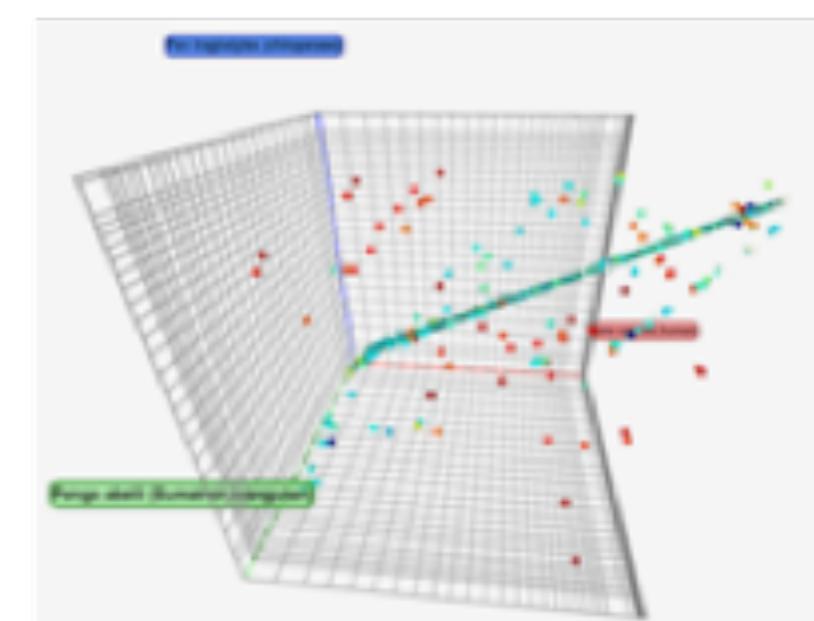
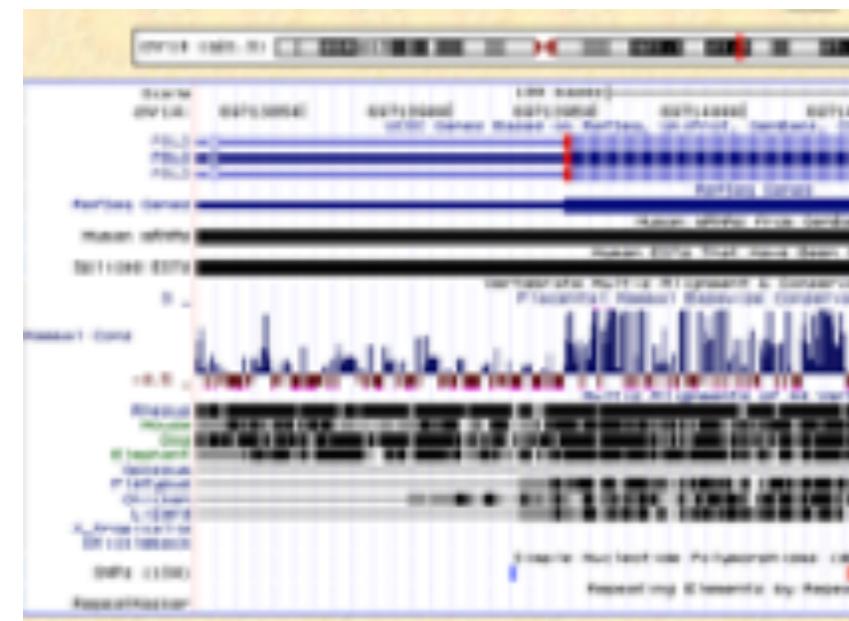


Computational Genomics

Introduction to Regular Expressions



Introduction to Regular Expressions

Regular expression

文 A 58 languages ▾

Article Talk

Read Edit View history

From Wikipedia, the free encyclopedia

"*Regex*" redirects here. For the comic book, see *Re:Gex*.

".*" redirects here. For the C++ operator, see [Pointer \(computer science\)](#) § [Pointer-to-member](#).

A **regular expression** (shortened as **regex** or **regexp**;^[1] sometimes referred to as **rational expression**^{[2][3]}) is a sequence of **characters** that specifies a **search pattern** in **text**. Usually such patterns are used by **string-searching algorithms** for "find" or "find and replace" operations on **strings**, or for **input validation**. Regular expression techniques are developed in **theoretical computer science** and **formal language theory**.

The concept of regular expressions began in the 1950s, when the American mathematician [Stephen Cole Kleene](#) formalized the concept of a **regular language**. They came into common use with [Unix](#) text-processing utilities. Different **syntaxes** for writing regular expressions have existed since the 1980s, one being the [POSIX](#) standard and another, widely used, being the [Perl](#) syntax.

Regular expressions are used in **search engines**, in search and replace dialogs of **word processors** and **text editors**, in **text processing** utilities such as [sed](#) and [AWK](#), and in **lexical analysis**. Regular expressions are supported in many programming languages.

In the beginning God created the heaven and the earth.
And the earth was without form, and void; and darkness was upon the face of the deep. And the Spirit of God moved upon the face of the waters.
And God said, Let there be light: and there was light.
And God saw the light, that it was good: and God divided the light from the darkness.

Blue highlights show the match results of the regular expression pattern `/h[aeiou]+/g` (the letter *h* followed by one or more vowels)

Introduction to Regular Expressions

Regular Expressions Syntax

The screenshot shows the homepage of the Rexegg website. At the top, there's a navigation bar with links to "Fundamentals", "Black Belt Program", "Regex in Action", "Humor & More", and "Ask Rex". The main content area features a large orange T-Rex logo on the left and the text "The world's most tyrannosauical regex tutorial" on the right. A complex regular expression pattern is displayed above the title "REX EGG". The central section is titled "Quick-Start: Regex Cheat Sheet" and contains several paragraphs of text explaining the purpose and use of the cheat sheet. On the right side, there's a sidebar with a "must-read" section about RegexBuddy 4, a "Ask Rex" section with a T-Rex icon, and a search bar.

