Benchmarking

Michiel Coghe 4 juni 2019

Benchmarking of the different functions in the bendDNA package

In this R Markdown file all the different functions included in the bendDNA package are benchmarked. This will be done by looking at every function individually, testing and timing the functions for different parameter values.

The testing is done x times (depending on which function) and summary statistics are given in a summary table. These are then plotted using violin plots comparing every different parameter value. The CRAN microbenchmark package is used for making both the summary table and plot.

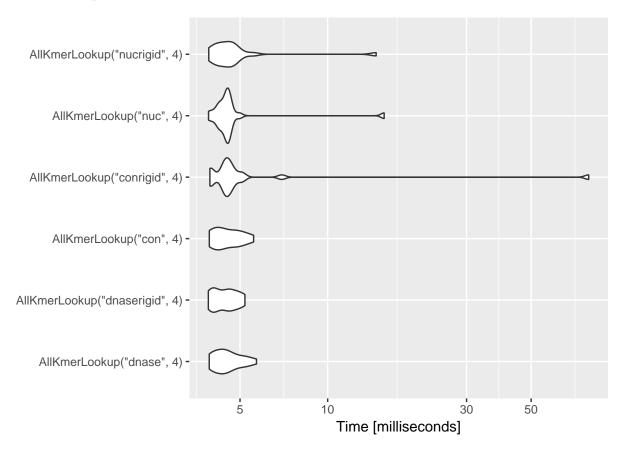
AllKmerLookup

Function for producing all possible k-mers and calculating their average bendability coefficients.

Usage: AllKmerLookup(scale, k, sequence.out = F)

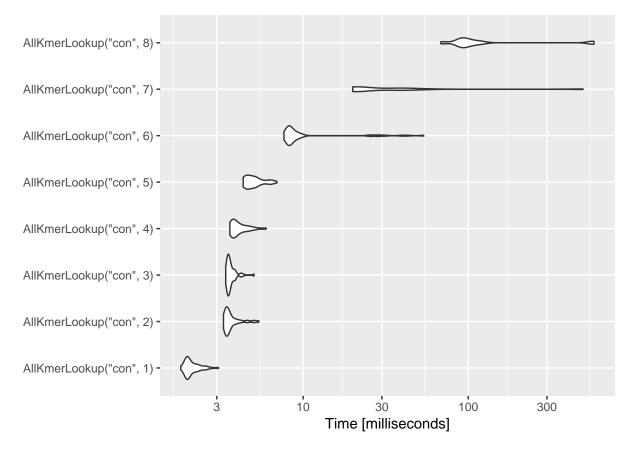
Parameters that can be used include: - Scale - Number of consecutive trinucleotides - Whether to output the sequences

Different lookup tables



```
FALSE Unit: milliseconds
FALSE
                                                                  median
                                 expr
                                          min
                                                     lq
                                                            mean
FALSE
            AllKmerLookup("dnase", 4) 3.919518 4.130370 4.541840 4.437062
FALSE AllKmerLookup("dnaserigid", 4) 3.891980 4.055586 4.444487 4.499967
              AllKmerLookup("con", 4) 3.921678 4.100402 4.544324 4.376587
FALSE
FALSE
         AllKmerLookup("conrigid", 4) 3.940036 4.348240 8.309063 4.507796
FALSE
              AllKmerLookup("nuc", 4) 3.897380 4.266707 4.987922 4.500237
         AllKmerLookup("nucrigid", 4) 3.904939 4.193274 5.001934 4.467840
FALSE
FALSE
             uq
                     max neval cld
FALSE 4.813680 5.694882
                            20
FALSE 4.766163 5.198125
FALSE 4.909522 5.566373
                             20
FALSE 4.751855 78.813120
                             20
FALSE 4.583659 15.632712
                             20
FALSE 4.741866 14.676996
```

