

Swiss Institute of  
Bioinformatics

## 2- Function large scale

Marc Robinson-Rechavi

SIB course Biodiversity bioinformatics 2023



*Unil*  
UNIL | Université de Lausanne  
Département d'écologie  
et évolution



@marcrr@ecoevo.social

[www.bgee.org](http://www.bgee.org)

---

# Evolution of gene expression

---

# Discussion

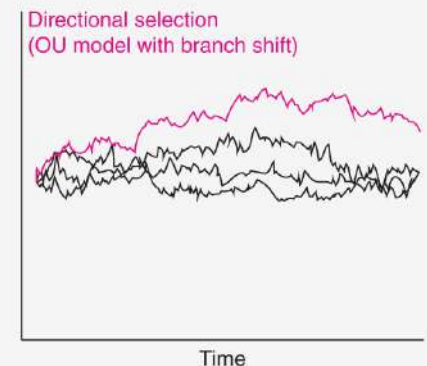
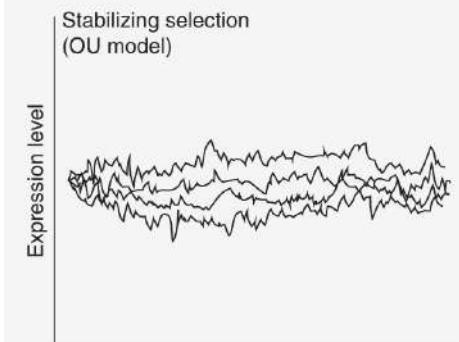
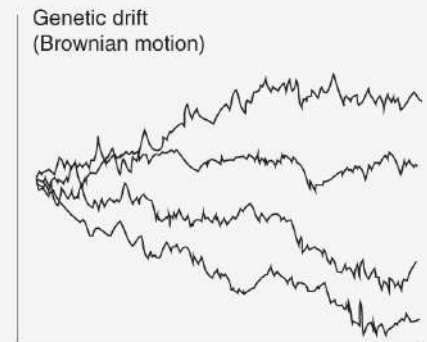
---

Describe in 2 words what you expect of the evolution of gene expression

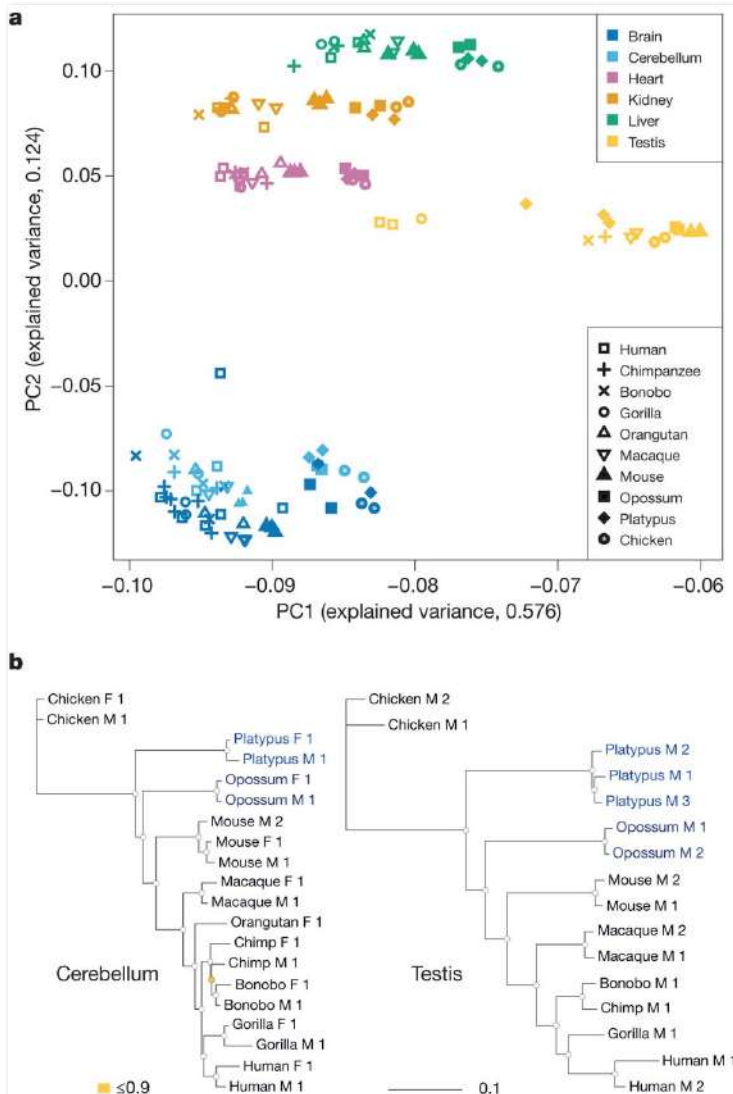
---

# Gene expression is a quantitative trait

- Expression level
- Tissue-specificity
- Ratio of expression between conditions, e.g., male/female

