Extension in model organisms
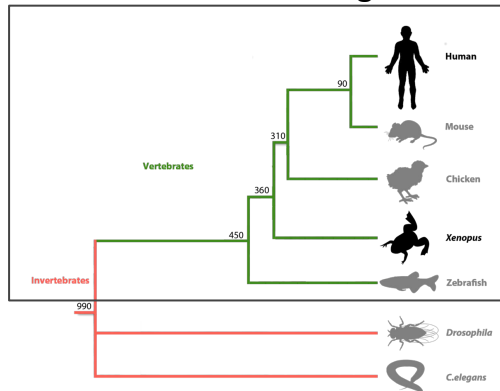
N terminal

DLG4_fly	S	T	E	D	I	T	R	E	F
DLG4_zebrafish	L	D	D	D	I	S	R	E	F
DLG4_dog	G	E	E	D	I	P	R	E	F
DLG4_human	G	E	E	D	I	P	R	E	F
DLG4_mouse	G	E	E	D	I	P	R	E	F

C terminal

DLG4_fly	Y	R	P	E	E	Y	N	R	F	E	A
DLG4_zebrafish	Y	R	P	E	E	Y	S	R	F	E	A
DLG4_dog	Y	K	P	E	E	Y	S	R	F	E	A
DLG4_human	Y	K	P	E	E	Y	S	R	F	E	A
DLG4_mouse	Y	K	P	E	E	Y	S	R	F	E	A

Taxonomical range

- Going from human to zebrafish as opposed to human to fly
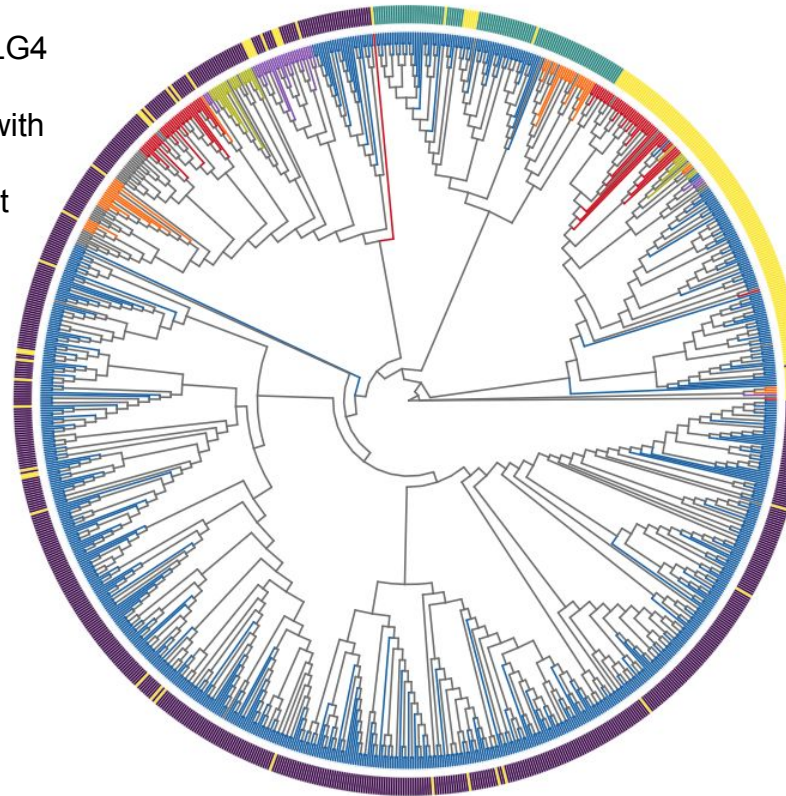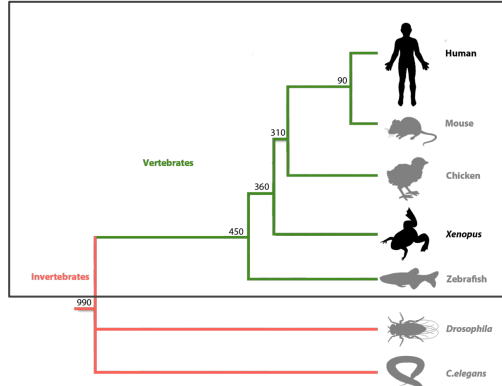 - Fly gene annotation is of poor quality

*: In the insects/worm clade, only one DLG4 homolog was found in oriental fruit fly (issues with accessing this genome)

Full result from protein blast

Protein Blast against human DLG4
- 1000 sequences
- Removed 2 sequences with insertion in N terminal extension for subsequent analysis