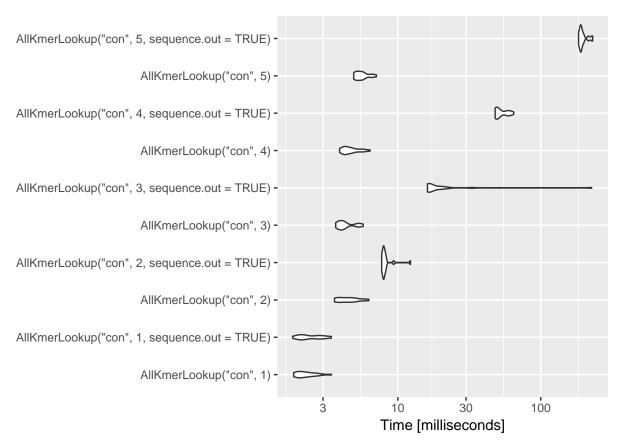
Number of consecutive trinucleotides



FALSE Unit: milliseconds						
FALSE	ex	pr	min	lq	mean	median
FALSE	AllKmerLookup("con",	1)	1.815862	1.974608	2.134628	2.018884
FALSE	AllKmerLookup("con",	2)	3.292633	3.407642	3.691139	3.523192
FALSE	AllKmerLookup("con",	3)	3.396843	3.491336	3.659703	3.570978
FALSE	AllKmerLookup("con",	4)	3.598786	3.751592	4.112475	3.908718
FALSE	AllKmerLookup("con",	5)	4.334202	4.557742	5.086885	4.974856
FALSE	AllKmerLookup("con",	6)	7.654911	8.103611	12.662260	8.409225
FALSE	AllKmerLookup("con",	7)	19.961514	20.532243	56.451853	21.567062

```
FALSE AllKmerLookup("con", 8) 68.306180 89.619734 161.829188 97.077831
                         max neval cld
FALSE
               uq
                                 50 a
FALSE
         2.239185
                    3.076112
FALSE
         3.764011
                    5.397909
                                 50 a
FALSE
         3.714336
                    5.043159
                                 50 a
FALSE
         4.410335
                    5.978357
                                50 a
FALSE
         5.372531
                    6.948113
                                 50 a
FALSE
         9.413537 53.739875
                                 50 a
        39.966221 496.320223
FALSE
                                 50
                                    b
FALSE 116.783675 578.312584
                                 50
```

Sequence Out



```
## Unit: milliseconds
##
                                              expr
                                                          min
                                                                       lq
##
                          AllKmerLookup("con", 1)
                                                                2.005926
                                                     1.868238
##
    AllKmerLookup("con", 1, sequence.out = TRUE)
                                                     1.835840
                                                                2.127955
                          AllKmerLookup("con", 2)
##
                                                     3.605805
                                                                3.846354
                         2, sequence.out = TRUE)
##
    AllKmerLookup("con",
                                                     7.712686
                                                                7.870891
                                                                3.914658
##
                          AllKmerLookup("con", 3)
                                                     3.678699
##
    AllKmerLookup("con", 3, sequence.out = TRUE)
                                                    16.118669
                                                               16.392156
##
                          AllKmerLookup("con", 4)
                                                     3.915738
                                                                4.220271
##
    AllKmerLookup("con", 4, sequence.out = TRUE)
                                                    47.972101
                                                               48.881380
##
                          AllKmerLookup("con", 5)
                                                     4.911411
                                                                5.130902
##
    AllKmerLookup("con", 5, sequence.out = TRUE) 183.746992 189.793923
##
          mean
                   median
                                   uq
                                              max neval cld
```

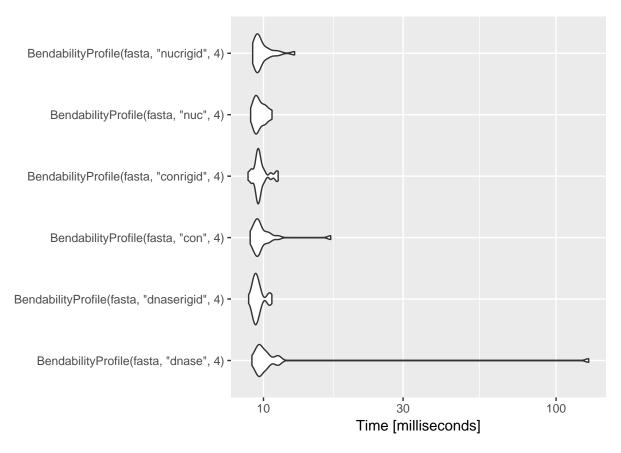
```
##
     2.297176
               2.175200
                          2.498362
                                    3.439499
                                                20 a
##
     2.485430
               2.178441
                          2.851761 3.428701
                                                20 a
##
     4.536792
               4.491058
                          4.911950
                                    6.273712
                                                20 a
##
     8.262844
               7.983742
                          8.166515 12.263948
                                                20 a
     4.356151
               4.140629
                         4.746186
                                    5.732139
                                                20 a
##
                                               20 b
##
    28.872542 16.747715 19.931816 226.734788
               4.522645
##
     4.717757
                          5.022101
                                    6.401141
                                                20 a
##
    53.043822 49.896492 58.689892 64.674998
                                                20 c
##
     5.626416
               5.468642
                          5.790184
                                   7.092820
                                                20 a
## 197.329989 191.534730 199.069231 230.663484
                                                20
                                                     d
```

BendabilityProfile

Splits sequence into k-mers with two-nucleotide overlap and calculates average bendability per k-mer according to the chosen scale.

