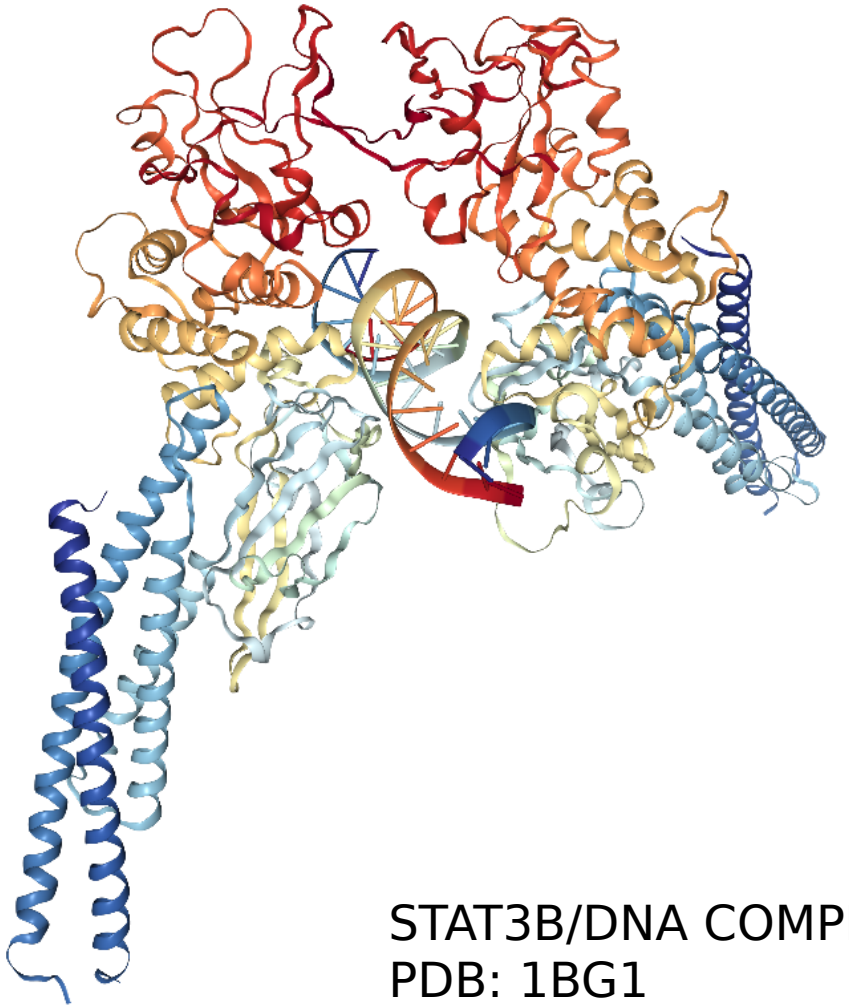
TFClass – a classification of transcription factors

Jürgen Dönitz, Edgar Wingender

Transcription factors, the proteins for regulation of transcription

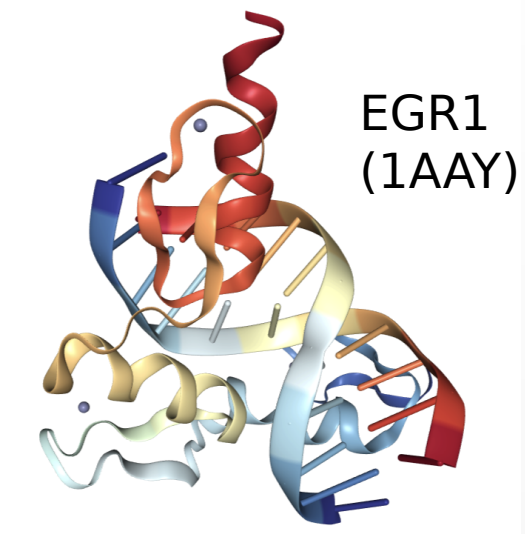


STAT3B/DNA COMPLEX
PDB: 1BG1

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

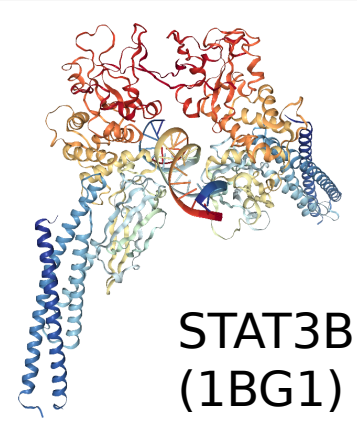
TFClass: a hierarchy of TFs according their DNA binding domain (DBD)

- 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


EGR1
(1AAY)

