



# Phylogenomics and tree building with ferlins

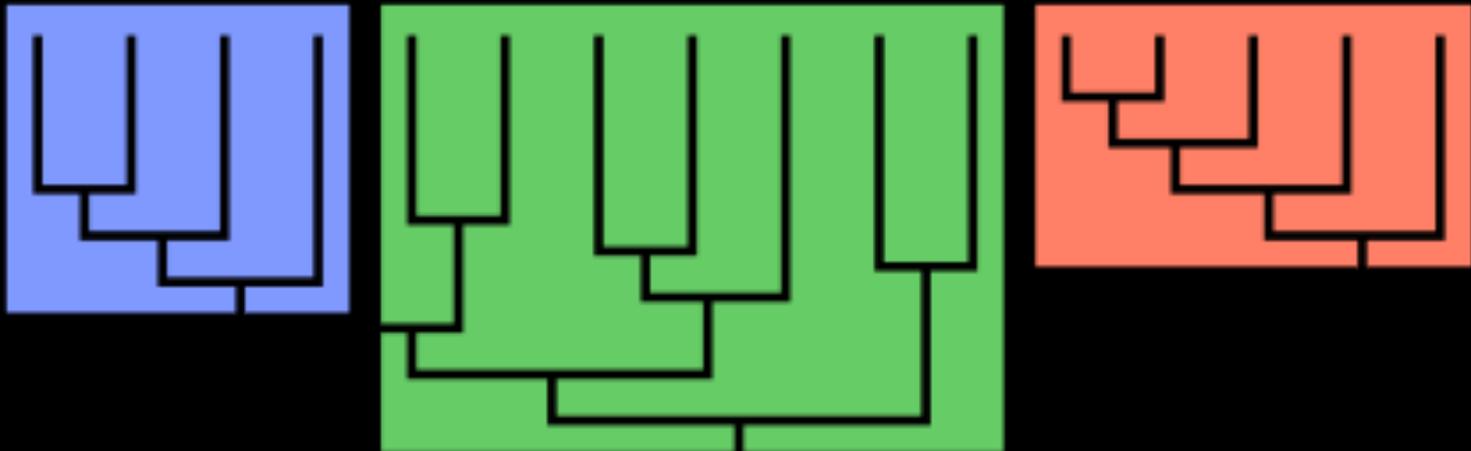
Bailey Marquardt and Troy Meikle

# Phylogenomics

Four methods for creating a phylogenetic tree

About ferlin

Paper discussion



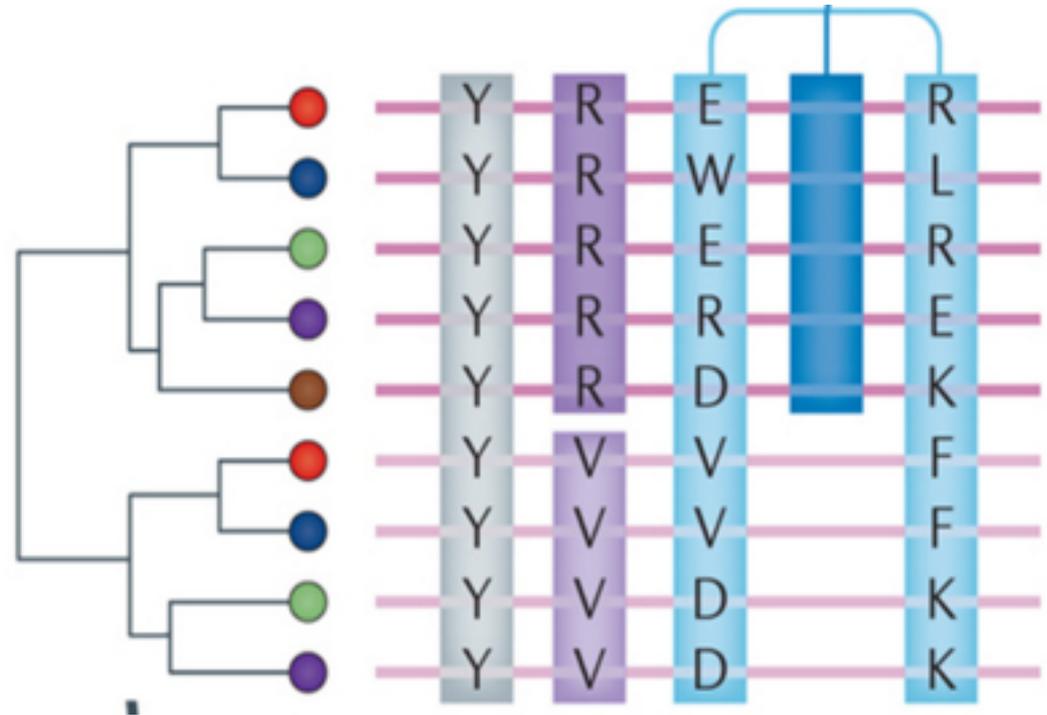
# What is Phylogenomics?

Bridges evolution fields and genomics, a term used to refer to the **use of genome data** in evolutionary reconstructions.



# Functional phylogenomics:

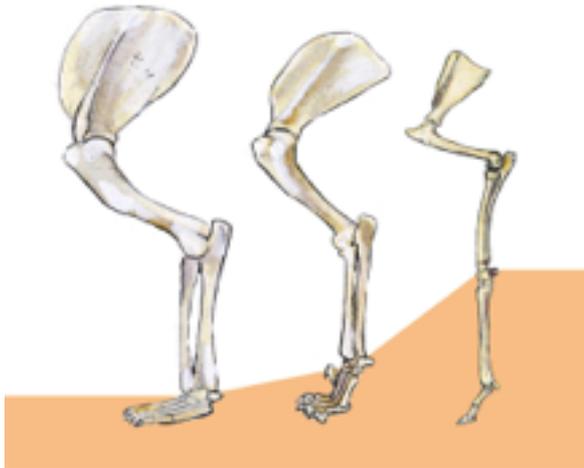
Uses common ancestry to infer protein function.



**Why is phylogenomics important?**

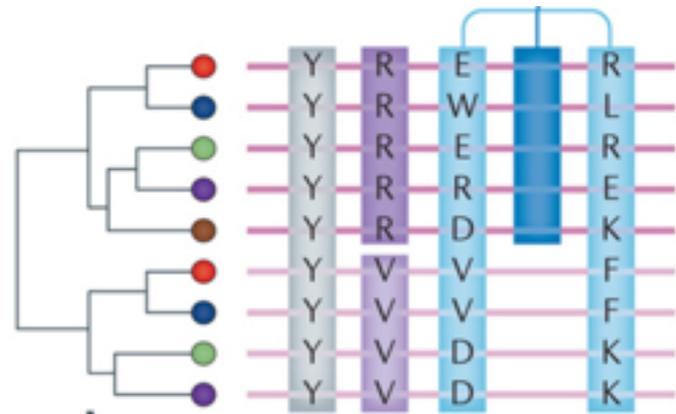
# Why is phylogenomics important?

## Evolution of complex features



<http://www.johnmurlaws.com/wp-content/uploads/2014/02/forelimb-comparison.jpg>

## Protein function



[http://www.nature.com/nrg/journal/v14/n4/fig\\_tab/nrg3414\\_F1.html](http://www.nature.com/nrg/journal/v14/n4/fig_tab/nrg3414_F1.html)

## Predictions about fossils



<http://www.weareteachers.com/images/default-source/blog-images/dinosaur--am-p--paleontology-in-the-classroom.jpg?sfvrsn=0>