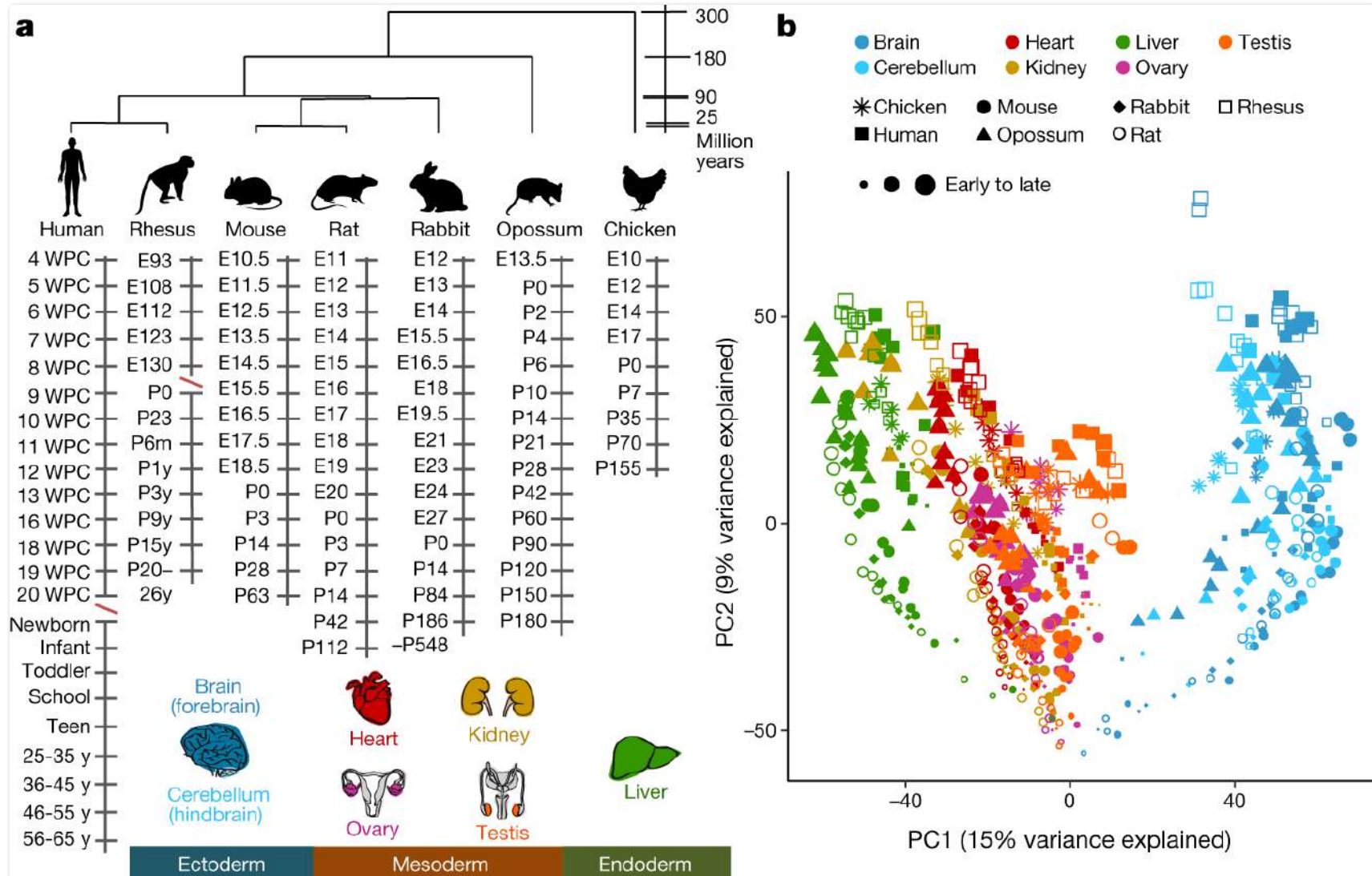


# Evolution of gene expression in mammals



Gene expression clusters by homologous organ, not by species

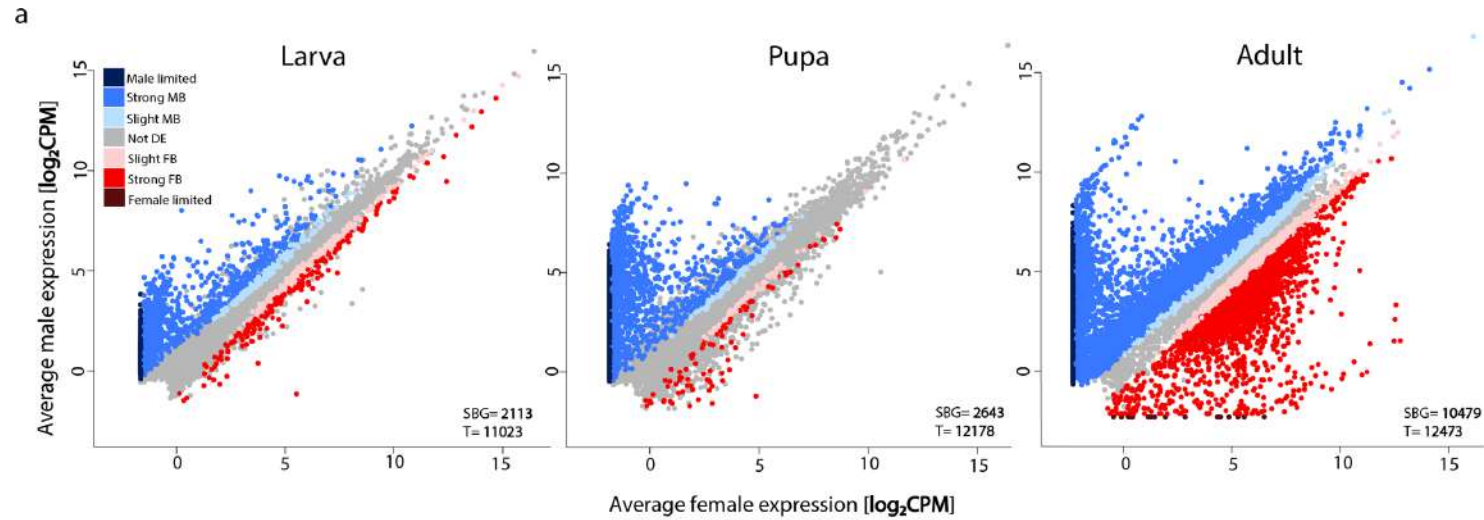
# Evolution of gene expression in mammals



