

Bioinformatics new scoring function

The function operates on the standard DNA alphabet ACGT. The system simulates my “made-up” biological conditions for sequence alignment.

The steps are as follows

1. Throughout the process, each time there are consecutive Cs or Ts, halve the score each time there is another i.e. (0.5x, 0.25x, 0.125x...)
2. Check through in steps of 2 using the normal scoring system until you match an AG/AC/AT/AA with itself
3. Check through the rest of the string in steps of 2 following the rules outlined below

RULES:

1. Any CA/GA/TA/AA returns you back to step 2
2. All positive point values are doubled for matches

Step 1 simulates the fact that in my system, consecutive Cs and Ts aren't a desirable DNA alignment

Steps 2-3 simulate a system in which we prefer to have A's matched in odd positions. It works by looping in twos until it finds an A in the first position (i.e. an odd position). Then it increases the value of any matches proceeding an odd A that don't cause an A to be in an odd position

The result is a string where large strings of C's and T's don't matter and the majority of the matches will have A's matched in odd numbered slots.

This is an example of the scoring matrix. The diagonal will be all positive values as will some of the matches where 1 is correct but the other is not.

[illegible]