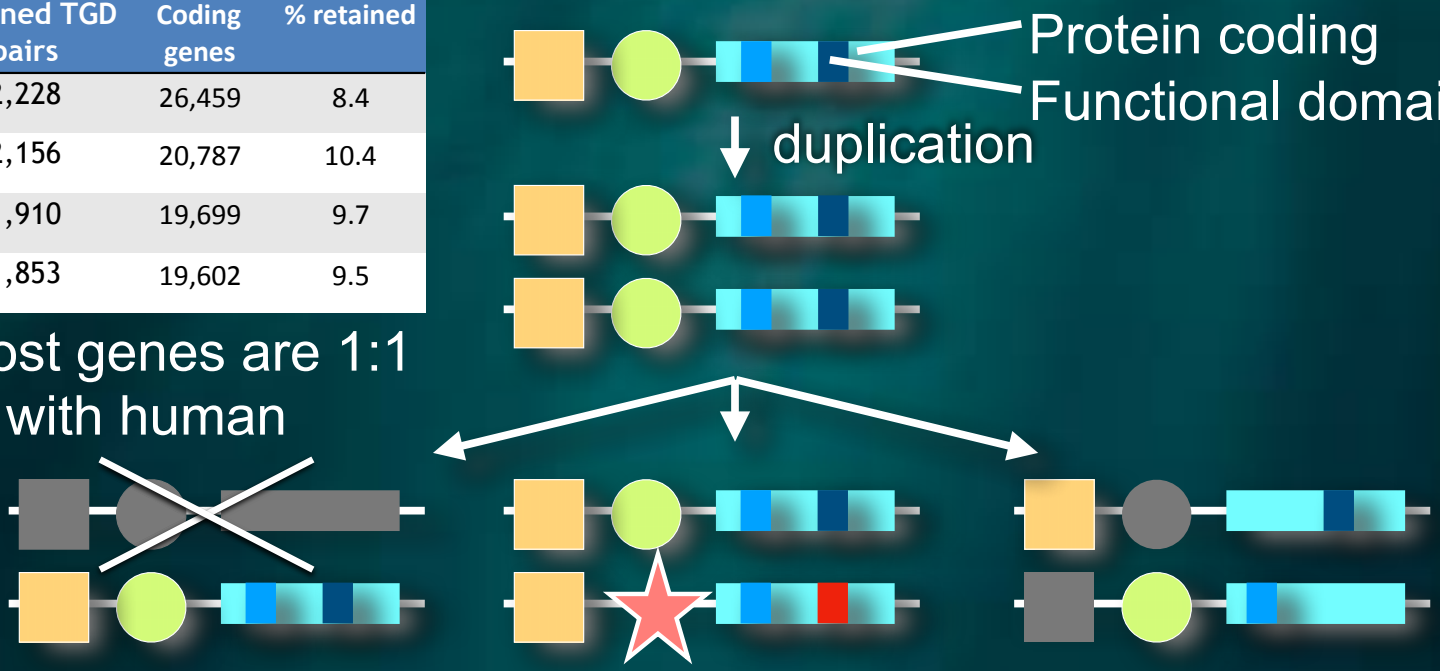
Alan Force

Force…Postlethwait 1999 *Genetics*

- Genome duplication complicates connectivity

Species	Retained TGD pairs	Coding genes	% retained
Zebrafish	2,228	26,459	8.4
Stickleback	2,156	20,787	10.4
Medaka	1,910	19,699	9.7
Tetraodon	1,853	19,602	9.5

Most teleost genes are 1:1 orthologs with human



nonfunctionalization

Gene loss: most frequent fate

neofunctionalization

New functions evolve from ancestral genes

subfunctionalization

Reciprocal loss of ancestral functions

We don't yet know the relative importance of these fates.

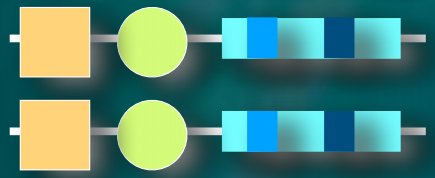
Alan Force

Force…Postlethwait 1999 *Genetics*

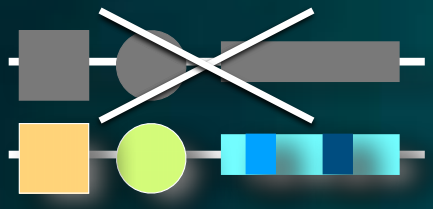
- Genome duplication complicates connectivity

Species	Retained TGD pairs	Coding genes	% retained
Zebrafish	2,228	26,459	8.4
Stickleback	2,156	20,787	10.4
Medaka	1,910	19,699	9.7
Tetraodon	1,853	19,602	9.5

Problem: All can occur in a lineage-specific fashion.

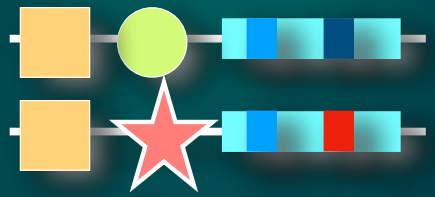


What does this mean for toxicology?



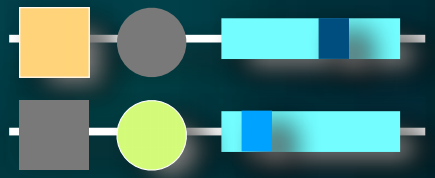
nonfunctionalization

Gene loss: most frequent fate



neofunctionalization

New functions evolve from ancestral genes



subfunctionalization

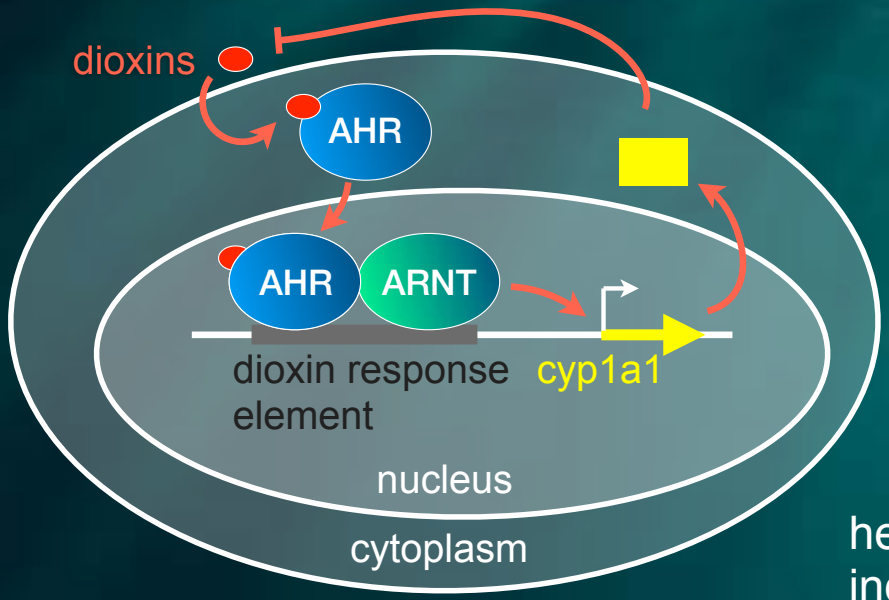
Reciprocal loss of ancestral functions

We don't yet know the relative importance of these fates.

Alan Force

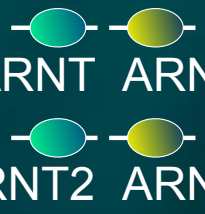
Force…Postlethwait 1999 *Genetics*

What does this mean for toxicology?



heterodimer with ligand-bound AHR, promotes expression of xenobiotic metabolism genes

heterodimer with CLOCK, promotes expression of circadian rhythm genes



heterodimer with hypoxia-inducible factor 1alpha, promotes expression of oxygen-responsive genes.

heterodimer with CLOCK & HIF1alpha

neofunctionalization **sub**functionalization

To see if this is **sub**functionalization or **neo**functionalization, must look at a pre-duplication outgroup.

Amphioxus

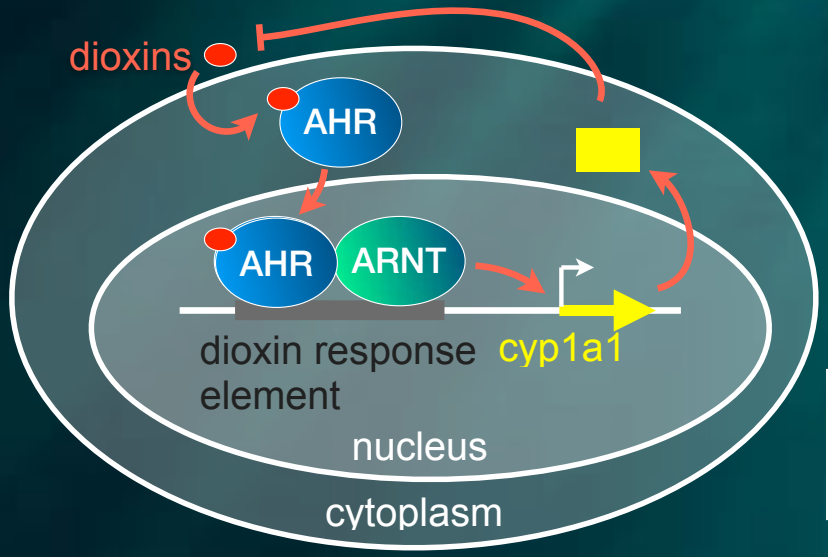
What about AHR?