Taxonomical range

Clade
- DLG1
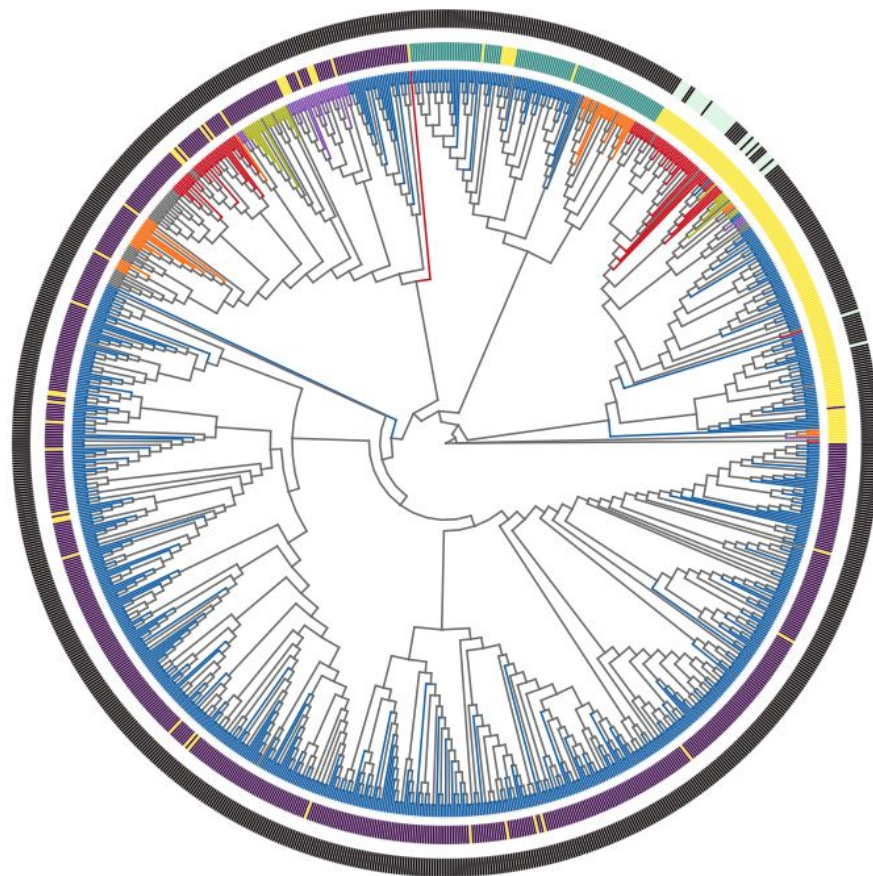- DLG2
- DLG4

Species Class
- Actinopteri
- Amphibia
- Aves
- Chondrichthyes
- Cladistia
- Lepidosauria
- Mammalia
- NA

Blue: fishes
Red: mammals
Orange: birds
Green: reptiles
Purple: frogs

N terminal extension conserved in mammals in dlg4 clade

998 dlg sequences
GEEDIP

Clade
- DLG1
- DLG2
- DLG4

N-terminal Extension
- FALSE
- TRUE

Species Class
- Actinopteri
- Amphibia
- Aves
- Chondrichthyes
- Cladistia
- Lepidosauria
- Mammalia
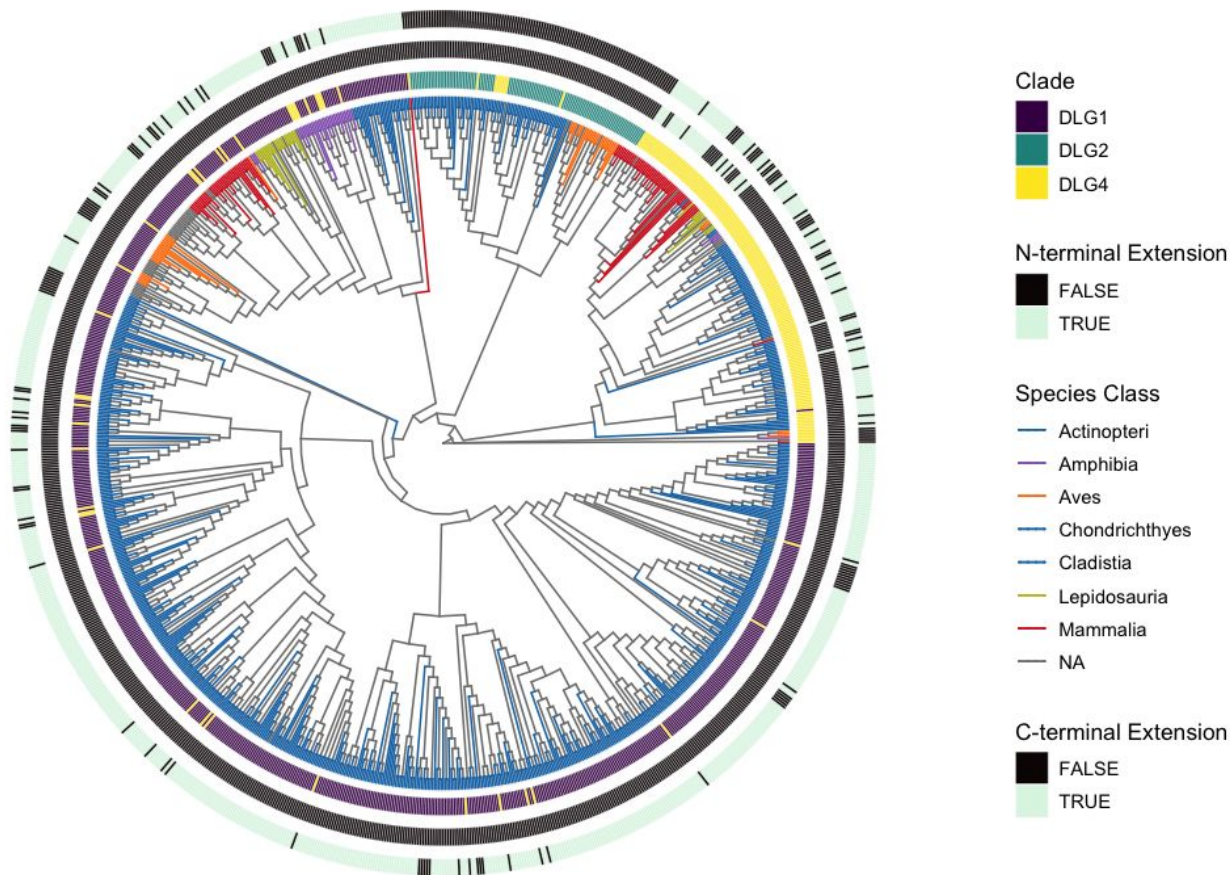- NA