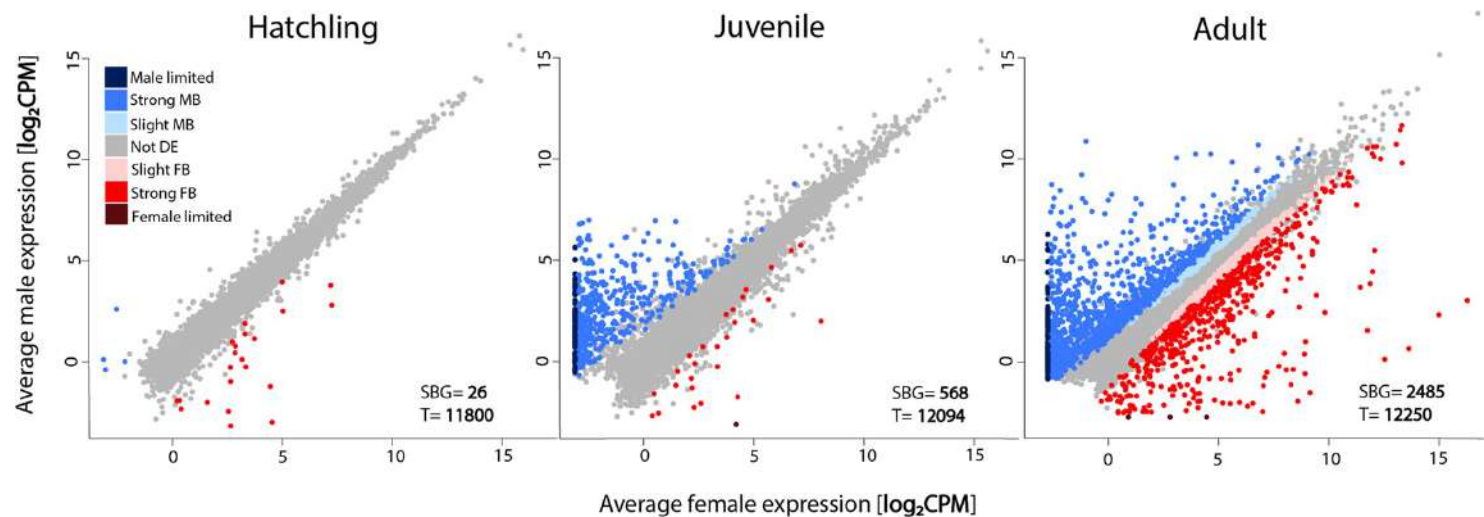
# Sex-biased gene expression



*Drosophila melanogaster*

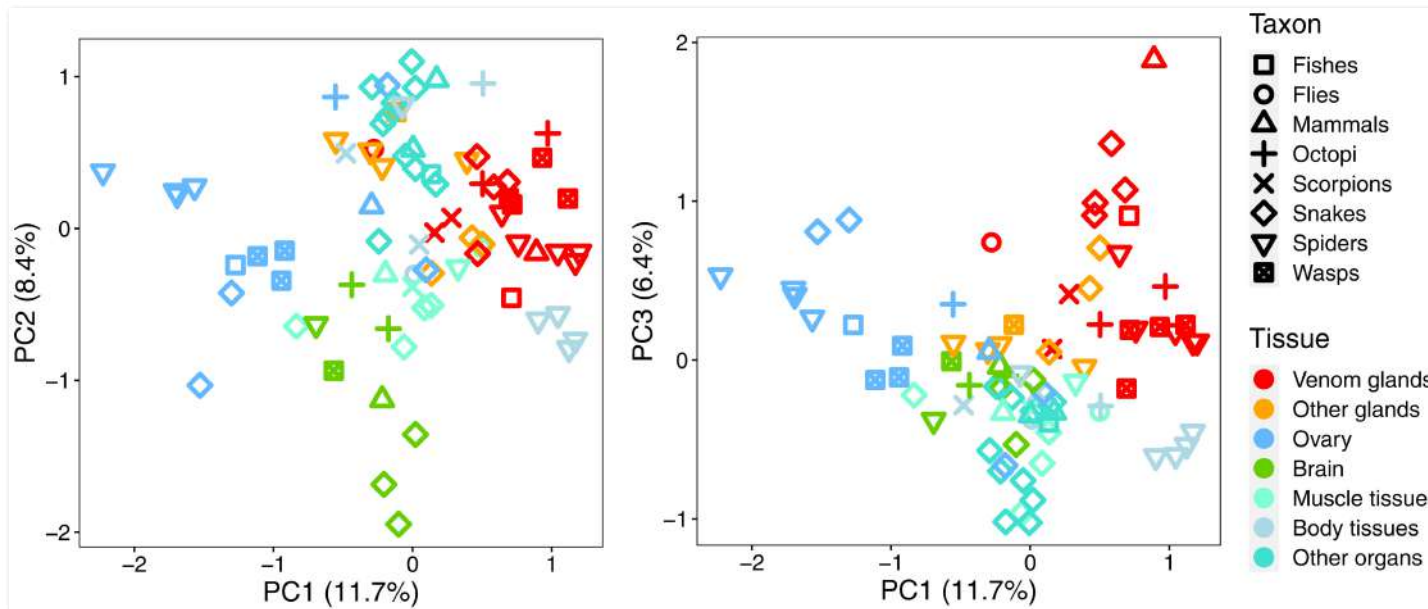


*Timema californicum*



# Convergent evolution of gene expression in venom glands

---





---

**Gene expression is  
informative for  
organismal evolution**

---

---

**"Ortholog conjecture"**

---

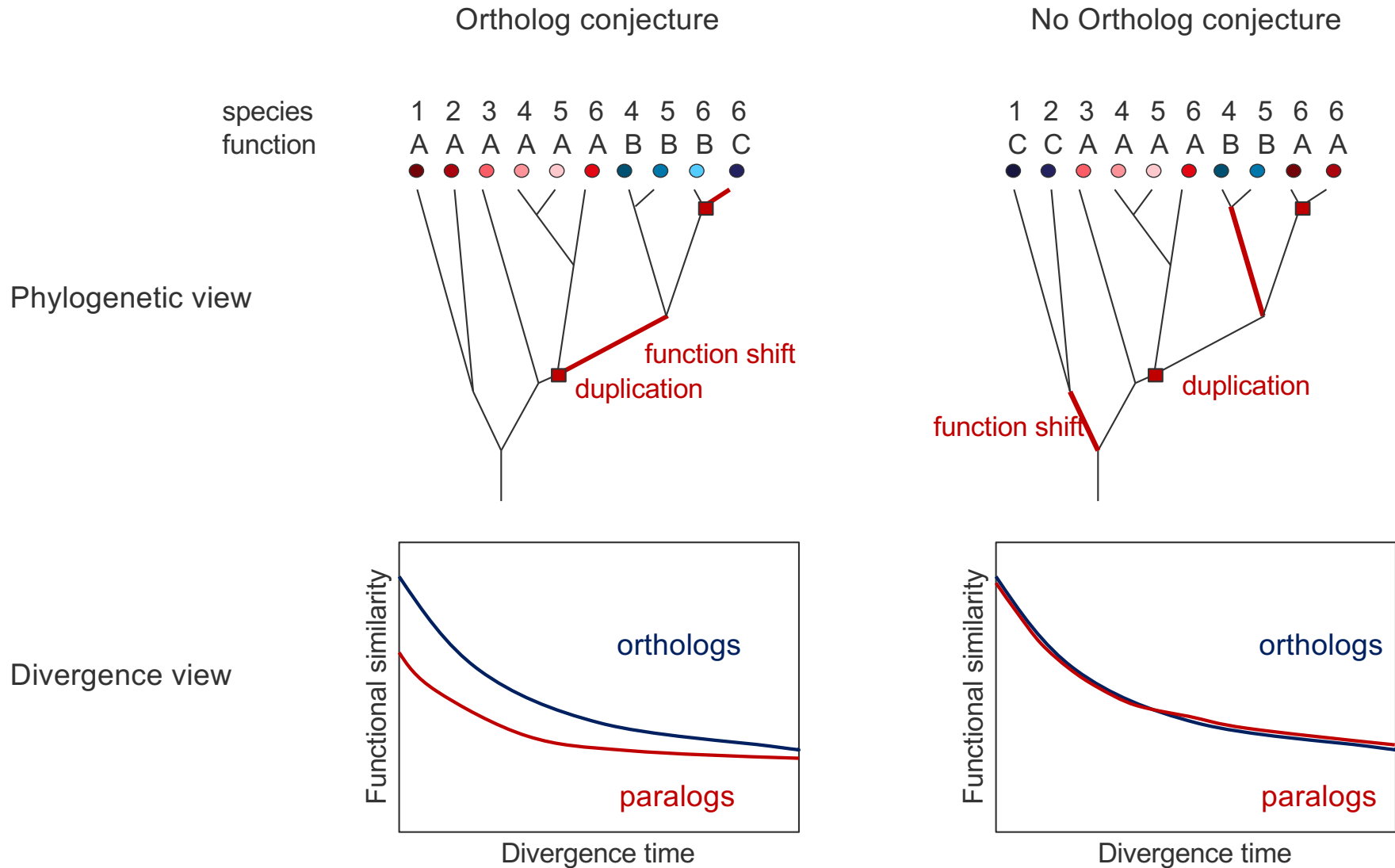
# Poll

---

To transfer annotations to a non model organism, you trust most:

1. protein domains
  2. orthologous genes
  3. genes with a Blast hit
-

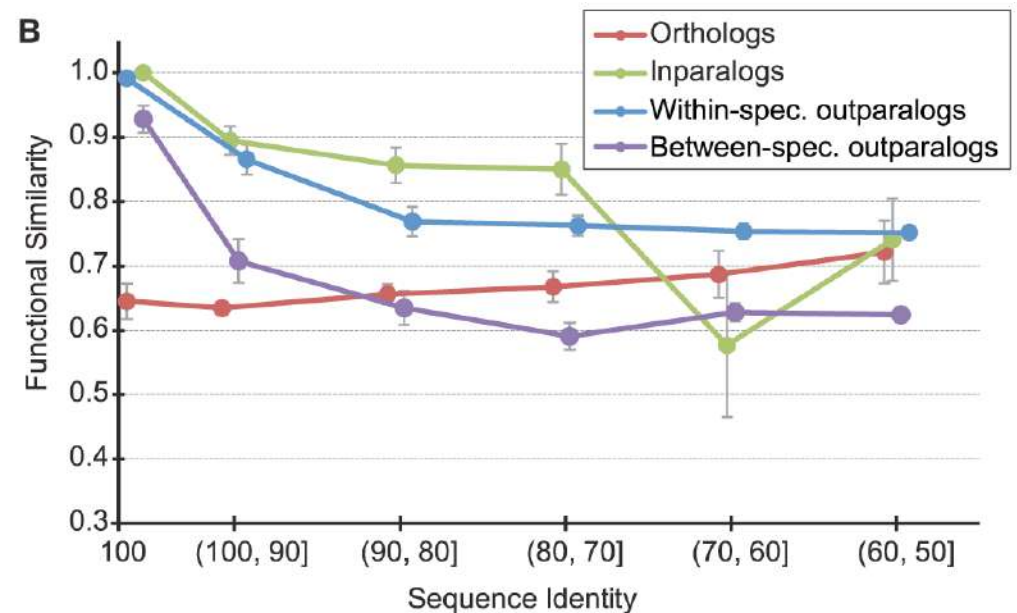
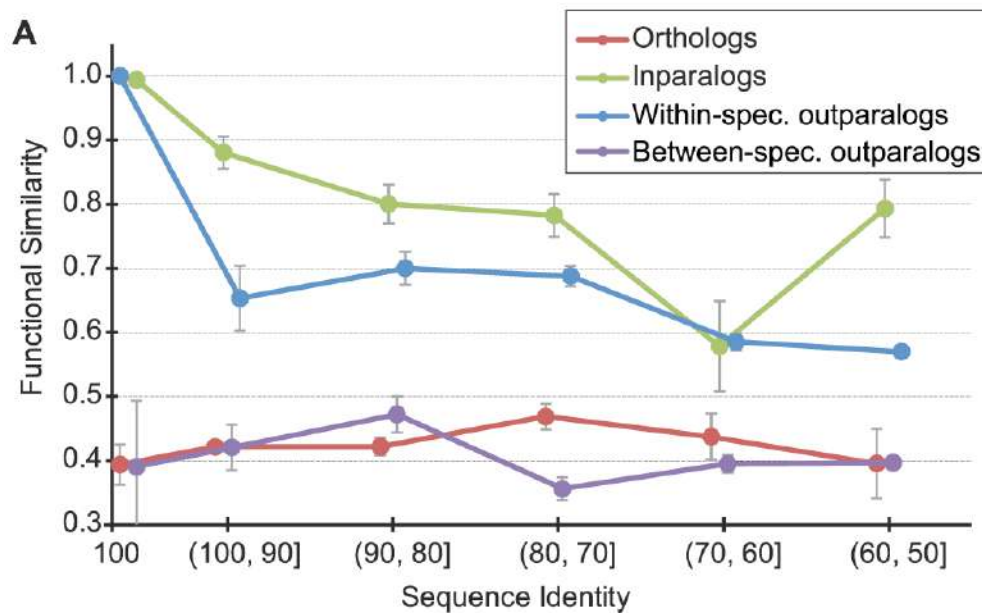
# Orthology and function



# GO similarity of orthologs and paralogs: Surprise!

Human-mouse

(A) Biological Process ontology, (B) Molecular Function ontology

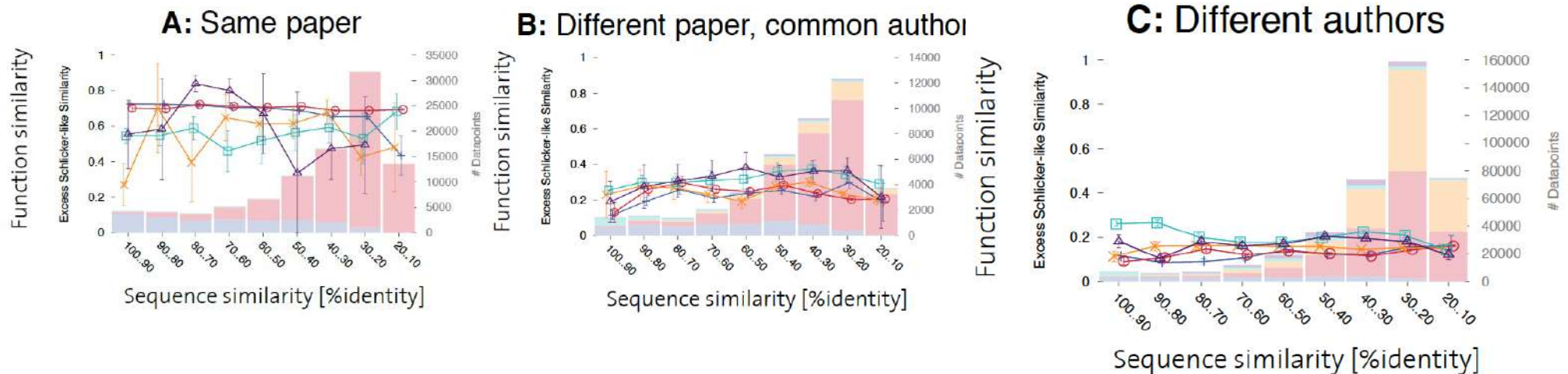


# GO similarity of orthologs and paralogs: Bias...

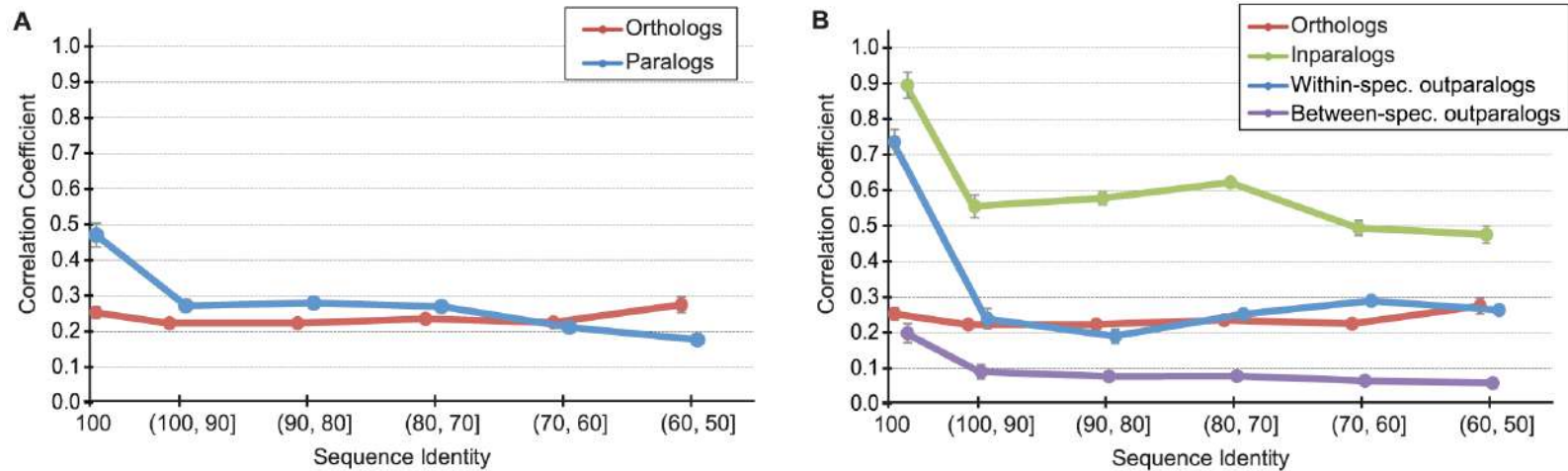
Dataset	Same-Species Paralogs	Different-Species Paralogs	1:1 Orthologs	Other Orthologs
Same Publications	1573	18	154	44
Different Publications, Same Authors	613	382	874	434
No Common Author	2492	20719	13149	11296
All Experimental Annotations	3312	20766	13309	11371