# How are trees made?

## 1. Homologous proteins



<https://www.google.com/search?q=blast&source>



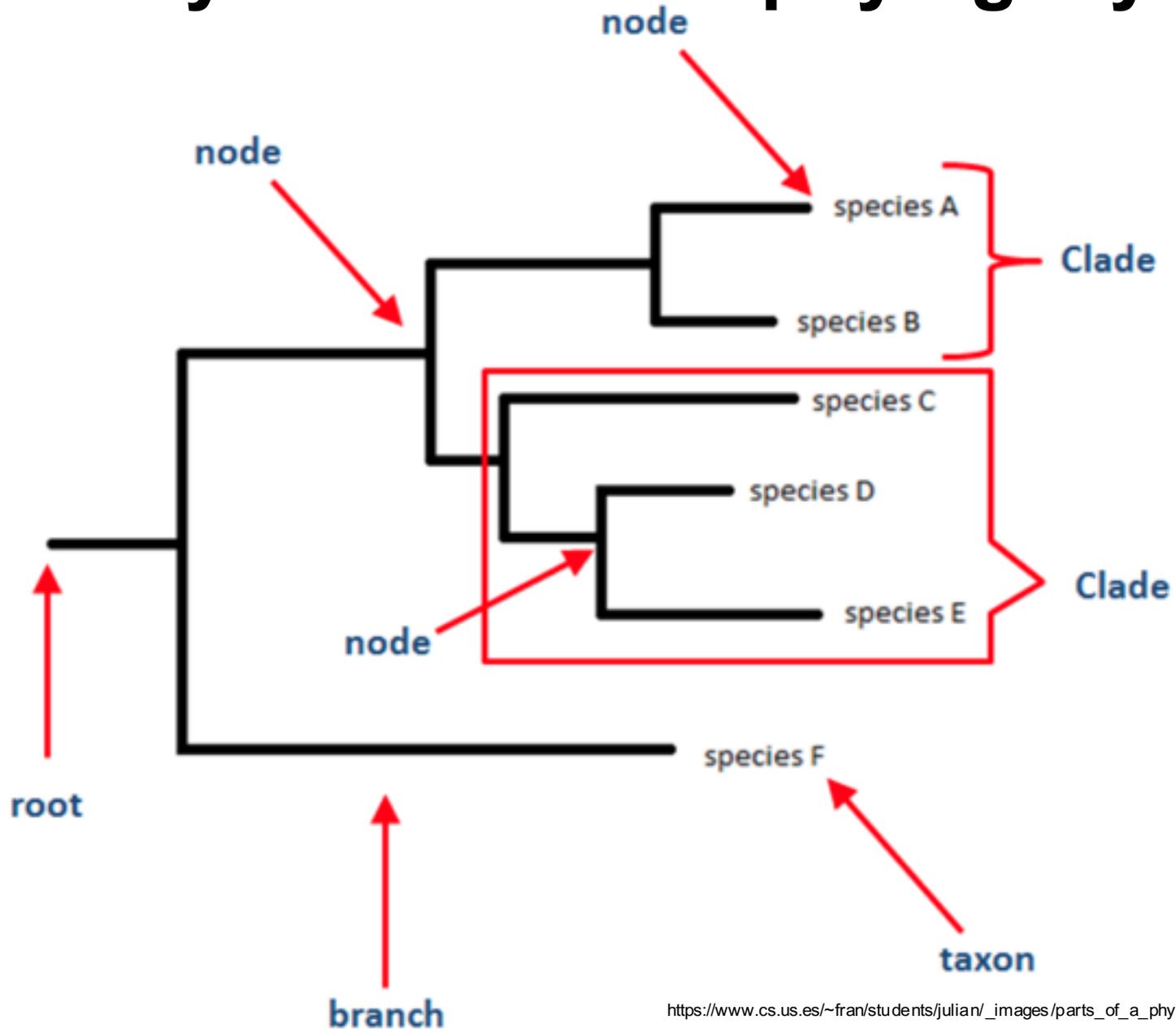
[http://i1.trekearth.com/photos/89575/tree\\_complex\\_1.jpg](http://i1.trekearth.com/photos/89575/tree_complex_1.jpg)

## 2. Sequence alignment

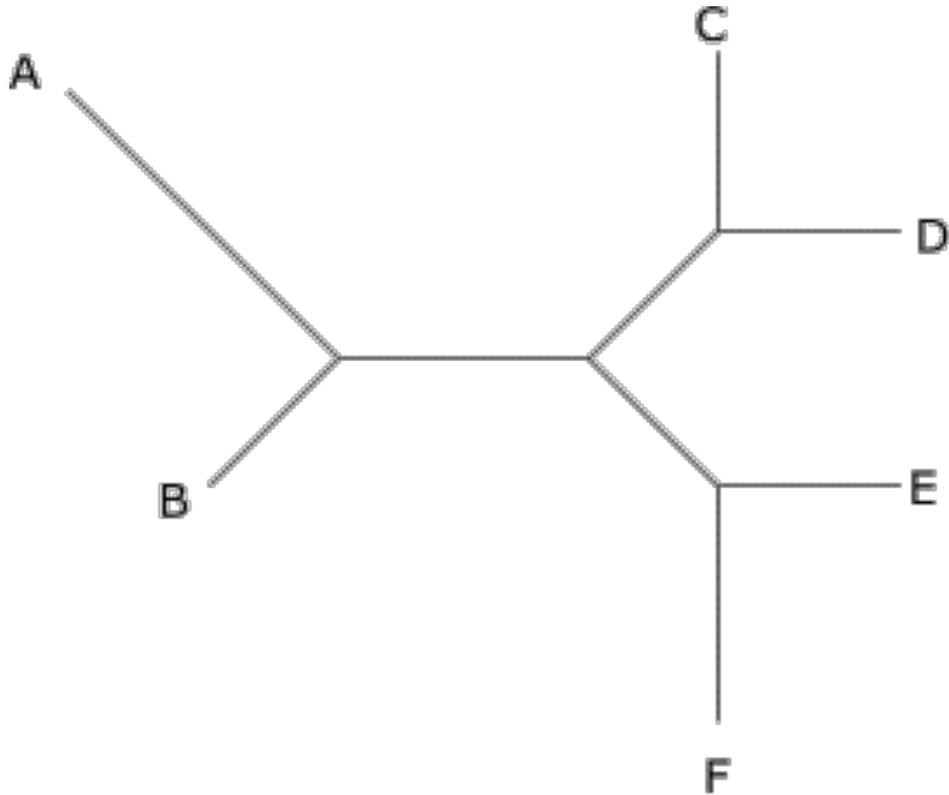
Xenopus	VATVESLGSIVSRTARLSVAGLP-RFTSQPESSSVYVGDIVLNCDVSP-ELVPPVHWEQ
Bombyx	EAT-NSYGTARSRNATLHVAVLREEFRLEPESKQSAQGDTVVLECLPPRGSPPEPNVYWK
P.	VAR-NYLGEAVSRNASLEVALLRDDFRQNPTDVVVAAGEPAILECQPPRGHPEPTIYWK
Danio	VAR-NYLGEAVSHNASLEVAILRDDFRQNPVDMVAAGEPAVMECQPPRGHPEPTISWKK
M.	VAR-NYLGEAVSHNASLEVAILRDDFRQNPSDVMVAVGEPVMECQPPRGHPEPTISWKK
Canis	VAR-NYLGEAVSHNASLEVAILRDDFRQNPSDVMVAVGEPVMECQPPRGHPEPTISWKK
Homo	VAR-NYLGEAVSHNASLEVAILRDDFRQNPSDVMVAVGEPVMECQPPRGHPEPTISWKK
Macaca	VAR-NYLGEAVSHNASLEVAILRDDFRQNPSDVMVAVGEPVMECQPPRGHPEPTISWKK
Equus	VAR-NYLGEAVSHNASLEVAILRDDFRQNPSDVMVAVGEPVMECQPPRGHPEPTISWKK
Loxodonta	VAR-NYLGEAVSHNASLEVAILRDDFRQNPSDVMVAVGEPVMECQPPRGHPEPTISWKK
Aquila	VAR-NYLGEAVSHNASLEVAILRDDFRQNPSDVMVAVGEPVMECQPPRGHPEPTISWKK
Alligator	VAR-NYLGEAVSHNASLEVAILRDDFRQNPSDVMVAVGEPVMECQPPRGHPEPTISWKK
Chelonia	-----MVAVGEPVMECQPPRGHPEPTISWKK



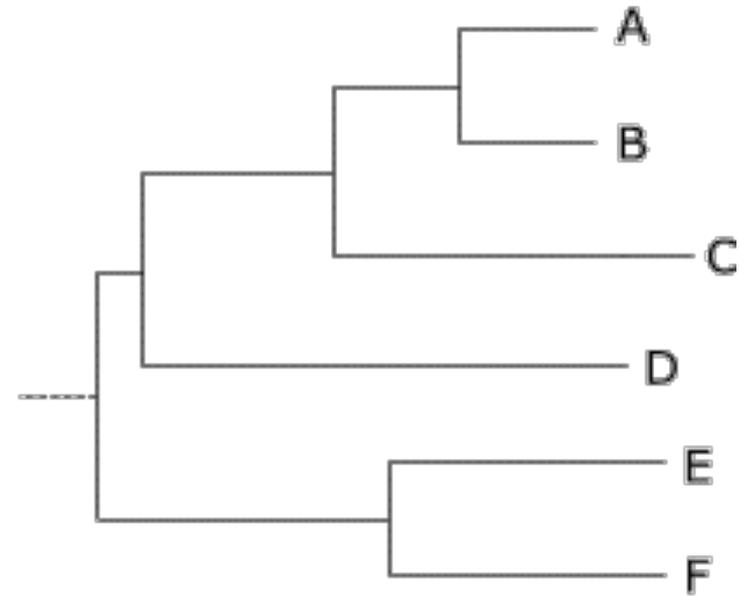
# How do you visualize a phylogeny?