Blue: fishes
Red: mammals
Orange: birds
Green: reptiles
Purple: frogs

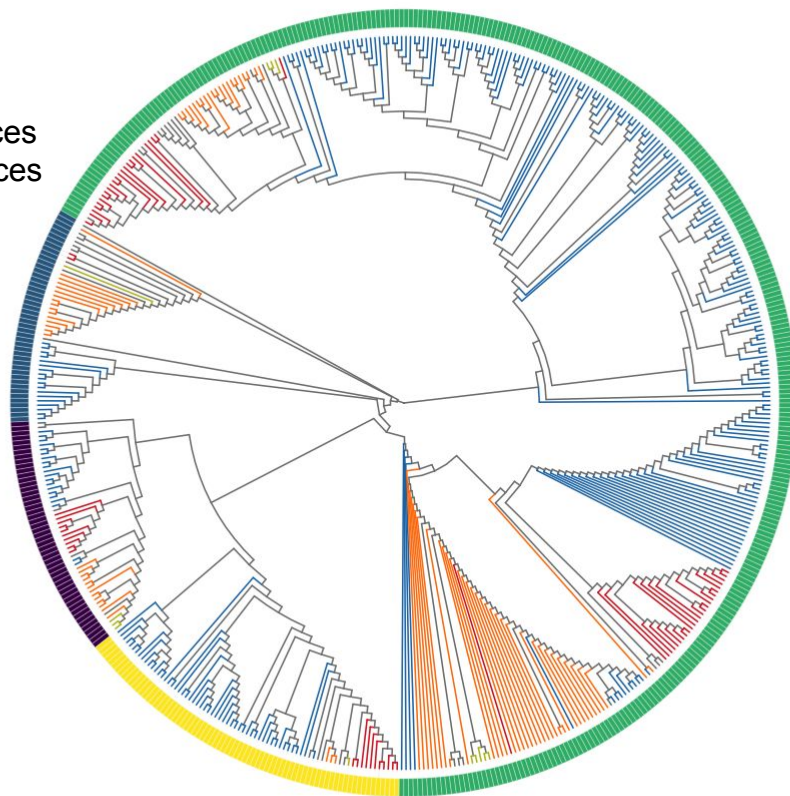# C terminal extension conserved in dlg1&dlg4 but not conserved in dlg2

998 dlg sequences
EYSRFEA

Clade
- DLG1
- DLG2
- DLG4

N-terminal Extension
- FALSE
- TRUE

Species Class
- Actinopteri
- Amphibia
- Aves
- Chondrichthyes
- Cladistia
- Lepidosauria
- Mammalia
- NA

C-terminal Extension
- FALSE
- TRUE

Results from cleaned-up dlg sequences

516 dlg sequences
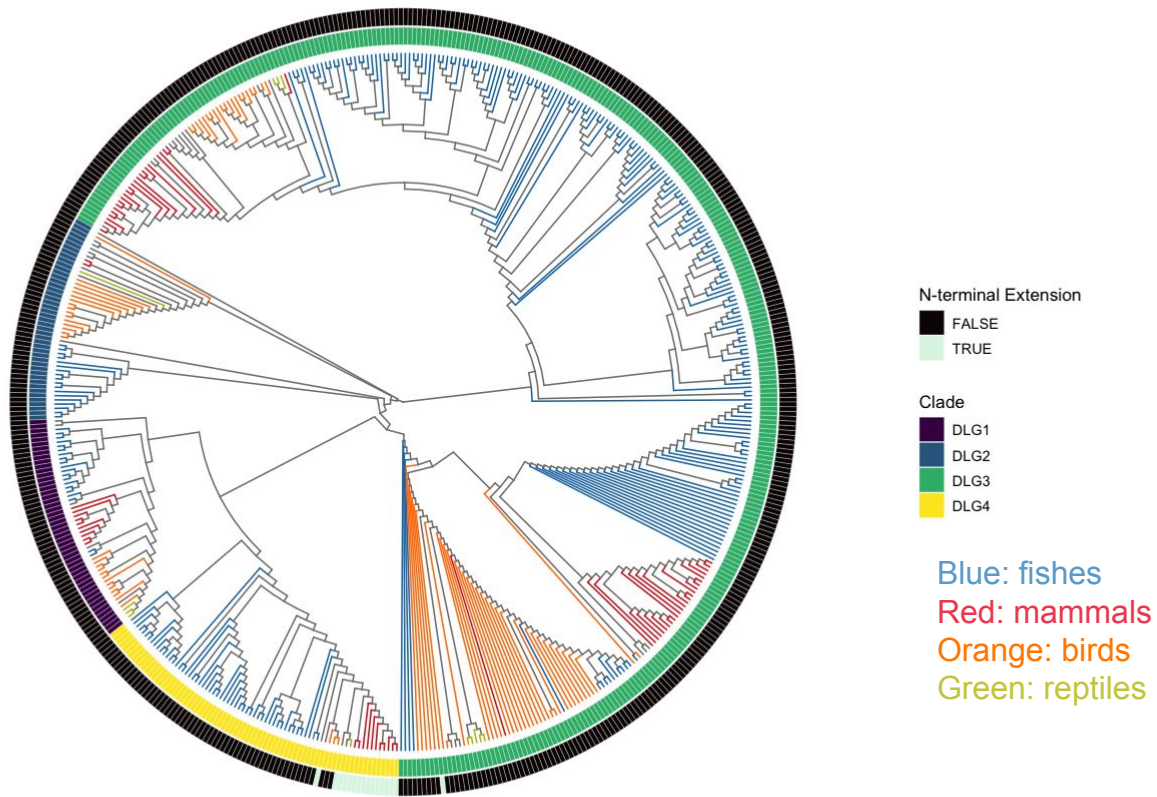- Removed redundant sequences
- Removed low quality sequences

