

CHALMERS UNIVERSITY OF TECHNOLOGY

Examination in Bioinformatics, MVE360

Monday 11 March 2013, 08:30-12:30

Solutions

Updated 2013-09-23

Question 1. a) catattt

4 p

- b) a
- c) at, t
- d) t
- e) gccggtt
- f) tt
- g) ag
- h) gcccgcgccc

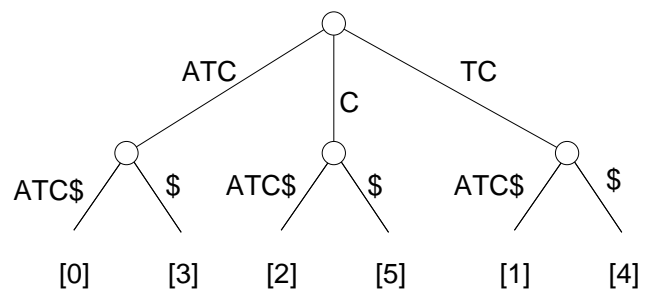
Question 2. a)

6 p

All suffixes

```
ATCATC$ [0]
TCATC$ [1]
CATC$ [2]
ATC$ [3]
TC$ [4]
C$ [5]
```

Suffix tree



b) `#!/usr/bin/perl -w`

```
print "Please type in the sequence: ";
$s = <STDIN>;
chomp($s);

for ( $i=0 ; $i<length($s) ; ++$i ) {
    print substr($s, $i) . "\n";
}
```

Question 3. a) `#!/usr/bin/perl -w`

12 p

```
$sequence = " ";

while ( <> ) {
    if ( /^ / ) {
        s/[^a-z]//g;
        $sequence .= $_
    }
    if ( /^FT   CDS *(\d+)..(\d+)/ ) {
        $startpos = $1;
        $endpos = $2;
    }
}

$cs = substr($sequence, $startpos, $endpos - $startpos + 1);
b) $cs =~ s/(...)/\1 /g;
@triplets = split(" ", $cs);

foreach $codon ( @triplets ) {
    print "$codon\n";
}
c) foreach $codon ( @triplets ) {
    if ( $codon =~ /ac./ ) {
        $thr_count{$codon} += 1;
    }
}

foreach $v ( values(%thr_count) ) {
    if ( $v > $max_count ) {
        $max_count = $v;
    }
}

foreach $c ( keys(%thr_count) ) {
    if ( $thr_count{$c} == $max_count ) {
        print "$c\n";
    }
}
```

Question 4. Optimal score is 7, giving 1 global alignment:

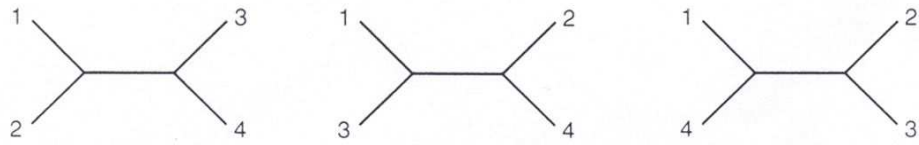
4 p

Q-FN
NGYE

Score matrix should be given in full.

Question 5. a)

4 p



- b) An informative site is a column in the multiple alignment that may be used as evidence for a particular tree. The site should be such that the number of molecular changes are different for different trees. From a practical point of view the informative site is such that at least two characters are present in at least two copies each. First and third columns are informative sites in the example given.