# What types of trees are there?

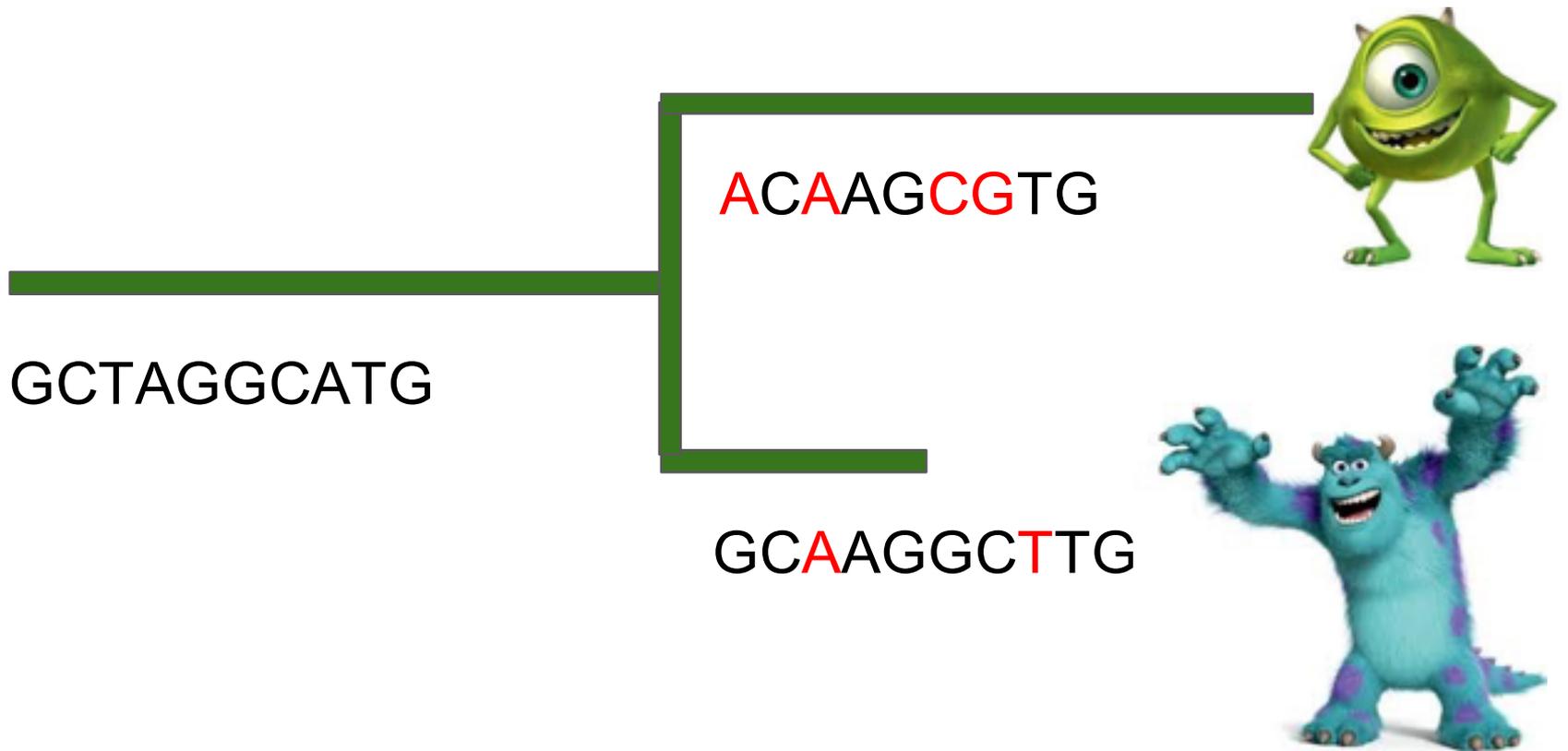


unrooted tree topology



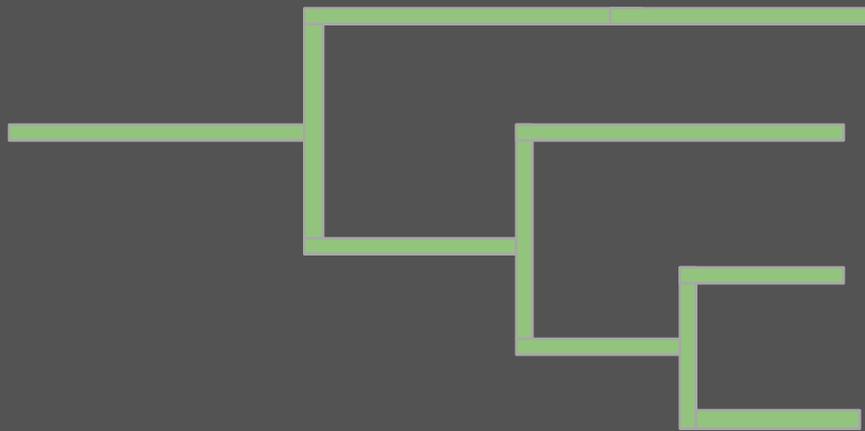
rooted tree topology

# What do branch lengths mean?



branch length= base substitutions per site

# What are the four methods of making a tree?



**Method 1: Distance matrix**

**Method 2: Maximum parsimony**

**Method 3: Maximum Likelihood**

**Method 4: Bayesian Inference**

# How does distance matrix work?

Species 1: GCTTAGAUCGAUTGG

Species 2: GCTTGGAGUTAUTGG

Species 3: TAUGAUCTCGAUTGA

Species 4: TAUGAUCAGGAUTCG

GCTTAGAUCGAUTGG

vs

GCTTAGAUCGAUTGG

vs

GCTTAGAUCGAUTGG

vs

GCTTGGAGUTAUTGG

TAUGAUCTCGAUTGA

TAUGAUCAGGAUTCG

GCTTGGAGUTAUTGG

vs

GCTTGGAGUTAUTGG

vs

TAUGAUCTCGAUTGA

vs

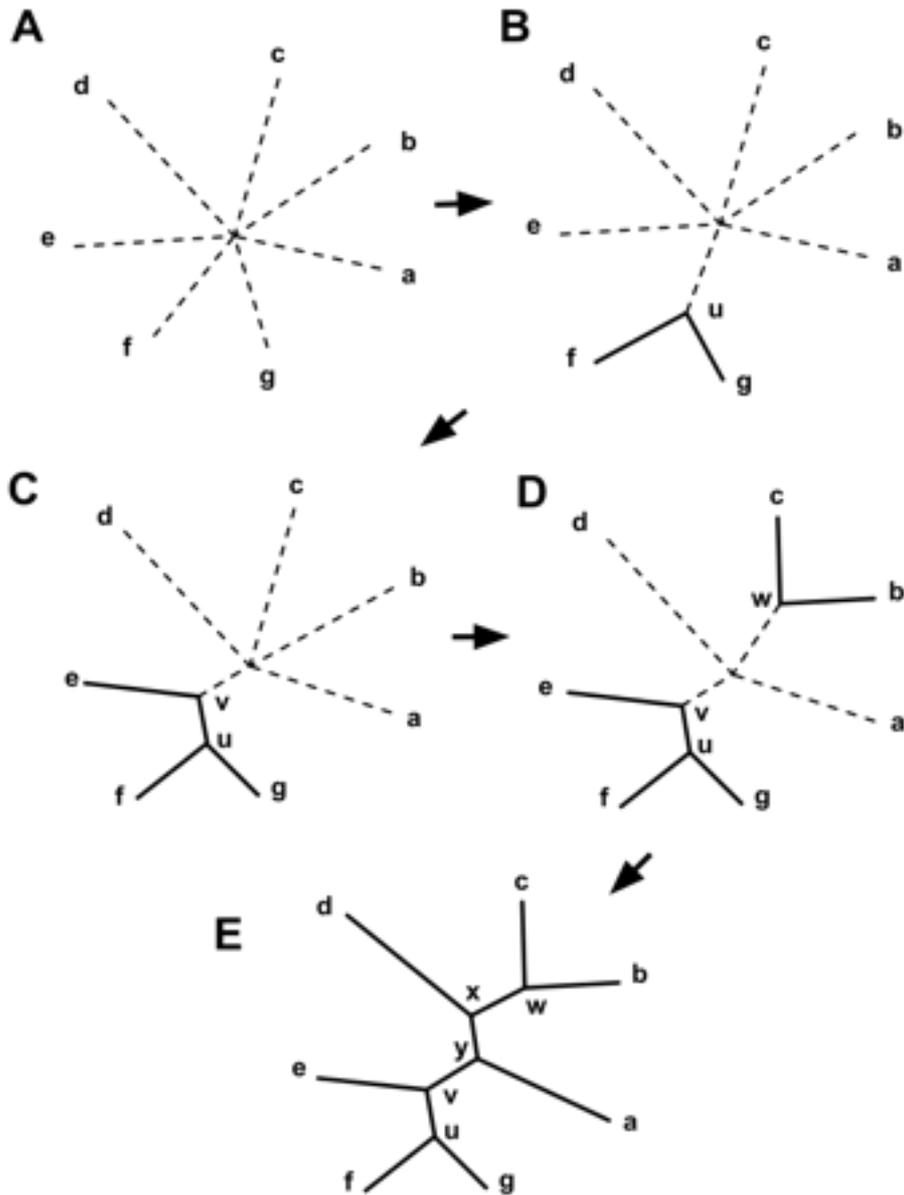
TAUGAUCTCGAUTGA