47%

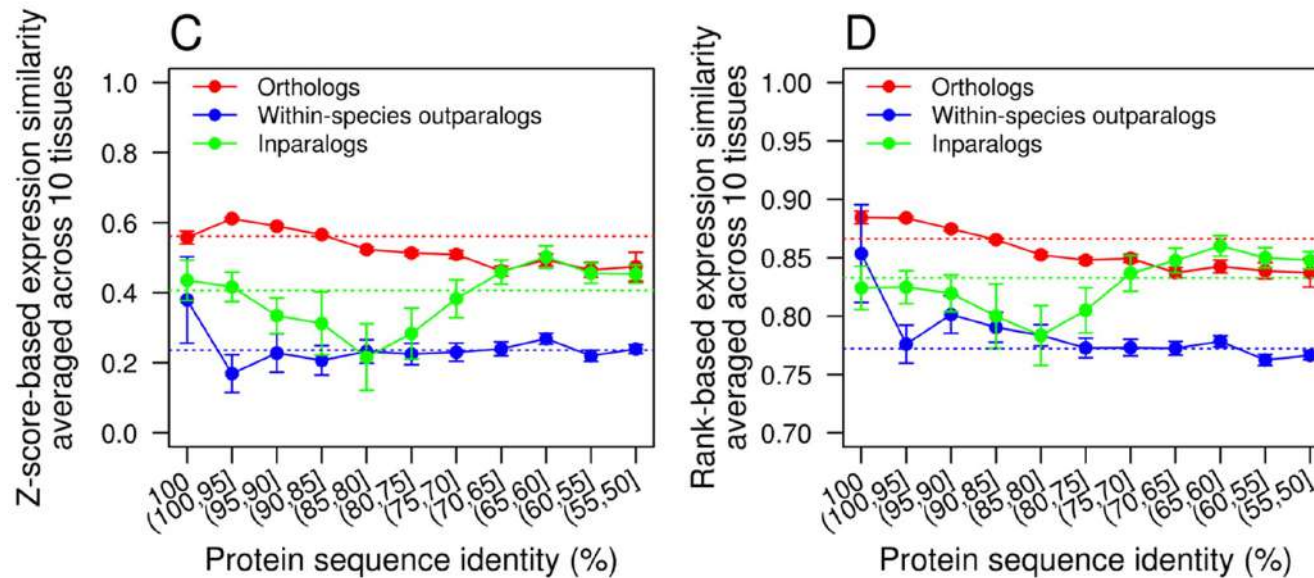
1%



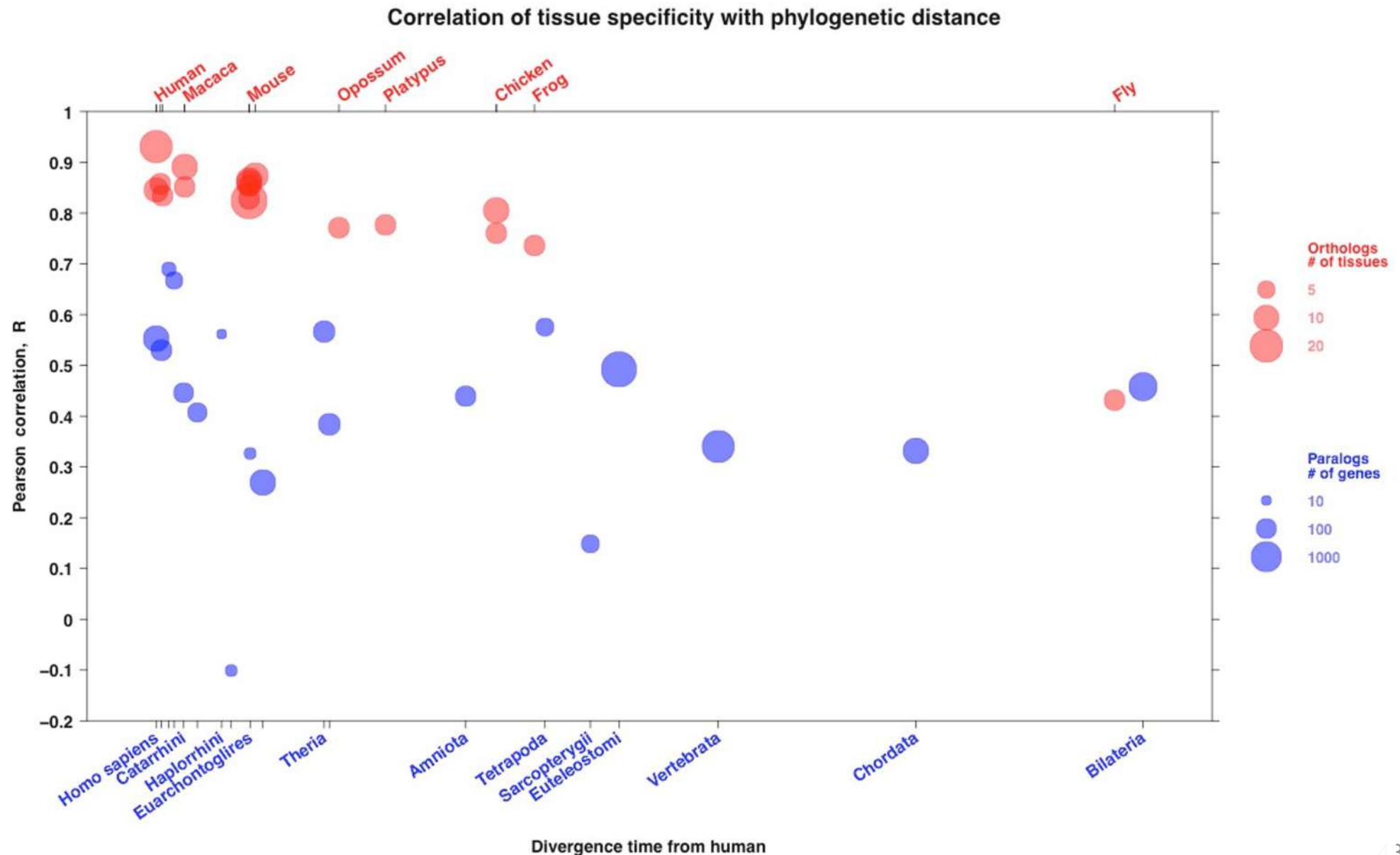
# Expression similarity of orthologs and paralogs



Nehrt et al <https://doi.org/10.1371/journal.pcbi.1002073>



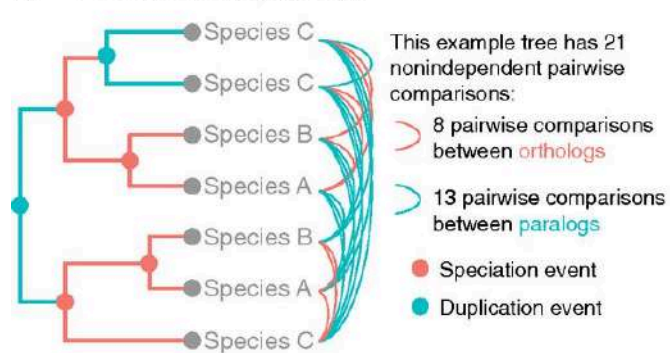
# Pairwise correlation of tissue-specificity of expression of orthologs and paralogs



