Introduction to "universalmotif"

Benjamin Jean-Marie Tremblay*

5 May 2026

Installation

```
if (!requireNamespace("BiocManager", quietly = TRUE))
  install.packages("BiocManager")
BiocManager::install("universalmotif")
```

Overview

For a brief explanation of sequence motifs, see the Introduction to sequence motifs vignette. This broadly covers the different 'types' of motif representation, and establishes the nomenclature used by the `universalmotif` package.

The capabilities of the `universalmotif` package can be divided into several general categories. These are briefly demonstrated in the following vignettes:

- Motif import, export, and manipulation
- Sequence scanning and manipulation
- Motif comparisons and P-values

Choosing a function

For ordinary DNA/RNA workflows, the `_lite` functions provide a smaller interface. The original functions remain necessary for features outside that interface; `_lite` does not mean a drop-in replacement for every argument or output type.

Task	DNA/RNA starting point	Use the original function when you need
Scan sequences	`scan_sequences_lite()`	Amino-acid/custom alphabets, higher-order or gapped motifs, non-P-value thresholds, exhaustive P-values, or q-values
Compare motifs	`compare_motifs_lite()`	Other alphabets, non-PCC metrics, higher-order comparisons, IC filters, or alternative score aggregation
Enrichment	`enrich_motifs_lite()`	Scanning options outside the lite interface; check the counting and statistical-test defaults before switching
Align logos or draw trees	`view_motifs_lite()`, `motif_tree_lite()`	Other alphabets or comparison options available only in the original functions
Merge motifs	`merge_motifs_lite()`, `merge_similar_lite()`	Other alphabets or the broader original comparison and clustering options

*benjamin.tremblay@uwaterloo.ca

`motif_pvalue()` is shared by both scanner interfaces. Its dynamic method uses an integerised score distribution; see the P-value vignette for inclusive thresholds, unattainable cutoffs, and background rebasing.

CWMs can be plotted explicitly with `use.type = "CWM"`; they are not probability matrices or ordinary scanning PWMs. See the contribution-weight section of the sequence-motif introduction before converting them.

Use `saveRDS()`/`readRDS()` for lossless storage of complete motif objects. The native `write_motifs()` format retains gap definitions and, by default, higher-order matrices, but rounds matrix entries. External motif formats preserve different subsets of metadata and warn when gaps or higher-order matrices cannot be represented.