Ascidians

Appendicularians

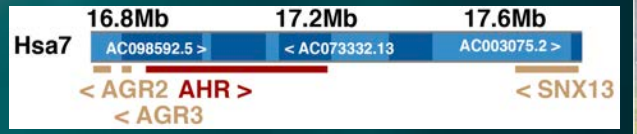
http://marineinvaders.myspecies.info/taxonomy/term/70 evodevogenomics-unibarcelona.weebly.com/oikopleura www2.le.ac.uk/departments/geology/people/goodall-r/personal/images/lancelet.

What does this mean for toxicology?



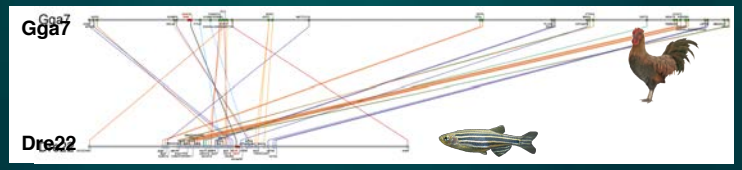
What about AHR?

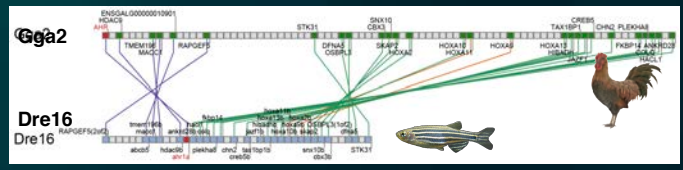
one AHR gene



three AHR-related genes

Conserved syntenies show that chicken and zebrafish AHR-related genes are orthologs

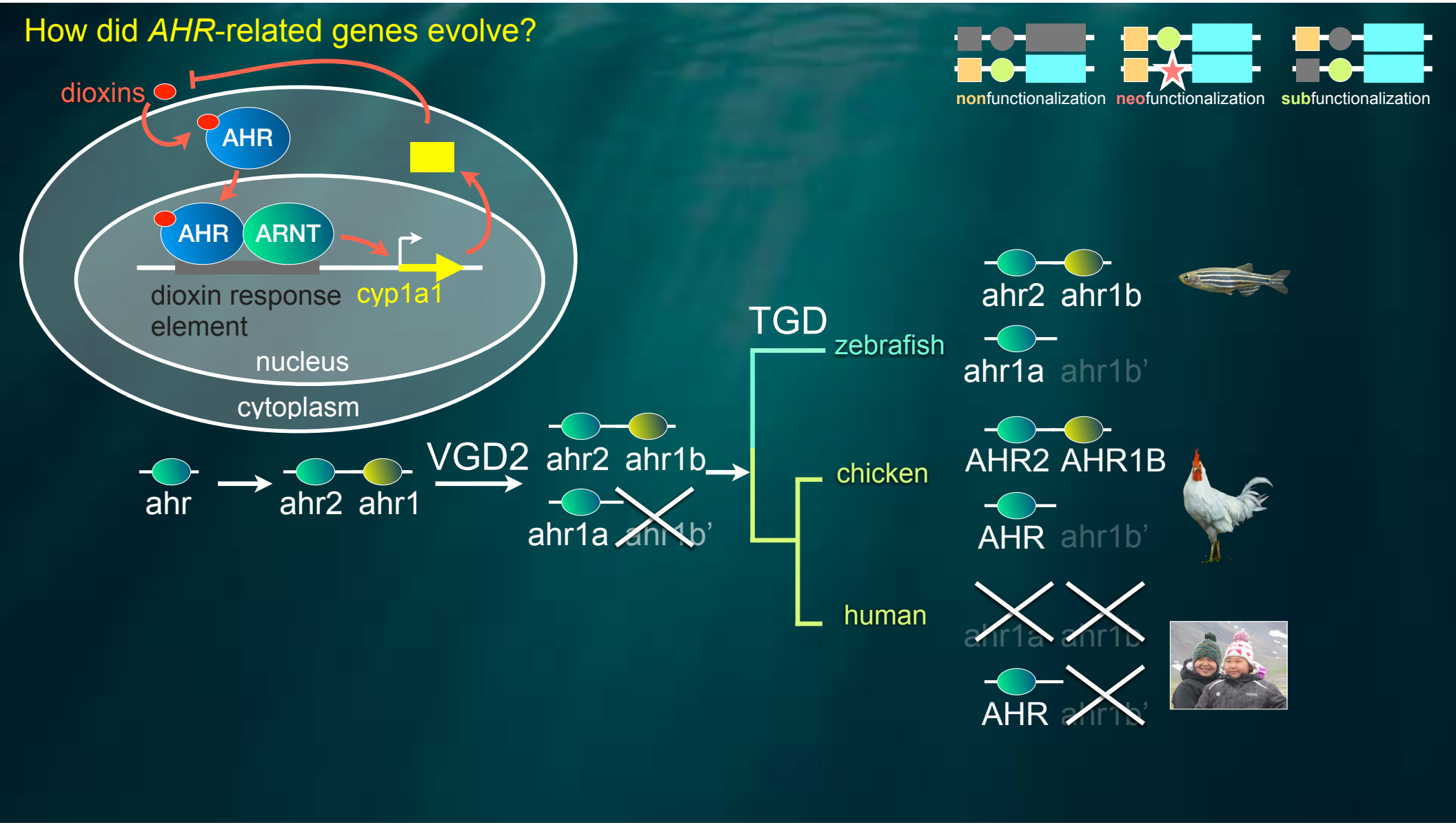




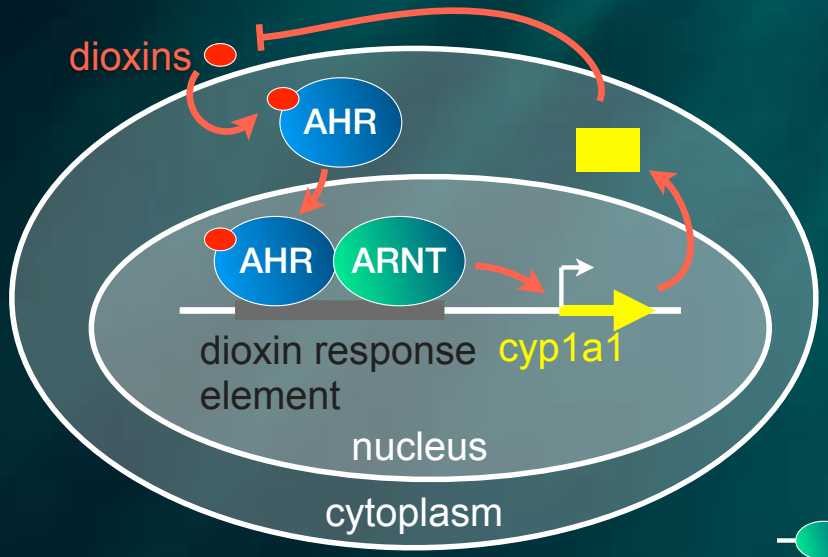
three AHR-related genes

How did *AHR*-related genes evolve?

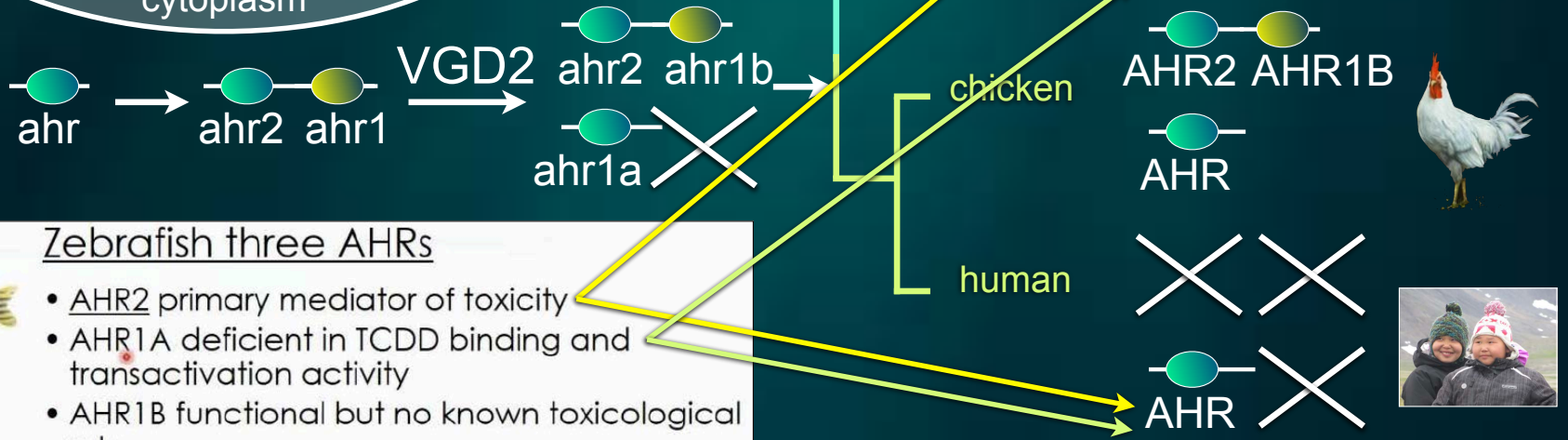
Karchner 2005 Biochem J. 392:153


How did AHR-related genes evolve?
nonfunctionalization
neofunctionalization
subfunctionalization
dioxins
AHR
AHR
ARNT
dioxin response element
cyp1a1
nucleus
cytoplasm
ahr
ahr2 ahr1
VGD2
ahr2 ahr1b
ahr1a ahr1b'
TGD
zebrafish
chicken
human
ahr2 ahr1b
ahr1a ahr1b'
AHR2 AHR1B
AHR ahr1b'
ahr1a ahr1b
AHR ahr1b'

How did *AHR*-related genes evolve?