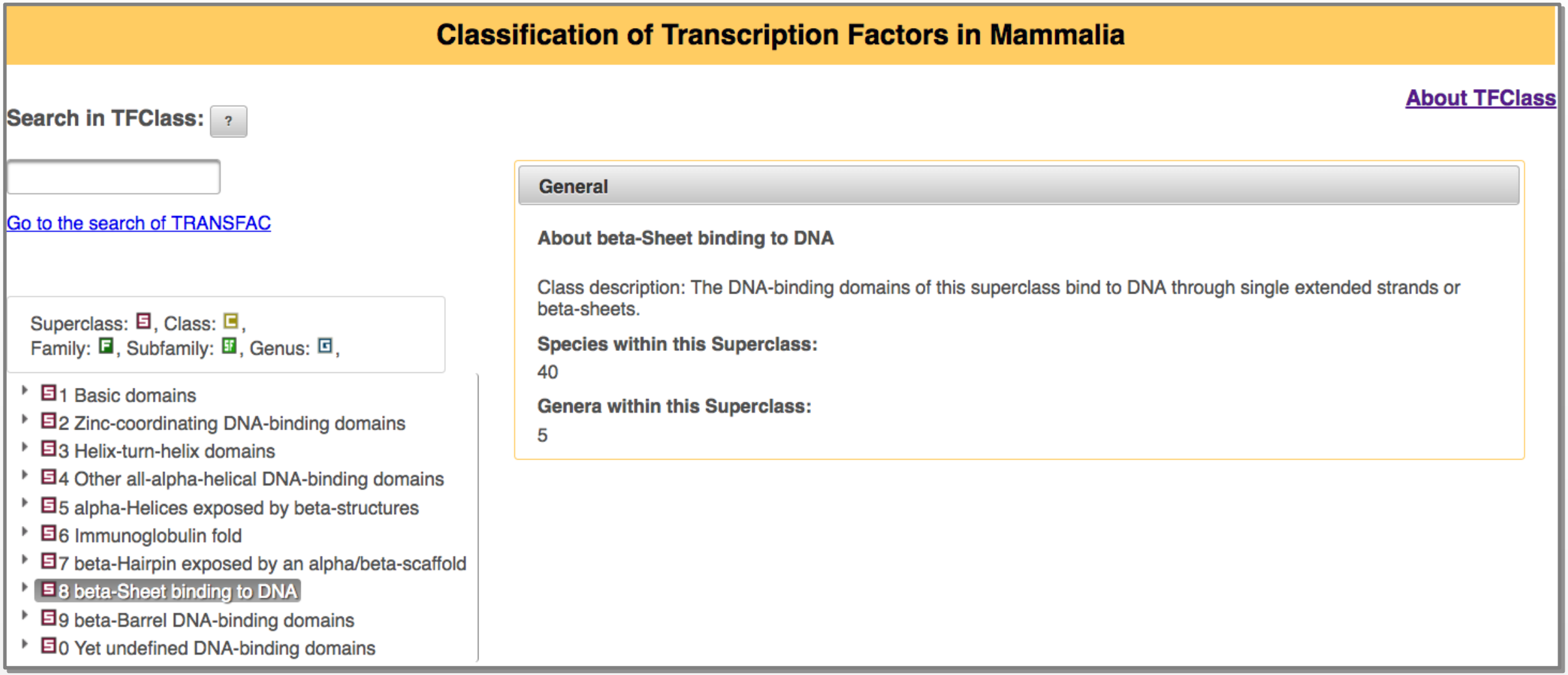
CDX2
(5LTY)

MEF2A
(1C7U)

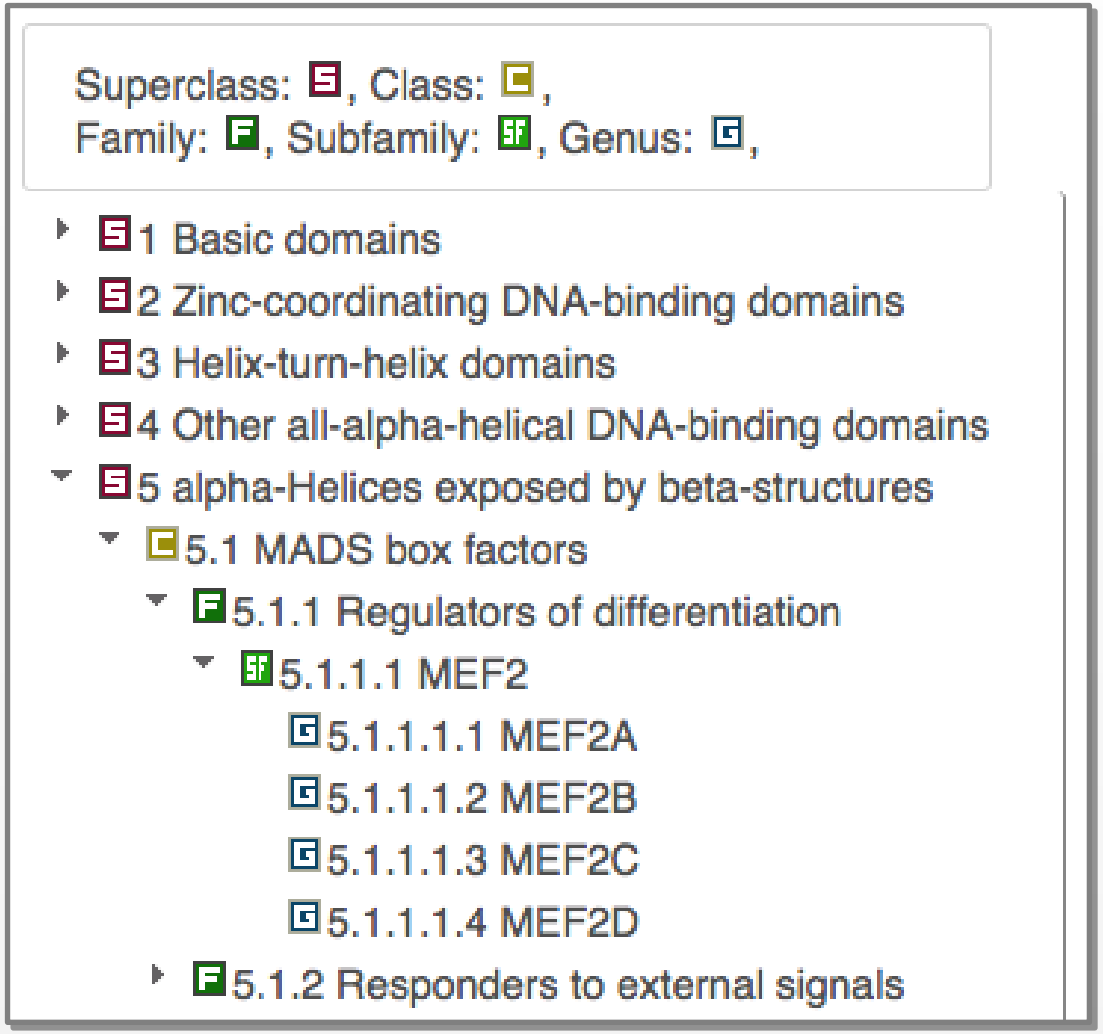

STAT3B
(1BG1)