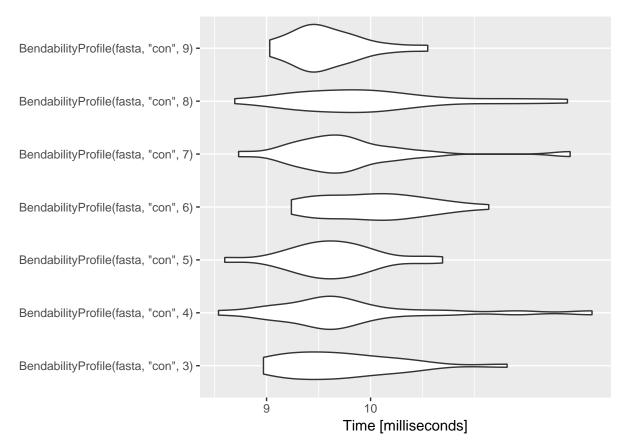
Parameters that can be used include: - Scale - Number of consecutive trinucleotides

Scale



```
## Unit: milliseconds
##
                                          expr
                                                    min
                                                                       mean
##
         BendabilityProfile(fasta, "dnase", 4) 9.130602 9.634377 15.962650
   BendabilityProfile(fasta, "dnaserigid", 4) 8.896802 9.239132
##
           BendabilityProfile(fasta, "con", 4) 9.007492 9.419476 10.090880
##
##
      BendabilityProfile(fasta, "conrigid", 4) 8.855766 9.464292 9.752816
           BendabilityProfile(fasta, "nuc", 4) 9.037729 9.366831 9.724036
##
     BendabilityProfile(fasta, "nucrigid", 4) 9.184597 9.444045 9.949656
##
##
      median
                             max neval cld
   9.773955 10.373842 129.20961
##
                                    20
   9.443504 9.696742
                       10.68351
                                    20
##
##
   9.602520 10.142203
                        16.97557
                                    20
                                         a
                       11.22562
   9.613589 9.873576
                                    20
   9.569853 10.066879 10.69484
                                    20
                                         a
   9.606840 10.204297 12.77420
                                    20
```

Number of consecutive trinucleotides



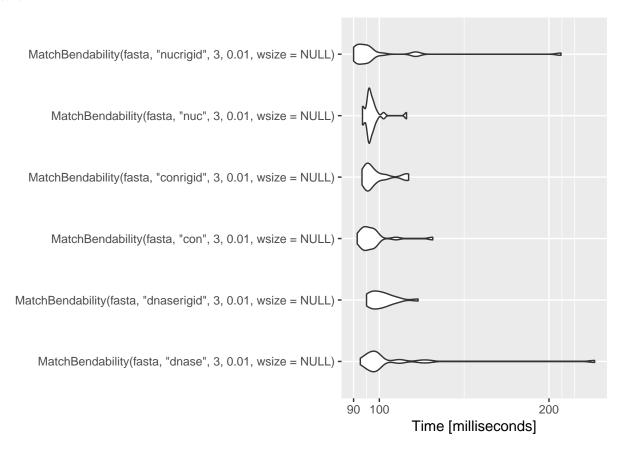
```
## Unit: milliseconds
##
                                                                       median
                                   expr
                                             \min
   BendabilityProfile(fasta, "con", 3) 8.971856 9.247501
                                                           9.718826 9.665964
##
   BendabilityProfile(fasta, "con", 4) 8.572290 9.373040
##
                                                           9.841017
                                                                     9.651656
##
   BendabilityProfile(fasta, "con", 5) 8.626825 9.372501 9.644825 9.626548
##
   BendabilityProfile(fasta, "con", 6) 9.229413 9.560134 10.035616 10.003435
##
   BendabilityProfile(fasta, "con", 7) 8.748855 9.406787 9.780651 9.695662
   BendabilityProfile(fasta, "con", 8) 8.714838 9.416237 10.024223 9.907054
##
   BendabilityProfile(fasta, "con", 9) 9.029090 9.349283 9.580004 9.482651
##
##
                   max neval cld
           uq
   10.073089 11.48317
##
                          20
##
   10.021254 12.51449
                               a
    9.900304 10.75640
                          20
##
   10.338206 11.27151
##
                          20
                               a
##
    9.942691 12.24019
                          20
##
   10.385992 12.20671
                          20
                               a
    9.778814 10.59603
                          20
##
```

MatchBendability

Calculates bendability profile of query sequence and finds all other sequences which have similar (or same) profile.

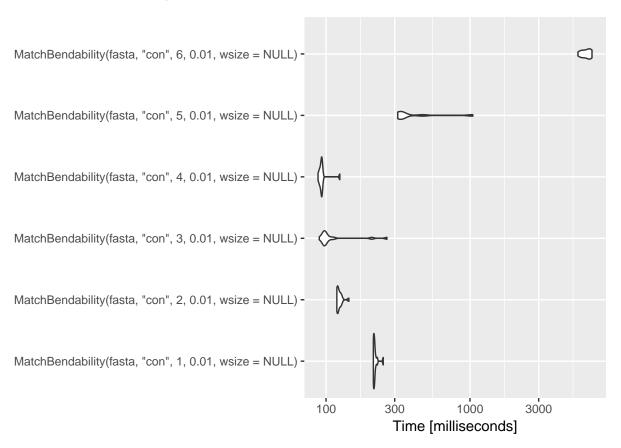
Parameters that can be used include: - Scale - Number of consecutive trinucleotides - Tolerance - wsize - output.list - random.out