The 'primary mediator of toxicity' for teleosts **is not** the ortholog of human *AHR*



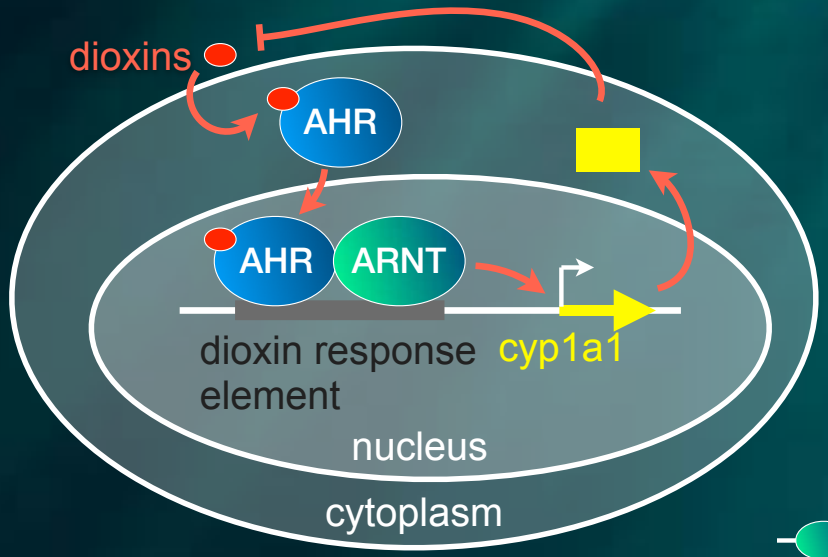
Zebrafish three AHRs

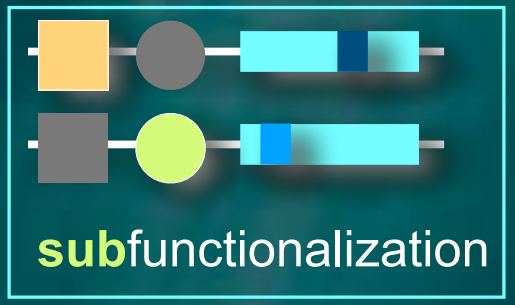
- AHR2 primary mediator of toxicity
- AHR1A deficient in TCDD binding and transactivation activity
- AHR1B functional but no known toxicological roles

Robert Tanguay Oregon State University 2018 10 25

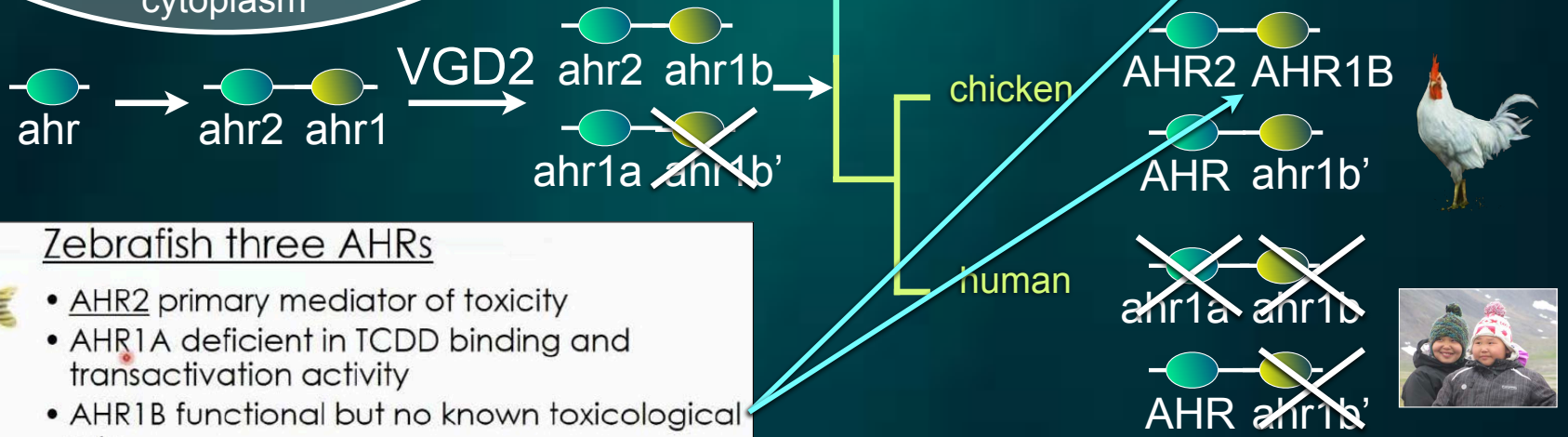
Zebrafish ortholog of human AHR is an 'incipient pseudogene'. Karchner 2005 Biochem J. 392:153

How did *AHR*-related genes evolve?





Study of *ahr1b* may reveal sub-functions of human *AHR* gene

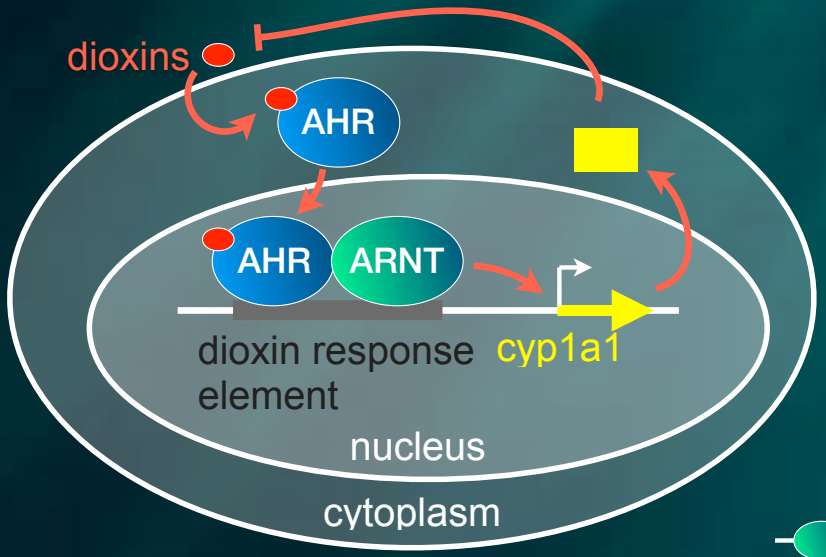


Zebrafish three AHRs

- AHR2 primary mediator of toxicity
- AHR1A deficient in TCDD binding and transactivation activity
- AHR1B functional but no known toxicological roles

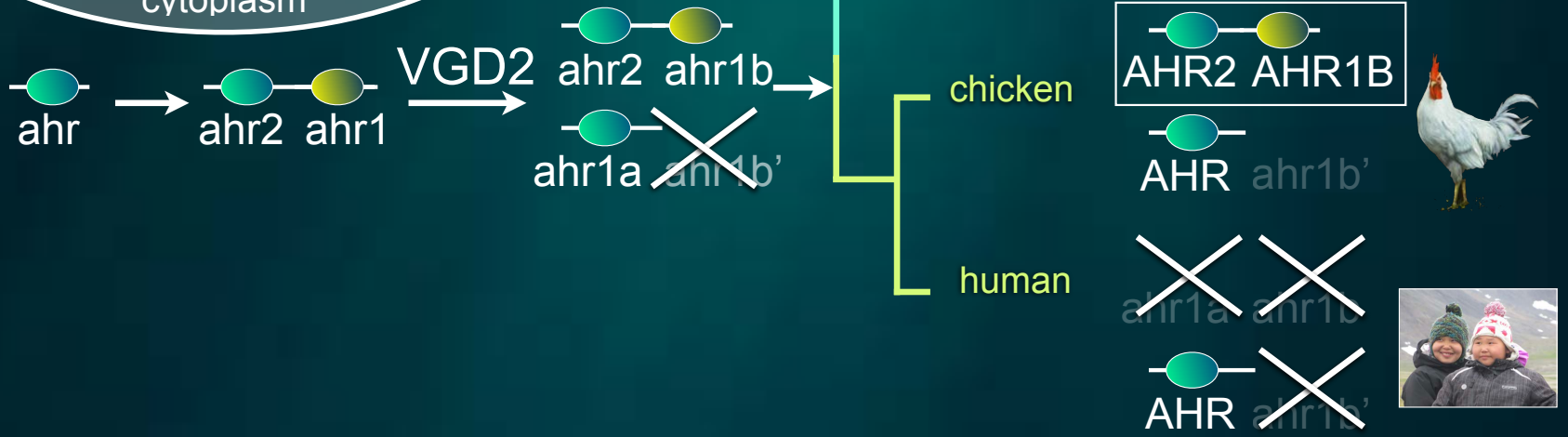
Robert Tanguay Oregon State University 2018 10 25

How did *AHR*-related genes evolve?



