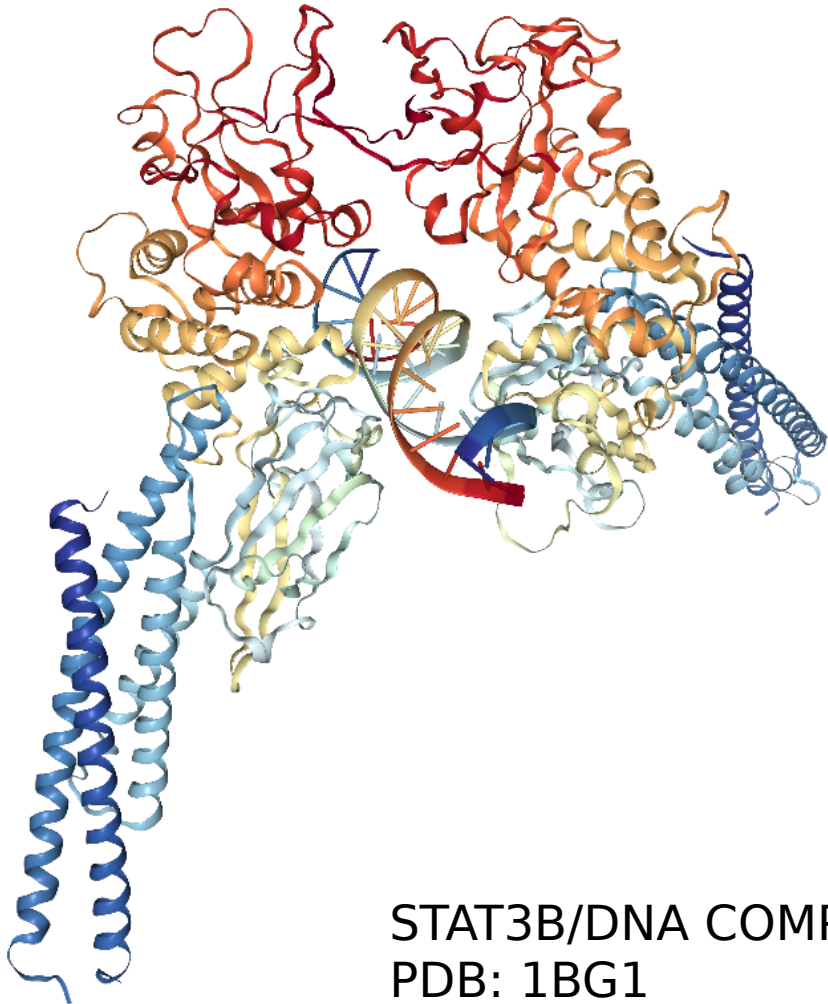


# TFClass - a classification of transcription factors

Jürgen Dönitz, Edgar Wingender

## Transcription factors, the proteins for regulation of transcription

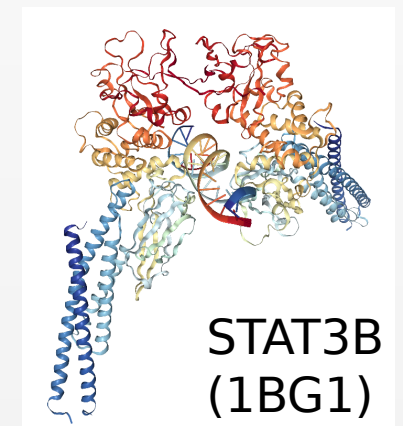
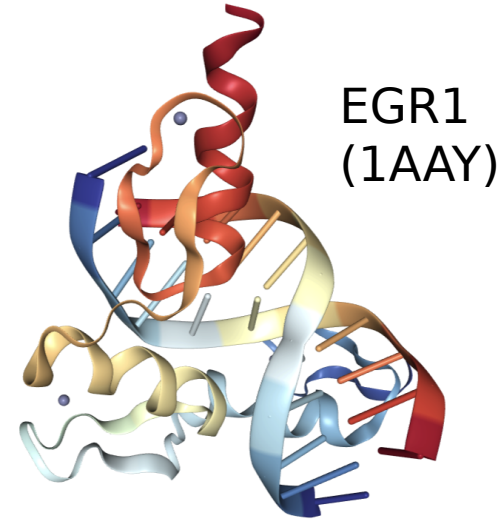


STAT3B/DNA COMPLEX  
PDB: 1BG1

„**Transcription factors** are proteins that regulate transcription by binding to specific sequence elements in regulatory genome regions such as promoters, enhancers etc.“