Clade
- DLG1
- DLG2
- DLG3
- DLG4

Blue: fishes
Red: mammals
Orange: birds
Green: reptiles

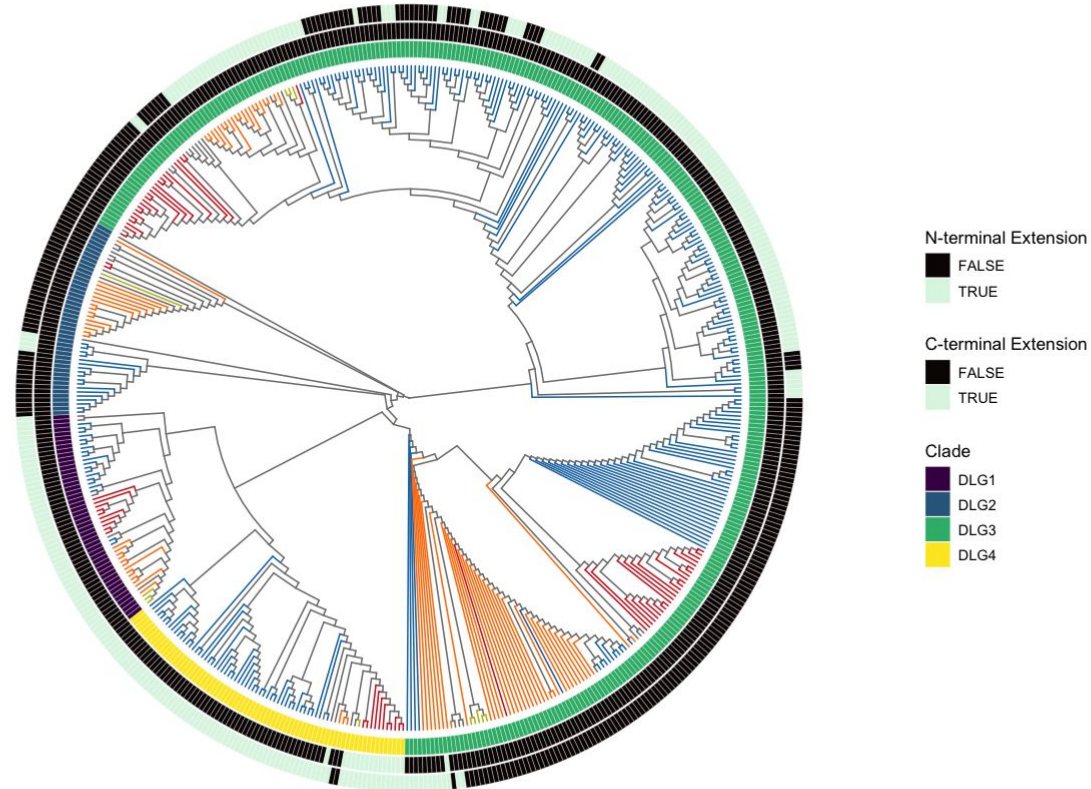# N terminal extension conserved in mammals in DLG4 clade
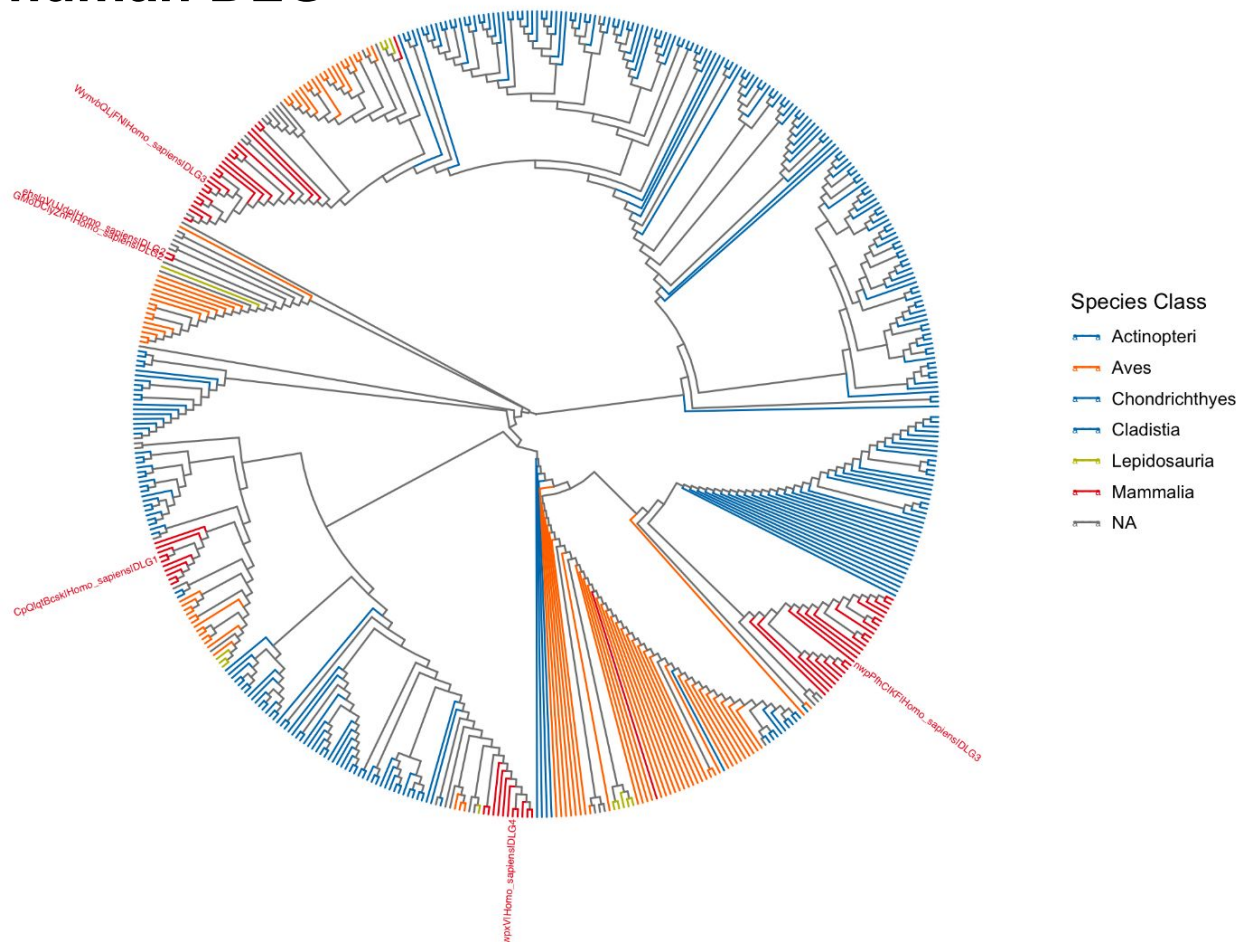
516 dlg sequences
GEEDIP

N-terminal Extension
- FALSE
- TRUE

Clade
- DLG1
- DLG2
- DLG3
- DLG4

Blue: fishes
Red: mammals
Orange: birds
Green: reptiles

C terminal extension conserved in dlg1& dlg4, but not conserved in dlg2 and subclade of dlg3