TAUGAUCAGGAUTCG

TAUGAUCAGGAUTCG

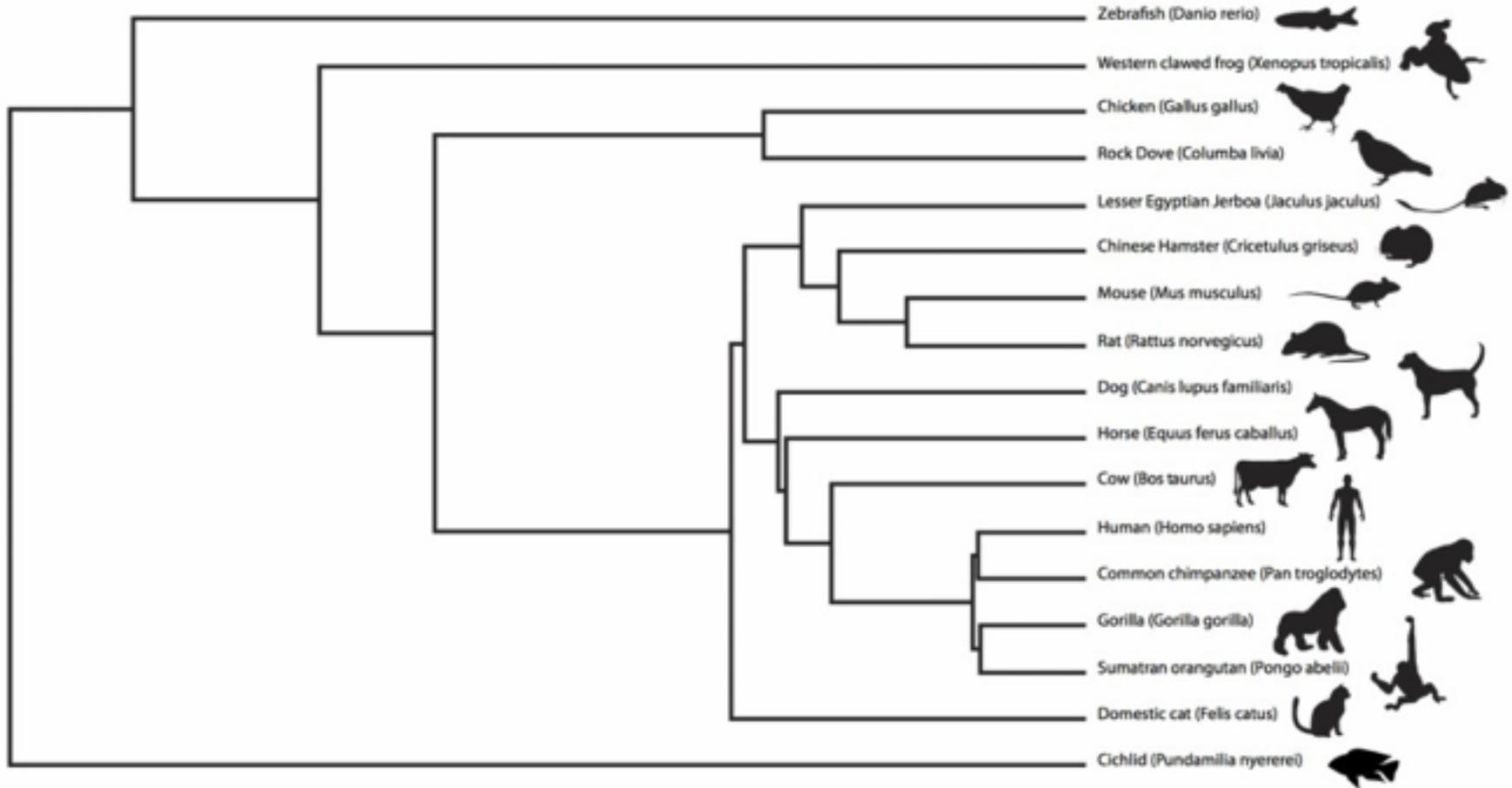
**Percent Identity:** what percentage of the two sequences are identical?

**BLOSUM:** two amino acids match, how likely is this to happen?



**What is neighbor joining?**

# What is Average distance?



**Similarity score determines which species are most closely related.**

**Assumes both species diverged from common ancestor equally.**

# Method 2: Maximum parsimony

- Prefers the minimum number of changes per nucleotide site.

G	C	T	T	A	G	A	U	C	G	A	U	T	G	G
G	C	T	T	G	G	A	G	U	T	A	U	T	G	G
T	A	U	G	A	U	C	T	C	G	A	U	T	G	A
T	A	U	G	A	U	C	A	G	G	A	U	T	C	G

**Which sites are informative?**

# Method 3: Maximum likelihood

Includes branch lengths

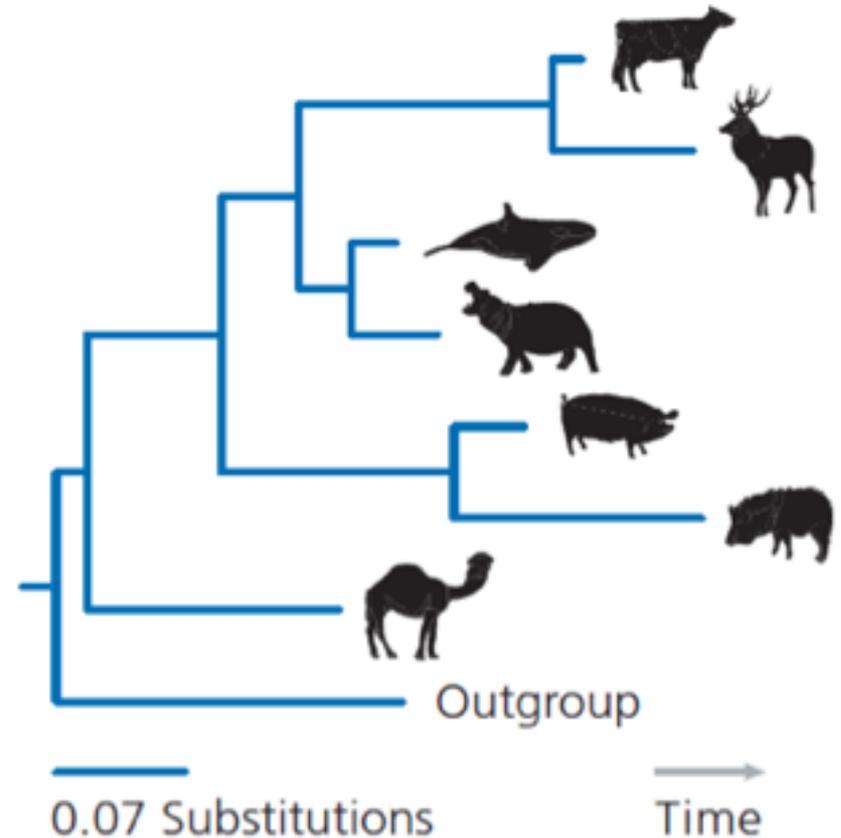
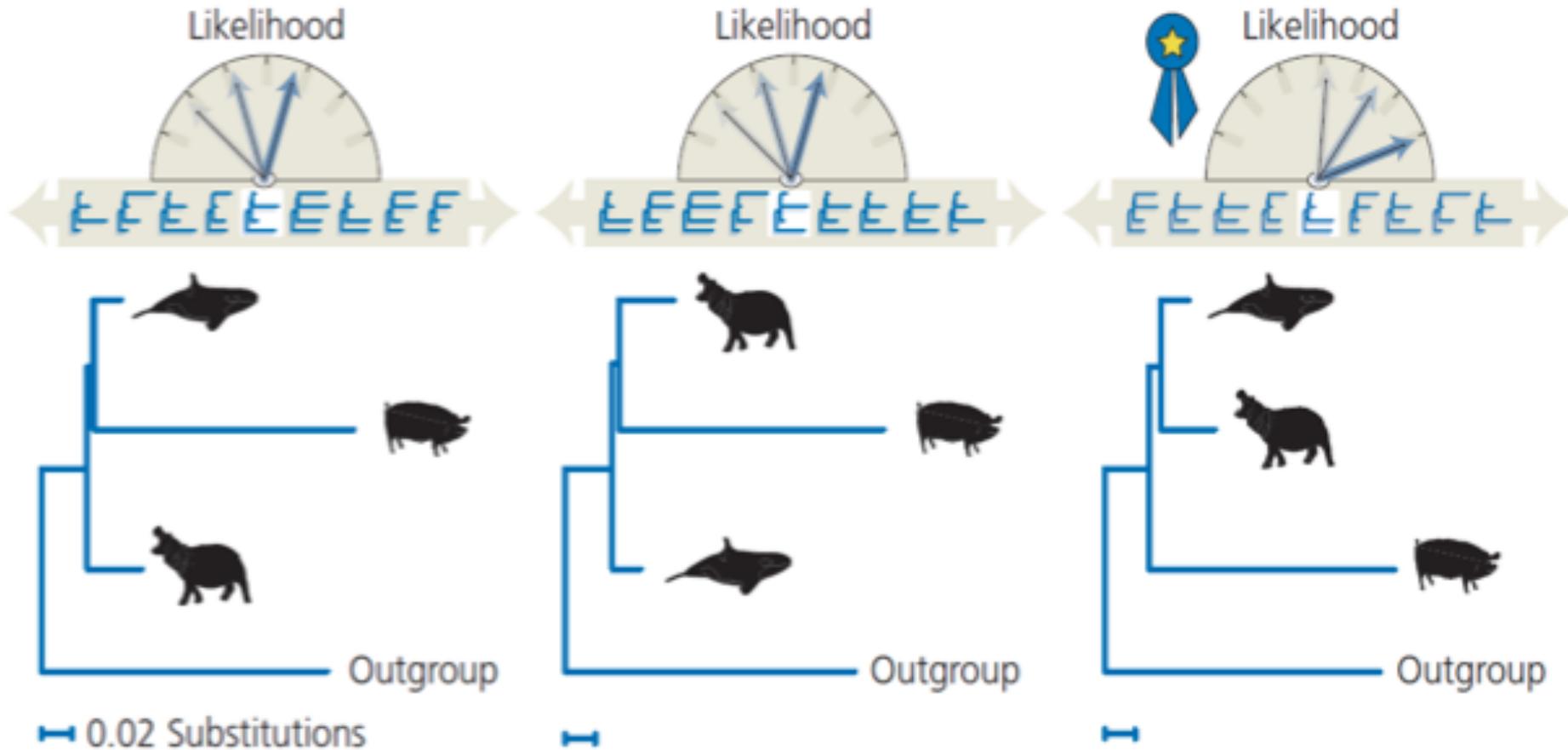