## TFClass: a hierarchy of TFs according their DNA binding domain (DBD)

- Proteins are built by modules, protein domains
- Most TF have at least one DBD
- The DBDs can be sorted into a few groups, e.g.
  - Zinc finger
  - Helix-turn-helix
  - alpha-helices exposed by beta-structures
  - Immunoglobulin fold



# TFClass: The top level Superclasses

## Classification of Transcription Factors in Mammalia

[About TFClass](#)

Search in TFClass:

[Go to the search of TRANSFAC](#)

Superclass: , Class: ,  
Family: , Subfamily: , Genus: ,

- 1 Basic domains
- 2 Zinc-coordinating DNA-binding domains
- 3 Helix-turn-helix domains
- 4 Other all-alpha-helical DNA-binding domains
- 5 alpha-Helices exposed by beta-structures
- 6 Immunoglobulin fold
- 7 beta-Hairpin exposed by an alpha/beta-scaffold
- 8 beta-Sheet binding to DNA
- 9 beta-Barrel DNA-binding domains
- 0 Yet undefined DNA-binding domains

### General

#### About beta-Sheet binding to DNA

Class description: The DNA-binding domains of this superclass bind to DNA through single extended strands or beta-sheets.

#### Species within this Superclass:

40

#### Genera within this Superclass:

5

# The five levels of TFClass

Superclass: **S**, Class: **C**,  
 Family: **F**, Subfamily: **SF**, Genus: **G**,

- ▶ **S1** Basic domains
- ▶ **S2** Zinc-coordinating DNA-binding domains
- ▶ **S3** Helix-turn-helix domains
- ▶ **S4** Other all-alpha-helical DNA-binding domains
- ▼ **S5** alpha-Helices exposed by beta-structures
  - ▼ **C5.1** MADS box factors
    - ▼ **F5.1.1** Regulators of differentiation
      - ▼ **SF5.1.1.1** MEF2
        - G5.1.1.1.1** MEF2A
        - G5.1.1.1.2** MEF2B
        - G5.1.1.1.3** MEF2C
        - G5.1.1.1.4** MEF2D
    - ▶ **F5.1.2** Responders to external signals

- Names of the levels similar to the biological taxonomy
- Levels 1 - 4 are abstract, Genus relates to a gene
- Mammalian orthologs of human TFs
- Pattern of the IDs similar to the EC numbers
- One number for each level, separated by a dot

# The species reports

MEF2A

[Go to the search of TRANSFAC](#)


Superclass: , Class: ,  
 Family: , Subfamily: , Genus:

- ▶  1 Basic domains
- ▶  2 Zinc-coordinating DNA-binding domains
- ▶  3 Helix-turn-helix domains
- ▶  4 Other all-alpha-helical DNA-binding domains
- ▶  5 alpha-Helices exposed by beta-structures
  - ▼  5.1 MADS box factors
    - ▼  5.1.1 Regulators of differentiation
      - ▼  5.1.1.1 MEF2
        - 5.1.1.1.1 MEF2A
        - 5.1.1.1.2 MEF2B
        - 5.1.1.1.3 MEF2C
        - 5.1.1.1.4 MEF2D
- ▶  5.1.2 Responders to external signals
- ▶  5.3 SAND domain factors
- ▶  6 Immunoglobulin fold
- ▶  7 beta-Hairpin exposed by an alpha/beta-scaffold
- ▶  8 beta-Sheet binding to DNA
- ▶  9 beta-Barrel DNA-binding domains
- ▶  0 Yet undefined DNA-binding domains

**General**

Species within this Genus:  
30

LOGO plot of the DNA binding sequence



**Species specific reports**

Jump to [human](#), [mouse](#) or [rat](#)  
 Hint: You can rearrange the panels by drag-and-drop. The new arrangement will persist during your session.


**Bos taurus**

Synonyms: [cattle](#), [domestic cow](#)

**Outlinks**

- Ensembl gene: [ENSBTAG00000010649](#)
- UniProt: [A2VDZ3](#)

**Isoforms and DNA binding domain**



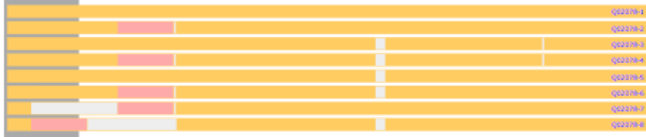
**Homo sapiens**

Synonyms: [human](#), [man](#)

**Outlinks**

- BioGPS: [ENSG00000068305](#)
- Ensembl gene: [ENSG00000068305](#)
- ProteinAtlas: [ENSG00000068305 \( with antibody\)](#)
- TRANSFAC: [T09508](#)
- UniProt: [Q02078](#)

**Isoforms and DNA binding domain**



Protein expression pattern: [Show table](#)

**Callithrix jacchus**

# The DBD in the Isoforms

MEF2A

[Go to the search of TRANSFAC](#)

