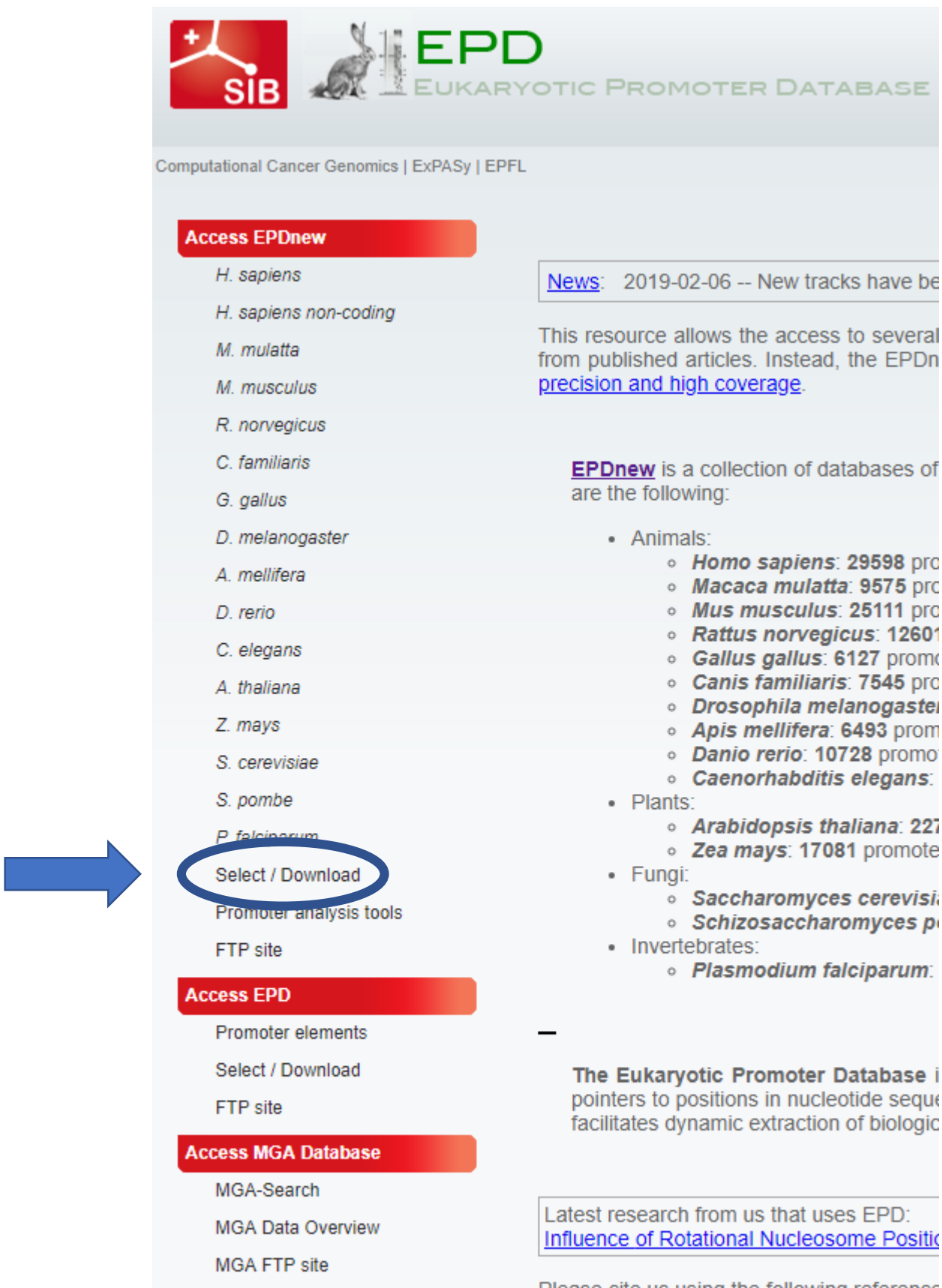


Navigate to <https://epd.epfl.ch//index.php>

Then click on Select / Download



The screenshot shows the EPD (Eukaryotic Promoter Database) website. At the top, there is a header with the SIB logo, a rabbit icon, and the text 'EPD EUKARYOTIC PROMOTER DATABASE'. Below the header, there is a navigation bar with links to 'Computational Cancer Genomics', 'ExPASy', and 'EPFL'.