Images taken from RCSB PDB, IDs in parentheses

TFClass: The top level Superclasses

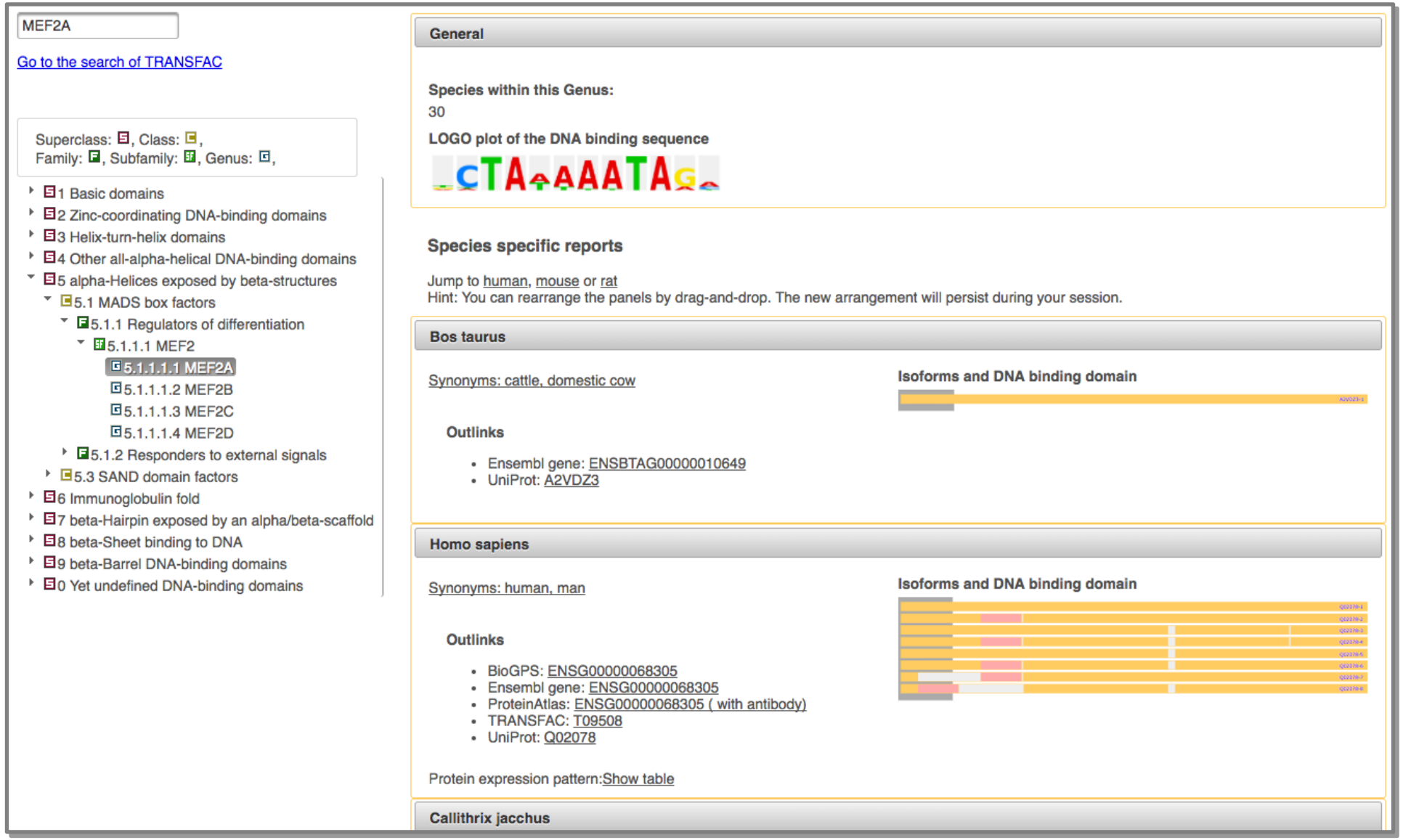


The five levels of TFClass

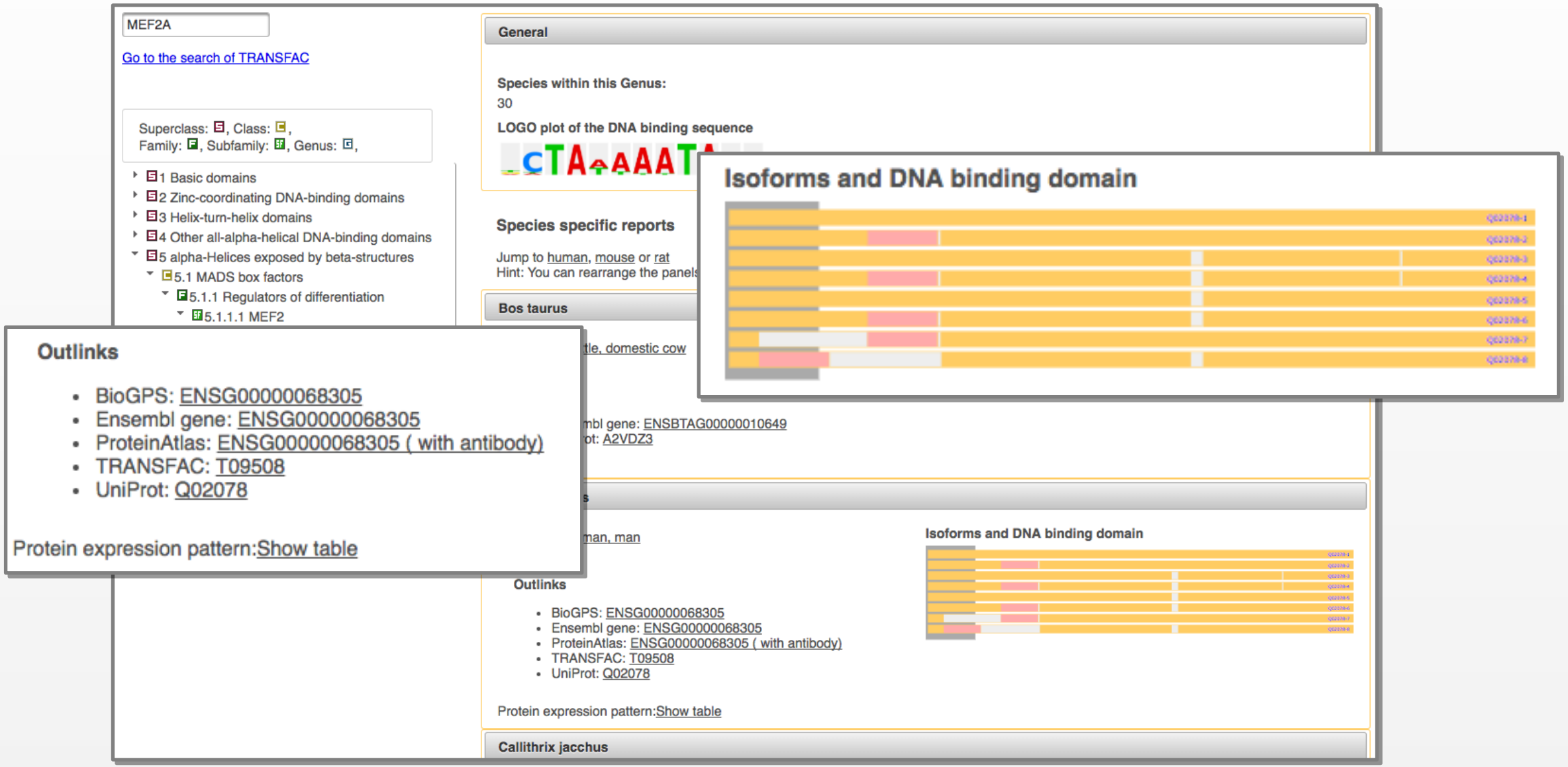


- Names of the levels similar to the biological taxonomy
- Levels 1 – 4 are abstract, Genus relates to a gene
- Mammalian orthologs of human TFs
- Pattern of the IDs similar to the EC numbers
- One number for each level, separated by a dot