Superclass: , Class: ,  
Family: , Subfamily: , Genus:

- ▶  1 Basic domains
- ▶  2 Zinc-coordinating DNA-binding domains
- ▶  3 Helix-turn-helix domains
- ▶  4 Other all-alpha-helical DNA-binding domains
- ▼  5 alpha-Helices exposed by beta-structures
  - ▶  5.1 MADS box factors
    - ▶  5.1.1 Regulators of differentiation
      - ▶  5.1.1.1 MEF2

**General**

Species within this Genus:  
30

LOGO plot of the DNA binding sequence

**Species specific reports**

Jump to [human](#), [mouse](#) or [rat](#)  
Hint: You can rearrange the panels

**Bos taurus**

...le, domestic cow

Ensembl gene: [ENSBTAG00000010649](#)  
...ot: [A2VDZ3](#)

**Outlinks**

- BioGPS: [ENSG00000068305](#)
- Ensembl gene: [ENSG00000068305](#)
- ProteinAtlas: [ENSG00000068305 \( with antibody\)](#)
- TRANSFAC: [T09508](#)
- UniProt: [Q02078](#)

Protein expression pattern:[Show table](#)

**Isoforms and DNA binding domain**

**Callithrix jacchus**

**Outlinks**

- BioGPS: [ENSG00000068305](#)
- Ensembl gene: [ENSG00000068305](#)
- ProteinAtlas: [ENSG00000068305 \( with antibody\)](#)
- TRANSFAC: [T09508](#)
- UniProt: [Q02078](#)

Protein expression pattern:[Show table](#)

**Outlinks**

- BioGPS: [ENSG00000068305](#)
- Ensembl gene: [ENSG00000068305](#)
- ProteinAtlas: [ENSG00000068305 \( with antibody\)](#)
- TRANSFAC: [T09508](#)
- UniProt: [Q02078](#)

Protein expression pattern:[Show table](#)

**Isoforms and DNA binding domain**

**Isoforms and DNA binding domain**

# General information and LOGO plot of DNA binding site

MEF2A

[Go to the search of TRANSFAC](#)


Superclass: , Class: ,  
Family: , Subfamily: , Genus:

- ▶  1 Basic domains
- ▶  2 Zinc-coordinating DNA-binding domains
- ▶  3 Helix-turn-helix domains
- ▶  4 Other all-alpha-helical DNA-binding domains
- ▶  5 alpha-Helices exposed by beta-structures
  - ▶  5.1 MADS box factors
    - ▶  5.1.1 Regulators of differentiation
      - ▶  5.1.1.1 MEF2
        - 5.1.1.1.1 MEF2A
        - 5.1.1.1.2 MEF2B
        - 5.1.1.1.3 MEF2C
        - 5.1.1.1.4 MEF2D
      - ▶  5.1.2 Responders to external signals
    - ▶  5.3 SAND domain factors
  - ▶  6 Immunoglobulin fold
  - ▶  7 beta-Hairpin exposed by an alpha/beta-scaffold
  - ▶  8 beta-Sheet binding to DNA
  - ▶  9 beta-Barrel DNA-binding domains
  - ▶  0 Yet undefined DNA-binding domains

**General**

Species within this Genus:  
30

LOGO plot of the DNA binding sequence



**Species specific reports**

Jump to [human](#), [mouse](#) or [Bos taurus](#)

Hint: You can rearrange the order of the species

**Bos taurus**

Synonyms: [cattle](#), [domestic](#)


**Outlinks**

- Ensembl gene: [ENSG000000068305](#)
- UniProt: [A2VDZ3](#)

**General**

Species within this Genus:  
30

LOGO plot of the DNA binding sequence



**Homo sapiens**

Synonyms: [human](#), [man](#)

**Outlinks**

- BioGPS: [ENSG000000068305](#)
- Ensembl gene: [ENSG000000068305](#)
- ProteinAtlas: [ENSG000000068305 \( with antibody\)](#)
- TRANSFAC: [T09508](#)
- UniProt: [Q02078](#)

Protein expression pattern: [Show table](#)

**Callithrix jacchus**



# Consensus sequence

- ▶  1 Basic domains
- ▶  2 Zinc-coordinating DNA-binding domains
- ▶  3 Helix-turn-helix domains
- ▶  4 Other all-alpha-helical DNA-binding domains
- ▶  5 alpha-Helices exposed by beta-structures
  - ▼  5.1 MADS box factors
    - ▼  5.1.1 Regulators of differentiation
      - ▼  5.1.1.1 MEF2
        - 5.1.1.1.1 MEF2A
        - 5.1.1.1.2 MEF2B
        - 5.1.1.1.3 MEF2C
        - 5.1.1.1.4 MEF2D
    - ▶  5.1.2 Responders to external signals