Scale



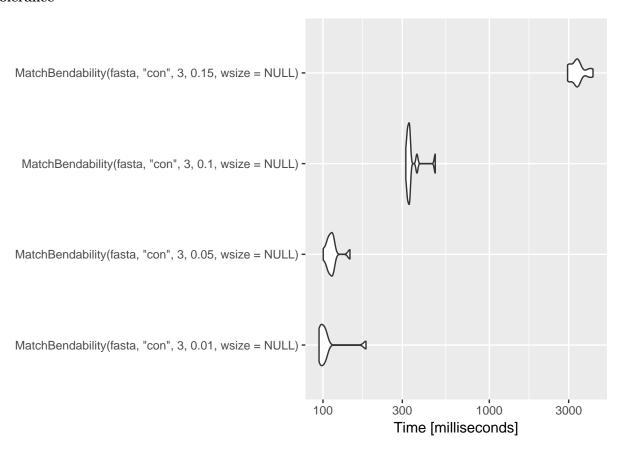
```
## Unit: milliseconds
##
                                                             expr
                                                                       min
         MatchBendability(fasta, "dnase", 3, 0.01, wsize = NULL) 92.51658
##
   MatchBendability(fasta, "dnaserigid", 3, 0.01, wsize = NULL) 94.89831
##
##
           MatchBendability(fasta, "con", 3, 0.01, wsize = NULL) 91.35352
      MatchBendability(fasta, "conrigid", 3, 0.01, wsize = NULL) 93.14832
##
##
           MatchBendability(fasta, "nuc", 3, 0.01, wsize = NULL) 93.33569
##
      MatchBendability(fasta, "nucrigid", 3, 0.01, wsize = NULL) 90.00904
##
          lq
                  mean
                          median
                                                 max neval cld
                                        uq
   95.59377 107.79314
                        98.57809 103.62314 241.0360
##
                                                        20
   97.05569 100.96857 100.11398 103.40824 117.1417
##
                                                        20
                                                             а
                       95.41694 98.13425 124.3916
##
   93.74119 97.41295
                                                        20
##
   95.04707 98.80212
                        96.14722 100.22009 112.7022
                                                        20
                                                             а
   95.55678 97.13291
                        96.07109
                                  97.41557 111.7502
                                                        20
##
                                                             a
   91.48095 101.79910
                        94.43827 96.40991 210.0697
                                                        20
                                                             а
```

Number of consecutive sequences



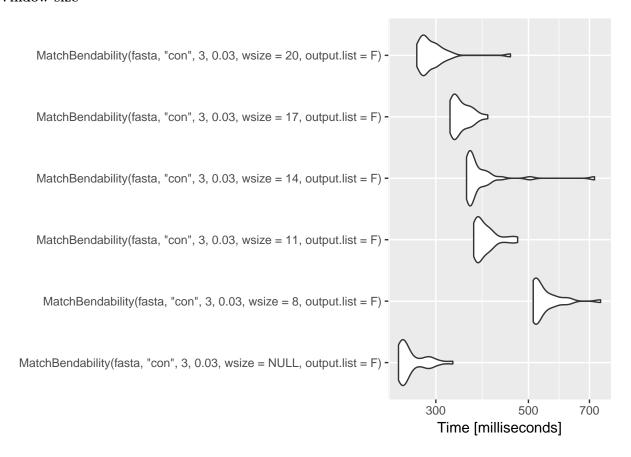
```
## Unit: milliseconds
##
                                                                 min
                                                     expr
   MatchBendability(fasta, "con", 1, 0.01, wsize = NULL)
##
                                                           213.17010
   MatchBendability(fasta, "con", 2, 0.01, wsize = NULL)
##
                                                           118.89705
   MatchBendability(fasta, "con", 3, 0.01, wsize = NULL)
                                                            89.76066
##
   MatchBendability(fasta, "con", 4, 0.01, wsize = NULL)
##
                                                            88.00581
##
   MatchBendability(fasta, "con", 5, 0.01, wsize = NULL)
                                                           314.94795
   MatchBendability(fasta, "con", 6, 0.01, wsize = NULL) 5620.04823
##
##
            lq
                     mean
                              median
                                             uq
                                                      max neval cld
##
     214.02511
                218.98269 216.34286
                                      218.38172
                                                 249.6871
                                                             15 ab
##
     119.94537
               123.76159 121.10006 125.90240
                                                 144.1236
                                                             15 a
##
     96.04112 117.32280
                            99.31945 104.19306
                                                 264.9813
                                                             15 a
##
     91.31789
                94.42945
                            93.07327
                                       93.76252 124.6805
                                                             15 a
     326.96000 443.49845 332.88355 403.64166 1046.6390
##
                                                             15
   6033.04322 6430.73386 6513.14677 6781.04809 7039.0110
                                                             15
```