- Ohnologs gone missing and lineage-specific evolution

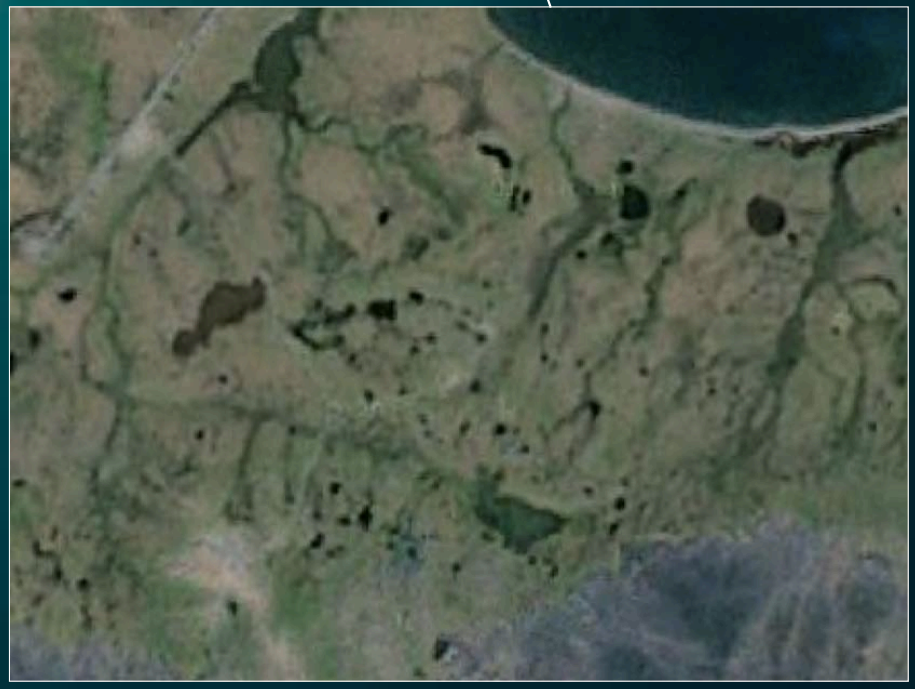
Postlethwait 2007 J Exp Zool B Mol Dev Evol. 308:563



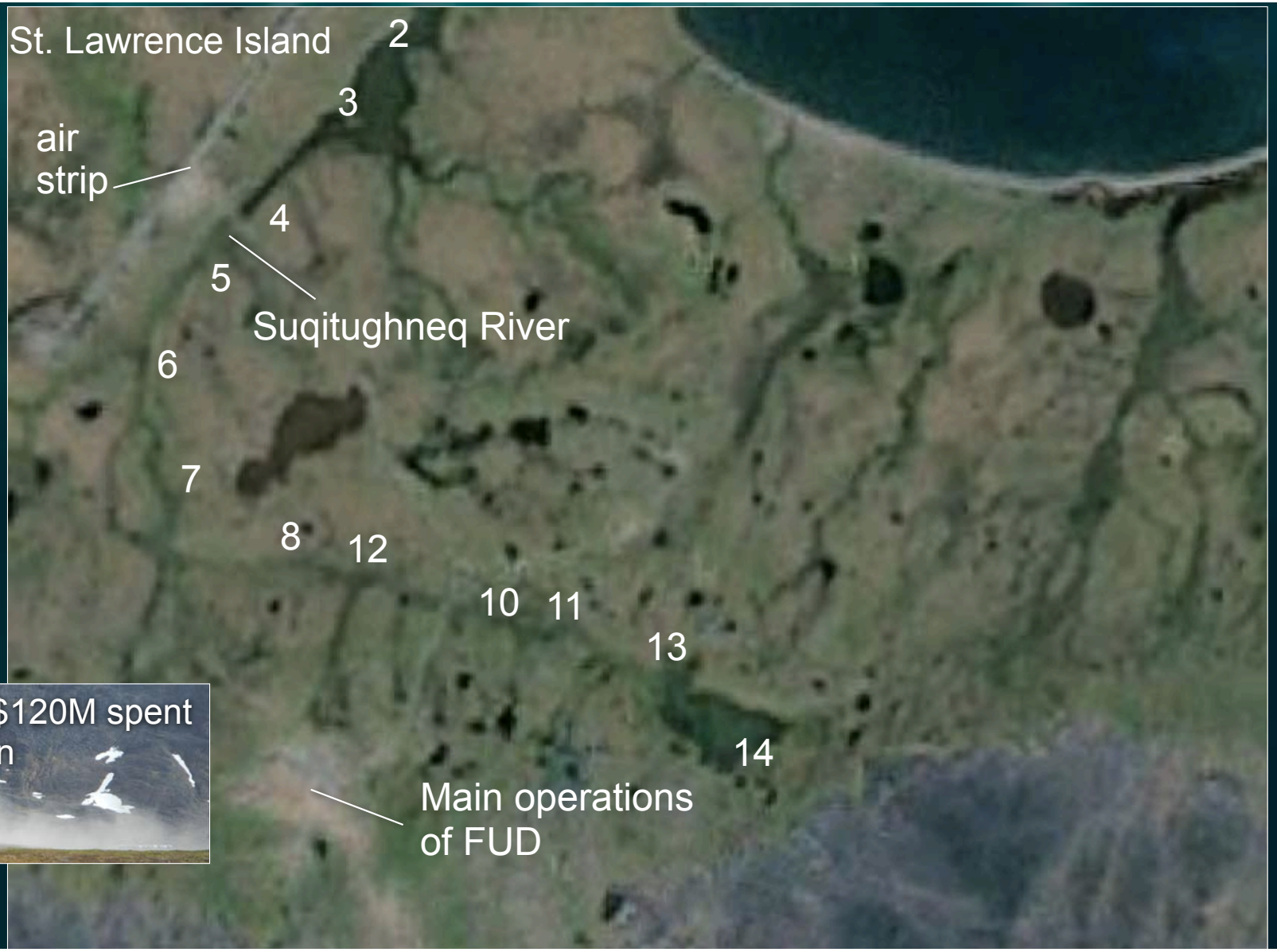
How to connect teleost genomes to human biology?

- Genome duplication in vertebrates
- Genome duplication complicates connectivity
- Ohnologs gone missing and lineage-specific evolution
- Application: St. Lawrence Island

• Application: St. Lawrence Island



• Application: St. Lawrence Island

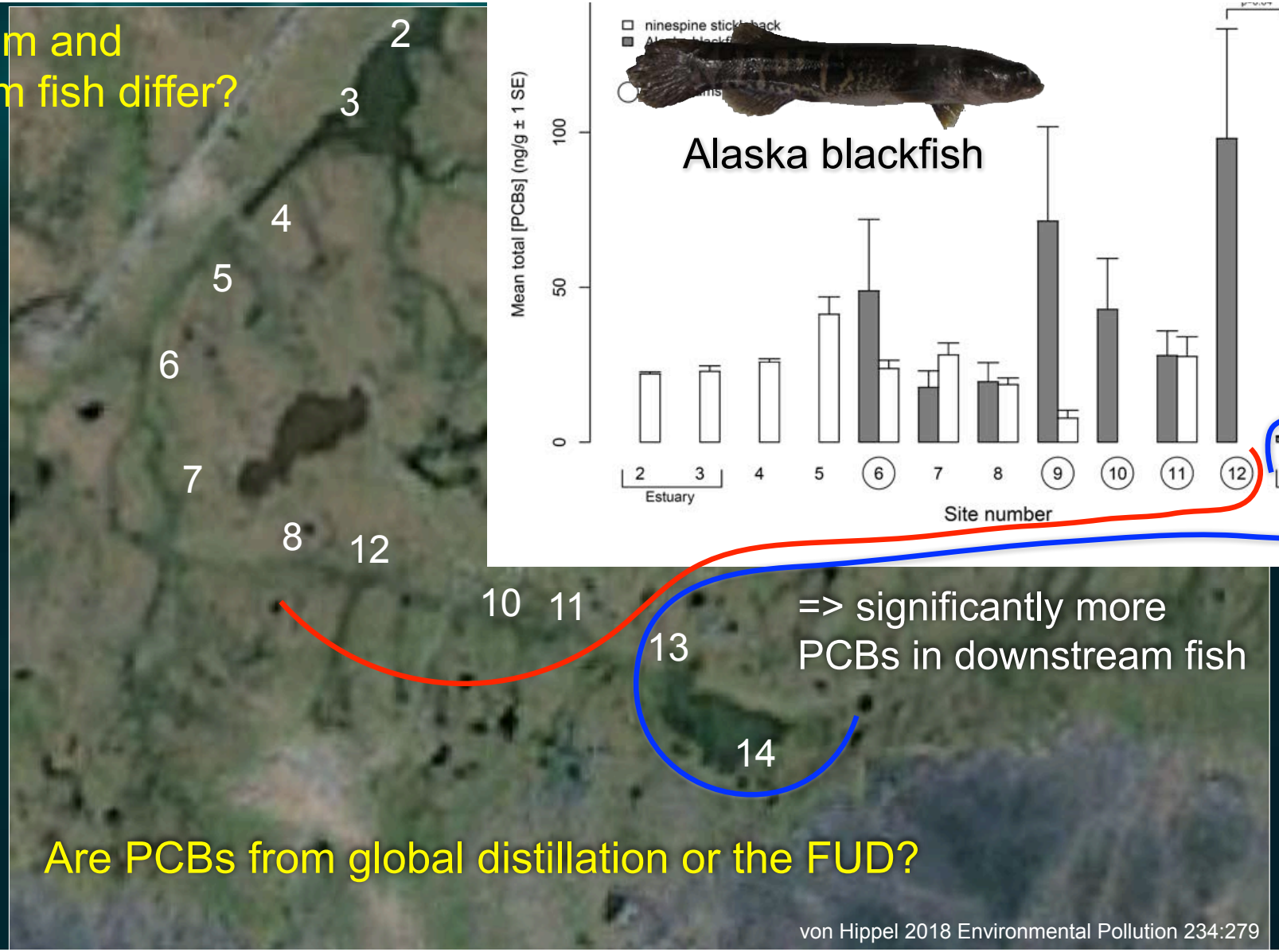


• Application: St. Lawrence Island

2
3
air strip
4
5
Suqitughneq River
6
7
8
12
10
11
13
14
Main operations of FUD

Do upstream and downstream fish differ?