General

**Consensus binding sequence**  
CTAAAATAG

For: [human](#), [mouse](#), [cow](#), [dog](#)

Species within this Subfamily:  
37

Genera within this Subfamily:  
4

LOGO plot of the DNA binding domain

FASTA file of the LogoPlot

FASTA DNA binding domain FASTA file

---

Phylogenetic tree Mammalia

webPRANK and iTOL      Phylogeny.fr and iTOL      Phylogeny.fr and iTOL

General

Species within this Genus:  
30

LOGO plot of the DNA binding sequence

LOGO plot of the DNA binding sequence

LOGO plot of the DNA binding sequence

MEF2D MEF2B

# The upper levels of the hierarchy

- ▶  1 Basic domains
- ▶  2 Zinc-coordinating DNA-binding domains
- ▶  3 Helix-turn-helix domains
- ▶  4 Other all-alpha-helical DNA-binding domains
- ▼  5 alpha-Helices exposed by beta-structures
  - ▼  5.1 MADS box factors
    - ▼  5.1.1 Regulators of differentiation
      - ▼  5.1.1.1 MEF2
        - 5.1.1.1.1 MEF2A
        - 5.1.1.1.2 MEF2B
        - 5.1.1.1.3 MEF2C
        - 5.1.1.1.4 MEF2D
      - ▶  5.1.2 Responders to external signals

**General**

**Consensus binding sequence**  
CTAAAATAG

**Predicted binding sites**  
For: [human](#), [mouse](#), [cow](#), [dog](#)

**Species within this Subfamily:**  
37

**Genera within this Subfamily:**  
4

**LOGO plot of the DNA binding domain**

5.1.1.1\_MEF2

weblogo.berkeley.edu

**FASTA file of the LogoPlot**

	DNA binding domain	Whole protein
FASTA	FASTA file	FASTA file
<b>Phylogenetic tree</b> Mammalia <input type="checkbox"/>	 webPRANK and iTOL	 Phylogeny.fr and iTOL

# LogoPlot of the DBD

- 1 Basic domains
- 2 Zinc-coordinating DNA-binding domains
- 3 Helix-turn-helix domains
- 4 Other all-alpha-helical DNA-binding domains
- ▾  5 alpha-Helices exposed by beta-structures
  - ▾  5.1 MADS box factors
    - ▾  5.1.1 Regulators of differentiation
      - ▾  5.1.1.1 MEF2
        - 5.1.1.1.1 MEF2A
        - 5.1.1.1.2 MEF2B
        - 5.1.1.1.3 MEF2C