516 dlg sequences
EYSRFEA

Location of human DLG
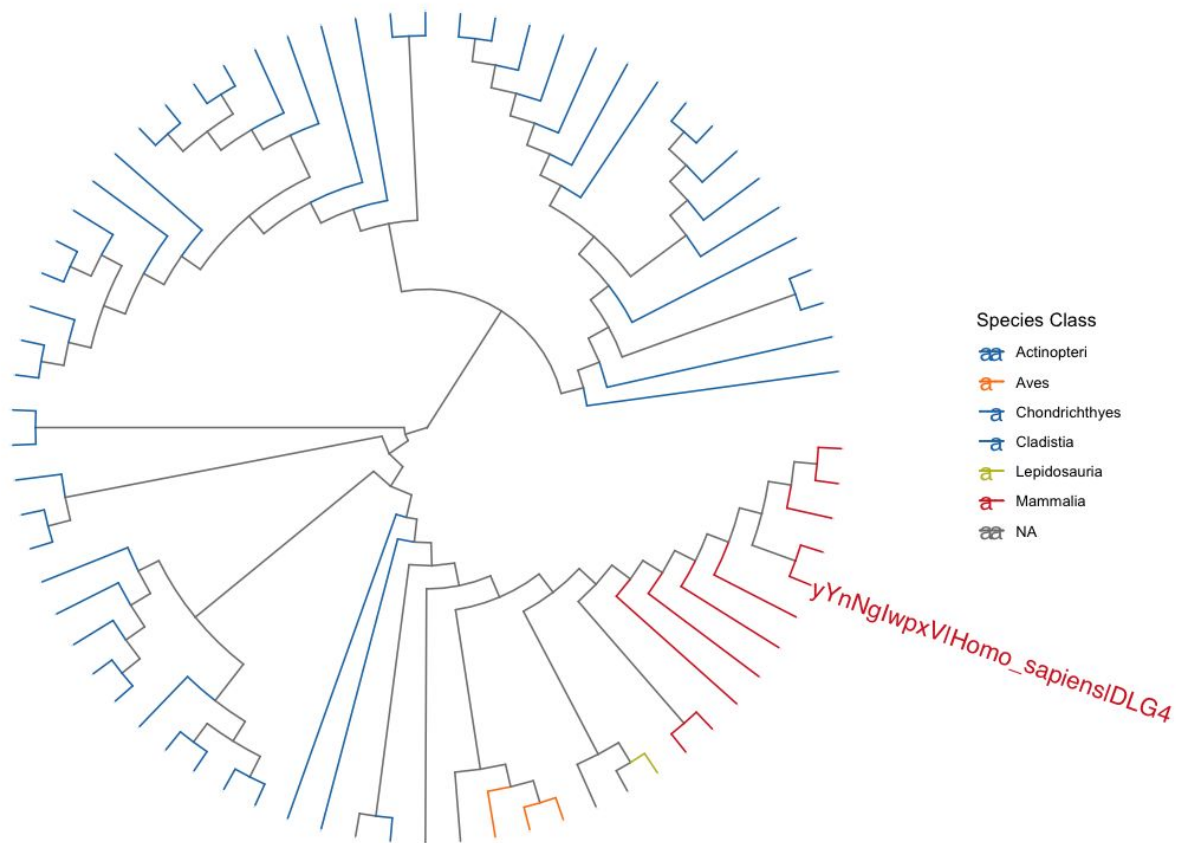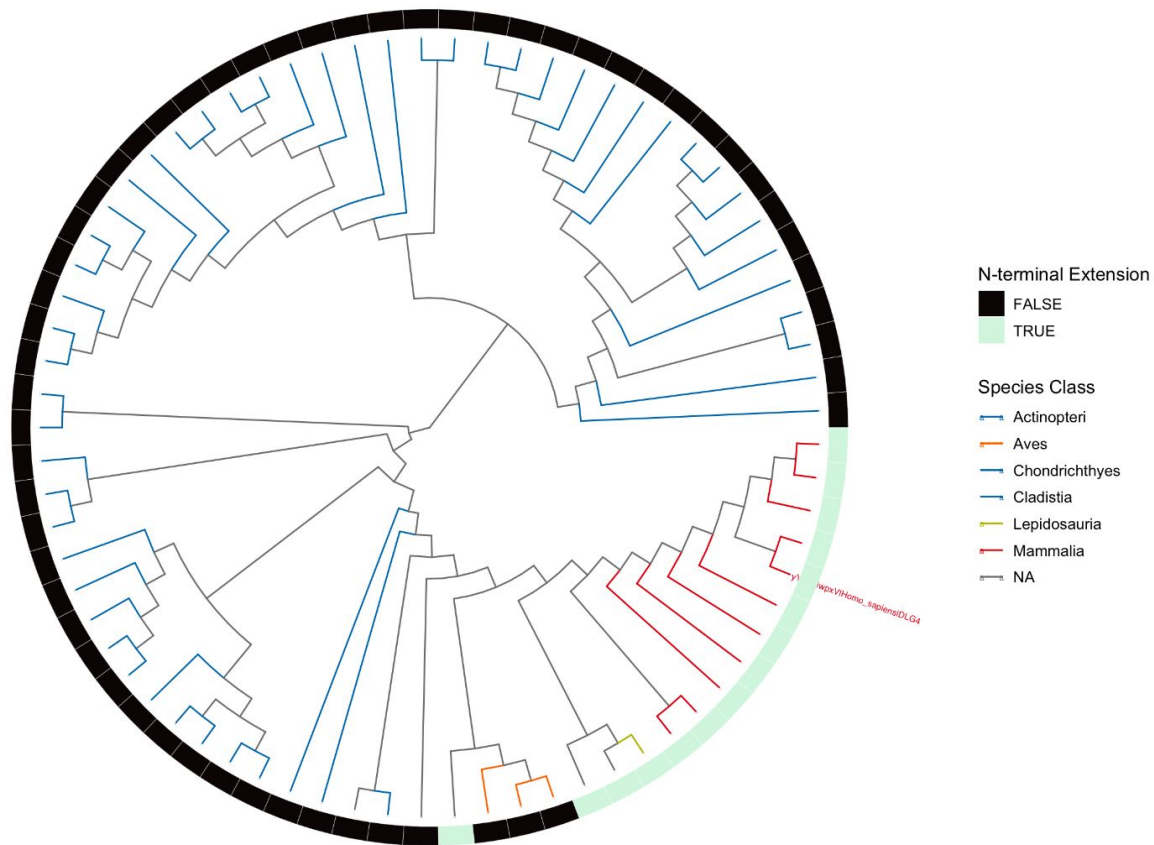
Zoom in on DLG4 tree