Do upstream and downstream fish differ?

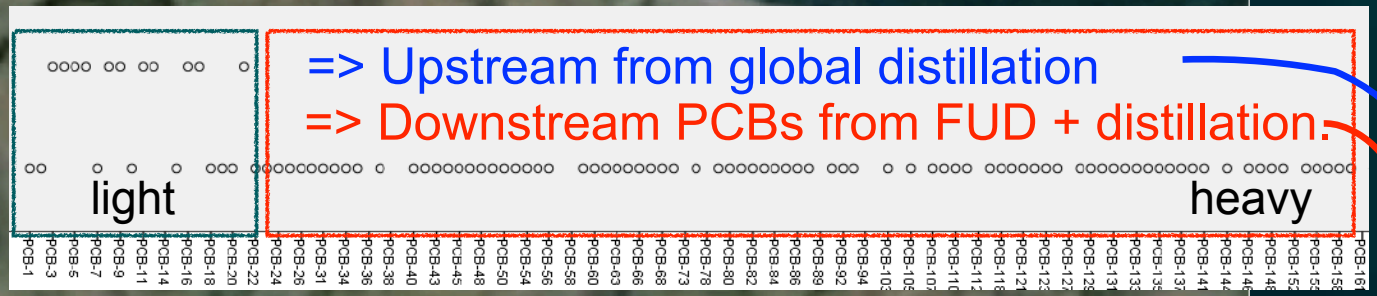


von Hippel 2018 Environmental Pollution 234:279

Are PCBs from global distillation or the FUD?

Greater upstream
Greater down stream

3
6
7
8
12
10
11
13
14



=> Upstream from global distillation
=> Downstream PCBs from FUD + distillation.

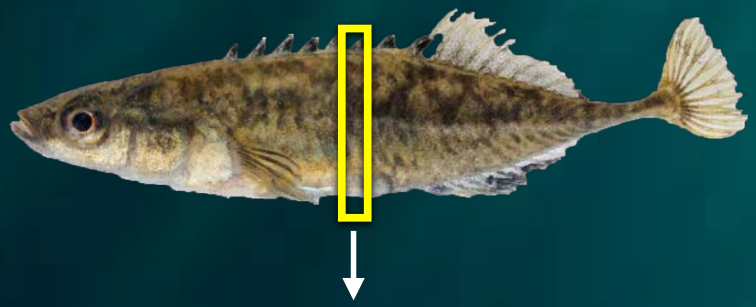

Danielle Dillon

Main operations of FUD

What are the biological effects?

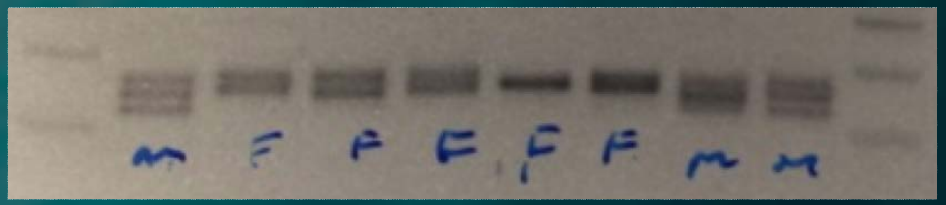
What are the biological effects?

Gene expression differences?



BIOO NEXTflex qRNA-seq

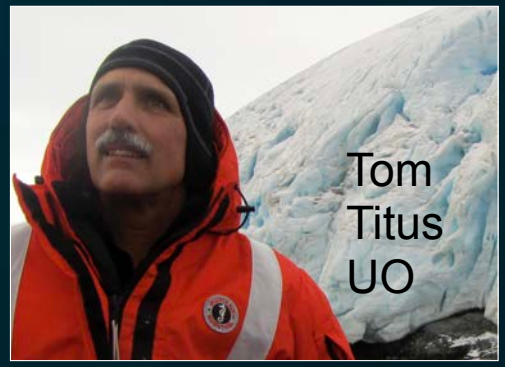
sex genotype PCR



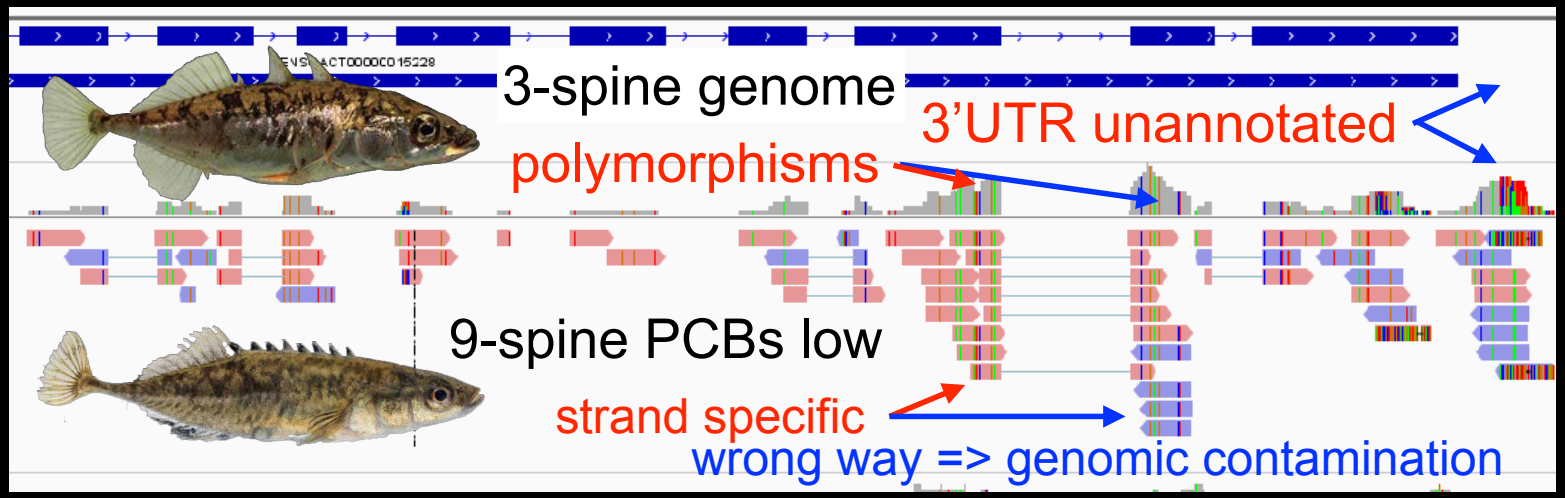
4 females from highly contaminated sites
4 females from less contaminated sites