The main content area is divided into three sections: 'Access EPDnew', 'Access EPD', and 'Access MGA Database'.

Access EPDnew section:

- Species list: *H. sapiens*, *H. sapiens non-coding*, *M. mulatta*, *M. musculus*, *R. norvegicus*, *C. familiaris*, *G. gallus*, *D. melanogaster*, *A. mellifera*, *D. rerio*, *C. elegans*, *A. thaliana*, *Z. mays*, *S. cerevisiae*, *S. pombe*, *P. falciparum*.
- Links: 'Select / Download' (circled in blue with an arrow pointing to it), 'Promoter analysis tools', and 'FTP site'.

Access EPD section:

- Links: 'Promoter elements', 'Select / Download', and 'FTP site'.

Access MGA Database section:

- Links: 'MGA-Search', 'MGA Data Overview', and 'MGA FTP site'.

News section:

News: 2019-02-06 -- New tracks have been added to the EPDnew database. This resource allows the access to several tracks of promoter elements from published articles. Instead, the EPDnew database provides [precision and high coverage](#).

EPDnew is a collection of databases of promoter elements. The following are the following:

- Animals:
 - Homo sapiens*: 29598 promoters
 - Macaca mulatta*: 9575 promoters
 - Mus musculus*: 25111 promoters
 - Rattus norvegicus*: 12601 promoters
 - Gallus gallus*: 6127 promoters
 - Canis familiaris*: 7545 promoters
 - Drosophila melanogaster*: 12601 promoters
 - Apis mellifera*: 6493 promoters
 - Danio rerio*: 10728 promoters
 - Caenorhabditis elegans*: 12601 promoters
- Plants:
 - Arabidopsis thaliana*: 227 promoters
 - Zea mays*: 17081 promoters
- Fungi:
 - Saccharomyces cerevisiae*: 12601 promoters
 - Schizosaccharomyces pombe*: 12601 promoters
- Invertebrates:
 - Plasmodium falciparum*: 12601 promoters

The Eukaryotic Promoter Database is a database of promoter elements. It provides pointers to positions in nucleotide sequences and facilitates dynamic extraction of biological information.

Latest research from us that uses EPD: [Influence of Rotational Nucleosome Positioning on Transcription](#)

Please cite us using the following reference:

Choose any combination of the four highlighted boxes and click 'select'

Select / Download tool

Use this tool to **select** promoters based on promoter name / ID or **liftOver** them to a different assembly or use them to perform full genome analysis

Database

Restrict the selection to the following IDs:

Enter one ID per line

Promoters with the following characteristics:

tags

samples

Additional options:

☐ Select only the most representative promoter for a gene

Note how many promoters are selected (circled)
Choose the start and end base to extract then hit 'submit' to get the fasta

Database:		Selection Parameters	
Database:	human_epdnew	TATA-box:	with
Assembly:	hg38	Initiator:	with
		CCAAT-box:	with
		GC-box:	without
		Marked as:	all
		Average expression:	
		Expressed in:	

Results: 120 promoters selected

[SGA file](#) [FPS file](#) [BED file](#)

LiftOver options

Sequence Extraction Tool (FASTA format)		Downstream Analysis	
From: <input type="text" value="50"/>	To: <input type="text" value="120"/>	Motif Enrichment	<input type="button" value="OProf"/> ?
	<input type="button" value="Submit"/>	Motif Discovery	<input type="button" value="FindM"/> ?
		Chromatin analysis	<input type="button" value="ChIP-Cor"/> ?