73 DLG4 sequences
- Filter for DLG4 clade

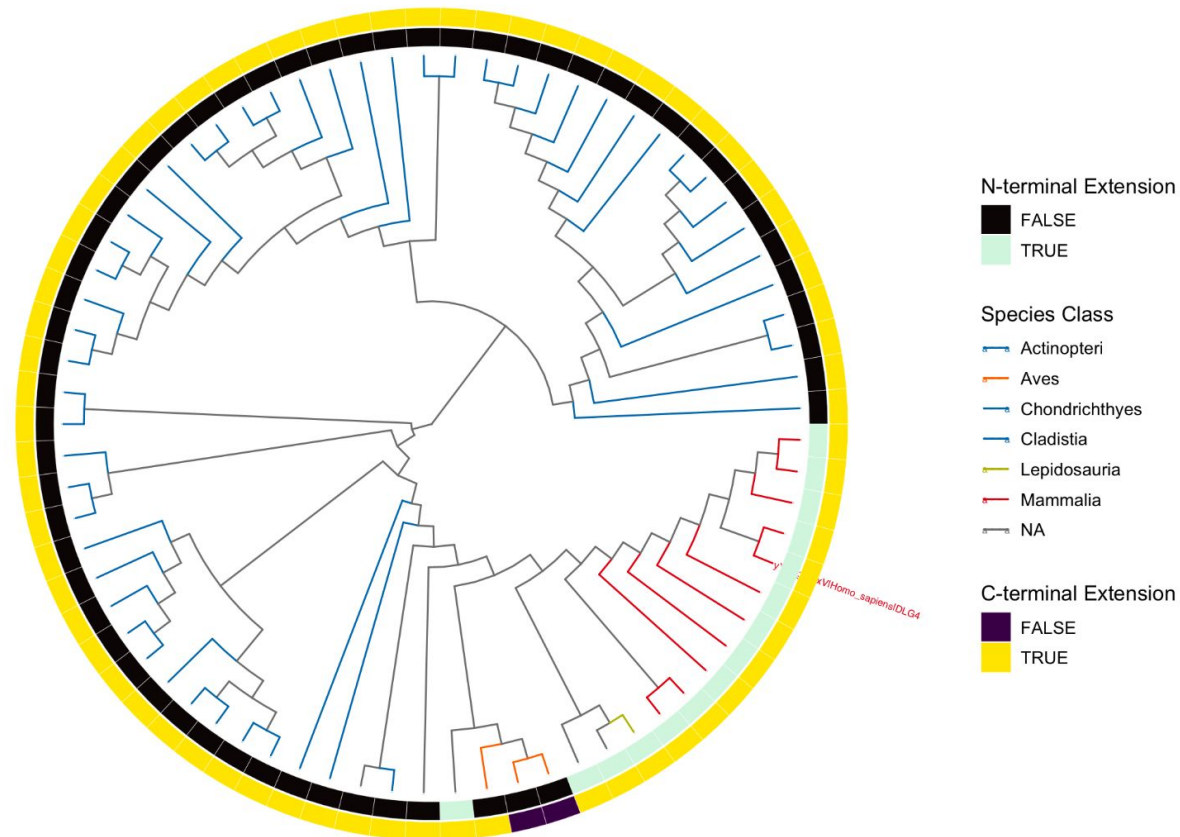# N terminal extension conserved in mammals