Strand-specific: guards against gDNA

Unique molecular indexes: guards against PCR duplicates



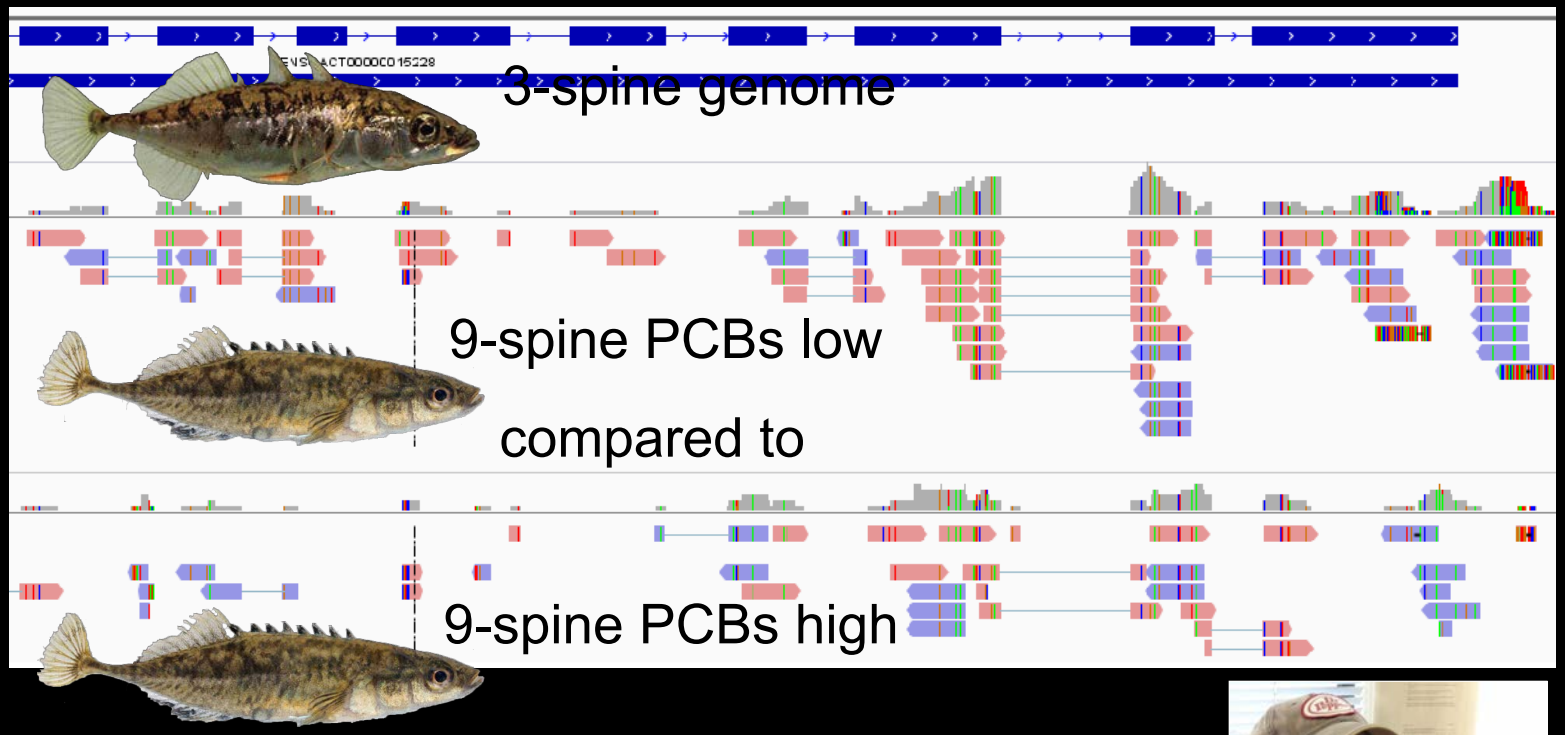
Aligned 9-spine reads to 3-spine genome



Aligned 9-spine reads to 3-spine genome

brca2

paired-end reads



DESeq2: Finds differentially expressed genes

DESeq2: Finds differentially expressed genes

	Differentially expressed genes
Females: contamin/uncontam	2312
Males: contamin/uncontam	642

What do these genes do?

DAVID 6.7: Database for Annotation, Visualization and Integrated Discovery

What do these genes do?

brca2 (fancd1) = down
cdc23 = down
fbxo5 = down
h2ax = down
haus6 = down
mad2l1 = down
rab11fip4a = down
llgl2 = down
ccnb1 = down
most tgf betas down

DAVID 6.7:

	Annotation Cluster 1	Enrichment Score: 15.17		Count	P_Value	Benjamini
☐	UP_KEYWORDS	Cell cycle	RT	135	1.9E-20	5.2E-18
☐	UP_KEYWORDS	Cell division	RT	91	1.2E-17	1.7E-15
☐	UP_KEYWORDS	Mitosis	RT	69	3.7E-16	3.0E-14
☐	GOTERM_BP_DIRECT	cell division	RT	83	1.1E-13	2.6E-10
☐	GOTERM_BP_DIRECT	mitotic nuclear division	RT	63	1.5E-11	2.3E-8

cell cycle genes

What do these genes do?

brca2 (fancd1) = down
cdc23 = down
fbxo5 = down
h2ax = down
haus6 = down
mad2l1 = down
rab11fip4a = down
llgl2 = down
ccnb1 = down
most tgf betas down

mcm3,4,6,7
cdc45
hsc70
pcna
rbbp4
tert
terf1, terf2

Annotation Cluster 1	Enrichment Score: 15.17			Count	P_Value	Benjamini
UP_KEYWORDS	Cell cycle	RT		135	1.9E-20	5.2E-18
UP_KEYWORDS	Cell division	RT		91	1.2E-17	1.7E-15
UP_KEYWORDS	Mitosis	RT		69	3.7E-16	3.0E-14
GOTERM_BP_DIRECT	cell division	RT		83	1.1E-13	2.6E-10
GOTERM_BP_DIRECT	mitotic nuclear division	RT		63	1.5E-11	2.3E-8
Annotation Cluster 2	**Enrichment Score: 11.74**			**Count**	**P_Value**	**Benjamini**
UP_KEYWORDS	DNA replication	RT		38	1.5E-15	9.8E-14
GOTERM_BP_DIRECT	G1/S transition of mitotic cell cycle	RT				E-8
GOTERM_BP_DIRECT	DNA replication initiation	RT		19	1.3E-10	1.2E-7

DNA replication genes

What do these genes do?

brca2 (fancd1) = down
cdc23 = down
fbxo5 = down
h2ax = down
haus6 = down
mad2l1 = down
rab11fip4a = down
llgl2 = down
ccnb1 = down
most tgf betas down

mcm3,4,6,7
cdc45
hsc70
pcna
rbbp4
tert
terf1, terf2

Fanconi anemia genes =

Annotation Cluster 1	Enrichment Score: 15.17			Count	P_Value	Benjamini
UP_KEYWORDS	Cell cycle	RT		135	1.9E-20	5.2E-18
UP_KEYWORDS	Cell division	RT		91	1.2E-17	1.7E-15
UP_KEYWORDS	Mitosis	RT		69	3.7E-16	3.0E-14
GOTERM_BP_DIRECT	cell division	RT		83	1.1E-13	2.6E-10
GOTERM_BP_DIRECT	mitotic nuclear division	RT		63	1.5E-11	2.3E-8
Annotation Cluster 2	**Enrichment Score: 11.74**			**Count**	**P_Value**	**Benjamini**
UP_KEYWORDS	DNA replication	RT		38	1.5E-15	9.8E-14
GOTERM_BP_DIRECT	G1/S transition of mitotic cell cycle	RT		36	3.1E-11	3.7E-8
GOTERM_BP_DIRECT	DNA replication initiation	RT		19	1.3E-10	1.2E-7
Annotation Cluster 3	**Enrichment Score: 6.94**			**Count**	**P_Value**	**Benjamini**
UP_KEYWORDS	DNA damage	RT		67	4.4E-9	1.8E-7
UP_KEYWORDS	DNA repair			58	1.8E-8	5.4E-7
GOTERM_BP_DIRECT	DNA repair	RT		46	1.9E-5	5.7E-3

DNA repair genes

What do these genes do?

brca2 (fancd1) = down
cdc23 = down
fbxo5 = down
h2ax = down
haus6 = down
mad2l1 = down
rab11fip4a = down
llgl2 = down
ccnb1 = down
most tgf betas down

mcm3,4,6,7
cdc45
hsc70
pcna
rbbp4
tert
terf1, terf2

Fanconi anemia genes =

ercc4
msh6
pola1, pola2
prim1

	Annotation Cluster 1	Enrichment Score: 15.17			Count	P_Value	Benjamini
	UP_KEYWORDS	Cell cycle	RT		135	1.9E-20	5.2E-18
	UP_KEYWORDS	Cell division	RT		91	1.2E-17	1.7E-15
	UP_KEYWORDS	Mitosis	RT		69	3.7E-16	3.0E-14
	GOTERM_BP_DIRECT	cell division	RT		83	1.1E-13	2.6E-10
	GOTERM_BP_DIRECT	mitotic nuclear division	RT		63	1.5E-11	2.3E-8
	Annotation Cluster 2	**Enrichment Score: 11.74**			**Count**	**P_Value**	**Benjamini**
	UP_KEYWORDS	DNA replication	RT		38	1.5E-15	9.8E-14
	GOTERM_BP_DIRECT	G1/S transition of mitotic cell cycle	RT		36	3.1E-11	3.7E-8
	GOTERM_BP_DIRECT	DNA replication initiation	RT		19	1.3E-10	1.2E-7
	Annotation Cluster 3	**Enrichment Score: 6.94**			**Count**	**P_Value**	**Benjamini**
	UP_KEYWORDS	DNA damage	RT		67	4.4E-9	1.8E-7
	UP_KEYWORDS	DNA repair	RT		58	1.8E-8	5.4E-7
	GOTERM_BP_DIRECT	DNA repair	RT		46	1.9E-5	5.7E-3
	Annotation Cluster 4	**Enrichment Score: 6.39**			**Count**	**P_Value**	**Benjamini**
	UP_KEYWORDS	Centromere	RT		36	1.0E-8	3.6E-7
	UP_KEYWORDS	Chromosome			8		3.7E-7
	GOTERM_BP_DIRECT	sister chromatid cohesion			8		6.9E-6
	UP_KEYWORDS	Kinetochore			7		2.8E-6
	GOTERM_CC_DIRECT	kinetochore	RT		20	1.7E-4	9.4E-3
	GOTERM_CC_DIRECT	condensed chromosome kinetochore	RT		21	1.9E-4	9.8E-3

chromosome behavior genes

What about duplicated genes?

What about duplicated genes?