Now we need to align the sequences we got. Copy the entire fasta output and go to <https://www.genome.jp/tools-bin/clustalw>

Paste the output in the text box and select the DNA check box, then hit 'Execute Multiple Alignment'.

```
>FP000093 TUBB3_1 :+U EU:NC; range -499 to 100.
GGGCCAGCCTTTACCTACCTCCCCACCCAAACCGGCAAAAGCTCAGAGCACCTTGTCT
GCCAAAAGACAGGGAGCTGGGATGGTGCGGGTTGGTCTCTAAACCGGCGTGGGGAAAAA
GACCTCCGTACAAAGCCGACGGGTGGGGCTGTCGCAAGGGCGGAACCGAGAGGGTAGCT
GGGGGCGGGGTTCCAGGGCCAAGAGGGGCCATTGTCTCCCTGGAGCCCGGCGCCCCCA
CAGCCAGCTCCTCTGGGAGACAGCCCCCTCTTTCGAATGCGCGGGGCCCTCAGACCGCGC
CCGGCCAGCGCTGGGGGATCCTTGGCTGCGGGAGGGGCGCCGATTGCGCGCGCGGGC
GGGGACGCGCGGTGCGGAGCCTGCGGGCCGGGCGGGGCTCTGCGGCGGCGCTCCCGATT
GGCCACCCGCGGTGACATCAGCCGATGCGAAGGGCGGGGCGCGGCTATAAGAGCGCGCG
GCCGCGGTCCCGACCTCAGCAGCCAGCCCGGCCGCGCGCGCGCTCCGAGCCGCC
GCCAGACGCGCCAGTATGAGGGAGATCGTGACATCCAGGCCGCGCAGTGCGGCAACCA
>FP000052 STC1_1 :+U EU:NC; range -499 to 100.
ATGTACACACAGAGAAGATAGGGAGTTATTGCATTTGTAGCCTACAAAACAGAACCGAGA
ATGTGCTGTTAAATTAGAGTAAACTGCTGTAAGCAGGTTAAGTTCTCATCTAAAGAGA
TCACATTTCCCAACCATACCCCTGCTATCCATTTCCCCCAAGTGGCTCATTAGAAAAAA
GATGGCTAGATTTCAAAAAGCAACTTGGAGAGATTTCTATAGGATTTTCTTTAGTTCAA
TCAATACAGAGTTATCTCTTACTTCCACGAAATAGCTTTTTCACACATCTCTGCACACA
CAGTCACACACACATATAAACATTGGCAGCAGGTACTTTTAATTTGCTGGAAAATATTT
CTAAGAAGTCAAAAAGCTCCAGCTGAATTGCGATGCCCTCTTATTGGCTCACCAGACCA
TGAGGGACCTGATTGGTCTTGATCCTGAGGACCGATAAGAAGCGCTATAAAATCCCTGG
GTGCAGCTCTTGGGCCCCAGTTTGCAAAAGCCAGAGGTGCAAGAAGCAGCGACTGCAGC
AGCAGCAGCAGCAGCGGCGGTGGCAGCAGCAGCAGCAGCGGCGGCGAGCAGCAGCAGC
```



Multiple Sequence Alignment by CLUSTALW

ETE3	MAFFT	CLUSTALW	PRRN
<div>General Setting Parameters: Output Format: <input type="text" value="CLUSTAL"/> Pairwise Alignment: <input type="radio"/> FAST/APPROXIMATE <input checked="" type="radio"/> SLOW/ACCURATE Enter your sequences (with labels) below (copy & paste): <input type="radio"/> PROTEIN <input checked="" type="radio"/> DNA Support Formats: FASTA (Pearson), NBRF/PIR, EMBL/Swiss Prot, GDE, CLUSTAL, and GCG/MSF <div>>FP000093 MBD3L3_1 :+U EU:NC; range 0 to 100. ACTGCATTTTCCGGCAAGCCAAGGGTTGTCTGCATCTCAAGAGTGGGGTCAGCAAGAGAA ACTCTACGGCTATGGGAGAGCCTGCGTTACCTCTTTTCCG >FP000007 MYH4_1 :+U EU:NC; range 0 to 100.</div><div>Or give the file name containing your query Choose File No file chosen Execute Multiple Alignment Reset</div></div>			

Finally, we'll take our multiple sequence alignment and create a sequence Logo using WebLogo. Copy the multiple sequence alignment (like the example Below), then go to <https://weblogo.berkeley.edu/logo.cgi>