**General**

**Consensus binding sequence**  
CTAAAAATAG

**Predicted binding sites**  
For: [human](#), [mouse](#), [cow](#), [dog](#)

**Species within this Subfamily:**  
37

**Genera within this Subfamily:**  
4

**LOGO plot of the DNA binding domain**

5.1.1.1\_MEF2

LOGO plot of the DNA binding domain

5.1.1.1\_MEF2

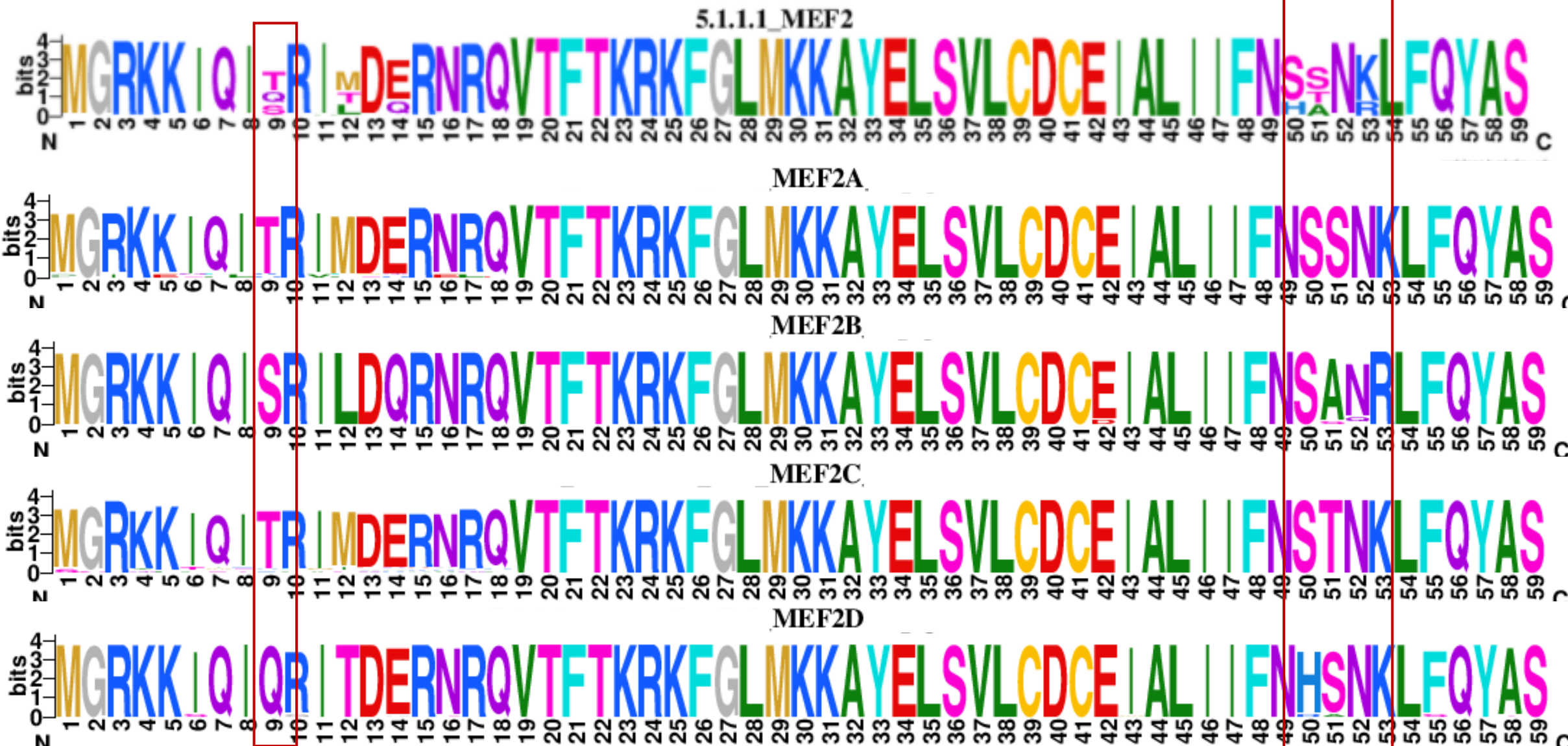


weblogo.berkeley.edu

[FASTA file of the LogoPlot](#)

# LogoPlot of the DBD

LOGO plot of the DNA binding domain



# The upper levels of the hierarchy

- ▶  1 Basic domains
- ▶  2 Zinc-coordinating DNA-binding domains
- ▶  3 Helix-turn-helix domains
- ▶  4 Other all-alpha-helical DNA-binding domains
- ▼  5 alpha-Helices exposed by beta-structures
  - ▼  5.1 MADS box factors
    - ▼  5.1.1 Regulators of differentiation
      - ▼  5.1.1.1 MEF2
        - 5.1.1.1.1 MEF2A
        - 5.1.1.1.2 MEF2B
        - 5.1.1.1.3 MEF2C
        - 5.1.1.1.4 MEF2D
      - ▶  5.1.2 Responders to external signals