Quick-Start: Regex Cheat Sheet

PROTECTED BY COPYSCAPE DO NOT COPY

The tables below are a reference to basic regex. While reading the rest of the site, when in doubt, you can always come back and look here. (If you want a bookmark, here's a direct link to the [regex reference tables](#)). I encourage you to print the tables so you have a cheat sheet on your desk for quick reference.

The tables are not exhaustive, for two reasons. First, every regex flavor is different, and I didn't want to crowd the page with overly exotic syntax. For a full reference to the particular regex flavors you'll be using, it's always best to go straight to the source. In fact, for some regex engines (such as Perl, PCRE, Java and .NET) you may want to check once a year, as their creators often introduce new features.

The other reason the tables are not exhaustive is that I wanted them to serve as a quick introduction to regex. If you are a complete beginner, you should get a firm grasp of basic regex syntax just by reading the examples in the tables. I tried to introduce features in a logical order and to keep out oddities that I've never seen in actual use, such as the "bell character". With these tables as a jumping board, you will be able to advance to mastery by exploring the other pages on the site.

How to use the tables

The tables are meant to serve as an accelerated regex course, and they are meant to be read slowly, one line at a time. On each line, in the leftmost column, you will find a new element of regex syntax. The next column, "Legend", explains what the element means (or encodes) in the regex syntax. The next two columns work hand in hand: the "Example" column gives a valid regular expression that uses the element, and the "Sample Match" column presents a text string that could be matched by the regular expression.

You can read the tables online, of course, but if you suffer from even the mildest case of online-ADD (attention deficit disorder), like most of us... Well then, I highly recommend you print them out. You'll be able to study them slowly, and to use them as a cheat sheet later, when you are reading the rest of the site or experimenting with your own regular

Introduction to Regular Expressions

Go to [regexr.com](https://regextester.com)

The screenshot shows the RegExr interface. On the left is a sidebar with links: Menu, Pattern Settings, My Patterns, Cheatsheet, RegEx Reference, Community Patterns, and Help. The main area has tabs for Expression, Text, and Tests (which is selected). The expression entered is `/([A-Z])\w+/g`. The text area contains two paragraphs of explanatory text. A status bar at the bottom right says "29 matches (0.2ms)".

RegExr is an online tool to **learn, build, & test** Regular Expressions (RegEx / RegExp).

- Supports **JavaScript & PHP/PCRE** RegEx.
- Results update in **real-time** as you type.
- Roll over** a match or expression for details.
- Validate patterns with suites of **Tests**.
- Save & share expressions with others.
- Use **Tools** to explore your results.
- Full **RegEx Reference** with help & examples.
- Undo & Redo with cmd-Z / Y in editors.
- Search for & rate **Community Patterns**.

Tools (Replace, List, Details, Explain)

Roll-over elements below to highlight in the Expression above. Click to open in Reference.

(Capturing group #1. Groups multiple tokens together and creates a capture group for extracting a substring or using a backreference.)

[Character set. Match any character in the set.]

A-Z Range. Matches a character in the range "A" to "Z" (char code 65 to 90). Case sensitive.

)

\w Word. Matches any word character (alphanumeric & underscore).

+ Quantifier. Match 1 or more of the preceding token.

Introduction to Regular Expressions

Select PCRE and Flags: global, multiline and extended

The screenshot shows the RegExr interface. On the left is a sidebar with links: Menu, Pattern Settings, My Patterns, Cheatsheet, RegEx Reference, Community Patterns, and Help. The main area has tabs for Expression, Text, and Tests. The Tests tab is active, showing the regular expression `/([A-Z])\w+/gm` applied to a block of text about RegExr. A sidebar on the right lists Expression Flags: global (checked), case insensitive (unchecked), multiline (checked), single line (dotall) (unchecked), extended (checked), and ungreedy (checked). Below the text area, there's a section titled 'Tools' with buttons for Replace, List, Details, and Explain.

RegExr is an online tool to **learn, build, & test** Regular Expressions (RegEx / RegExp).

- Supports **JavaScript & PHP/PCRE** RegEx.
- Results update in **real-time** as you type.
- Roll over** a match or expression for details.
- Validate patterns with suites of **Tests**.
- Save & share expressions with others.
- Use **Tools** to explore your results.
- Full **RegEx Reference** with help & examples.
- Undo & Redo** with cmd-Z / Y in editors.
- Search for & rate **Community Patterns**.

Tools

Roll-over elements below to highlight in the Expression above. Click to open in Reference.

(Capturing group #1. Groups multiple tokens together and creates a capture group for extracting a substring or using a backreference.

[Character set. Match any character in the set.

 A-Z Range. Matches a character in the range "A" to "Z" (char code 65 to 90). Case sensitive.

]

)

\w Word. Matches any word character (alphanumeric & underscore).

+ Quantifier. Match 1 or more of the preceding token.

Introduction to Regular Expressions

A Regular Expression is a pattern describing a certain amount of text

The () { } [] . * ? + ^ \$ are all special characters

\ can be used to “escape” a special character, allowing that special characters (i.e., () { } [] . * ? + ^ \$), to be searched for

chr1	65564	65573	ENST00000641515.2_cds_1_0_chr1_65565_f	0	+
chr1	69036	70008	ENST00000641515.2_cds_2_0_chr1_69037_f	0	+
chrX	284187	284314	ENST00000429181.6_cds_2_0_chrX_284188_f	0	+
chrX	288732	288787	ENST00000429181.6_cds_3_0_chrX_288733_f	0	+
chrY	284187	284314	ENST00000429181.6_cds_2_0_chrY_284188_f	0	+
chrY	288732	288787	ENST00000429181.6_cds_3_0_chrY_288733_f	0	+
chr10	48054	48114	ENST00000562809.1_cds_2_0_chr10_48055_r	0	-
chr10	48614	48725	ENST00000562809.1_cds_1_0_chr10_48615_r	0	-
chr11	168957	169052	ENST00000410108.5_cds_4_0_chr11_168958_r	0	-
chr11	180208	180404	ENST00000410108.5_cds_3_0_chr11_180209_r	0	-
chr12	66882	67436	ENST00000538872.6_cds_0_0_chr12_66883_f	0	+
chr12	99145	99214	ENST00000538872.6_cds_1_0_chr12_99146_f	0	+