Tolerance



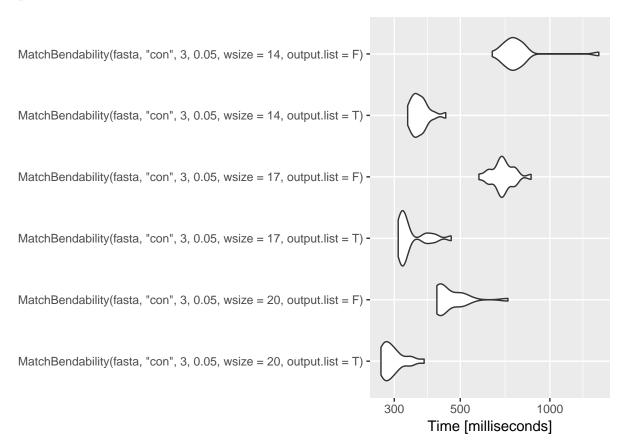
```
## Unit: milliseconds
##
                                                                        min
    MatchBendability(fasta, "con", 3, 0.01, wsize = NULL)
##
                                                                   94.72823
    MatchBendability(fasta, "con", 3, 0.05, wsize = NULL)
MatchBendability(fasta, "con", 3, 0.1, wsize = NULL)
##
                                                                 100.40961
##
                                                                 314.82052
    MatchBendability(fasta, "con", 3, 0.15, wsize = NULL) 2953.85780
##
##
                      mean
                                median
                                                          max neval cld
             lq
                                               uq
      95.39993
                 107.5737
                              99.94444 104.3010
##
                                                    181.4328
                                                                      a
##
     107.13310 113.6325 111.76481
                                        115.2027
                                                    145.3245
                                                                 10
                                                                      a
##
     322.60717 344.8043 330.04691 333.7000 473.2135
                                                                 10
                                                                      a
    3054.20801 3406.8254 3362.35606 3474.8965 4199.7447
                                                                 10
                                                                       b
```

Window size



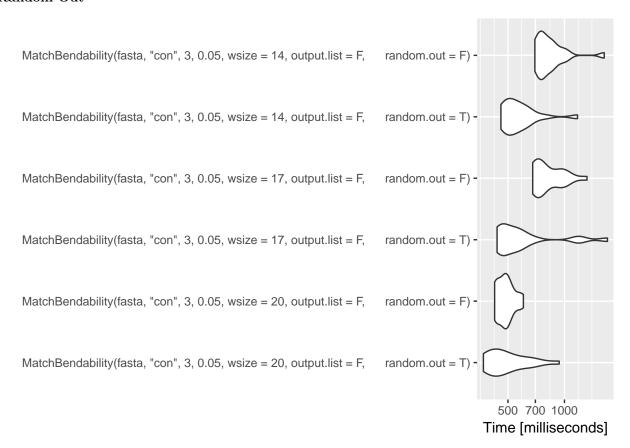
```
## Unit: milliseconds
##
   MatchBendability(fasta, "con", 3, 0.03, wsize = NULL, output.list = F)
##
       MatchBendability(fasta, "con", 3, 0.03, wsize = 8, output.list = F)
##
##
      MatchBendability(fasta, "con", 3, 0.03, wsize = 11, output.list = F)
##
      MatchBendability(fasta, "con", 3, 0.03, wsize = 14, output.list = F)
##
      MatchBendability(fasta, "con", 3, 0.03, wsize = 17, output.list = F)
      MatchBendability(fasta, "con", 3, 0.03, wsize = 20, output.list = F)
##
##
         min
                   lq
                          mean
                                 median
                                               uq
                                                       max neval
                                                                   cld
##
   244.1812 248.5392 262.5597 252.7967 271.9979 329.5461
   513.3709 515.8028 547.4515 530.0222 561.9142 744.3189
                                                              30
##
                                                                     е
   369.6090 375.5787 395.4637 385.9496 404.6659 471.0974
                                                              30
                                                                    d
   355.3656 360.8180 388.4465 365.5680 387.7665 719.8790
                                                              30
                                                                    d
##
   324.3923 330.8091 345.7495 339.4505 357.8996 399.5229
                                                              30
                                                                   С
   270.0865 279.0945 295.6261 285.1169 300.9691 452.8319
                                                              30
```

Output List



```
## Unit: milliseconds
##
   MatchBendability(fasta, "con", 3, 0.05, wsize = 20, output.list = T)
##
   MatchBendability(fasta, "con", 3, 0.05, wsize = 20, output.list = F)
##
   MatchBendability(fasta, "con", 3, 0.05, wsize = 17, output.list = T)
##
   MatchBendability(fasta, "con", 3, 0.05, wsize = 17, output.list = F)
##
##
   MatchBendability(fasta, "con", 3, 0.05, wsize = 14, output.list = T)
   MatchBendability(fasta, "con", 3, 0.05, wsize = 14, output.list = F)
##
##
         min
                   lq
                          mean
                                 median
                                              uq
                                                       max neval cld
##
   270.6281 277.8643 298.8592 290.6919 307.5168
                                                              20 a
                                                  377.9696
##
   417.1378 425.4082 467.1603 435.7913 493.8248
                                                  722.9810
                                                              20 b
   309.2946 317.1653 342.8729 323.4957 356.0437
                                                  466.0667
                                                              20 a
## 577.5178 667.0139 697.3933 688.5150 733.6146
                                                  864.6614
                                                              20
                                                                   С
   332.7561 350.3564 366.5674 362.1495 376.9642
##
                                                  446.7877
                                                              20 a
   640.3478 713.9341 785.1070 752.0359 802.7059 1460.2552
                                                              20
                                                                    d
```