Paste the multiple sequence alignment into the box at the top, then select DNA/RNA, increase the bitmap resolution and hit create logo

clustalw.aln

CLUSTAL 2.1 multiple sequence alignment

```

FP000009 -----ACTGCATTTTCC---GGCAAGCCAAGGGTGTCTGCATCTC
FP000008 -----ACTGCATTTTCC---GGCAAGCCAAGGGTGTCTGCATCTC
FP000005 -----ACTCTGCCCTTTG---GACGTGAGAGAGAGCGCACCTTTAC
FP000003 -----ACTCTGCCCTTTG---GACGTGAGAGAGAGCGCACCTTTAC
FP000001 -----ATCTGCTCTGACTCCCAGGGACGTGTCTGCTCCTGCGTGTGAC
FP000006 AGCACAGTTGAGTCTCCAGCCTTGACTCTTC--TCAAGAGCCTGTGACTTTCCTC---C
FP000002 -----ATCATCTTGGTCATCAACACAACCTTGCTTCTCTCCAGACTTGGCT
FP000004 -----AGCCTCTCCAGCCCCAGCAAGCGACCTGTCAGGCGGCGTGGACT
FP000007 -----ATCCTTCTCTCAAAATTCTGAAGGTATGTATATGTG
FP000010 -----AGCAGACAGAGAGAGGAGTGTCTGGGACAGACTGCTCTC

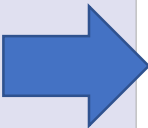
FP000009 AAGAGTGGG-GTCAGCAAGAGAACTCTACGGCTATGGGAGAGC-CTGCGTTACCTCTT
FP000008 AAGAGTGGG-GTCAGCAAGAGAACTCTACGGCTATGGGAGAGC-CTGCGTTACCTCTT
FP000005 TTGAGCT---TCAACATGGGAAAG-----GGAAATGAAGACCC-CGATCTCCACTGCTC
FP000003 TTGAGCT---TCAACATGGGAAAG-----GGAAATGAAGACTC-CGATCTCCACTGCTC
FP000001 CAGGGTGAGTGGCAACCTGGGATACCAGAGGGGTATGAGCAAGG-CAGAGGG-ATGG--
FP000006 CTGGACAAA-GGCATCATGAGTTGTC--AGATCTCTTGCAATC-TCGAGG-----
FP000002 TAAGGTA---CTGCGTTTACATACAGCAAAATTGCTATCATTT-TACATTATCTAATCT
FP000004 CAGACTC---CGGAGATGAAGCCCTGCTCCTGGCCGTCAGCC-TTGGCCTCATTGCTG
FP000007 GAAGAACAC--TTACTTTTACATTCTGATGATGATTTTTCATTTAAGGGATGTCTA
FP000010 GACAGAAGGTGCCAGGCTGGGG-GTGGCAGGCCCTGGGGGGGCTCTGGCCTGGGATGGAG

```

WEBLogo

· [about](#) · [create](#) · [examples](#) ·

Multiple Sequence Alignment



Upload Sequence Data: No file chosen

Image Format & Size

Image Format: Logo Size per Line: X

Advanced Logo Options

Sequence Type: ☐ amino acid ☒ DNA / RNA ☐ Automatic Detection

First Position Number: Logo Range: -

Small Sample Correction: ☒ Frequency Plot: ☐

Multiline Logo (Symbols per Line):

Advanced Image Options

Bitmap Resolution: ☒ 96 pixels/inch (dpi) ☐ Antialias Bitmaps: ☒

Title: Y-Axis Height: (bits)

Show Y-Axis: ☒ Y-Axis Label:

Show X-Axis: ☒ X-Axis Label:

Show Error Bars: ☐ Label Sequence Ends: ☒

Boxed / Boxed Shrink Factor: / Outline Symbols: ☐

Show fine print: ☒ Y-Axis Tic Spacing: (bits)

Colors

Color Scheme: ☒ Default ☐ Black & White ☐ Custom (See Below.)

Symbols	Color	RGB	Symbols	Color	RGB
KRH	green			purple	
DE	blue			orange	
AVLIPWFM	red			black	
	black		Other	black	

And there's a representative output! There are many parameters you can modify, for example if you have a very long sequence logo but you see something interesting, you can zoom in on that area by changing the Logo Range parameter. You can also modify the multiple alignment parameters to get slightly different alignment outputs. Play around and see how different combinations of promoters, alignment parameters and logo output parameters change your final image.