Introduction to Regular Expressions

Expression ↗ PCRE ▾ 🏴 Flags ▾

/([A-Z])\w+/gmx

Text Tests 16 matches (0.1ms)

```
chr1→|65564→|65573→|ENST00000641515.2_cds_1_0_chr1_65565_f→|0→|+→
chr1→|69036→|70008→|ENST00000641515.2_cds_2_0_chr1_69037_f→|0→|+→
chrX→|284187→|284314→|ENST00000429181.6_cds_2_0_chrX_284188_f→|0→|+→
chrX→|288732→|288787→|ENST00000429181.6_cds_3_0_chrX_288733_f→|0→|+→
chrY→|284187→|284314→|ENST00000429181.6_cds_2_0_chrY_284188_f→|0→|+→
chrY→|288732→|288787→|ENST00000429181.6_cds_3_0_chrY_288733_f→|0→|+→
chr10→|48054→|48114→|ENST00000562809.1_cds_2_0_chr10_48055_r→|0→|!-→
chr10→|48614→|48725→|ENST00000562809.1_cds_1_0_chr10_48615_r→|0→|!-→
chr11→|168957→|169052→|ENST00000410108.5_cds_4_0_chr11_168958_r→|0→|!-→
chr11→|180208→|180404→|ENST00000410108.5_cds_3_0_chr11_180209_r→|0→|!-→
chr12→|66882→|67436→|ENST00000538872.6_cds_0_0_chr12_66883_f→|0→|+→
chr12→|99145→|99214→|ENST00000538872.6_cds_1_0_chr12_99146_f→|0→|+→
```

Introduction to Regular Expressions

“\A” matches the beginning of a string
(but not an internal line)

Regex: \Achr1

Expression PCRE ▾ Flags ▾

/\Achr1/gmx

Text Tests 1 match (0.0ms)

```
chr1→65564→65573→ENST00000641515.2_cds_1_0_chr1_65565_f→0→+→
chr1→69036→70008→ENST00000641515.2_cds_2_0_chr1_69037_f→0→+→
chrX→284187→284314→ENST00000429181.6_cds_2_0_chrX_284188_f→0→+→
chrX→288732→288787→ENST00000429181.6_cds_3_0_chrX_288733_f→0→+→
chrY→284187→284314→ENST00000429181.6_cds_2_0_chrY_284188_f→0→+→
chrY→288732→288787→ENST00000429181.6_cds_3_0_chrY_288733_f→0→+→
chr10→48054→48114→ENST00000562809.1_cds_2_0_chr10_48055_r→0→!→
chr10→48614→48725→ENST00000562809.1_cds_1_0_chr10_48615_r→0→!→
chr11→168957→169052→ENST00000410108.5_cds_4_0_chr11_168958_r→0→!→
chr11→180208→180404→ENST00000410108.5_cds_3_0_chr11_180209_r→0→!→
chr12→66882→67436→ENST00000538872.6_cds_0_0_chr12_66883_f→0→+→
chr12→99145→99214→ENST00000538872.6_cds_1_0_chr12_99146_f→0→+→
```

Introduction to Regular Expressions

“\d” matches a digit class, same as [0-9]

Regex: chr\d

Expression /chr\d/gmx

PCRE Flags

Text Tests 16 matches (0.1ms)

```
chr1→65564→65573→ENST00000641515.2_cds_1_0_chr1_65565_f→0→+→
chr1→69036→70008→ENST00000641515.2_cds_2_0_chr1_69037_f→0→+→
chrX→284187→284314→ENST00000429181.6_cds_2_0_chrX_284188_f→0→+→
chrX→288732→288787→ENST00000429181.6_cds_3_0_chrX_288733_f→0→+→
chrY→284187→284314→ENST00000429181.6_cds_2_0_chrY_284188_f→0→+→
chrY→288732→288787→ENST00000429181.6_cds_3_0_chrY_288733_f→0→+→
chr10→48054→48114→ENST00000562809.1_cds_2_0_chr10_48055_r→0→!→
chr10→48614→48725→ENST00000562809.1_cds_1_0_chr10_48615_r→0→!→
chr11→168957→169052→ENST00000410108.5_cds_4_0_chr11_168958_r→0→!→
chr11→180208→180404→ENST00000410108.5_cds_3_0_chr11_180209_r→0→!→
chr12→66882→67436→ENST00000538872.6_cds_0_0_chr12_66883_f→0→+→
chr12→99145→99214→ENST00000538872.6_cds_1_0_chr12_99146_f→0→+→
```

Introduction to Regular Expressions

“\d” matches a digit class, same as [0-9]

Regex: chr\d+

Expression /chr\d+/gmx PCRE Flags

Text Tests 16 matches (0.1ms)

```
chr1→65564→65573→ENST00000641515.2_cds_1_0_chr1_65565_f→0→+→
chr1→69036→70008→ENST00000641515.2_cds_2_0_chr1_69037_f→0→+→
chrX→284187→284314→ENST00000429181.6_cds_2_0_chrX_284188_f→0→+→
chrX→288732→288787→ENST00000429181.6_cds_3_0_chrX_288733_f→0→+→
chrY→284187→284314→ENST00000429181.6_cds_2_0_chrY_284188_f→0→+→
chrY→288732→288787→ENST00000429181.6_cds_3_0_chrY_288733_f→0→+→
chr10→48054→48114→ENST00000562809.1_cds_2_0_chr10_48055_r→0→!→
chr10→48614→48725→ENST00000562809.1_cds_1_0_chr10_48615_r→0→!→
chr11→168957→169052→ENST00000410108.5_cds_4_0_chr11_168958_r→0→!→
chr11→180208→180404→ENST00000410108.5_cds_3_0_chr11_180209_r→0→!→
chr12→66882→67436→ENST00000538872.6_cds_0_0_chr12_66883_f→0→+→
chr12→99145→99214→ENST00000538872.6_cds_1_0_chr12_99146_f→0→+→
```