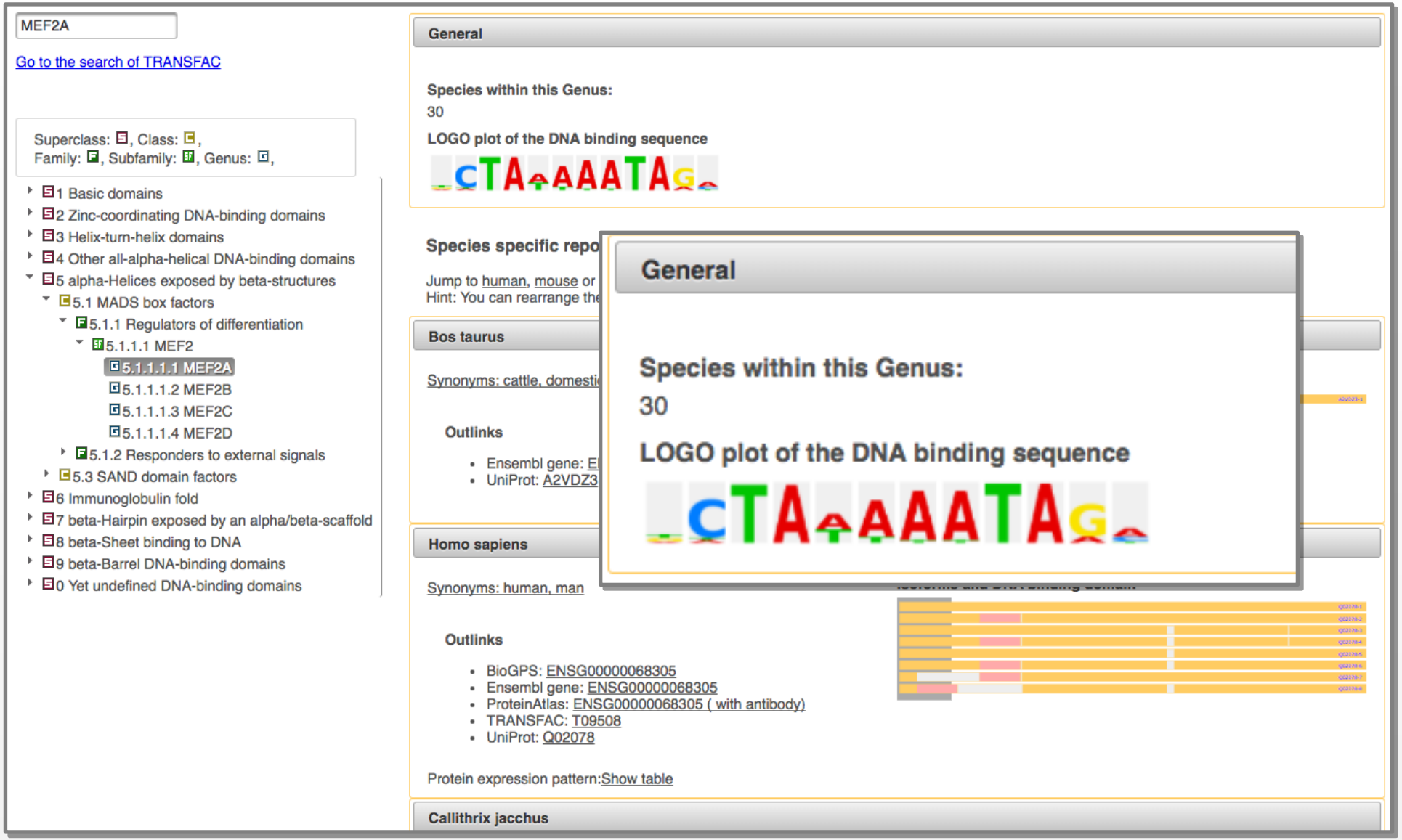
The species reports



The DBD in the Isoforms

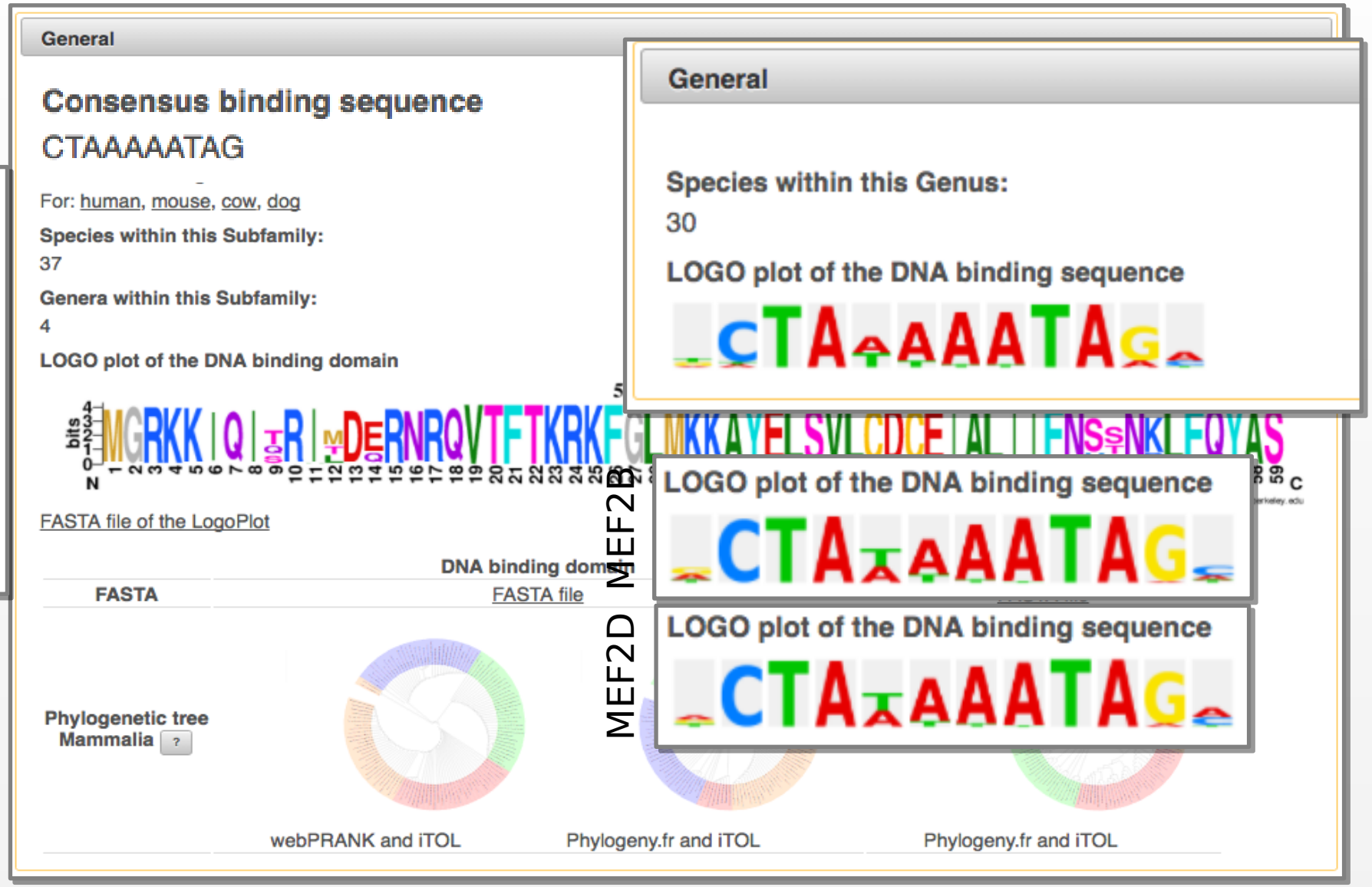


General information and LOGO plot of DNA binding site



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



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

FASTA file of the LogoPlot

	DNA binding domain	Whole protein
FASTA	FASTA file	FASTA file
Phylogenetic tree Mammalia ?	webPRANK and iTOL	Phylogeny.fr 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