73 DLG4 sequences
GEEDIP

C terminal extension mostly conserved in DLG4

73 DLG4 sequences
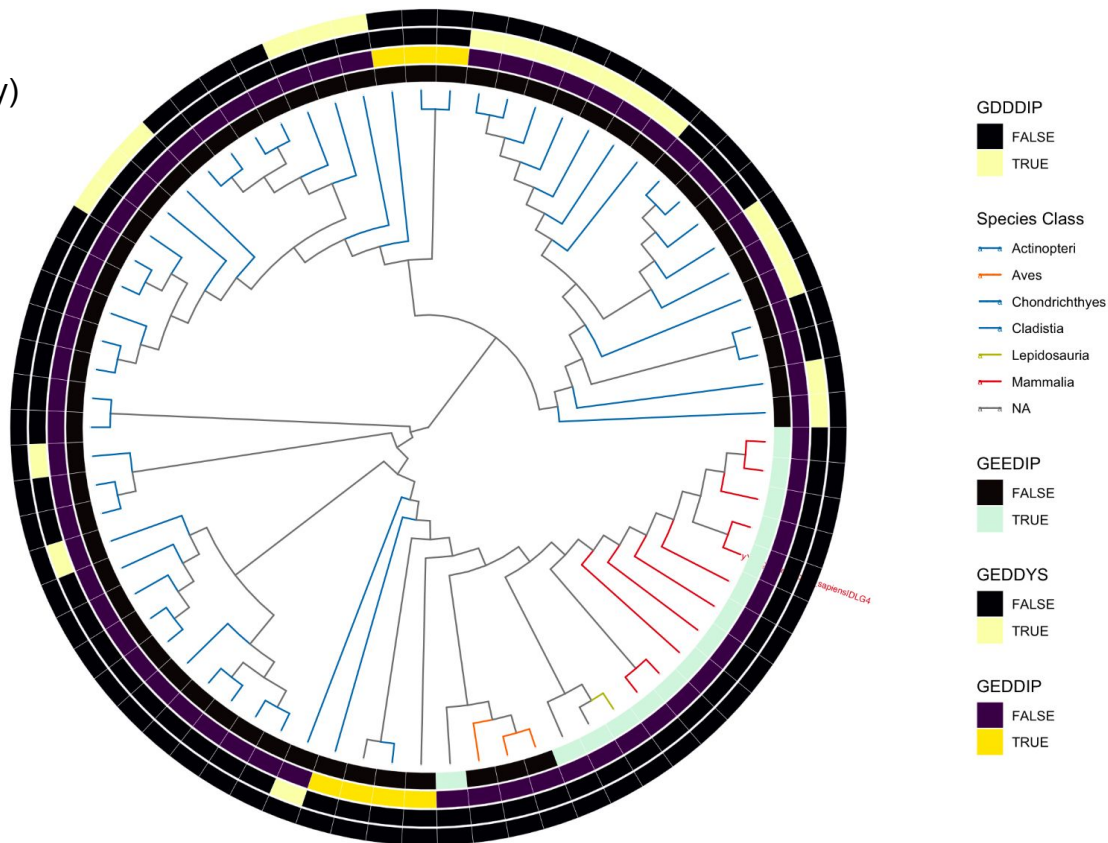EYSRFEA

N terminal extension variations
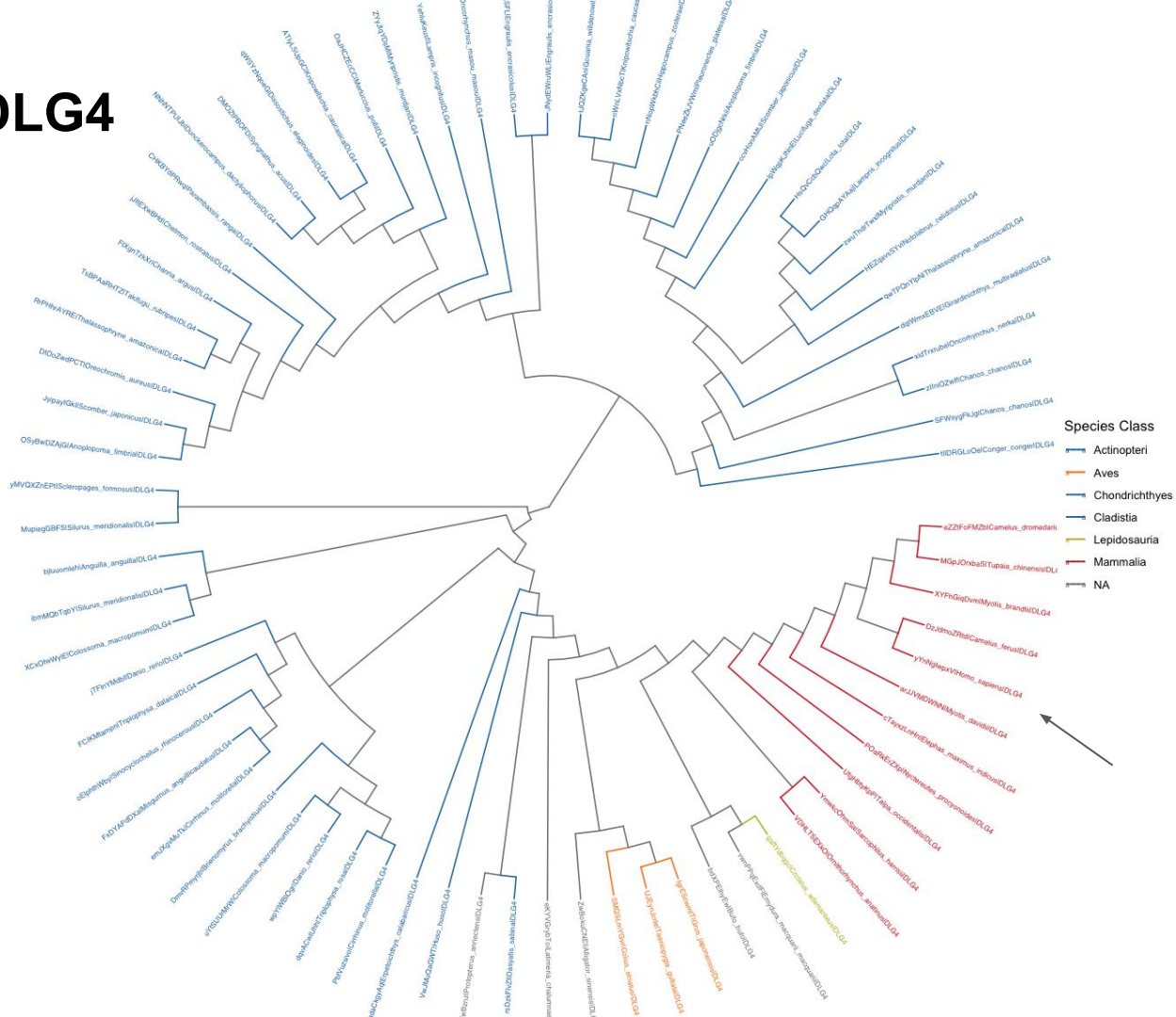
73 DLG4 sequences (Topiary)

```
GEEDIP (x13)
GEDDIP (x12)
GDDDIP (x9)
GEDDYS (x7)
```