Introduction to Regular Expressions

“\t” matches a tab

Regex: \t

Expression PCRE ▾ Flags ▾

/\t/gmx

Text Tests 60 matches (0.2ms)

```
chr1→65564→65573→ENST00000641515.2_cds_1_0_chr1_65565_f→0→+→
chr1→69036→70008→ENST00000641515.2_cds_2_0_chr1_69037_f→0→+→
chrX→284187→284314→ENST00000429181.6_cds_2_0_chrX_284188_f→0→+→
chrX→288732→288787→ENST00000429181.6_cds_3_0_chrX_288733_f→0→+→
chrY→284187→284314→ENST00000429181.6_cds_2_0_chrY_284188_f→0→+→
chrY→288732→288787→ENST00000429181.6_cds_3_0_chrY_288733_f→0→+→
chr10→48054→48114→ENST00000562809.1_cds_2_0_chr10_48055_r→0→-→
chr10→48614→48725→ENST00000562809.1_cds_1_0_chr10_48615_r→0→-→
chr11→168957→169052→ENST00000410108.5_cds_4_0_chr11_168958_r→0→-→
chr11→180208→180404→ENST00000410108.5_cds_3_0_chr11_180209_r→0→-→
chr12→66882→67436→ENST00000538872.6_cds_0_0_chr12_66883_f→0→+→
chr12→99145→99214→ENST00000538872.6_cds_1_0_chr12_99146_f→0→+→
```

Introduction to Regular Expressions

“\t” matches a tab

Regex: \t+

Expression PCRE Flags

/\t+/gmx

Text Tests 60 matches (0.2ms)

```
chr1→65564→65573→ENST00000641515.2_cds_1_0_chr1_65565_f→0→+→
chr1→69036→70008→ENST00000641515.2_cds_2_0_chr1_69037_f→0→+→
chrX→284187→284314→ENST00000429181.6_cds_2_0_chrX_284188_f→0→+→
chrX→288732→288787→ENST00000429181.6_cds_3_0_chrX_288733_f→0→+→
chrY→284187→284314→ENST00000429181.6_cds_2_0_chrY_284188_f→0→+→
chrY→288732→288787→ENST00000429181.6_cds_3_0_chrY_288733_f→0→+→
chr10→48054→48114→ENST00000562809.1_cds_2_0_chr10_48055_r→0→-→
chr10→48614→48725→ENST00000562809.1_cds_1_0_chr10_48615_r→0→-→
chr11→168957→169052→ENST00000410108.5_cds_4_0_chr11_168958_r→0→-→
chr11→180208→180404→ENST00000410108.5_cds_3_0_chr11_180209_r→0→-→
chr12→66882→67436→ENST00000538872.6_cds_0_0_chr12_66883_f→0→+→
chr12→99145→99214→ENST00000538872.6_cds_1_0_chr12_99146_f→0→+→
```

Introduction to Regular Expressions

“\d” matches a digit class, same as [0-9]

Regex: \t\d+\t

Expression ✖ PCRE ▾ 

/\t\d+\t/gmx

Text Tests 24 matches (0.1ms)

```
chr1→65564→65573→ENST00000641515.2_cds_1_0_chr1_65565_f→0→+→
chr1→69036→70008→ENST00000641515.2_cds_2_0_chr1_69037_f→0→+→
chrX→284187→284314→ENST00000429181.6_cds_2_0_chrX_284188_f→0→+→
chrX→288732→288787→ENST00000429181.6_cds_3_0_chrX_288733_f→0→+→
chrY→284187→284314→ENST00000429181.6_cds_2_0_chrY_284188_f→0→+→
chrY→288732→288787→ENST00000429181.6_cds_3_0_chrY_288733_f→0→+→
chr10→48054→48114→ENST00000562809.1_cds_2_0_chr10_48055_r→0→-→
chr10→48614→48725→ENST00000562809.1_cds_1_0_chr10_48615_r→0→-→
chr11→168957→169052→ENST00000410108.5_cds_4_0_chr11_168958_r→0→-→
chr11→180208→180404→ENST00000410108.5_cds_3_0_chr11_180209_r→0→-→
chr12→66882→67436→ENST00000538872.6_cds_0_0_chr12_66883_f→0→+→
chr12→99145→99214→ENST00000538872.6_cds_1_0_chr12_99146_f→0→+→
```

Introduction to Regular Expressions

“\D” matches a non-digit

Regex: \D

Expression PCRE Flags

/\D/gmx

Text Tests 344 matches (1.3ms)

```
chr1→65564→65573→ENST00000641515.2_cds_1_0_chr1_65565_f→0→+→
chr1→69036→70008→ENST00000641515.2_cds_2_0_chr1_69037_f→0→+→
chrX→284187→284314→ENST00000429181.6_cds_2_0_chrX_284188_f→0→+→
chrX→288732→288787→ENST00000429181.6_cds_3_0_chrX_288733_f→0→+→
chrY→284187→284314→ENST00000429181.6_cds_2_0_chrY_284188_f→0→+→
chrY→288732→288787→ENST00000429181.6_cds_3_0_chrY_288733_f→0→+→
chr10→48054→48114→ENST00000562809.1_cds_2_0_chr10_48055_r→0→-→
chr10→48614→48725→ENST00000562809.1_cds_1_0_chr10_48615_r→0→-→
chr11→168957→169052→ENST00000410108.5_cds_4_0_chr11_168958_r→0→-→
chr11→180208→180404→ENST00000410108.5_cds_3_0_chr11_180209_r→0→-→
chr12→66882→67436→ENST00000538872.6_cds_0_0_chr12_66883_f→0→+→
chr12→99145→99214→ENST00000538872.6_cds_1_0_chr12_99146_f→0→+→
```