246 TGD gene pairs had at least one copy that was differentially expressed.
21 pairs had both copies differentially expressed.

4 of those 21 pairs had one copy up, other copy down in contaminated fish.

Gene	baseMean	log2FC	pvalue	padj
camk1ga	27.345	-2.275	0.003	0.037
camk1gb	60.386	1.587	0.010	0.075
fabp10a	5935.662	2.457	0.000	0.007
fabp10b	32.048	-1.766	0.001	0.019
fam65a	91.364	1.367	0.002	0.028
fam65b	64.561	-1.156	0.009	0.074
kcnj2a	97.195	-2.508	0.001	0.012
kcnj2b	18.028	1.338	0.011	0.084

225 TGD pairs had only one copy DE.
Study of how they differ with toxicants may give insight into mechanisms.

neofunctionalization **sub**functionalization

TGD pairs provide that important advantage for analysis of toxicity.

Water at North East Cape is toxic to stickleback genome.

Water at North East Cape is toxic to stickleback genome.

DNA repair pathways: **low** expression in contaminated sites.

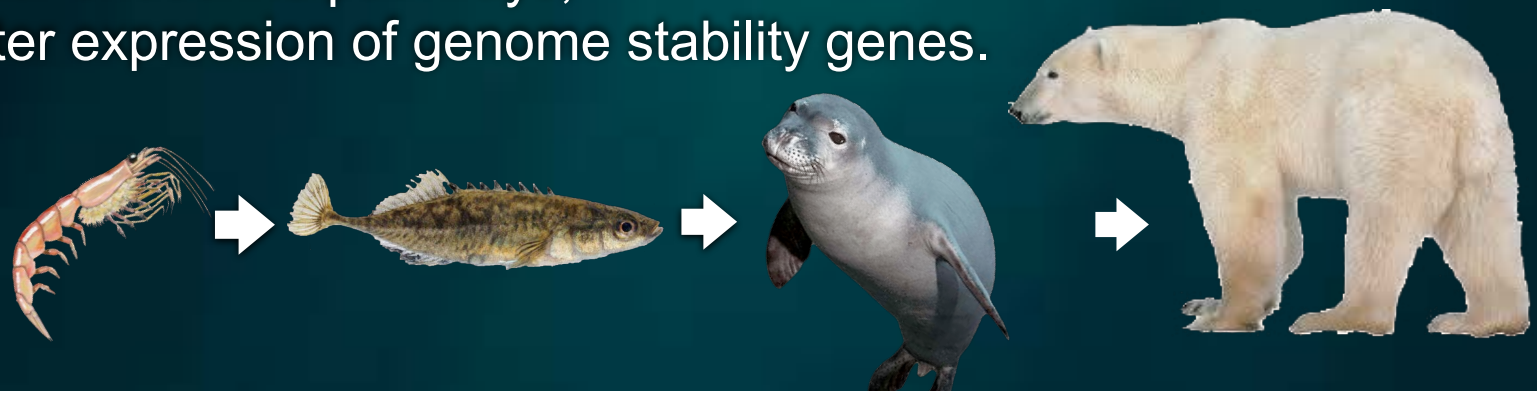
Cell cycle genes: **low** expression in contaminated sites.

DNA replication: **low** in contaminated sites.

Estrogenic effect on males.

Even after site remediation:
- contaminants persist,
- accumulate in fish,
- alter endocrine pathways,
- alter expression of genome stability genes.

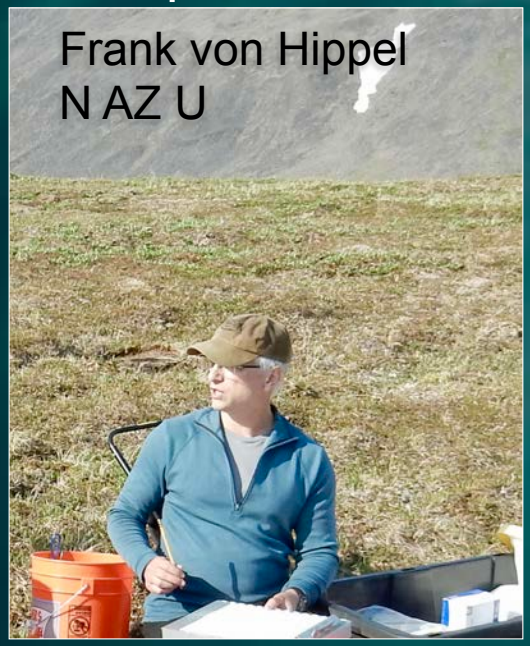
Health of indigenous people is put at risk.

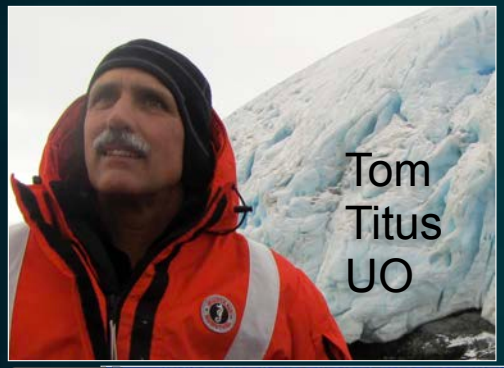


How to connect teleost genomes to human biology?

- Genome duplication in vertebrates
- Genome duplication complicates connectivity
- Ohnologs gone missing and lineage-specific evolution
- Application: St. Lawrence Island

Genome duplication and fish models for toxicology

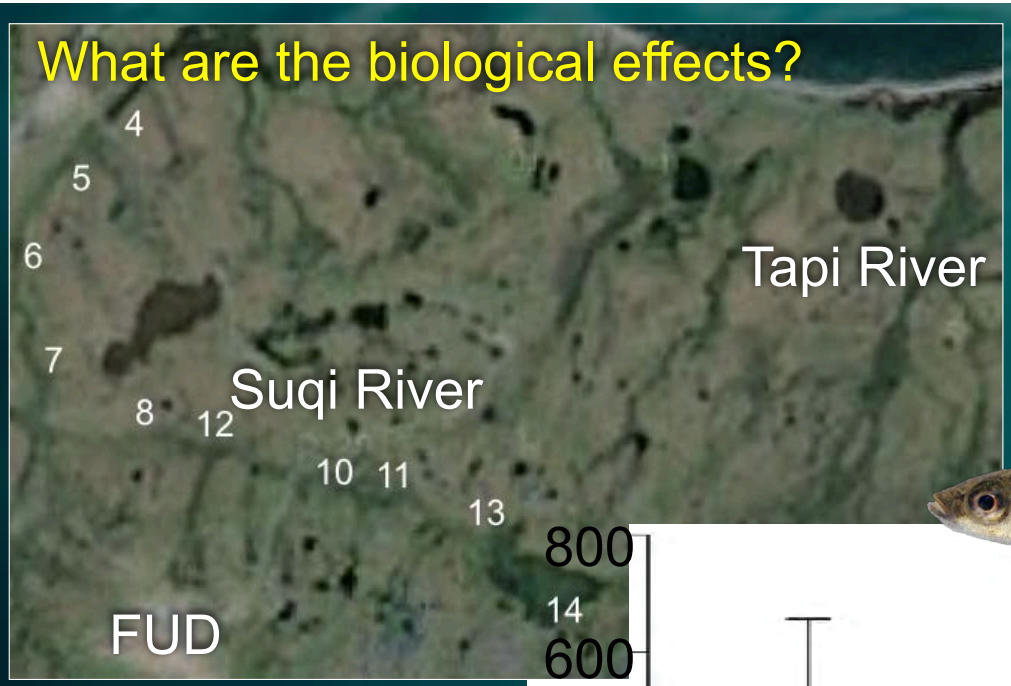




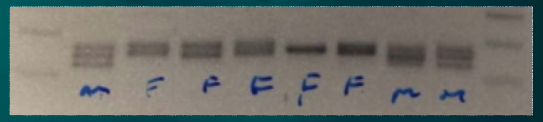
NIH DIVISION OF COMPARATIVE MEDICINE

NIH National Institute of Environmental Health Sciences
Your Environment. Your Health.

What are the biological effects?