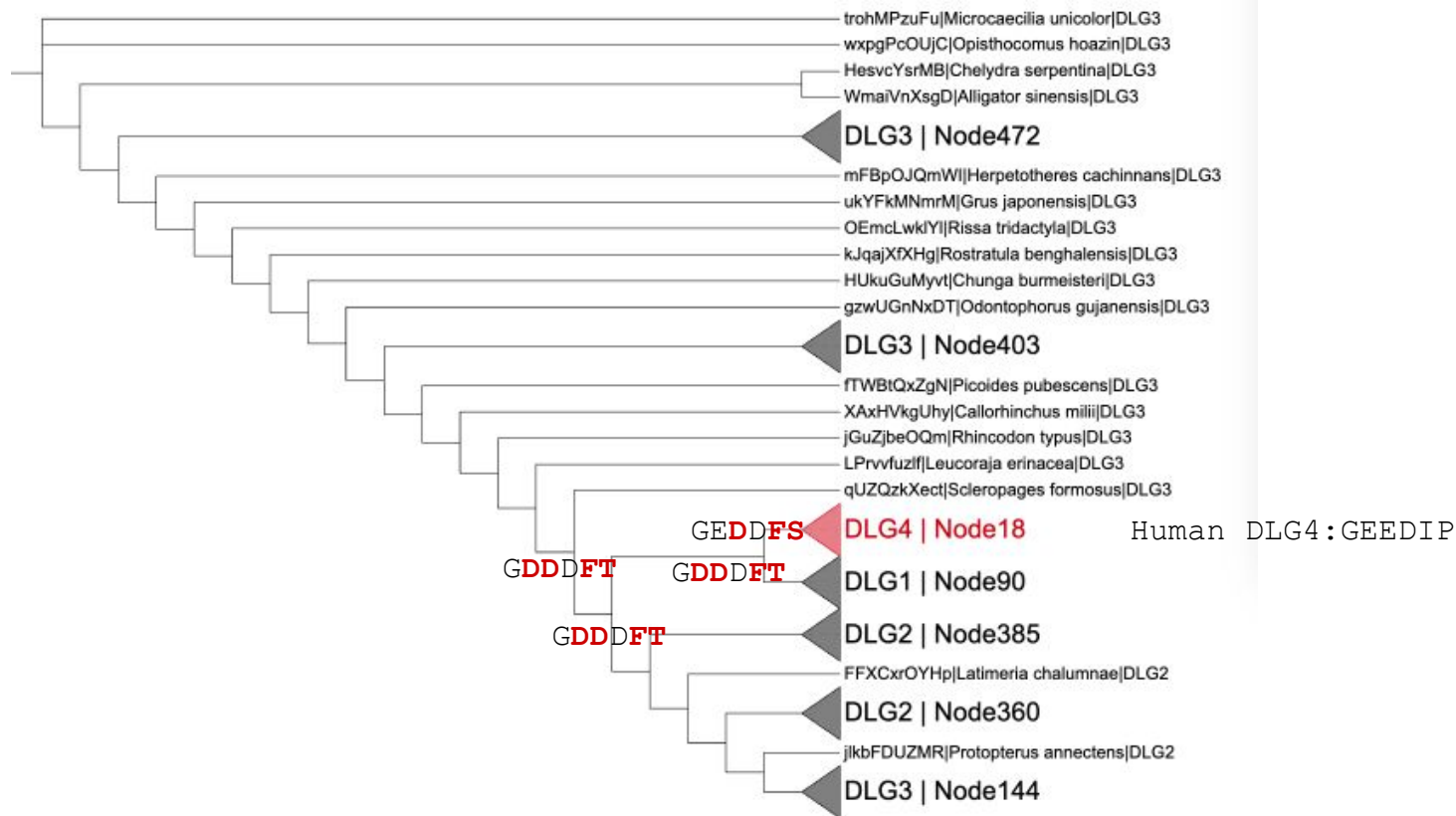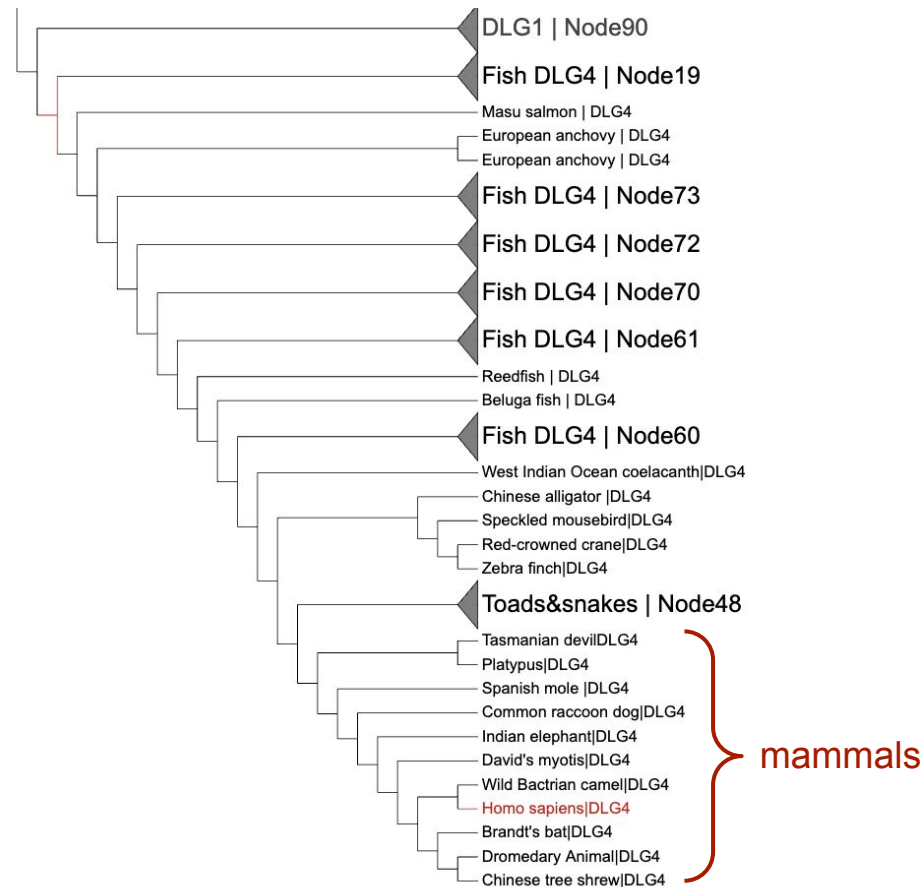
Location of human DLG4

Extension variations in reconstructed ancestral sequences
N-terminal

Extension variations in reconstructed ancestral sequences

N-terminal

Extension variations in reconstructed ancestral sequences
N-terminal