Introduction to Regular Expressions

“\s” matches a whitespace character

Regex: \s

Expression PCRE ▾ Flags ▾

/\s/gmx

Text Tests 72 matches (0.4ms)

```
chr1→65564→65573→ENST00000641515.2_cds_1_0_chr1_65565_f→0→+→
chr1→69036→70008→ENST00000641515.2_cds_2_0_chr1_69037_f→0→+→
chrX→284187→284314→ENST00000429181.6_cds_2_0_chrX_284188_f→0→+→
chrX→288732→288787→ENST00000429181.6_cds_3_0_chrX_288733_f→0→+→
chrY→284187→284314→ENST00000429181.6_cds_2_0_chrY_284188_f→0→+→
chrY→288732→288787→ENST00000429181.6_cds_3_0_chrY_288733_f→0→+→
chr10→48054→48114→ENST00000562809.1_cds_2_0_chr10_48055_r→0→-→
chr10→48614→48725→ENST00000562809.1_cds_1_0_chr10_48615_r→0→-→
chr11→168957→169052→ENST00000410108.5_cds_4_0_chr11_168958_r→0→-→
chr11→180208→180404→ENST00000410108.5_cds_3_0_chr11_180209_r→0→-→
chr12→66882→67436→ENST00000538872.6_cds_0_0_chr12_66883_f→0→+→
chr12→99145→99214→ENST00000538872.6_cds_1_0_chr12_99146_f→0→+→
```

Introduction to Regular Expressions

“\s” matches a whitespace character

Regex: \s+

Expression /**\s+**/gm

PCRE Flags

Text Tests 72 matches (0.4ms)

```
chr1→65564→65573→ENST00000641515.2_cds_1_0_chr1_65565_f→0→+→
chr1→69036→70008→ENST00000641515.2_cds_2_0_chr1_69037_f→0→+→
chrX→284187→284314→ENST00000429181.6_cds_2_0_chrX_284188_f→0→+→
chrX→288732→288787→ENST00000429181.6_cds_3_0_chrX_288733_f→0→+→
chrY→284187→284314→ENST00000429181.6_cds_2_0_chrY_284188_f→0→+→
chrY→288732→288787→ENST00000429181.6_cds_3_0_chrY_288733_f→0→+→
chr10→48054→48114→ENST00000562809.1_cds_2_0_chr10_48055_r→0→-→
chr10→48614→48725→ENST00000562809.1_cds_1_0_chr10_48615_r→0→-→
chr11→168957→169052→ENST00000410108.5_cds_4_0_chr11_168958_r→0→-→
chr11→180208→180404→ENST00000410108.5_cds_3_0_chr11_180209_r→0→-→
chr12→66882→67436→ENST00000538872.6_cds_0_0_chr12_66883_f→0→+→
chr12→99145→99214→ENST00000538872.6_cds_1_0_chr12_99146_f→0→+→
```

Introduction to Regular Expressions

“\S” matches anything BUT a whitespace

Regex: \S

Expression: /**\s**/gm

PCRE Flags

Text Tests 678 matches (2.9ms)

```
chr1→65564→65573→ENST00000641515.2_cds_1_0_chr1_65565_f→0→+→
chr1→69036→70008→ENST00000641515.2_cds_2_0_chr1_69037_f→0→+→
chrX→284187→284314→ENST00000429181.6_cds_2_0_chrX_284188_f→0→+→
chrX→288732→288787→ENST00000429181.6_cds_3_0_chrX_288733_f→0→+→
chrY→284187→284314→ENST00000429181.6_cds_2_0_chrY_284188_f→0→+→
chrY→288732→288787→ENST00000429181.6_cds_3_0_chrY_288733_f→0→+→
chr10→48054→48114→ENST00000562809.1_cds_2_0_chr10_48055_r→0→-→
chr10→48614→48725→ENST00000562809.1_cds_1_0_chr10_48615_r→0→-→
chr11→168957→169052→ENST00000410108.5_cds_4_0_chr11_168958_r→0→-→
chr11→180208→180404→ENST00000410108.5_cds_3_0_chr11_180209_r→0→-→
chr12→66882→67436→ENST00000538872.6_cds_0_0_chr12_66883_f→0→+→
chr12→99145→99214→ENST00000538872.6_cds_1_0_chr12_99146_f→0→+→
```

Introduction to Regular Expressions

“\w” matches an alphanumeric character

Regex: \w

Expression PCRE ▾ Flags ▾

/\w/gmx

Text Tests 654 matches (2.4ms)

```
chr1→65564→65573→ENST00000641515.2_cds_1_0_chr1_65565_f→0→+→
chr1→69036→70008→ENST00000641515.2_cds_2_0_chr1_69037_f→0→+→
chrX→284187→284314→ENST00000429181.6_cds_2_0_chrX_284188_f→0→+→
chrX→288732→288787→ENST00000429181.6_cds_3_0_chrX_288733_f→0→+→
chrY→284187→284314→ENST00000429181.6_cds_2_0_chrY_284188_f→0→+→
chrY→288732→288787→ENST00000429181.6_cds_3_0_chrY_288733_f→0→+→
chr10→48054→48114→ENST00000562809.1_cds_2_0_chr10_48055_r→0→!→
chr10→48614→48725→ENST00000562809.1_cds_1_0_chr10_48615_r→0→!→
chr11→168957→169052→ENST00000410108.5_cds_4_0_chr11_168958_r→0→!→
chr11→180208→180404→ENST00000410108.5_cds_3_0_chr11_180209_r→0→!→
chr12→66882→67436→ENST00000538872.6_cds_0_0_chr12_66883_f→0→+→
chr12→99145→99214→ENST00000538872.6_cds_1_0_chr12_99146_f→0→+→
```