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

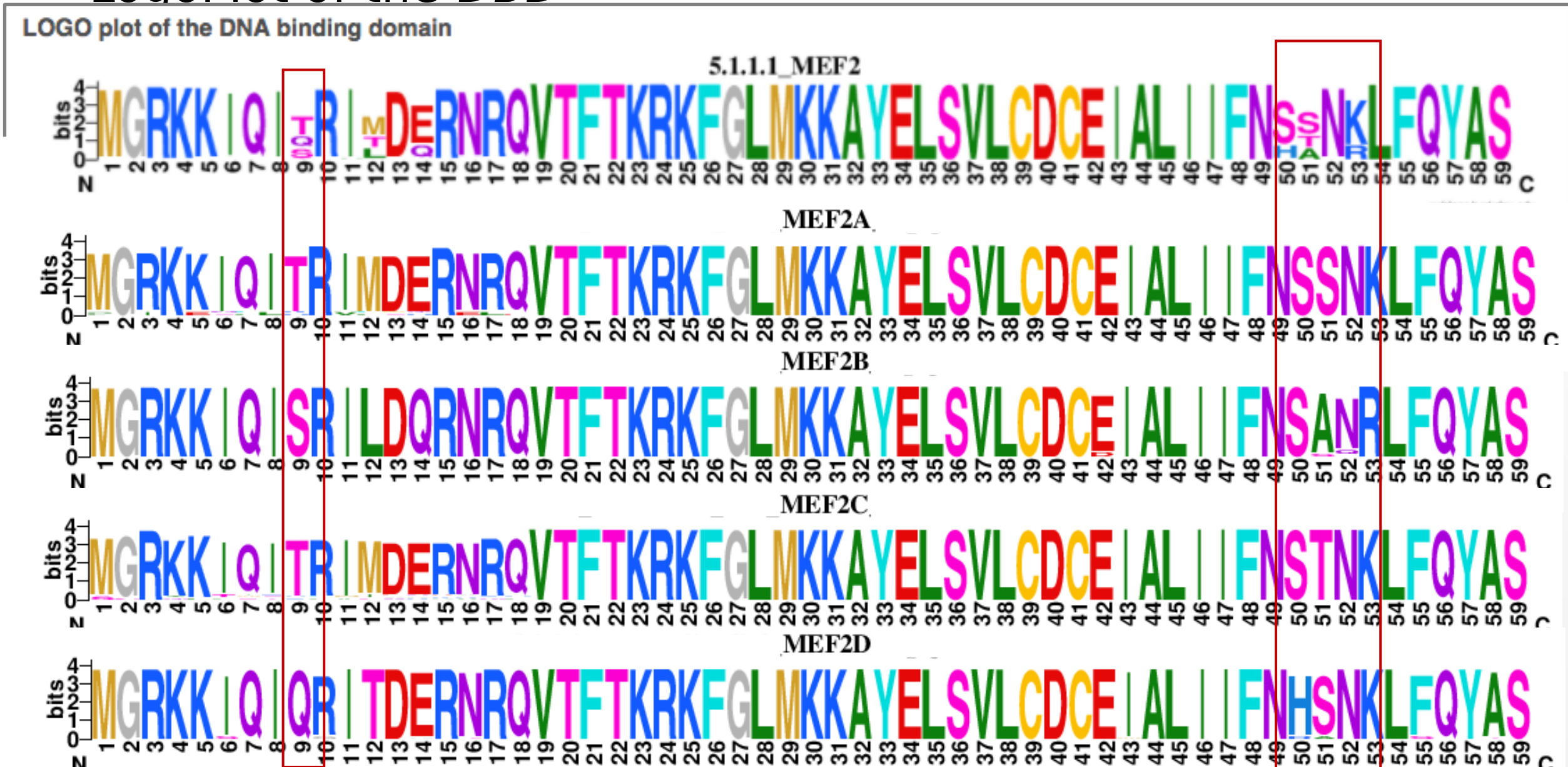
LOGO plot of the DNA binding domain

FASTA file of the LogoPlot

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

UMG

LogoPlot of the DBD



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