**General**

**Consensus binding sequence**  
CTAAAAATAG

**Predicted binding sites**  
For: [human](#), [mouse](#), [cow](#), [dog](#)

**Species within this Subfamily:**  
37

**Genera within this Subfamily:**  
4

**LOGO plot of the DNA binding domain**

5.1.1.1\_MEF2

weblogo.berkeley.edu

**FASTA file of the LogoPlot**

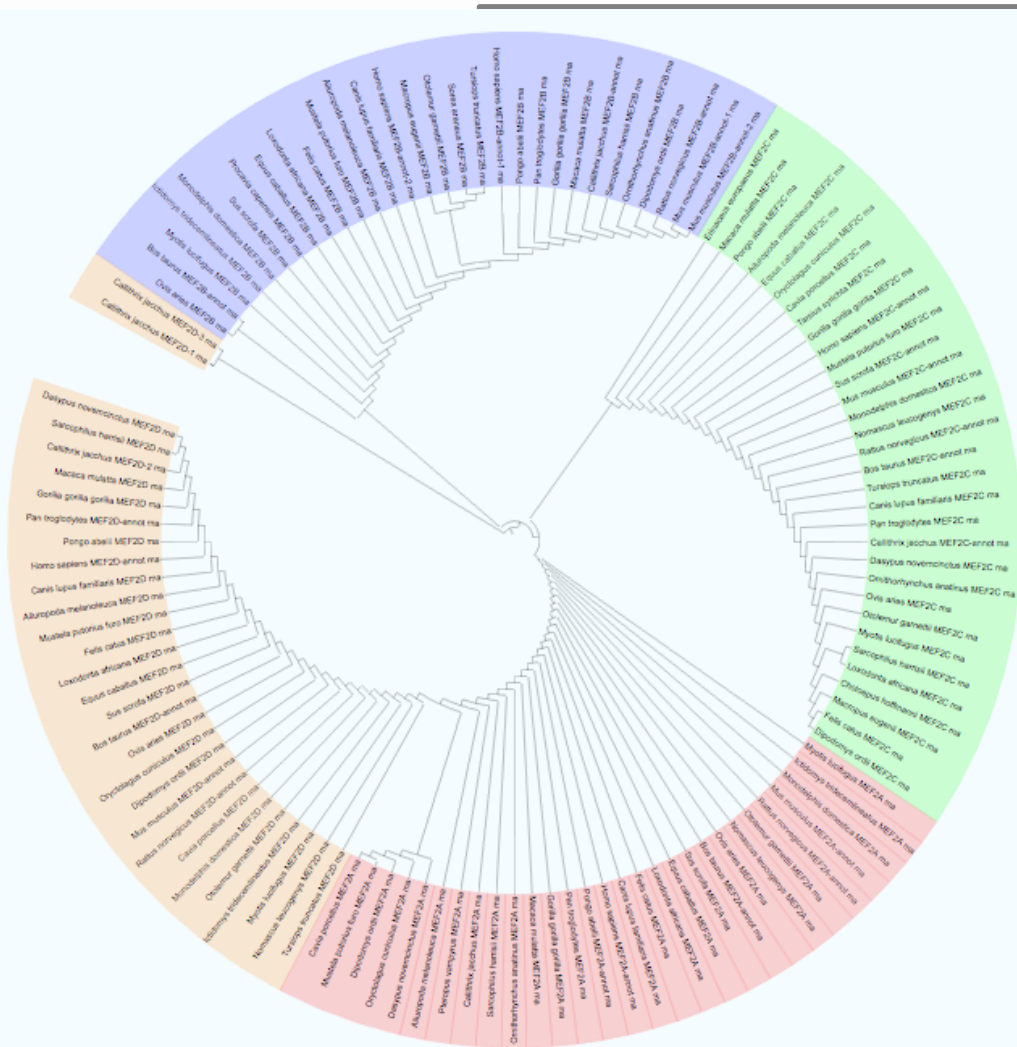
	DNA binding domain	Whole protein
FASTA	FASTA file	FASTA file
<b>Phylogenetic tree</b> Mammalia <input type="checkbox"/> ?	 webPRANK and iTOL	 Phylogeny.fr and iTOL