Introduction to Regular Expressions

“\w” matches an alphanumeric character

Regex: \w+

Expression PCRE ▾ Flags ▾

/\w+/gmx

Text Tests 72 matches (0.3ms)

```
chr1→|65564→|65573→|ENST00000641515.2_cds_1_0_chr1_65565_f→|0→|+→
chr1→|69036→|70008→|ENST00000641515.2_cds_2_0_chr1_69037_f→|0→|+→
chrX→|284187→|284314→|ENST00000429181.6_cds_2_0_chrX_284188_f→|0→|+→
chrX→|288732→|288787→|ENST00000429181.6_cds_3_0_chrX_288733_f→|0→|+→
chrY→|284187→|284314→|ENST00000429181.6_cds_2_0_chrY_284188_f→|0→|+→
chrY→|288732→|288787→|ENST00000429181.6_cds_3_0_chrY_288733_f→|0→|+→
chr10→|48054→|48114→|ENST00000562809.1_cds_2_0_chr10_48055_r→|0→|!-→
chr10→|48614→|48725→|ENST00000562809.1_cds_1_0_chr10_48615_r→|0→|!-→
chr11→|168957→|169052→|ENST00000410108.5_cds_4_0_chr11_168958_r→|0→|!-→
chr11→|180208→|180404→|ENST00000410108.5_cds_3_0_chr11_180209_r→|0→|!-→
chr12→|66882→|67436→|ENST00000538872.6_cds_0_0_chr12_66883_f→|0→|+→
chr12→|99145→|99214→|ENST00000538872.6_cds_1_0_chr12_99146_f→|0→|+→
```

Introduction to Regular Expressions

“\W” matches anything but an alphanumeric character
Regex: \W

Expression PCRE Flags

/\W/gmx

Text Tests 96 matches (0.4ms)

```
chr1→65564→65573→ENST00000641515.2_cds_1_0_chr1_65565_f→0→+→
chr1→69036→70008→ENST00000641515.2_cds_2_0_chr1_69037_f→0→+→
chrX→284187→284314→ENST00000429181.6_cds_2_0_chrX_284188_f→0→+→
chrX→288732→288787→ENST00000429181.6_cds_3_0_chrX_288733_f→0→+→
chrY→284187→284314→ENST00000429181.6_cds_2_0_chrY_284188_f→0→+→
chrY→288732→288787→ENST00000429181.6_cds_3_0_chrY_288733_f→0→+→
chr10→48054→48114→ENST00000562809.1_cds_2_0_chr10_48055_r→0→-→
chr10→48614→48725→ENST00000562809.1_cds_1_0_chr10_48615_r→0→-→
chr11→168957→169052→ENST00000410108.5_cds_4_0_chr11_168958_r→0→-→
chr11→180208→180404→ENST00000410108.5_cds_3_0_chr11_180209_r→0→-→
chr12→66882→67436→ENST00000538872.6_cds_0_0_chr12_66883_f→0→+→
chr12→99145→99214→ENST00000538872.6_cds_1_0_chr12_99146_f→0→+→
```

Introduction to Regular Expressions

“\Z” matches the end of a string
(but not a internal line)

Regex: \+\Z

Expression ↗ PCRE ▾ 🏴 Flags ▾

/\+\Z/gmx

Text Tests 1 match (0.0ms)

```
chr1→65564→65573→ENST00000641515.2_cds_1_0_chr1_65565_f→0→+→
chr1→69036→70008→ENST00000641515.2_cds_2_0_chr1_69037_f→0→+→
chrX→284187→284314→ENST00000429181.6_cds_2_0_chrX_284188_f→0→+→
chrX→288732→288787→ENST00000429181.6_cds_3_0_chrX_288733_f→0→+→
chrY→284187→284314→ENST00000429181.6_cds_2_0_chrY_284188_f→0→+→
chrY→288732→288787→ENST00000429181.6_cds_3_0_chrY_288733_f→0→+→
chr10→48054→48114→ENST00000562809.1_cds_2_0_chr10_48055_r→0→!→
chr10→48614→48725→ENST00000562809.1_cds_1_0_chr10_48615_r→0→!→
chr11→168957→169052→ENST00000410108.5_cds_4_0_chr11_168958_r→0→!→
chr11→180208→180404→ENST00000410108.5_cds_3_0_chr11_180209_r→0→!→
chr12→66882→67436→ENST00000538872.6_cds_0_0_chr12_66883_f→0→+→
chr12→99145→99214→ENST00000538872.6_cds_1_0_chr12_99146_f→0→+→
```

Introduction to Regular Expressions

“{n}” The preceding item is matched exactly n times
Regex: 6{2}

Expression /6{2}/gm

PCRE Flags

Text Tests 2 matches (0.1ms)

```
chr1→65564→65573→ENST00000641515.2_cds_1_0_chr1_65565_f→0→+→
chr1→69036→70008→ENST00000641515.2_cds_2_0_chr1_69037_f→0→+→
chrX→284187→284314→ENST00000429181.6_cds_2_0_chrX_284188_f→0→+→
chrX→288732→288787→ENST00000429181.6_cds_3_0_chrX_288733_f→0→+→
chrY→284187→284314→ENST00000429181.6_cds_2_0_chrY_284188_f→0→+→
chrY→288732→288787→ENST00000429181.6_cds_3_0_chrY_288733_f→0→+→
chr10→48054→48114→ENST00000562809.1_cds_2_0_chr10_48055_r→0→!-→
chr10→48614→48725→ENST00000562809.1_cds_1_0_chr10_48615_r→0→!-→
chr11→168957→169052→ENST00000410108.5_cds_4_0_chr11_168958_r→0→!-→
chr11→180208→180404→ENST00000410108.5_cds_3_0_chr11_180209_r→0→!-→
chr12→66882→67436→ENST00000538872.6_cds_0_0_chr12_66883_f→0→+→
chr12→99145→99214→ENST00000538872.6_cds_1_0_chr12_99146_f→0→+→
```