Figure 4.24. Herron, J.C. and Freeman S. (2014). *Evolutionary Analysis: Fifth Edition*. Glenview, IL: Pearson Education, Inc.

$$L(\text{tree}) = P(\text{data} \mid \text{tree, branch lengths, model})$$

# Which is the best tree?



# What is Bayesian Theory?

Bayesian Inference:  $P(\text{tree} | \text{data})$       Maximum Likelihood:  $P(\text{data} | \text{tree})$

$$P(\text{tree}/\text{data}) = \frac{P(\text{data}/\text{tree}) P(\text{tree})}{P(B)}$$

## Method 1: Distance matrix

fast

least likely to generate  
correct tree due to alignment  
gaps

## Method 2: Maximum parsimony

evolutionary conservative

slow

## Method 3: Maximum

Likelihood

branch length

Computationally demanding

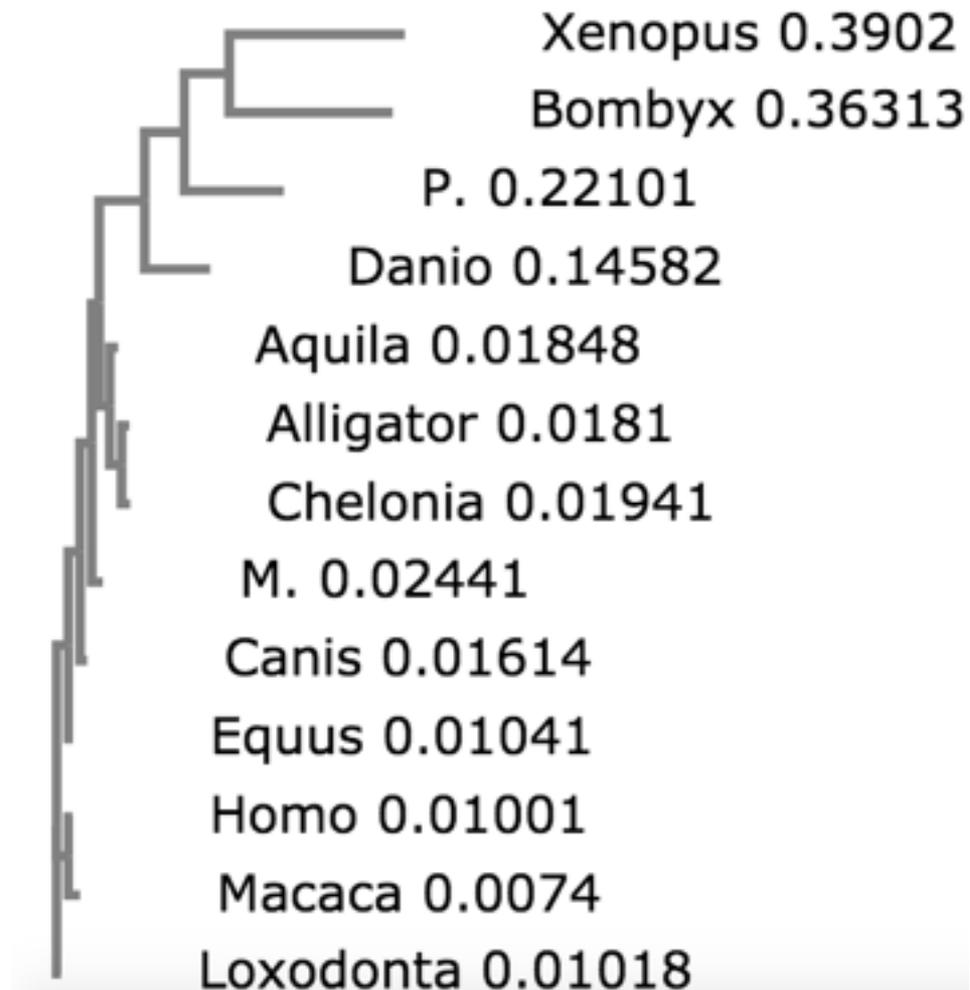
## Method 4: Bayesian Inference

easy to interpret

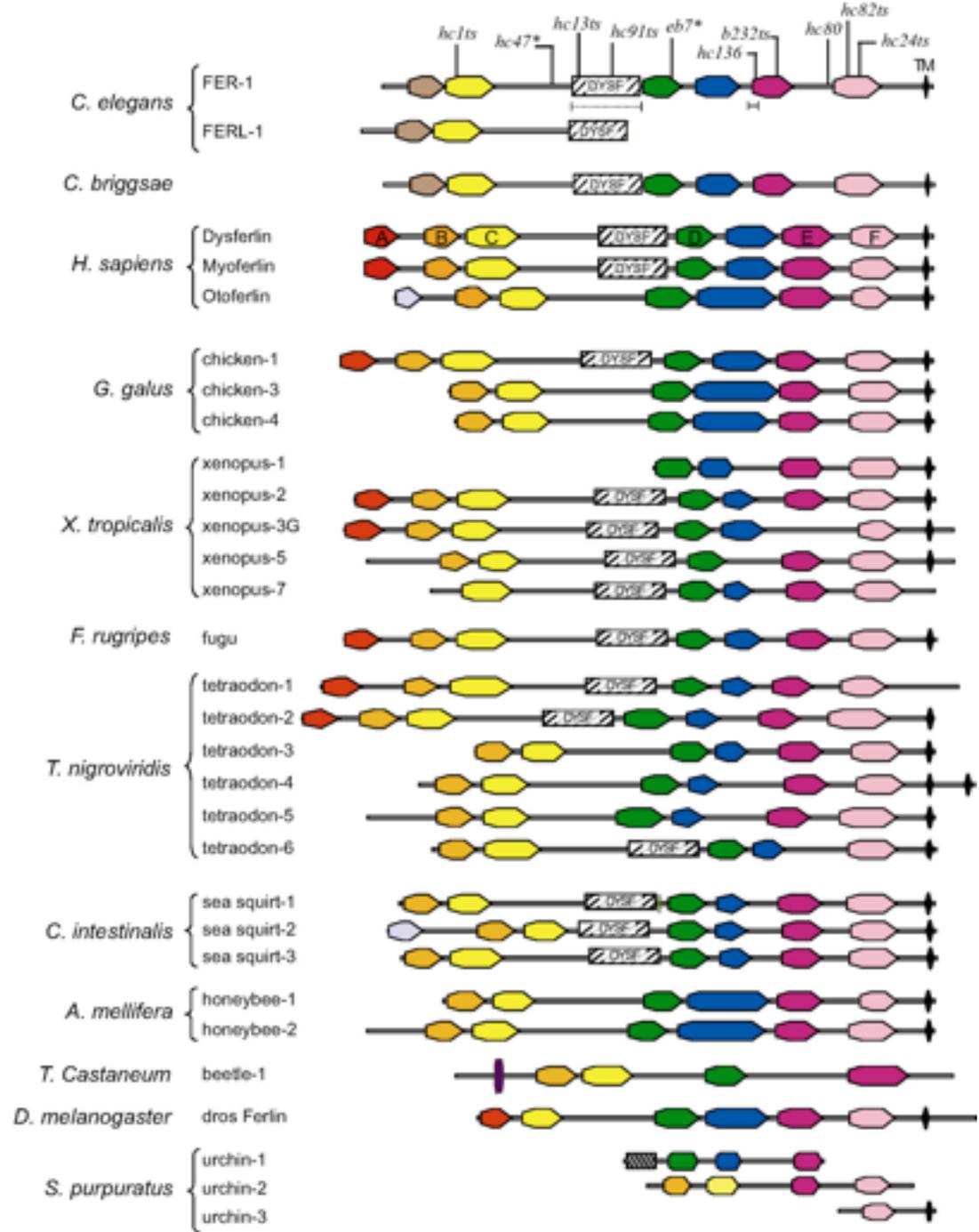
branch length

Computationally demanding

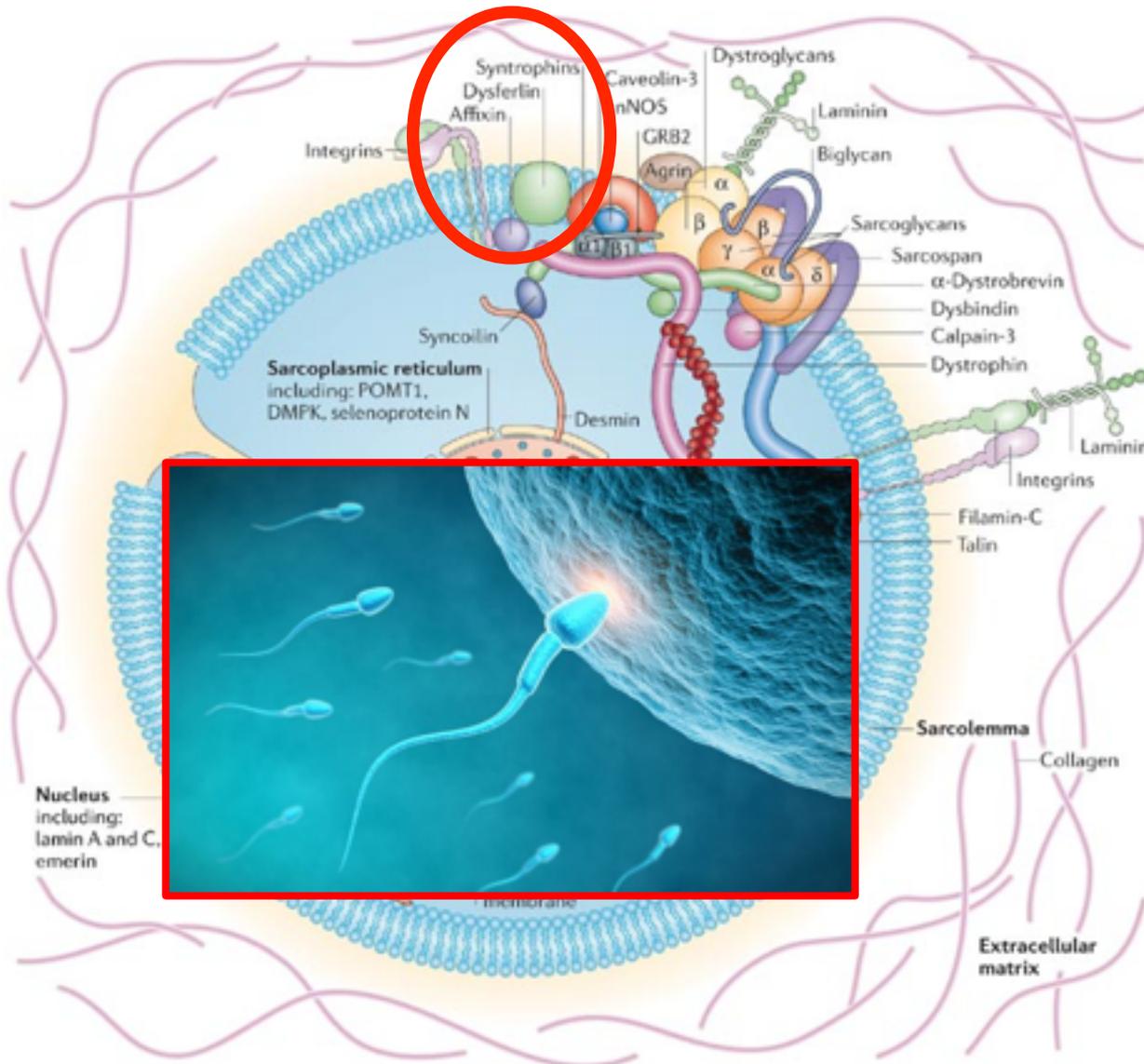
# My tree



# What are ferlins?



# Why are ferlins important?



Transmembrane protein which is key for **vesicle fusion**.

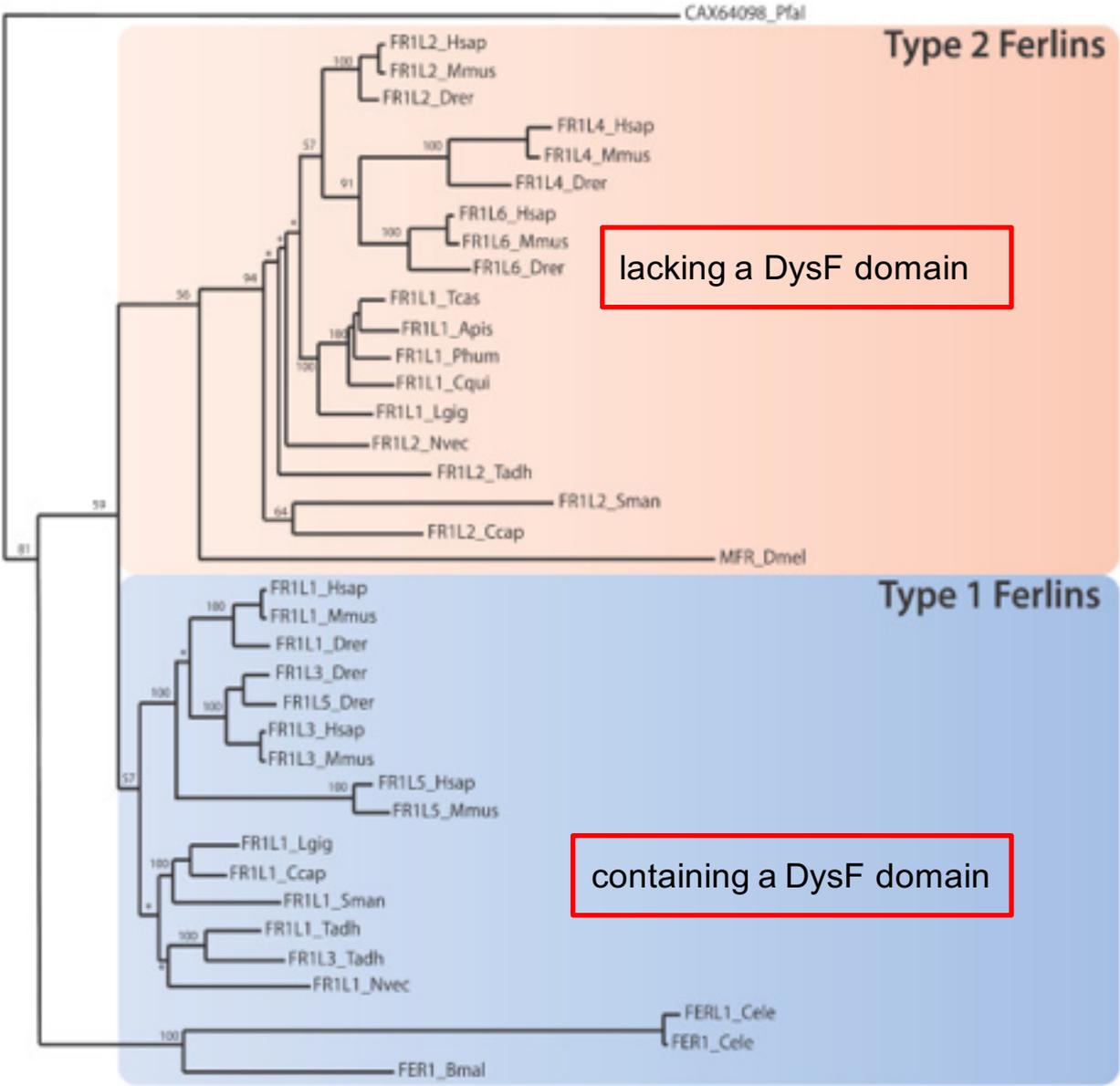
Mutant ferlin genes cause problems in reproduction.

Why?

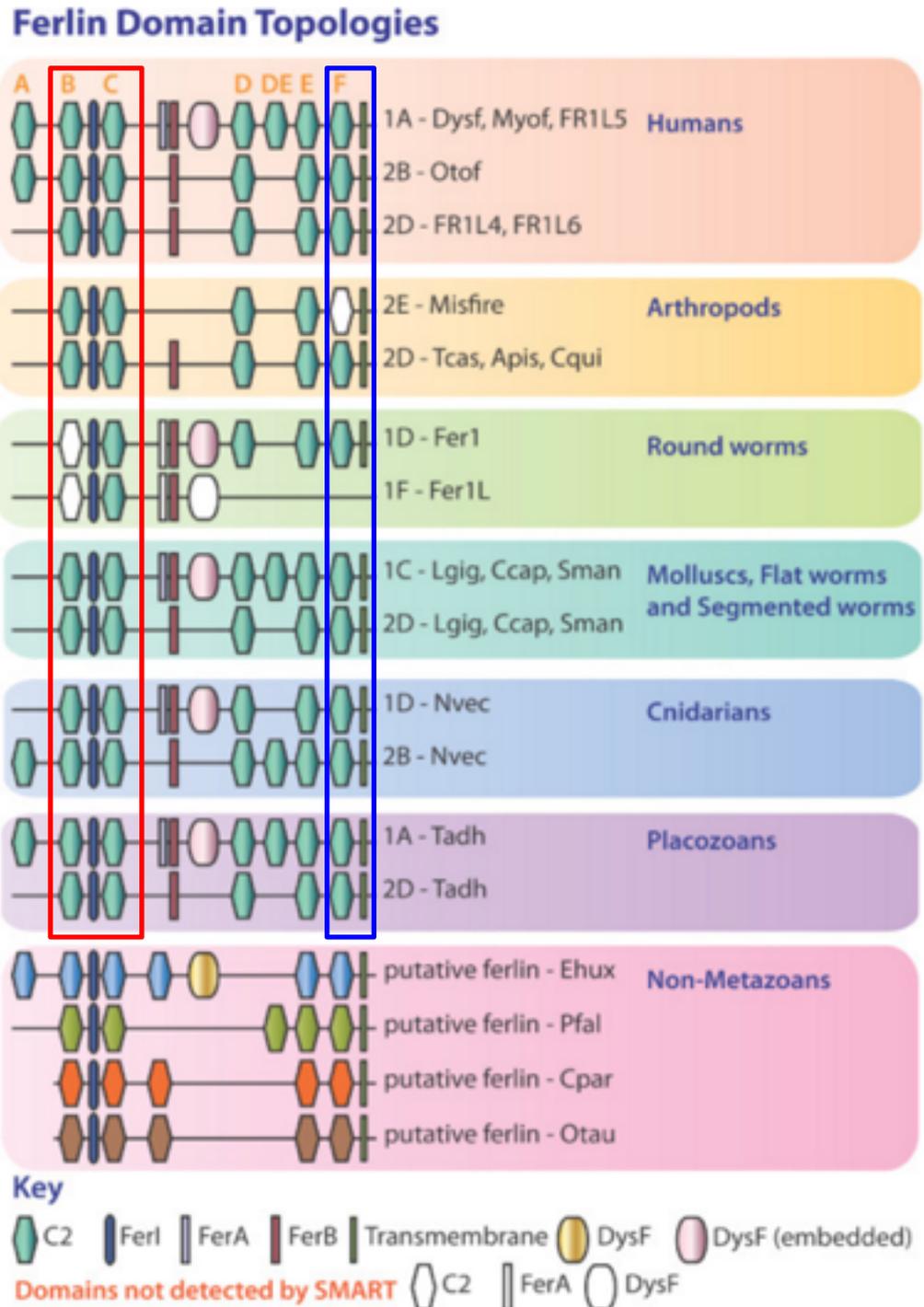
# Phylogenetic analysis of ferlin genes reveals ancient eukaryotic origins