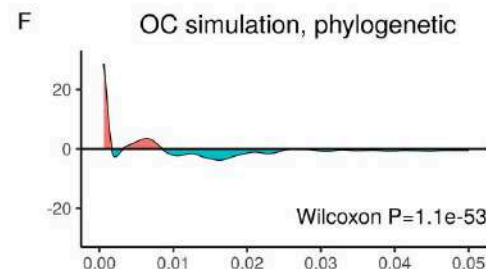
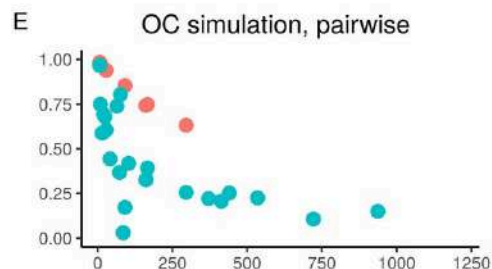
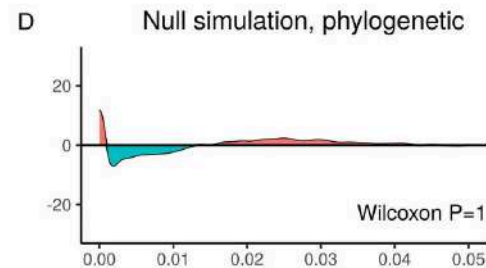
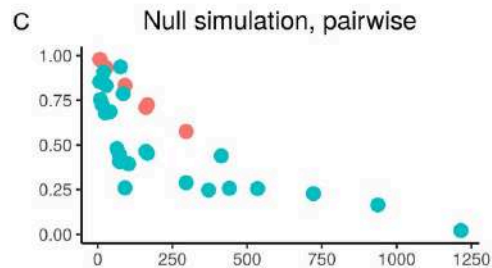
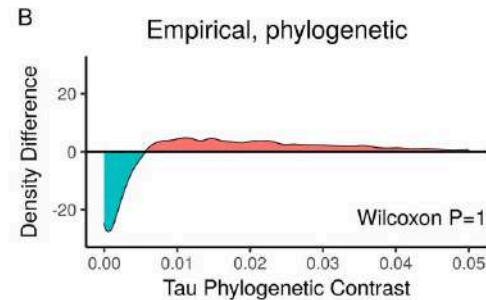
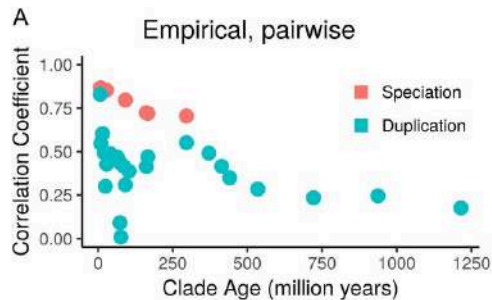
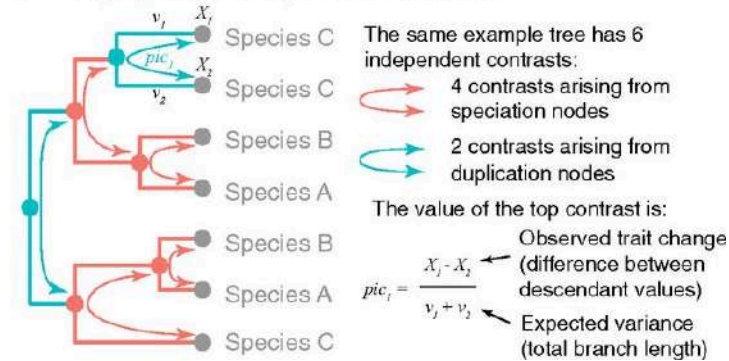


# Phylogenetic contrast of tissue-specificity of expression of orthologs and paralogs (1)

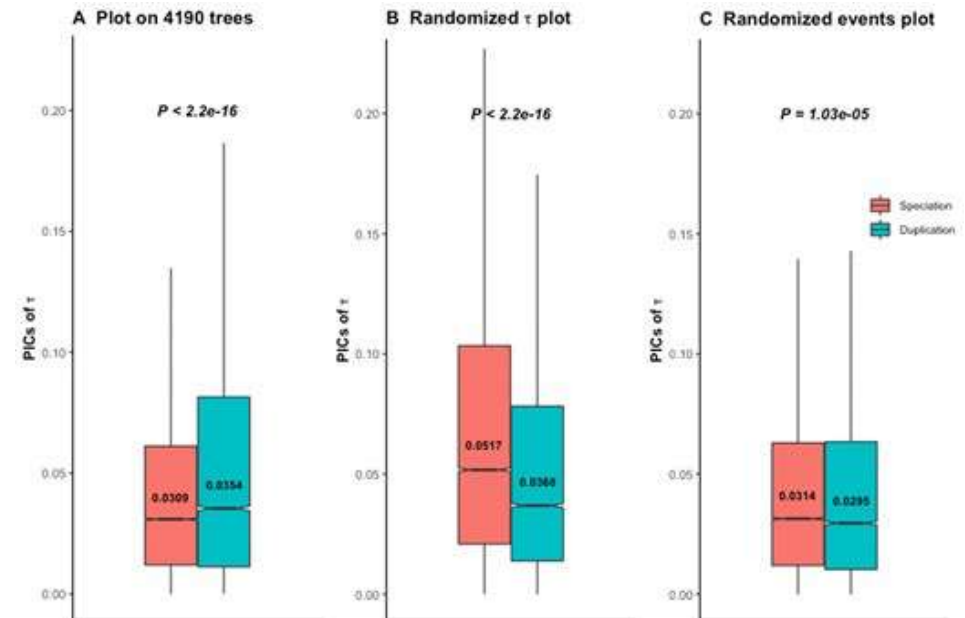
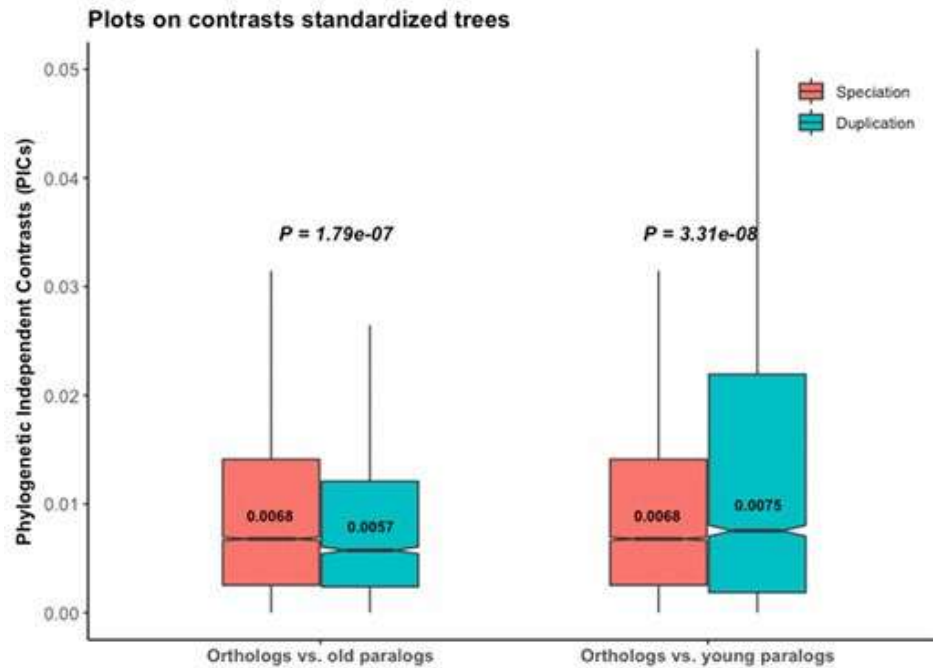
## A Pairwise comparisons