Introduction to Regular Expressions

“{n,}” The preceding item is matched n or more times
Regex: 6{1,}

Expression /6{1,}/gmx

PCRE Flags

Text Tests 28 matches (0.1ms)

```
chr1→65564→65573→ENST00000641515.2_cds_1_0_chr1_65565_f→0→+→
chr1→69036→70008→ENST00000641515.2_cds_2_0_chr1_69037_f→0→+→
chrX→284187→284314→ENST00000429181.6_cds_2_0_chrX_284188_f→0→+→
chrX→288732→288787→ENST00000429181.6_cds_3_0_chrX_288733_f→0→+→
chrY→284187→284314→ENST00000429181.6_cds_2_0_chrY_284188_f→0→+→
chrY→288732→288787→ENST00000429181.6_cds_3_0_chrY_288733_f→0→+→
chr10→48054→48114→ENST00000562809.1_cds_2_0_chr10_48055_r→0→!→
chr10→48614→48725→ENST00000562809.1_cds_1_0_chr10_48615_r→0→!→
chr11→168957→169052→ENST00000410108.5_cds_4_0_chr11_168958_r→0→!→
chr11→180208→180404→ENST00000410108.5_cds_3_0_chr11_180209_r→0→!→
chr12→66882→67436→ENST00000538872.6_cds_0_0_chr12_66883_f→0→+→
chr12→99145→99214→ENST00000538872.6_cds_1_0_chr12_99146_f→0→+→
```

Introduction to Regular Expressions

“[...]” creates a character class
Within the brackets, single characters can be placed
A dash (-) may be used to indicate a range such as a-z
Regex: [a-z]

Expression / [a-z]/gmx PCRE Flags

Text Tests 120 matches (0.5ms)

```
chr1→65564→65573→ENST00000641515.2_cds_1_0_chr1_65565_f→0→+→
chr1→69036→70008→ENST00000641515.2_cds_2_0_chr1_69037_f→0→+→
chrX→284187→284314→ENST00000429181.6_cds_2_0_chrX_284188_f→0→+→
chrX→288732→288787→ENST00000429181.6_cds_3_0_chrX_288733_f→0→+→
chrY→284187→284314→ENST00000429181.6_cds_2_0_chrY_284188_f→0→+→
chrY→288732→288787→ENST00000429181.6_cds_3_0_chrY_288733_f→0→+→
chr10→48054→48114→ENST00000562809.1_cds_2_0_chr10_48055_r→0→!-→
chr10→48614→48725→ENST00000562809.1_cds_1_0_chr10_48615_r→0→!-→
chr11→168957→169052→ENST00000410108.5_cds_4_0_chr11_168958_r→0→!-→
chr11→180208→180404→ENST00000410108.5_cds_3_0_chr11_180209_r→0→!-→
chr12→66882→67436→ENST00000538872.6_cds_0_0_chr12_66883_f→0→+→
chr12→99145→99214→ENST00000538872.6_cds_1_0_chr12_99146_f→0→+→
```

Introduction to Regular Expressions

“.” Matches any single character except a newline

“*” The preceding item will be matched zero or more times

“?” The preceding item is optional and matched at most once

“+” The preceding item will be matched one or more time

“^” has two meanings:

matches the beginning of a line or string
indicates negation in a character class

For example, [^...] matches every character except the ones inside brackets

“\$” matches the end of a line or string

“|” Separates alternate possibilities

“(. .)” groups a particular pattern

Regex: ^(chr.)\t([0-9]+)\t+([0-9]+)\t+ENST([0-9]+)

Introduction to Regular Expressions

Regex: `^(chr.)\t([0-9]+)\t+([0-9]+)\t+ENST([0-9]+)`

Expression PCRE ▾ Flags ▾

/`^(chr.)\t([0-9]+)\t+([0-9]+)\t+ENST([0-9]+)`/gm

Text Tests 6 matches (0.1ms)

```
chr1→65564→65573→ENST00000641515.2_cds_1_0_chr1_65565_f→0→+→
chr1→69036→70008→ENST00000641515.2_cds_2_0_chr1_69037_f→0→+→
chrX→284187→284314→ENST00000429181.6_cds_2_0_chrX_284188_f→0→+→
chrX→288732→288787→ENST00000429181.6_cds_3_0_chrX_288733_f→0→+→
chrY→284187→284314→ENST00000429181.6_cds_2_0_chrY_284188_f→0→+→
chrY→288732→288787→ENST00000429181.6_cds_3_0_chrY_288733_f→0→+→
chr10→48054→48114→ENST00000562809.1_cds_2_0_chr10_48055_r→0→!-→
chr10→48614→48725→ENST00000562809.1_cds_1_0_chr10_48615_r→0→!-→
chr11→168957→169052→ENST00000410108.5_cds_4_0_chr11_168958_r→0→!-→
chr11→180208→180404→ENST00000410108.5_cds_3_0_chr11_180209_r→0→!-→
chr12→66882→67436→ENST00000538872.6_cds_0_0_chr12_66883_f→0→+→
chr12→99145→99214→ENST00000538872.6_cds_1_0_chr12_99146_f→0→+→
```

