ruby-libssw

kojix2

24 May 2022

Summary

Ruby-libssw is the Ruby binding of libssw, a library that uses the Smith-Waterman algorithm to find the best pairwise alignment of two sequences. ruby-libssw was created using fiddle, the Ruby standard library. Ruby-libssw can be used to create local alignments of nucleotide and amino acid sequences in the Ruby language.

Code: https://github.com/kojix2/ruby-libssw

Statement of need

("SSW Library: An SIMD Smith-Waterman C/C++ Library for Use in Genomic Applications | PLOS ONE" n.d.)

Benchmark

Examples

```
require 'libssw'
ref_str = "AAAAAAAACGTTAAAAAAAA"
ref_int = SSW::DNA.to_int_array(ref_str)
# [0, 0, 0, 0, 0, 0, 0, 0, 1, 2, 3, 3, 0, 0, 0, 0, 0, 0, 0, 0, 0]
read str1 = "ACGTT"
read str2 = SSW::DNA.revcomp(read str1)
# "AACGT"
read_int1 = SSW::DNA.to_int_array(read_str1)
# [0, 1, 2, 3, 3]
read_int2 = SSW::DNA.to_int_array(read_str2)
# [0, 0, 1, 2, 3]
mat = SSW.create_scoring_matrix(SSW::DNA::Elements, 2, -2)
\# mat = [2, -2, -2, -2, 0,
       -2, 2, -2, -2, 0,
        -2, -2, 2, -2, 0,
#
       -2, -2, -2, 2, 0,
        0, 0, 0, 0,
profile1 = SSW.init(read_int1, mat)
       = SSW.align(profile1, ref_int, 3, 1, 1, 0, 0)
align1
pp align1.to_h
# {
```

```
=> 10.
# :score1
# :score2
                => 0,
               => 8,
# :ref_begin1
# :ref_end1
                => 12,
# :read_begin1 => 0,
# :read_end1
                => 4,
# :ref_end2
                => 0.
                => [80],
# :cigar
# :cigar_len
                => 1,
# :cigar_string => "5M"
# }
profile2 = SSW.init(read_int2, mat)
align2 = SSW.align(profile2, ref_int, 3, 1, 1, 0, 0)
pp align2.to_h
# {
# :score1
                => 10,
# :score2
                => O,
# :ref_begin1 => 7,
# :ref_end1
                => 11,
# :read_begin1 => 0,
# :read_end1
                => 4,
# :ref_end2
                => O,
                => [80],
# :cigar
# :cigar_len
                => 1,
# :cigar_string => "5M"
# }
puts SSW.build_path(read_str1, ref_str, align1)
# 5M
# ACGTT
# ||||
# ACGTT
```

Reference

"SSW Library: An SIMD Smith-Waterman C/C++ Library for Use in Genomic Applications | PLOS ONE." n.d. Accessed May 24, 2022. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0082138.