# Phylogenetic analysis

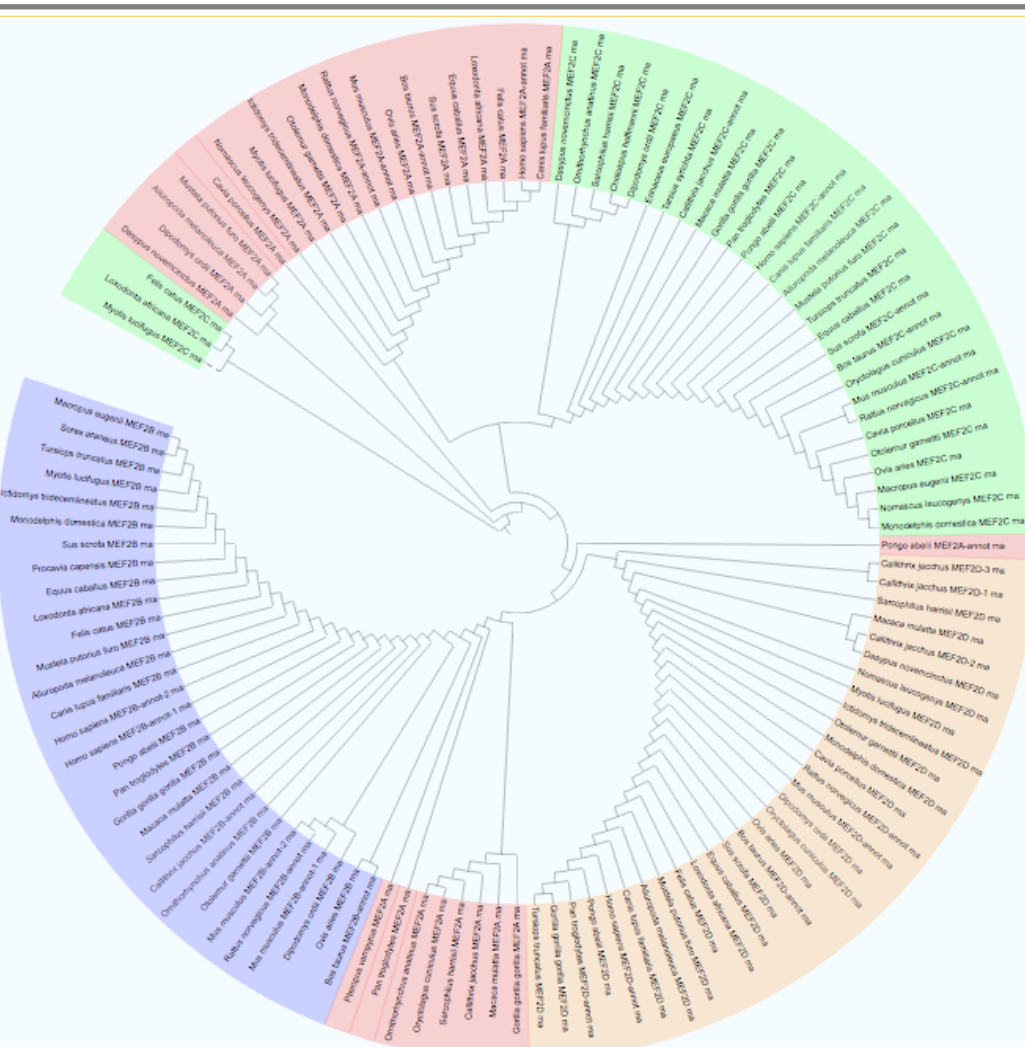
**Colored ranges**

- MEF2A
- MEF2B
- MEF2C
- MEF2D

- ▶ 2 Zinc-
- ▶ 3 Helix
- ▶ 4 Other
- ▼ 5 alpha
- ▼ 5.1 M
- ▼ 5.2
- ▼ 5.3
- ▼ 5.4
- ▶ 5.5