Introduction to Regular Expressions

Regex: `^(chr.*)\t([0-9]+)\t+([0-9]+)\t+ENST([0-9]+)`

Expression PCRE Flags

/`^(chr.*)\t([0-9]+)\t+([0-9]+)\t+ENST([0-9]+)`/gm

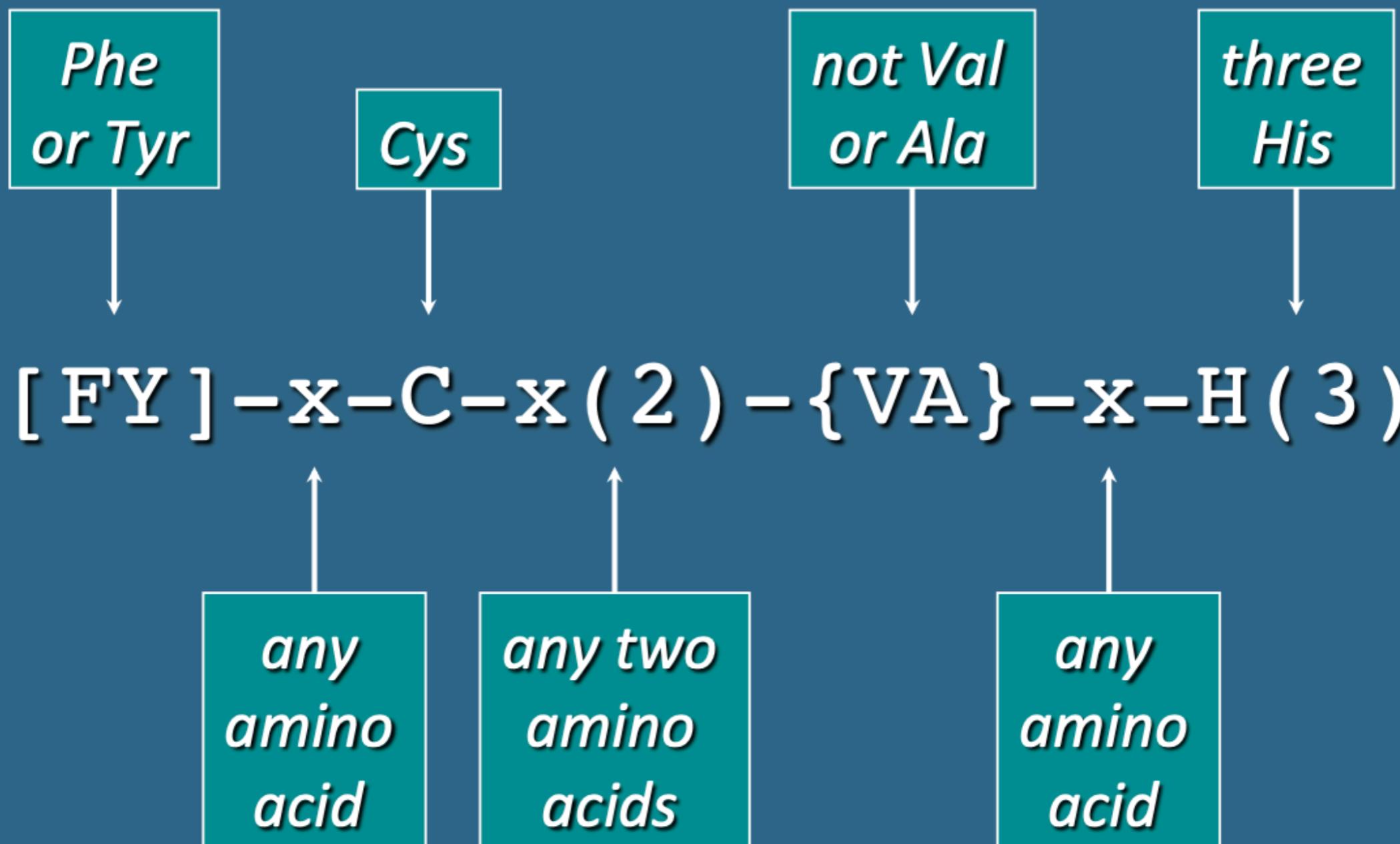
Text Tests 12 matches (0.3ms)

```
chr1→65564→65573→ENST00000641515.2_cds_1_0_chr1_65565_f→0→+→
chr1→69036→70008→ENST00000641515.2_cds_2_0_chr1_69037_f→0→+→
chrX→284187→284314→ENST00000429181.6_cds_2_0_chrX_284188_f→0→+→
chrX→288732→288787→ENST00000429181.6_cds_3_0_chrX_288733_f→0→+→
chrY→284187→284314→ENST00000429181.6_cds_2_0_chrY_284188_f→0→+→
chrY→288732→288787→ENST00000429181.6_cds_3_0_chrY_288733_f→0→+→
chr10→48054→48114→ENST00000562809.1_cds_2_0_chr10_48055_r→0→!→
chr10→48614→48725→ENST00000562809.1_cds_1_0_chr10_48615_r→0→!→
chr11→168957→169052→ENST00000410108.5_cds_4_0_chr11_168958_r→0→!→
chr11→180208→180404→ENST00000410108.5_cds_3_0_chr11_180209_r→0→!→
chr12→66882→67436→ENST00000538872.6_cds_0_0_chr12_66883_f→0→+→
chr12→99145→99214→ENST00000538872.6_cds_1_0_chr12_99146_f→0→+→
```

Introduction to Regular Expressions

Patterns as formulas:

Patterns



Introduction to Regular Expressions

HMM Logo Family: Piwi (PF02171)

HMM logo

HMM logos is one way of visualising profile HMMs. Logos provide a quick overview of the properties of an HMM in a graphical form. You can see a more detailed description of HMM logos and find out how you can interpret them [here](#).

If you find these logos useful in your own work, please consider citing the following article:

HMM Logos for visualization of protein families: B. Schuster-Böckler, J. Schultz, S. Rahmann
BMC Bioinformatics (2004) 5:7

