genotypst: A bioinformatics Typst package for sequence analysis and visualization

`genotypst` is a bioinformatics Typst package for biological sequence analysis and visualization. It provides functionality for reading FASTA files and generating common visualizations, such as multiple sequence alignments and sequence logos, for use in publications and presentations.

Loading sequence data

The `parse-fasta-file` function reads FASTA files and returns a dictionary where sequence identifiers are mapped to their corresponding sequences.

```
#let sequences = parse-fasta-file("/docs/data/dna.fna")
```

```
(
  seq_1: "AAGGGACACTGATTTTCTCCCACAGCTGGGCCGTGGACCGTAGTGTTTCAAGAAGCCCACACAC",
  seq_2: "GCAATGGAGACAACATAGCCAACTACCTACTAGATGGCCTAGATCTGCCGCA",
  seq_3: "GGAACTGGCGTTACAGACAGTTGTGAGCCACCACATGGGCCTGGGATTTAAATATTATAAAGCTCCTC",
)
```

Working with sequence data

`genotypst` provides a multiple functions to produce different visualizations of sequence data.

FASTA rendering

Use `render-fasta` to display sequences in the standard FASTA format.

```
#render-fasta(sequences, max-width: 50)
```

```
            >seq_1
            AAGGGACACTGATTTTCTCCCACAGCTGGGCCGTGGACCGTAGTGTTTCA
            AGAAGCCCACACAC
            >seq_2
            GCAATGGAGACAACATAGCCAACTACCTACTAGATGGCCTAGATCTGCCG
            CA
            >seq_3
            GGAACTGGCGTTACAGACAGTTGTGAGCCACCACATGGGCCTGGGATTTA
            AATATTATAAAGCTCCTC
```

In this example, `max-width` controls how many characters appear per line (default is 60).

Multiple sequence alignments

The `render-msa` function displays multiple sequence alignments with optional residue coloring and conservation bars.

In the example below:

- `colors: true` enables residue coloring based on biochemical properties.
- `conservation: true` adds conservation bars above the alignment.
- `start: 100` and `end: 160` limit the display to a specific region of interest (residues 100 to 160).

```
#let protein_msa = parse-fasta-file("/docs/data/msa.afa")

#render-msa(
  protein_msa,
  start: 100,
  end: 160,
  colors: true,
  conservation: true,
)
```


Figure 1. MSA visualization for positions 100–160, with residue coloring and conservation bars enabled.

Residue coloring represents amino acid physicochemical properties. The sequence alphabet (amino acid, DNA, or RNA) is determined automatically and a suitable color palette is applied.

The bars above the alignment indicate the degree of conservation at each column.

Sequence logos

Sequence logos[1] summarize conservation patterns within a sequence alignment and are commonly used to visualize binding sites, motifs, and functional domains. In a sequence logo, the total height of each stack represents the information content (in bits) at that position, while the height of individual letters reflects their relative frequencies.

In the example below, we visualize the same region as the MSA of the previous section (positions 100 to 160).

```
#render-sequence-logo(protein_msa, start: 100, end: 160)
```
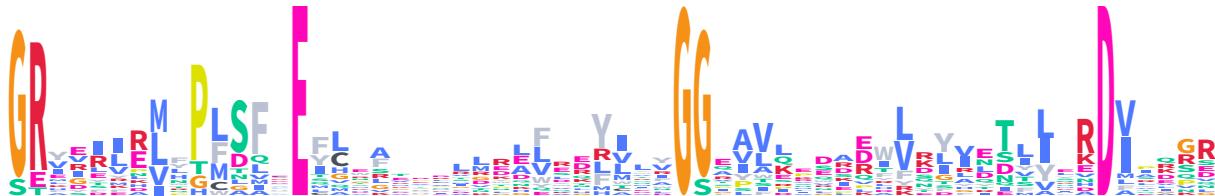

Figure 2. Sequence logo for positions 100–160, showing conservation and residue frequency.

Like `render-msa` , `render-sequence-logo` automatically applies the appropriate color palette based on the sequence alphabet.

Color palettes

`genotypst` uses predefined color palettes to assign colors to sequence residues.