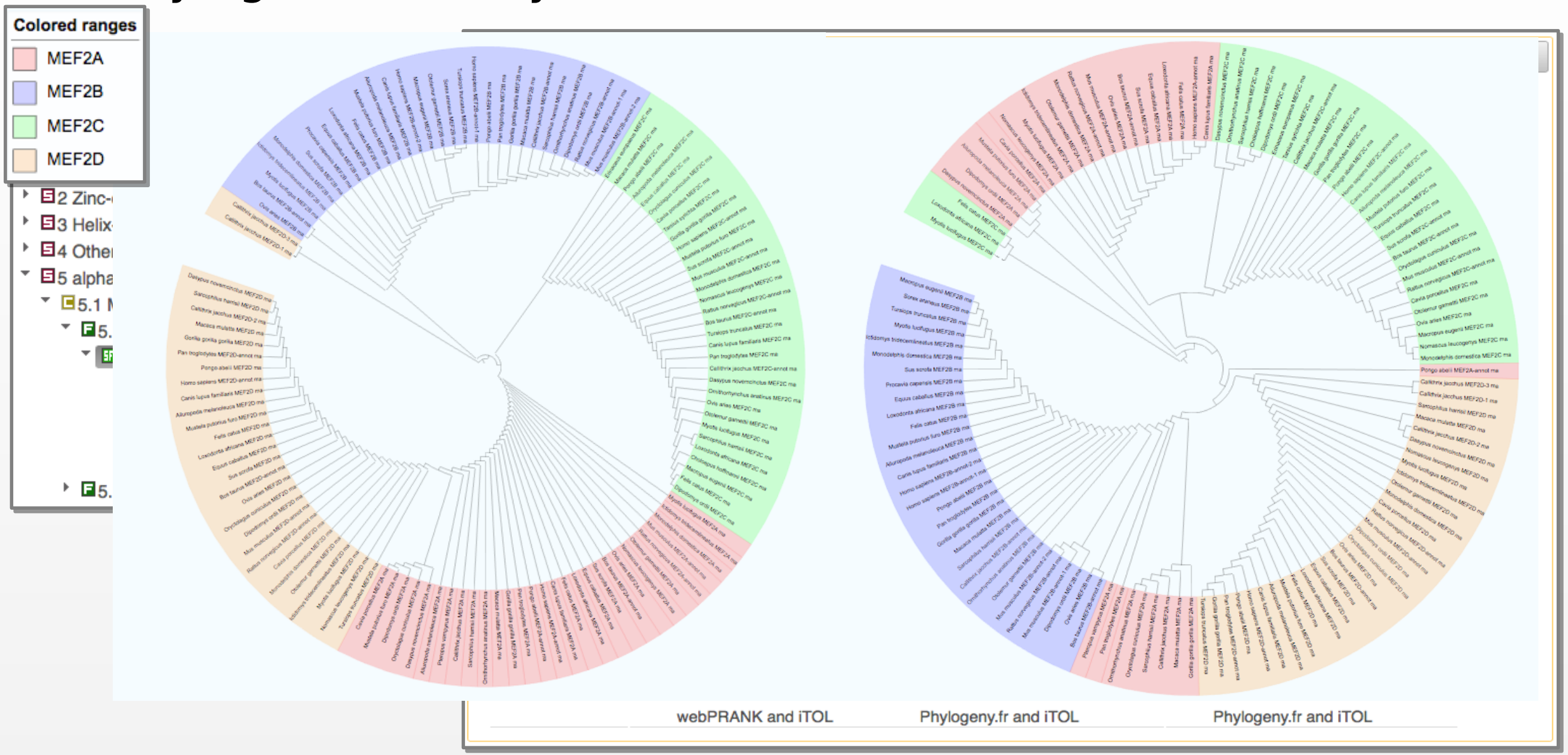
Genera within this Subfamily:
4

LOGO plot of the DNA binding domain

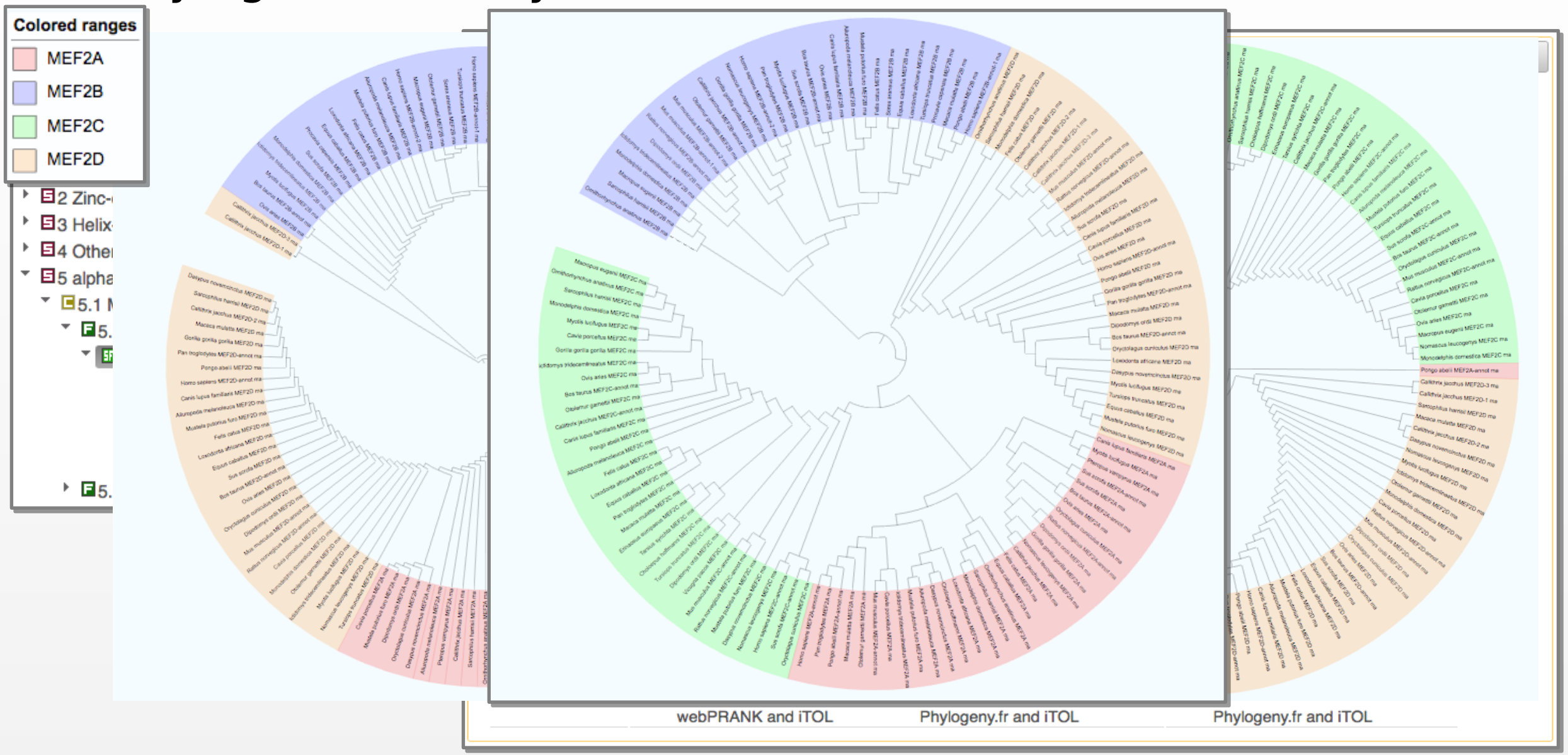
FASTA file of the LogoPlot

	DNA binding domain	Whole protein
FASTA	FASTA file	FASTA file
Phylogenetic tree Mammalia ?	webPRANK and iTOL	Phylogeny.fr and iTOL	Phylogeny.fr and iTOL