Random Out



```
## Unit: milliseconds
##
                                                                                            expr
##
   MatchBendability(fasta, "con", 3, 0.05, wsize = 20, output.list = F,
                                                                                random.out = T)
##
   MatchBendability(fasta, "con", 3, 0.05, wsize = 20, output.list = F,
                                                                                random.out = F)
   MatchBendability(fasta, "con", 3, 0.05, wsize = 17, output.list = F,
##
                                                                                random.out = T)
   MatchBendability(fasta, "con", 3, 0.05, wsize = 17, output.list = F,
                                                                                random.out = F)
   MatchBendability(fasta, "con", 3, 0.05, wsize = 14, output.list = F,
##
                                                                                random.out = T)
##
   MatchBendability(fasta, "con", 3, 0.05, wsize = 14, output.list = F,
                                                                                random.out = F)
##
         min
                   lq
                                 median
                                                        max neval cld
                          mean
                                               uq
##
   370.1209 409.2126 509.7598 472.7507 567.2708
                                                   937.5145
                                                                    а
   424.6685 442.3714 486.6252 484.5398 509.3196
                                                   604.2903
##
                                                                20
                                                                    а
   437.6301 465.1447 645.3722 539.9174 609.3383 1694.5142
                                                                20
                                                                    a
##
   677.4220 710.4433 830.8216 763.5334 938.7872 1318.0219
                                                                20
   458.7838 491.0835 587.6018 543.2046 627.1972 1173.5770
                                                                20
                                                                    a
   695.6092 744.5702 853.0123 790.6196 873.9008 1628.3651
                                                                20
                                                                     b
```

ClusterAndConsensus

Intended to present output of function MatchBendability in a more human-friendly format. The output can contain very large numbers of highly similar sequences, so clustering and calculating a consensus can help provide a clearer picture. DISCLAIMER: due to the nature of bendability problem and limitations of IUPAC extended genetic alphabet, output of this function is not intended for further machine use. Not all sequences which can be inferred from consensus output will have satisfied the initial bendability requirements (for example AA/TT will be marked as WW, same as AT/TA).

Parameters that can be used include: - Cutoff - sample.seq

Cutoff

```
## Time difference of 0.54 secs
##
## Time difference of 0.12 secs
##
##
_____
##
## Time difference of 0.16 secs
## -----
##
## Time difference of 0.25 secs
##
## Time difference of 0.16 secs
## Time difference of 0.29 secs
##
##
______
## Time difference of 0.25 secs
##
## Time difference of 0.17 secs
##
## Time difference of 0.13 secs
## Time difference of 0.32 secs
## ------
##
## Time difference of 0.25 secs
##
##
## Time difference of 0.29 secs
##
```

```
## Time difference of 0.27 secs
## ------
##
## Time difference of 0.26 secs
##
## Time difference of 0.29 secs
## Time difference of 0.13 secs
## Time difference of 0.33 secs
## Time difference of 0.12 secs
## Time difference of 0.12 secs
## Time difference of 0.12 secs
##
## Time difference of 0.25 secs
## Time difference of 0.18 secs
##
## Time difference of 0.12 secs
##
## Time difference of 0.27 secs
##
## Time difference of 0.12 secs
## Time difference of 0.27 secs
##
```

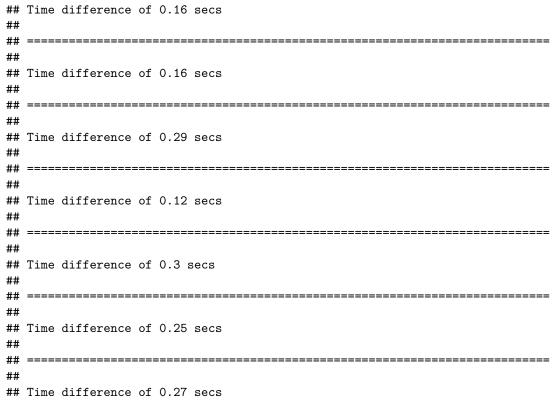
```
## Time difference of 0.27 secs
##
## Time difference of 0.17 secs
##
## Time difference of 0.27 secs
##
## Time difference of 0.3 secs
##
## Time difference of 0.26 secs
## Time difference of 0.25 secs
## Time difference of 0.17 secs
## -----
## Time difference of 0.27 secs
## Time difference of 0.32 secs
## -----
##
## Time difference of 0.16 secs
##
## Time difference of 0.26 secs
## Time difference of 0.29 secs
## Time difference of 0.12 secs
##
```