Amino acid palette

Amino acids are colored according to their physicochemical properties. Grouping residues by color helps reveal the chemical nature of conserved positions (e.g., whether a position is consistently hydrophobic or charged), which is often important for understanding protein structure, function, and evolution.

Hydrophobic

Alanine	Ala	A	#4d78ff
Histidine	His	H	#4d78ff
Isoleucine	Ile	I	#4d78ff
Leucine	Leu	L	#4d78ff
Methionine	Met	M	#4d78ff
Valine	Val	V	#4d78ff

Polar

Serine	Ser	S	#00c990
Threonine	Thr	T	#00c990
Glutamine	Gln	Q	#00c990
Asparagine	Asn	N	#00c990

Aromatic

Phenylalanine	Phe	F	#bac1d2
Tryptophan	Trp	W	#bac1d2
Tyrosine	Tyr	Y	#bac1d2

Negatively charged

| Aspartic acid | Asp | D | #ff07b8 |
| Glutamic acid | Glu | E | #ff07b8 |

Positively charged

| Lysine | Lys | K | #e51f3e |
| Arginine | Arg | R | #e51f3e |

Cysteine

| Cysteine | Cys | C | #494e5b |

Glycine

| Glycine | Gly | G | #f59116 |

Proline

| Proline | Pro | P | #dce100 |

Nucleic acid palettes

The DNA and RNA palettes assign a distinct color to each nucleotide.

DNA palette		
Adenine	A	`#00c990`
Cytosine	C	`#4d78ff`
Guanine	G	`#ff07b8`
Thymine	T	`#f59116`

RNA palette		
Adenine	A	`#00c990`
Cytosine	C	`#4d78ff`
Guanine	G	`#ff07b8`
Uracil	U	`#f59116`

Customizing visualizations

Font selection

By default, `render-fasta` and `render-msa` inherit the monospaced font used for raw text in your document. To use a different font, wrap the rendering function in a `context` block with a custom font for raw text.

```
#let dna_msa = (
  "seq1": "AGTCTCAAGATAACTTTCGAAACAAACTTC",
  "seq2": "AGTTTCCAAGTGGATTTGGAATTGAACTTT",
  "seq3": "ACTCT-CGGATGGATTCGGATACAAACTTT",
  "seq4": "AGTCT---GATTGATGTGGATACAAACTTC",
  "seq5": "AGTCT--GGGTGGATTTGG-AACAAATTTT",
  "seq6": "CAGTGCTCCCTGGTGGTGG-ACCATCTTAC",
  "seq7": "AGTCTCAAGACGGATACTG--ATGCCCTAT",
)

#context {
  show raw: set text(font: "Maple Mono")
  render-msa(dna_msa)
}
```

```
seq1    AGTCTCAAGATAACTTTCGAAACAAACTTC
seq2    AGTTTCCAAGTGGATTTGGAATTGAACTTT
seq3    ACTCT-CGGATGGATTCGGATACAAACTTT
seq4    AGTCT---GATTGATGTGGATACAAACTTC
seq5    AGTCT--GGGTGGATTTGG-AACAAATTTT
seq6    CAGTGCTCCCTGGTGGTGG-ACCATCTTAC
seq7    AGTCTCAAGACGGATACTG--ATGCCCTAT
```

Default document font for raw text

```
seq1    AGTCTCAAGATAACTTTCGAAACAAACTTC
seq2    AGTTTCCAAGTGGATTTGGAATTGAACTTT
seq3    ACTCT-CGGATGGATTCGGATACAAACTTT
seq4    AGTCT---GATTGATGTGGATACAAACTTC
seq5    AGTCT--GGGTGGATTTGG-AACAAATTTT
seq6    CAGTGCTCCCTGGTGGTGG-ACCATCTTAC
seq7    AGTCTCAAGACGGATACTG--ATGCCCTAT
```

Custom font (Maple Mono)

Sequence logos are rendered using the default document font, rather than the monospaced font for raw text. To specify a custom font for sequence logos, use a `show text` rule instead.

```
#context {
  show text: set text(font: "Maple Mono")
  render-sequence-logo(dna_msa)
}
```


Default document font

Custom font (New Computer Modern)

Bibliography

1. Schneider, T. D. & Stephens, R. Sequence logos: a new way to display consensus sequences. *Nucleic Acids Research* **18,** 6097–6100 (1990).