Angela Lek<sup>1,2</sup>, Monkol Lek<sup>1,2</sup>, Kathryn N North<sup>1,2</sup>, Sandra T Cooper<sup>1,2\*</sup>

# Figure 1: What is the relationship between ferlins?



# Figure 2: What domains are conserved in Ferlin homologues?



C2 domains are structural motifs that mediate lipid and protein binding as well as coordinate calcium ions.

# Ferlin Domain Topologies

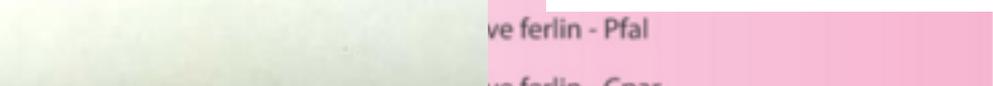
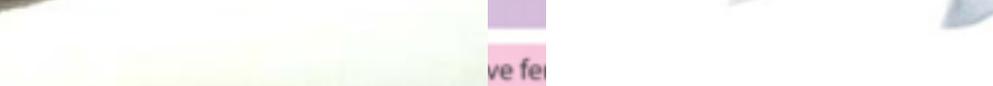
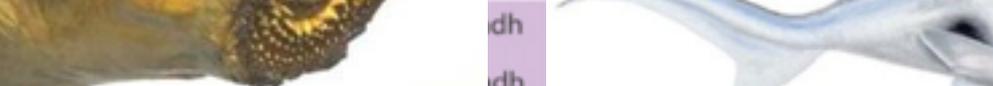
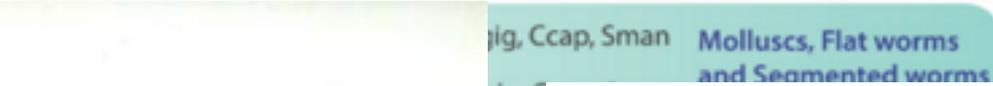
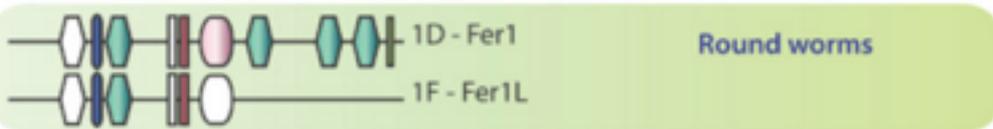
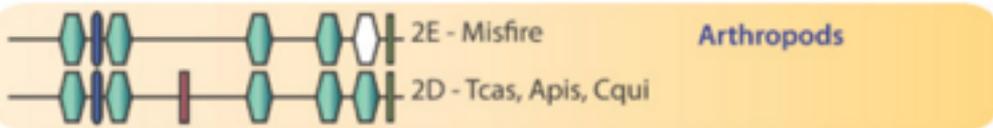
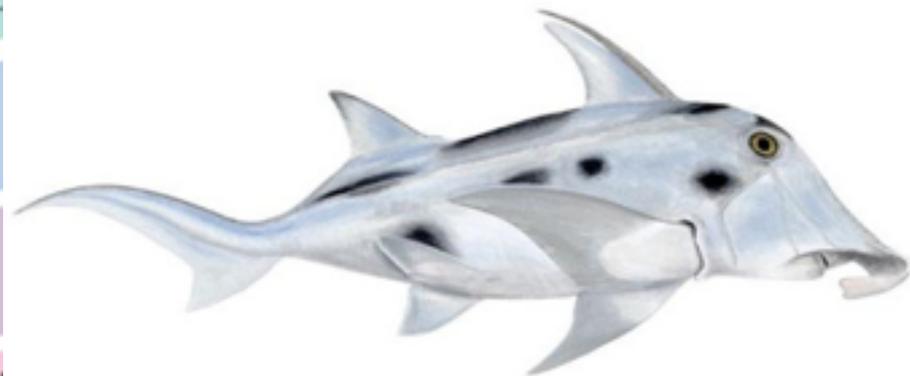


Table 2: H

s?

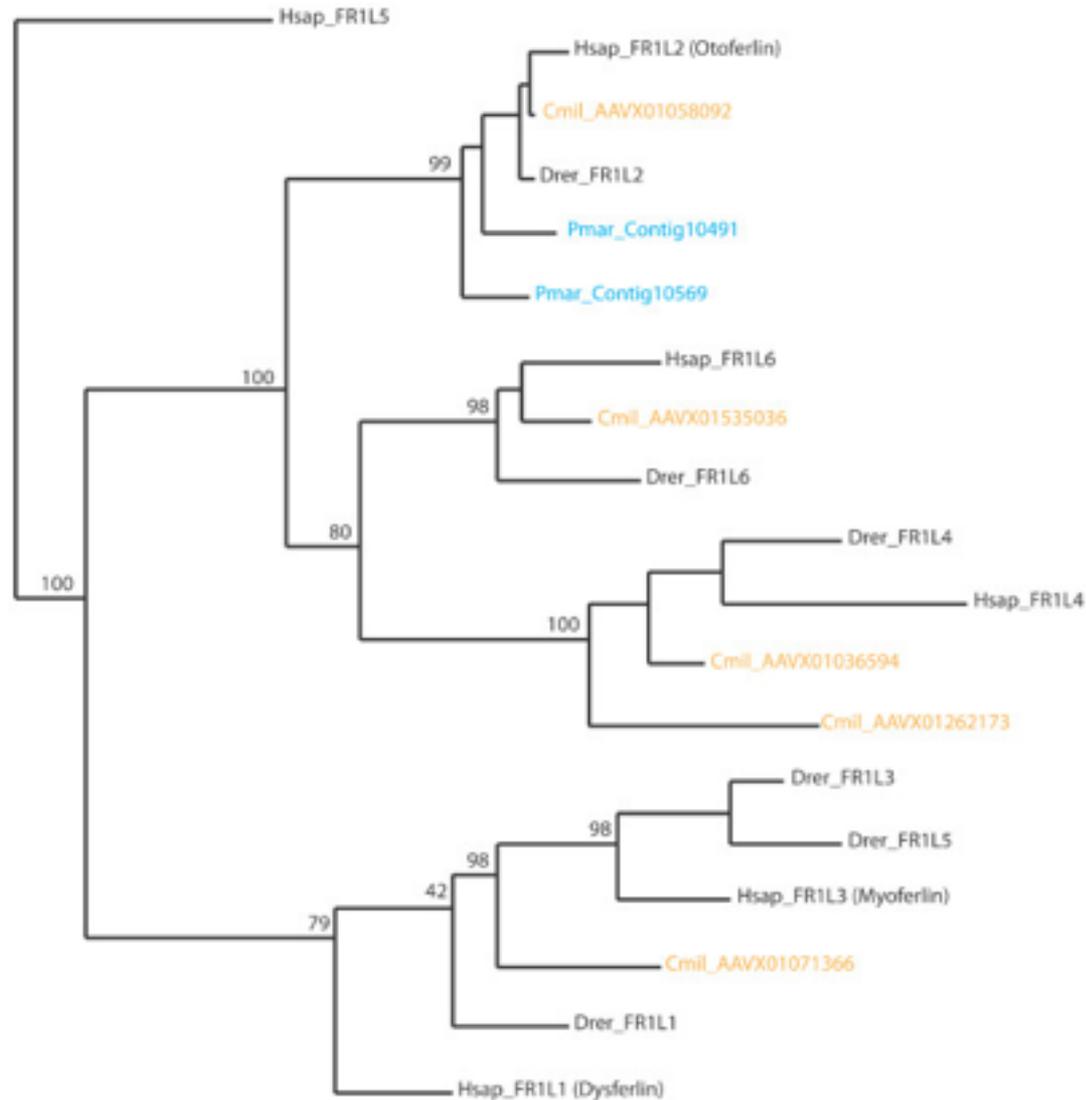
Type	Ferlin
1	FER1L1_Tadh
1	FER1L3_Tadh
1	FER1L1_Nvec
1	FER1L1_Ccap