Phylogenetic analysis

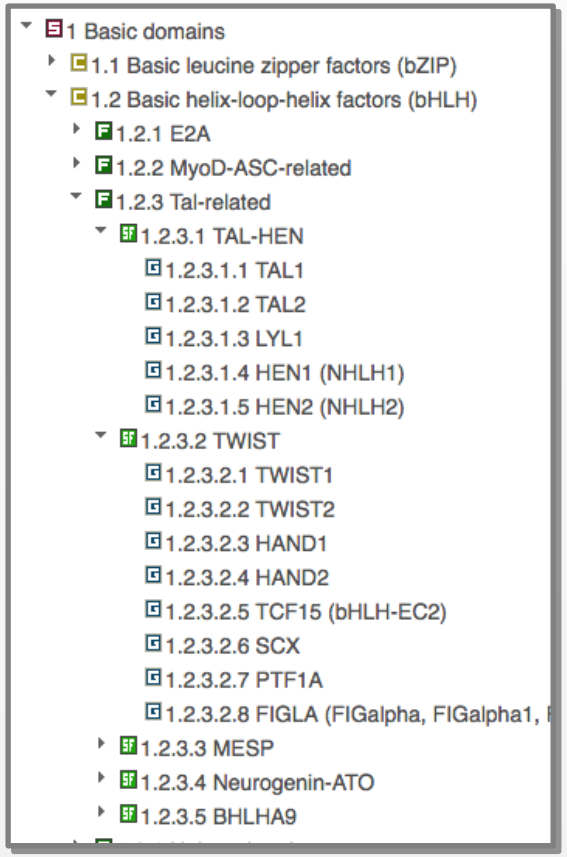


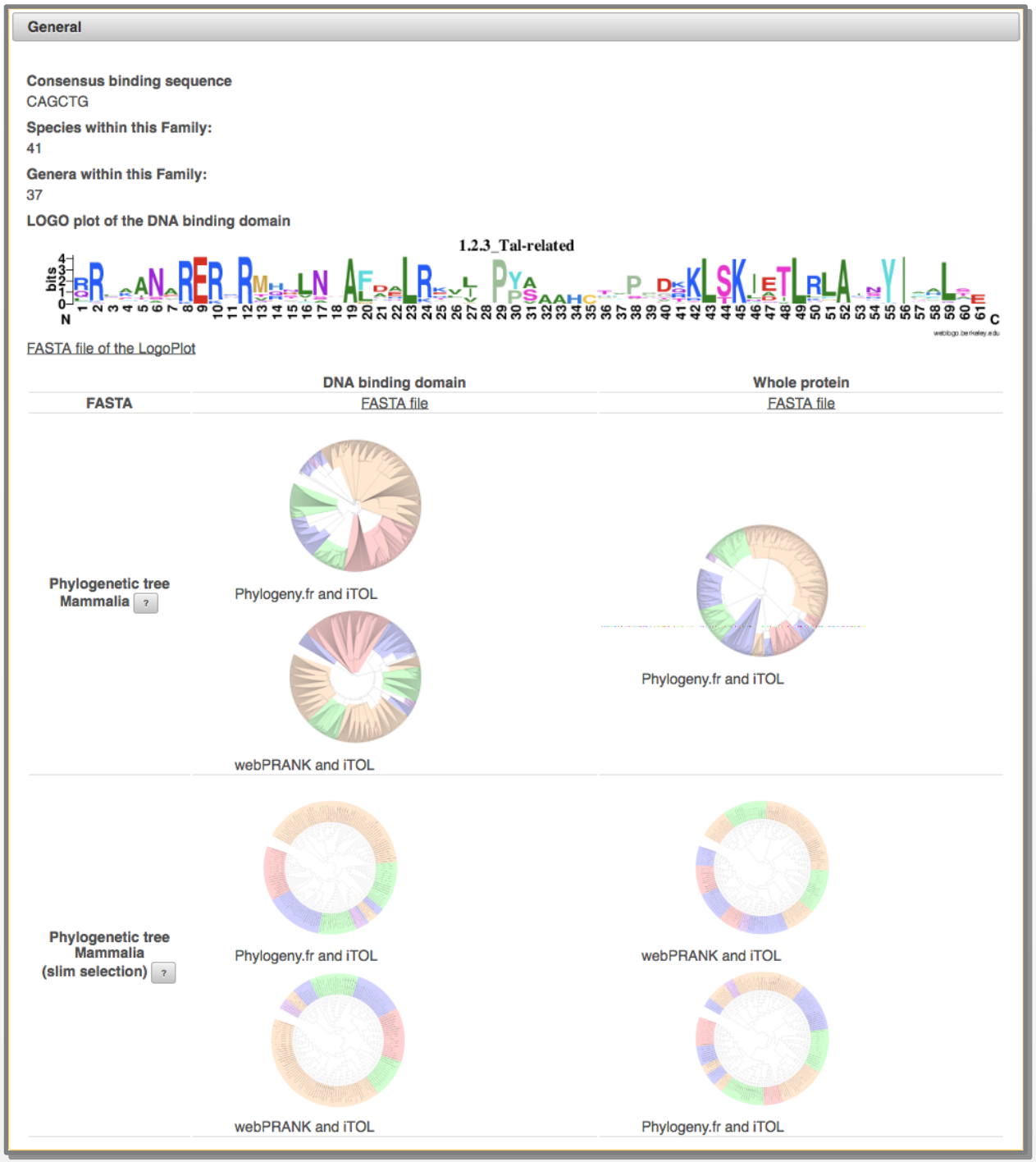
Phylogenetic analysis



Upstream nodes

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



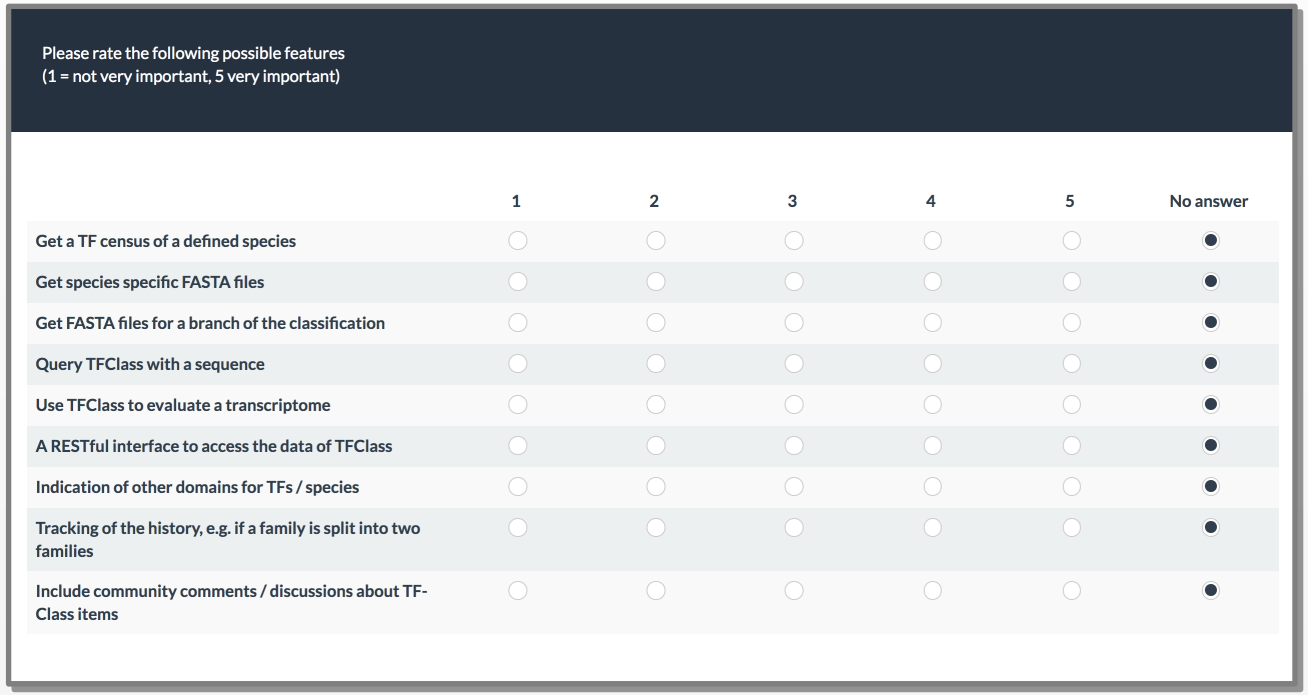


Future plans

- 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	○	○	○	○	○	●
Get species specific FASTA files	○	○	○	○	○	●
Get FASTA files for a branch of the classification	○	○	○	○	○	●
Query TFClass with a sequence	○	○	○	○	○	●
Use TFClass to evaluate a transcriptome	○	○	○	○	○	●
A RESTful interface to access the data of TFClass	○	○	○	○	○	●
Indication of other domains for TFs / species	○	○	○	○	○	●
Tracking of the history, e.g. if a family is split into two families	○	○	○	○	○	●
Include community comments / discussions about TF-Class items	○	○	○	○	○	●

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

Thank you!

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