

## BioBIKE Pattern Matching (using MATCHES-OF-PATTERN)

### Character sets and some special characters:

[ <i>abc</i> ]	Set of characters
[~ <i>abc</i> ]	Set of excluded characters
[ <i>a-z</i> ]	Set of characters from first character to last
[~ <i>a-z</i> ]	Set of excluded characters from first character to last
*	Any character
#	Any digit (equivalent to [0-9])
~#	Any non-digit (equivalent to [~0-9])
\$	Any word character (letters and digits) (equivalent to [0-9a-z])
~\$	Any non-word character (equivalent to [~0-9a-z])
^ or ~@	Any space character (space, tab, and newline)
@ or ~^	Any non-space character
'	Either ' or "

### Examples:

(MATCHES-OF-PATTERN "C\*\*\*C\*\*\*C\*\*\*C" IN candidate-gene)

*Looks for iron-sulfer cofactor binding site in sequence of candidate gene*

(MATCHES-OF-PATTERN "##-\$\$\$-####"

IN "LOCUS ANGLNA 2225 bp DNA linear BCT 12-SEP-1993")

*Looks for the date within a locus line of a GenBank file*

(MATCHES-OF-PATTERN "[~ACGT]" IN (SEQUENCE-OF Cw?0002))

*Looks for nonstandard nucleotides within a gene sequence*

### Repetition symbols

{ <i>n</i> }	Previous element must be present exactly <i>n</i> number of times
{ <i>n</i> ,}	Previous element must be present at least <i>n</i> number of times
{ <i>m</i> , <i>n</i> }	Previous element may be present anywhere from <i>m</i> to <i>n</i> number of times
?	Previous element may be present or absent (equivalent to {0,1})
..	Previous element may be present 1 or any number of times (choose minimum number of times)
...	Previous element may be present 1 or any number of times (choose maximum number of times) (equivalent to {1,})
?..	Previous element may be absent or present any number of times (choose minimum length that satisfies the rest of the pattern)
?...	Previous element may be absent or present any number of times (choose maximum length that satisfies the rest of the pattern) (equivalent to {0,})

### Example:

```
(MATCHES-OF-PATTERN " [acgt]..." IN  
" 1021 accacgaagt tgctactggt ggtcagtgcg agctaggctt cgcctttggt")
```

Looks for blocks of nucleotides (any length), preceded by a space

```
(MATCHES-OF-PATTERN "am_{0,5}a $..." IN "I am not a crook (or am I?)")
```

Looks “am” within 5 characters of “ a “, followed by a sequence of letters of any length

### Other special symbols

- ~ Negation
- \ Escape (the character that follows is to be interpreted literally)
- \# Pound sign (because # itself is special)
- \\$ Dollar sign (because \$ itself is special)
- \\* Asterisk (because \* itself is special)
- \^ Carat (because ^ itself is special)
- ( ) Group (to be considered a single element in pattern matching)
- ( ) Remember these elements
- | Or