## B Phylogenetic independent contrasts



# Phylogenetic contrast of tissue-specificity of expression of orthologs and paralogs (2)

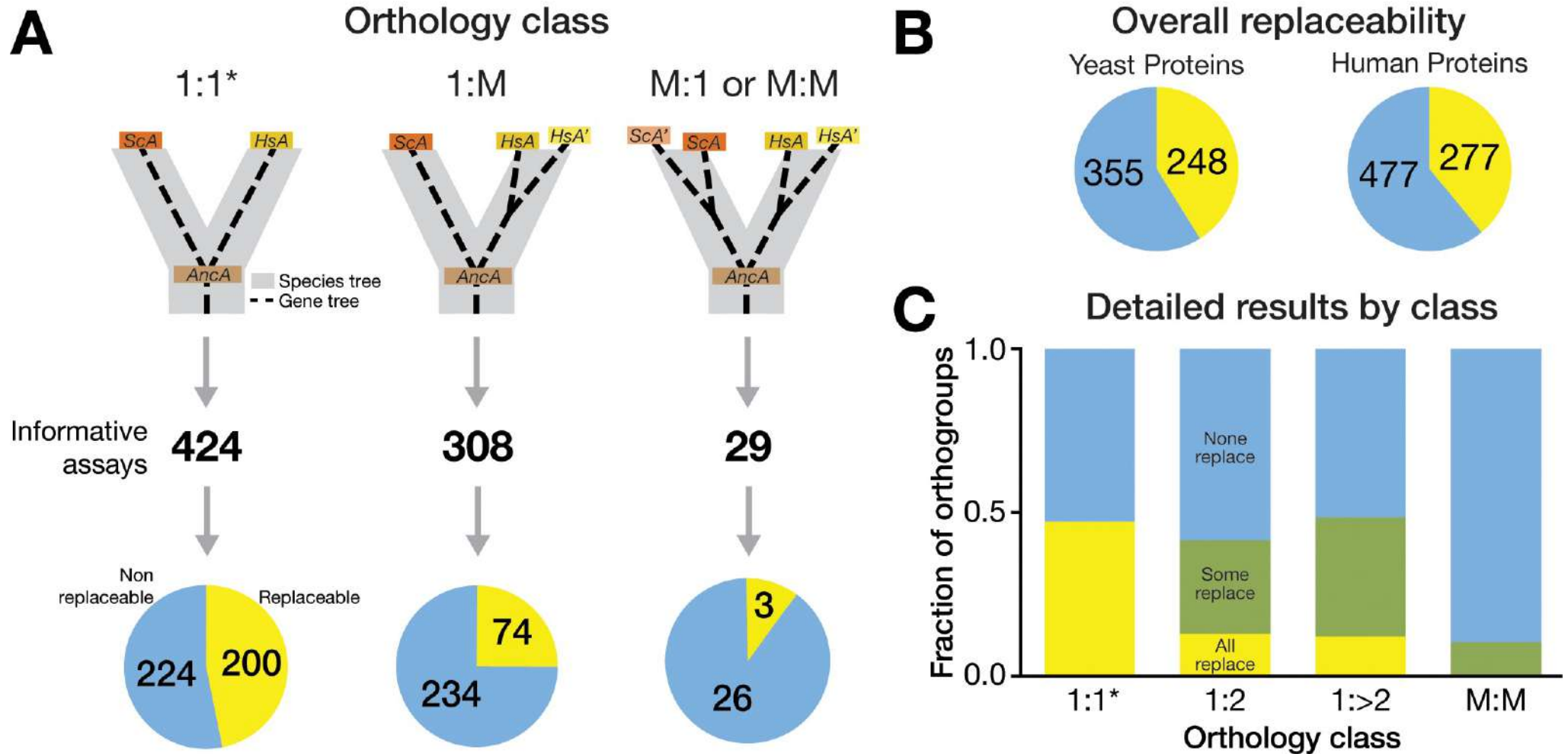


restricted to trees passing diagnostic tests for phylogenetic independence

after branch length standardization

randomized trees have similar results to the first phylogenetic tests (and to old paralogs in trees passing diagnostic tests)

# Humanization of yeast genes: functional divergence between paralogs



# Phylogenomics of function is hard

---

<b>Evidence gene A</b>	<b>Evidence gene A'</b>	<b>Apparent conclusion</b>	<b>Relevance</b>
Experiment X: function x	Homology transfer: function x	Conserved function	No: circular reasoning
Experiment X: function x	Experiment Y: function y	Different function	No: experiments cannot be compared
Experiment X: function x	Experiment X: function x	Conserved function	Yes: evolutionary conservation
Experiment X: function x	Experiment X: function x'	Different function	Yes: evolutionary change

# Poll

---

Do you believe the ortholog conjecture?

---

---

# Gene list enrichment

---

# Discussion

---

One scenario where you need the function of a list of genes

---

# General principle of gene list enrichment

---

Biologically relevant terms = GO terms

List of interesting genes

Associations gene – term = GO annotations

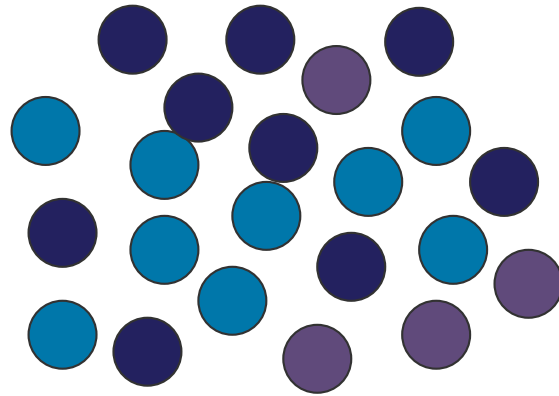
Are the genes in the list more associated to certain GO terms than expected by chance?

---



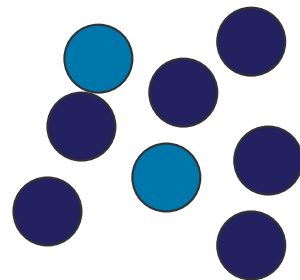
# Over-representation test

Reference  
gene list



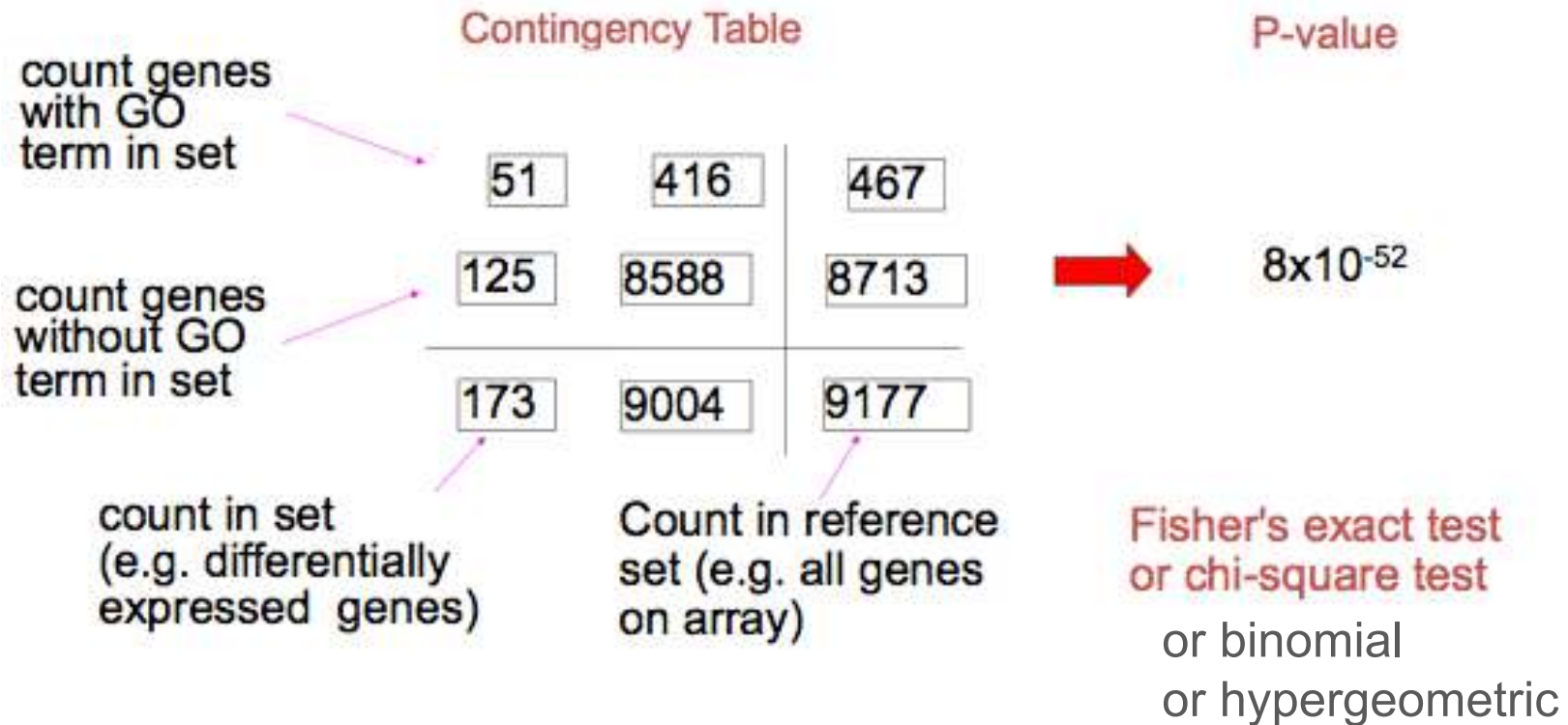
Genes annotated with a given GO term  
Genes **not** annotated with a given GO term