	C2E	C2F
2	79	76.9
3	80	51.4
	78	76.2
5	75	80.8
1	76	76.9



	75	70.2
	89	90.8
	66	

# Figure 3: How did ferlins expand during evolution?



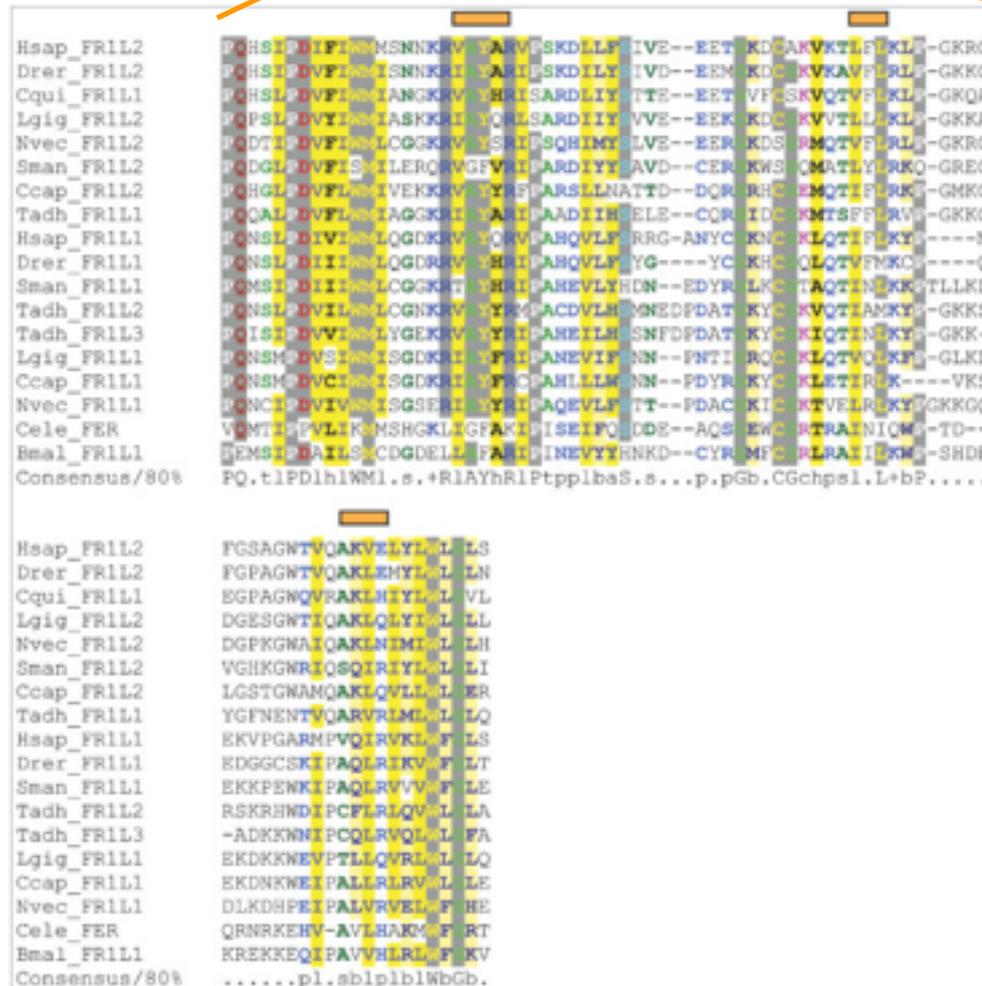
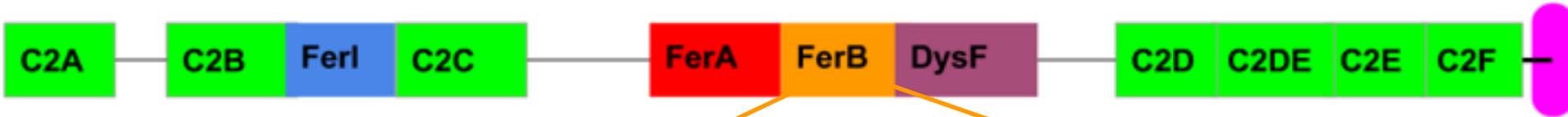
## Table 3: How are calcium-coordinating residues conserved?

	Dysferlin	Paralogs	Orthologs
C2A	No	N/A	N/A
C2B	No	N/A	N/A
C2C	Yes	No	No
C2D	Yes	No	Yes (1,2)
C2E	Yes	Yes	Yes (1)
C2F	Yes	Yes	Yes(1)

Classical calcium-sensitive C2 domains have 5 aspartates involved in binding calcium ions



# Figure 5: How is the FerB domain conserved?



# Figure 7: How is the DysF domain conserved?



outer

```

Scer_PEX30  FITYDNT-- .---WKKPSK EDSFSKY--- TARRRQVETAELV
Nvec_FR1L1  WEYIVD-- .-LSYGVYSPV QKAYHL---SARRRQVERRDLK
Tadh_FR1L1  WEYTLS--- .---SGYSS VEKNYHL---YARRRQVERRKLT
Tadh_FR1L2  WEYTLG--- .---SGYVP VEKNYHL---EARRRQVVKRRLK
Cele_Fer    WVYSTNGVF .---FGSGVL TDREKVVHNN FARRRQVVKRRLK
Sman_FR1L1  FEYIVVEA- .---STGGYVAA EKLYHM---FARRRRLISRLI LG
Lgig_FR1L1  WEYCVEA- .-TIGGYGPV EKTYHL---CARRRRLASRTLT
Ccap_FR1L1  WEYCVEV- .-TMGGYGPV EKTYHM---CARRRQVVSRLKLM
Drer_FR1L1  WEYGITLPP DRR-PKSNVP SEKMYHT---NARRRRLILRRRD
Drer_FR1L3  WEYGITIPP DEK-PKSNVA CEKMYHV---HARRRRLMIPRKKI
Drer_FR1L5  WEYGVITIPP DDK-PRSNVA AEKMYHI---HARRRRLVLSRRKT
Hsap_FR1L1  WEYSITIPP ERK-PKSNVP AEKMYYT---HARRRRLVLSRRRD
Hsap_FR1L3  WEYGITIPP DHK-PKSNVA AEKMYHT---HARRRRLVLSRRKRD
Hsap_FR1L5  SWEYGVGIPP .SGLPQVWSP VEKTYHS---CARRRRLAVRFRN
GWpY*ssh.. ....hssbsb pc+bpch... hRRRQVWRp. +bp
  
```

inner

```

Scer_PEX30  FITYDNT-- .---WKKPSK EDSFSKY--- TARRRQVETAELV
Nvec_FR1L1  WEYIVP-- .---TMKFHA KERKMDF---VARRRRLMVKL IPE
Tadh_FR1L1  VWEYAKMF-- .---ITKFHS NRAKSDL---VARRRRLHNLK AV
Tadh_FR1L1  WQYATSF-- .---SSKFHP KSSMTDF---VARRRRLIKLK AV
Cele_Fer    NWEYSPS-- .-KKEGPYHDL EDRTDR---IARRRQVVEVENK
Sman_FR1L1  AWEYVNF-- .---DSKFHTKERKVDL---VARRRRLHNLK AV
Lgig_FR1L1  WEYVPLF-- .---NMKFHAIERTMDL---VARRRRLHNLK AV
Ccap_FR1L1  WEYVPLF-- .---NMKFHAKERKMDL---VARRRRLHNLK AV
Drer_FR1L1  WEYVSLF-- .---GWKFHL KQKTDL---FARRRRLHNLK AV
Drer_FR1L1  WEYVSLI-- .---GWKFHR QQRSSDT---FARRRRLHNLK AV
Drer_FR1L1  WEYVSLI-- .---GWKFHR KERSSDT---FARRRRLHNLK AV
Hsap_FR1L1  WEYVSLF-- .---GWKFHL EYRKTDA---FARRRRLHNLK AV
Hsap_FR1L3  WEYVSLI-- .---GWKFHW QQRSSDT---FARRRRLHNLK AV
Hsap_FR1L5  WEY-DTF-- .---GSKFHL NPQPQSR---FARRRRLHNLK AV
GWpY*ssh.. ....ssabp ccbpch...h RRRcWbRph+ bp
  
```

# Conclusions



By finding homologs and sequencing alignment data, specific conserved domains were identified among ferlin genes across eukaryotic evolution:

multiple tandem **C2** domains

a single C-terminal **transmembrane domain**

a **FerB** domain

a highly conserved N-terminal motif consisting of **Fer1** tightly sandwiched between two **C2** domains

**Future studies?**

# Questions?