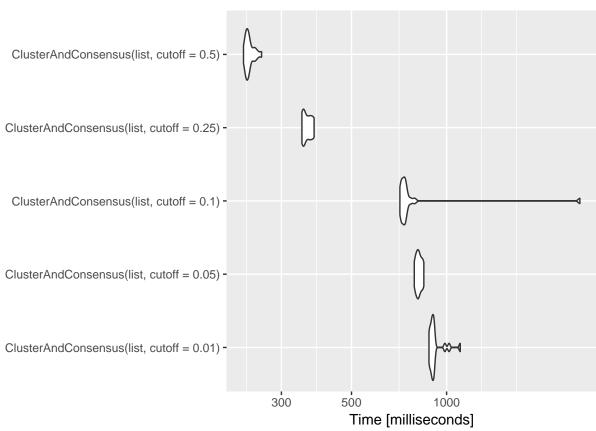
```
## Time difference of 0.12 secs
##
## Time difference of 0.25 secs
##
## Time difference of 0.28 secs
## Time difference of 0.29 secs
## Time difference of 0.3 secs
## Time difference of 0.13 secs
## Time difference of 0.31 secs
## Time difference of 0.24 secs
##
## Time difference of 0.3 secs
## Time difference of 0.24 secs
##
## Time difference of 0.12 secs
##
## Time difference of 0.27 secs
##
## Time difference of 0.3 secs
## Time difference of 0.27 secs
##
```

```
## Time difference of 0.27 secs
##
## Time difference of 0.16 secs
##
## Time difference of 0.12 secs
##
## Time difference of 0.25 secs
##
## Time difference of 0.29 secs
## Time difference of 0.29 secs
## Time difference of 0.13 secs
## -----
## Time difference of 0.12 secs
## Time difference of 0.32 secs
## -----
##
## Time difference of 0.17 secs
##
## Time difference of 0.3 secs
## Time difference of 0.17 secs
## Time difference of 0.16 secs
##
```

```
## Time difference of 0.27 secs
## ------
##
## Time difference of 0.29 secs
##
## Time difference of 0.25 secs
## Time difference of 0.16 secs
## Time difference of 0.25 secs
## Time difference of 0.29 secs
## Time difference of 0.18 secs
## Time difference of 0.37 secs
##
## Time difference of 0.18 secs
## Time difference of 0.52 secs
##
## Time difference of 0.12 secs
##
## Time difference of 0.32 secs
##
## Time difference of 0.16 secs
## Time difference of 0.27 secs
##
```

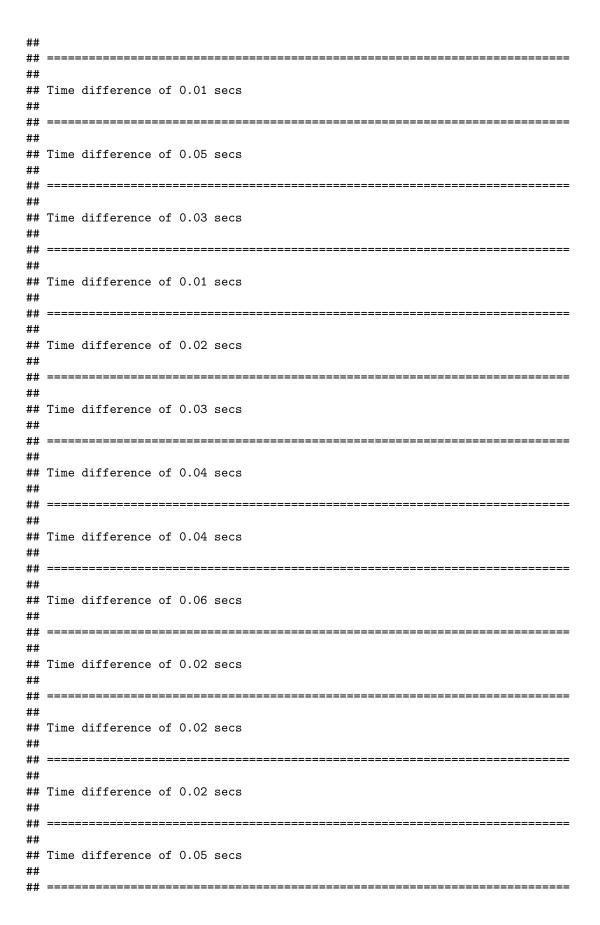
```
## Time difference of 0.3 secs
##
## Time difference of 0.29 secs
##
## Time difference of 0.13 secs
##
## Time difference of 0.12 secs
##
## Time difference of 0.27 secs
## Time difference of 0.32 secs
## Time difference of 0.29 secs
## -----
## Time difference of 0.16 secs
## -----
## Time difference of 0.12 secs
## -----
##
## Time difference of 0.3 secs
##
## Time difference of 0.26 secs
## Time difference of 0.17 secs
## Time difference of 0.16 secs
##
```