Your **gene list**  
of interest



Is the given annotation class over- or  
under- represented compared to a  
reference?

# Over (under) representation test example



Displaying only results with P<0.05; [click here to display all results](#)

	Homo sapiens (REF)	upload_1 (▼ Hierarchy NEW! ?)				
GO biological process complete	#	#	expected	Fold Enrichment	+/-	P value
<a href="#">ganglioside catabolic process</a>	6	9	.06	> 100	+	2.01E-13
↳ <a href="#">ceramide catabolic process</a>	13	9	.13	68.29	+	1.99E-10
↳ <a href="#">sphingolipid catabolic process</a>	21	10	.21	46.97	+	2.83E-10
↳ <a href="#">sphingolipid metabolic process</a>	17	17	1.41	12.06	+	1.16E-09
↳ <a href="#">membrane lipid metabolic process</a>	186	21	1.89	11.14	+	6.73E-12
↳ <a href="#">cellular lipid metabolic process</a>	929	33	9.42	3.50	+	3.54E-06
↳ <a href="#">lipid metabolic process</a>	8981	14	11.89	3.20	+	1.89E-06
↳ <a href="#">organic substance metabolic process</a>	9021	14	91.04	1.59	+	4.74E-10
↳ <a href="#">metabolic process</a>	8021	15	100.07	1.55	+	6.62E-11
↳ <a href="#">primary metabolic process</a>	8	15	1.53	1.53	+	9.92E-07
↳ <a href="#">single-organism metabolic process</a>	4451	112	45.12	2.48	+	5.40E-20
↳ <a href="#">single-organism process</a>	12867	12	1.55	1.55	+	2.30E-22
↳ <a href="#">cellular metabolic process</a>	8449	13	1.53	1.53	+	7.10E-09
↳ <a href="#">cellular process</a>	14439	201	146.37	1.37	+	3.04E-17
↳ <a href="#">single-organism cellular process</a>	11573	185	117.32	1.58	+	6.15E-20
↳ <a href="#">organonitrogen compound metabolic process</a>	1828	1	1.33E-18	1.33E-18		1.33E-18
↳ <a href="#">nitrogen compound metabolic process</a>	5539	96	56.15	1.71	+	2.43E-05
↳ <a href="#">membrane lipid catabolic process</a>	24	10	.24	41.10	+	1.05E-09
↳ <a href="#">cellular lipid catabolic process</a>	151	13	1.53	8.49	+	5.91E-05
↳ <a href="#">lipid catabolic process</a>	250	13	2.53	5.13	+	1.73E-02
↳ <a href="#">organic substance catabolic process</a>	1538	49	15.59	3.14	+	4.67E-09

# in the reference list

# observed

#expected

Fold enrichment

Overrepresent (+) or under-representation (-)

P value

---

# TopAnat

---

**Bgee**

# GO-like enrichment of expression in anatomy

---

For each **anatomical structure**:

	Gene list	Other genes
Expressed	$n_1$	$n_3$
Not expressed	$n_2$	$n_4$

Fisher / Hypergeometric test

Deconvolution of ontology graph (topGO)

---

## Gene list

113 genes entered, 98 in zebrafish, 15 not found in Bgee



ENSDARG00000021163  
ENSDARG00000006110  
ENSDARG000000014986  
ENSDARG00000007329  
ENSDARG00000031894  
ENSDARG00000017803  
ENSDARG00000027199  
ENSDARG00000024894

## Background

[Bgee data for zebrafish](#) Custom data

## Analysis options

### Development stages

- embryo stage
- post-embryonic stage

### Expression types

Present

### With data types:

- RNA-Seq
- Affymetrix data
- In situ hybridization
- EST

## Advanced options

### Data quality:

All High confidence

### Decorrelation type:

No decorrelation Elim **Weight** Parent-child

Node size: 20

Nb of nodes: 20

FDR threshold: 0.2

p-value threshold: 1

Start a new job

Email

ral fin phenotype in ZFIN

TopAnat request successful. Found 119 records, from 2 analyses with results, over 2 analyses launched.

## zebrafish genes with pectoral fin phenotype in ZFIN

View:

All (119) [Download Job Archive](#)

embryo stage, expression type "Present" (82) [Download Job Archive](#)

post-embryonic stage, expression type "Present" (37) [Download Job Archive](#)

Search:



Show:

20 rows

[Download Selected 119 Results](#)

Anat Entity Id	Anat Entity Name	Annotated	Significant	Expected	Fold Enric...	P Value	Fdr
<a href="#">UBERON:0000151</a>	pectoral fin	1124	53	7.43	7.13	1.91e-32	1.41e-29
<a href="#">UBERON:0004357</a>	paired limb/fin bud	337	34	2.23	15.25	1.36e-31	5.01e-29
<a href="#">UBERON:0003051</a>	ear vesicle	816	35	5.39	6.49	5.6e-20	1.37e-17
<a href="#">UBERON:0002539</a>	pharyngeal arch	1140	45	7.53	5.98	1.43e-16	2.62e-14
<a href="#">UBERON:2000040</a>	median fin fold	128	15	0.85	17.65	4.58e-15	6.72e-13
<a href="#">UBERON:0004185</a>	endodermal part of digestive tract	976	32	6.45	4.96	7.95e-15	9.73e-13
<a href="#">UBERON:0003933</a>	cranial cartilage	95	12	0.53	22.64	1.46e-13	8.22e-11
<a href="#">UBERON:0010312</a>	immature eye	1366	35	9.03	3.88	4.63e-13	4.86e-11
<a href="#">UBERON:0000151</a>	pectoral fin	462	20	2.56	7.81	6.43e-13	1.81e-10

# Pitfalls of enrichment analyses

---

- Background is critical
- Multiple testing
- Non independence of terms

---

**Function provides  
insight into genome &  
species evolution**

---