webPRANK and iTOL



Phylogeny.fr and iTOL

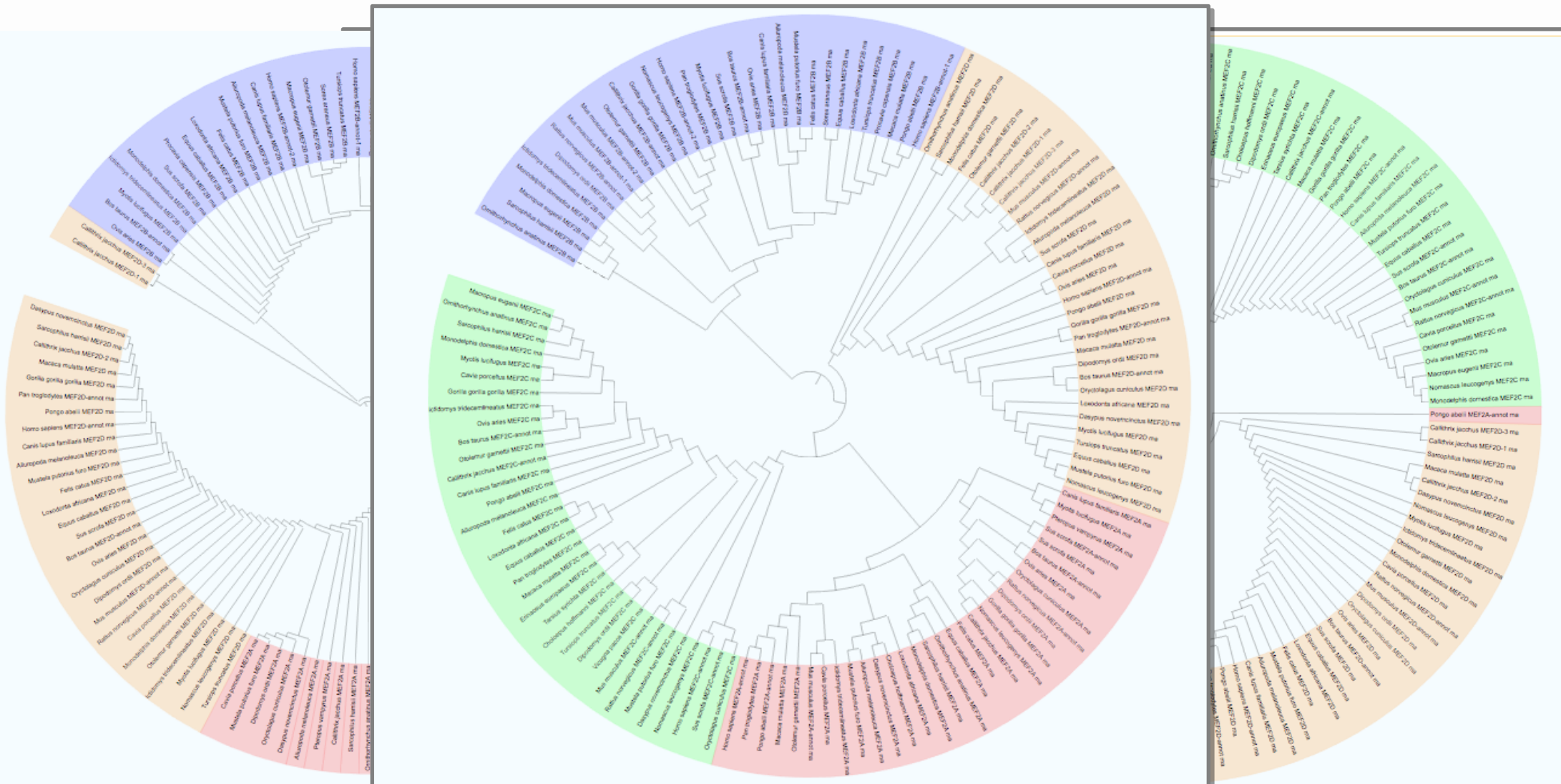
Phylogeny.fr and iTOL

# Phylogenetic analysis

**Colored ranges**

- MEF2A
- MEF2B
- MEF2C
- MEF2D

- ▶ 2 Zinc-
- ▶ 3 Helix
- ▶ 4 Other
- ▼ 5 alpha
- ▼ 5.1 M
- ▼ 5.2
- ▶ 5.3



webPRANK and iTOL

Phylogeny.fr and iTOL

Phylogeny.fr and iTOL

# Upstream nodes

- Phylogenetic trees, LOGO plot combined from the lower level
- Slim selection to control amount data

- ▾ 1 Basic domains
  - 1.1 Basic leucine zipper factors (bZIP)
  - ▾ 1.2 Basic helix-loop-helix factors (bHLH)
    - 1.2.1 E2A
    - 1.2.2 MyoD-ASC-related
    - ▾ 1.2.3 Tal-related
      - ▾ 1.2.3.1 TAL-HEN
        - ▢ 1.2.3.1.1 TAL1
        - ▢ 1.2.3.1.2 TAL2
        - ▢ 1.2.3.1.3 LYL1
        - ▢ 1.2.3.1.4 HEN1 (NHLH1)
        - ▢ 1.2.3.1.5 HEN2 (NHLH2)
      - ▾ 1.2.3.2 TWIST
        - ▢ 1.2.3.2.1 TWIST1
        - ▢ 1.2.3.2.2 TWIST2
        - ▢ 1.2.3.2.3 HAND1
        - ▢ 1.2.3.2.4 HAND2
        - ▢ 1.2.3.2.5 TCF15 (bHLH-EC2)
        - ▢ 1.2.3.2.6 SCX
        - ▢ 1.2.3.2.7 PTF1A
        - ▢ 1.2.3.2.8 FIGLA (FIGalpha, FIGalpha1, ...)
      - 1.2.3.3 MESP
      - 1.2.3.4 Neurogenin-ATO
      - 1.2.3.5 BHLHA9

**General**

Consensus binding sequence  
CAGCTG

Species within this Family:  
41

Genera within this Family:  
37

LOGO plot of the DNA binding domain

1.2.3 Tal-related

FASTA file of the LogoPlot

FASTA	DNA binding domain FASTA file	Whole protein FASTA file
	<p>Phylogenetic tree Mammalia ?</p> <p>Phylogeny.fr and iTOL</p>	<p>Phylogeny.fr and iTOL</p>
	<p>webPRANK and iTOL</p>	
	<p>Phylogenetic tree Mammalia (slim selection) ?</p> <p>Phylogeny.fr and iTOL</p>	<p>webPRANK and iTOL</p>
	<p>webPRANK and iTOL</p>	<p>Phylogeny.fr and iTOL</p>



# Future plans

- Recent publications:  
Wingender et al. Nucleic Acids Res (2013, 2015, 2018)

- Adding more species
- Data based backend
- More functions
- Funding to achieve
  - long time operating
  - new features
 ☐ Please fill out the survey

Please rate the following possible features  
(1 = not very important, 5 very important)

	1	2	3	4	5	No answer
Get a TF census of a defined species	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
Get species specific FASTA files	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
Get FASTA files for a branch of the classification	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
Query TFClass with a sequence	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
Use TFClass to evaluate a transcriptome	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
A RESTful interface to access the data of TFClass	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
Indication of other domains for TFs / species	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
Tracking of the history, e.g. if a family is split into two families	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
Include community comments / discussions about TF-Class items	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>

- <http://tfclass.bioinf.med.uni-goettingen.de>

Thank you!

TFClass is available at  
<http://tfclass.bioinf.med.uni-goettingen.de>