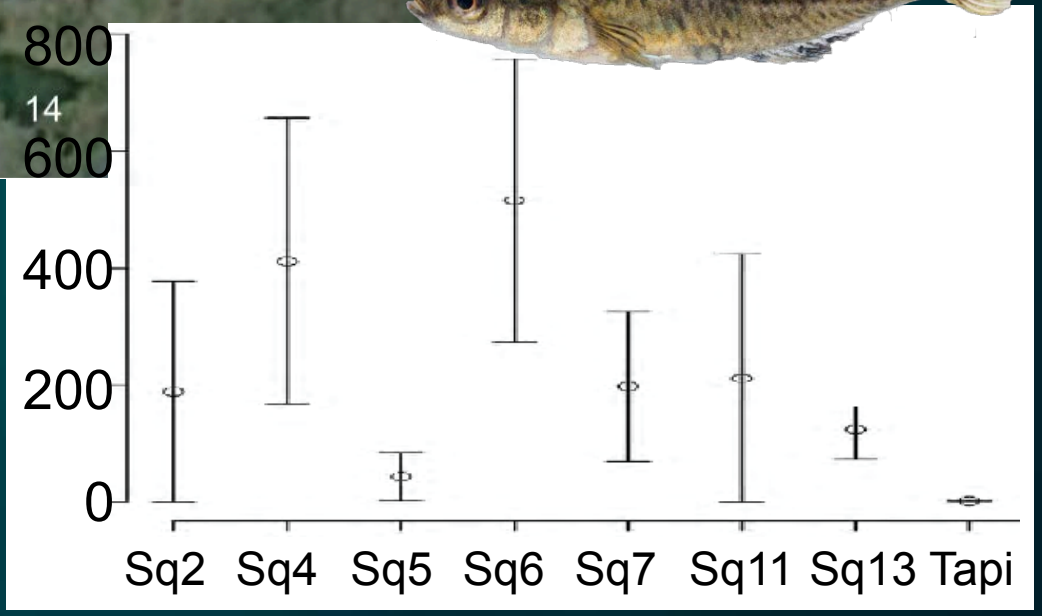


sex genotype PCR



Ninespine stickleback

[vitellogenin] in males (µg/g)

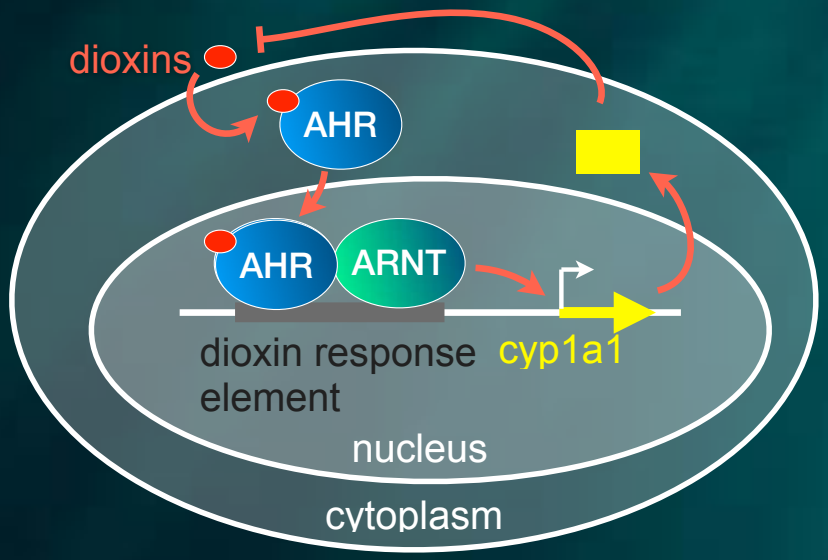


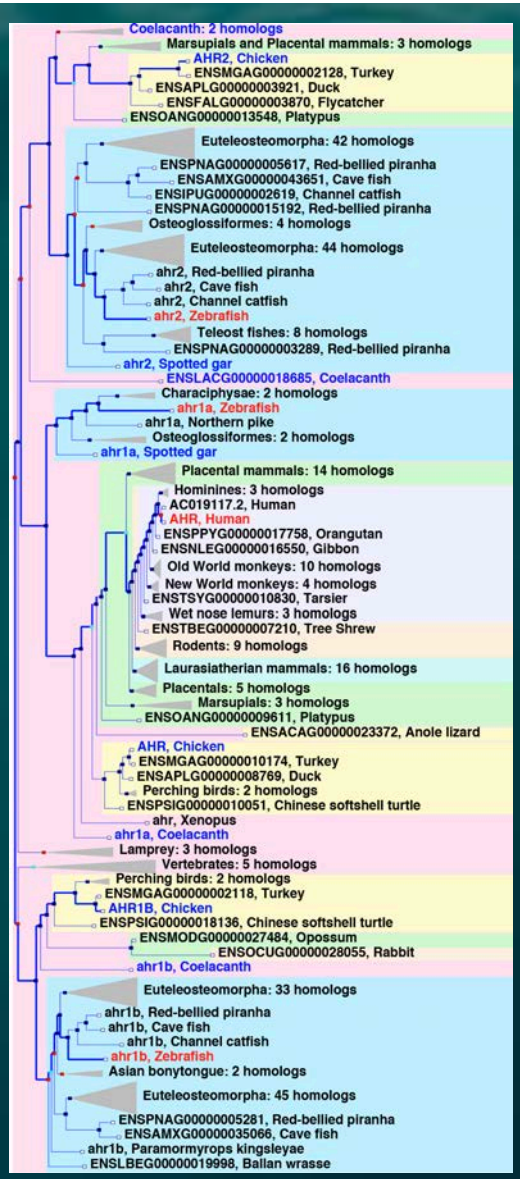
=> estrogenic effect on Suqi males

What is the genetic response?

What does this mean for toxicology?

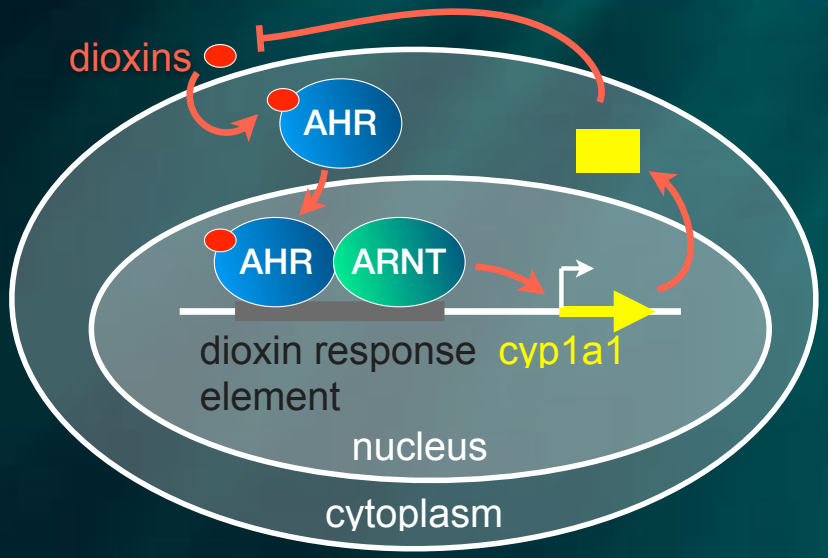
What about AHR?

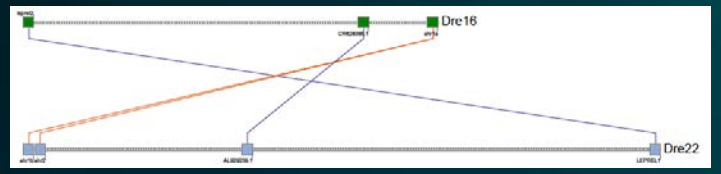


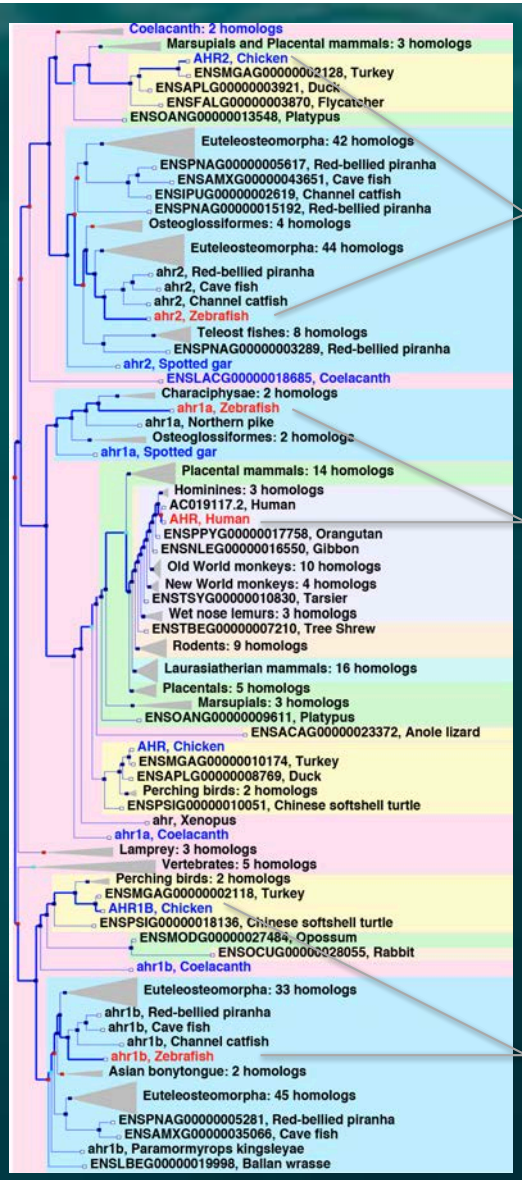


Karchner 2005 Biochem J. 392:153

What does this mean for toxicology?

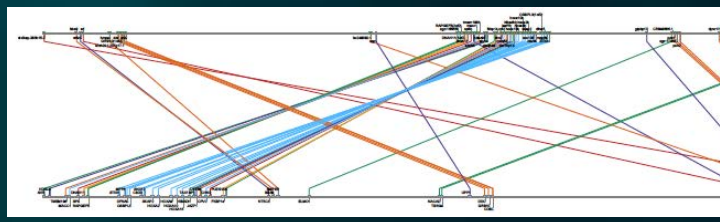






zebrafish and chicken have *ahr2*

zebrafish *ahr1a* appears to be ortholog of human *AHR*



zebrafish and chicken have *ahr1b*

Karchner 2005 Biochem J. 392:153