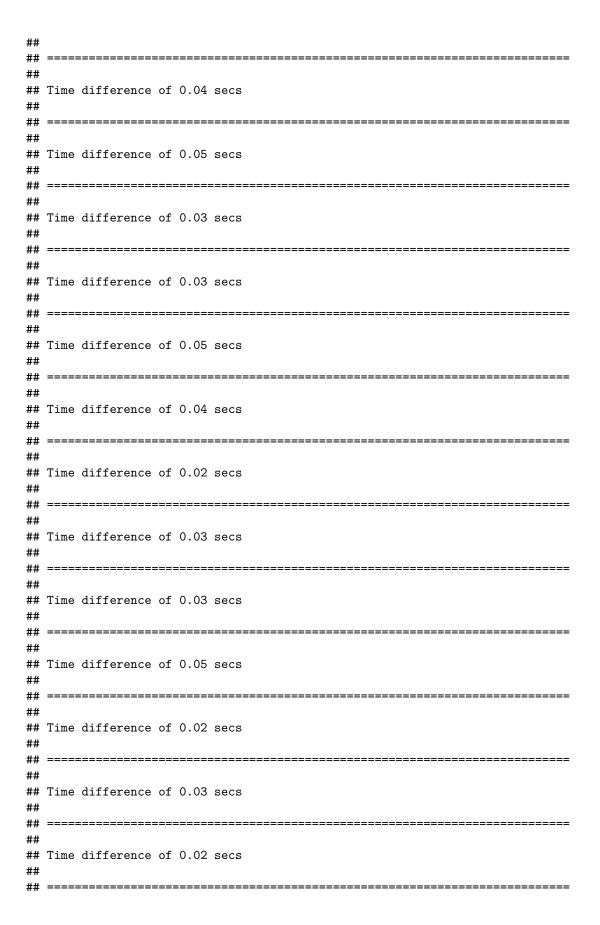


Unit: milliseconds

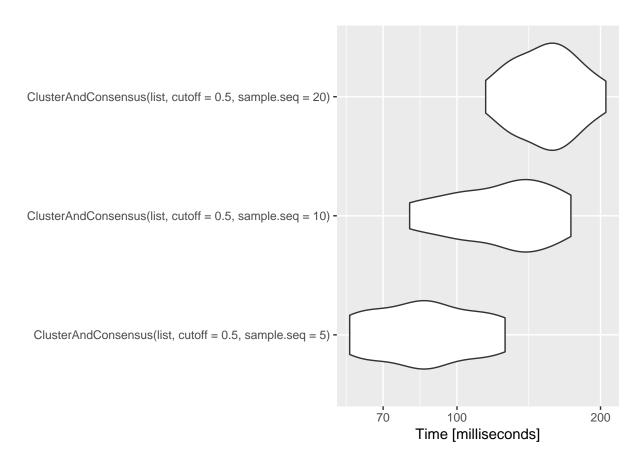
```
##
                         min
                     expr
                            lq
## ClusterAndConsensus(list, cutoff = 0.01) 879.4863 893.0848 920.1664
## ClusterAndConsensus(list, cutoff = 0.05) 790.1144 803.7156 817.3361
  ClusterAndConsensus(list, cutoff = 0.1) 711.4730 718.9919 830.8027
## ClusterAndConsensus(list, cutoff = 0.25) 348.3419 351.7409 362.2762
  ClusterAndConsensus(list, cutoff = 0.5) 227.5318 231.8292 237.6747
##
  median
         uq
              max neval cld
## 903.0080 913.3711 1104.8399
                  20
## 815.4793 831.3965 847.1934
                  20
                 20 b
## 735.3835 743.0333 2641.9503
## 359.8545 371.7647 380.6694
                 20 a
## 235.5897 240.7449 259.6222
                 20 a
Sample.seq
## Time difference of 0.07 secs
## Time difference of 0.02 secs
## Time difference of 0.05 secs
##
## Time difference of 0.01 secs
##
##
## Time difference of 0.06 secs
##
## Time difference of 0.06 secs
## Time difference of 0.03 secs
##
## Time difference of 0.01 secs
## Time difference of 0.05 secs
## Time difference of 0.04 secs
```



```
## Time difference of 0.04 secs
## Time difference of 0.02 secs
## Time difference of 0.03 secs
## Time difference of 0.02 secs
## Time difference of 0.05 secs
## ------
##
## Time difference of 0.04 secs
## Time difference of 0.03 secs
## Time difference of 0.05 secs
## Time difference of 0.05 secs
## Time difference of 0.08 secs
## -----
## Time difference of 0.03 secs
## Time difference of 0.03 secs
## Time difference of 0.04 secs
##
## Time difference of 0.03 secs
```



```
## Time difference of 0.04 secs
## Time difference of 0.03 secs
## Time difference of 0.03 secs
## Time difference of 0.02 secs
## Time difference of 0.05 secs
## -----
##
## Time difference of 0.04 secs
## Time difference of 0.06 secs
## Time difference of 0.01 secs
## Time difference of 0.02 secs
## Time difference of 0.01 secs
```



```
## Unit: milliseconds
##
                                                        expr
                                                                  min
##
    ClusterAndConsensus(list, cutoff = 0.5, sample.seq = 5)
                                                             59.56813
    ClusterAndConsensus(list, cutoff = 0.5, sample.seq = 10) 79.56527
    ClusterAndConsensus(list, cutoff = 0.5, sample.seq = 20) 114.86360
##
##
                 mean
                         median
                                              max neval cld
          lq
                                       uq
     68.6882 87.92426 86.12624 106.1674 126.1475
##
  103.1893 126.69119 131.33648 147.4662 173.3567
                                                      20 b
## 134.0257 155.34378 154.00721 170.6218 205.2474
                                                      20
```