PLNT4610 BIOINFORMATICS

MID-TERM EXAMINATION

3:00 p.m. to 4:00 p.m. Monday, October 26, 2009

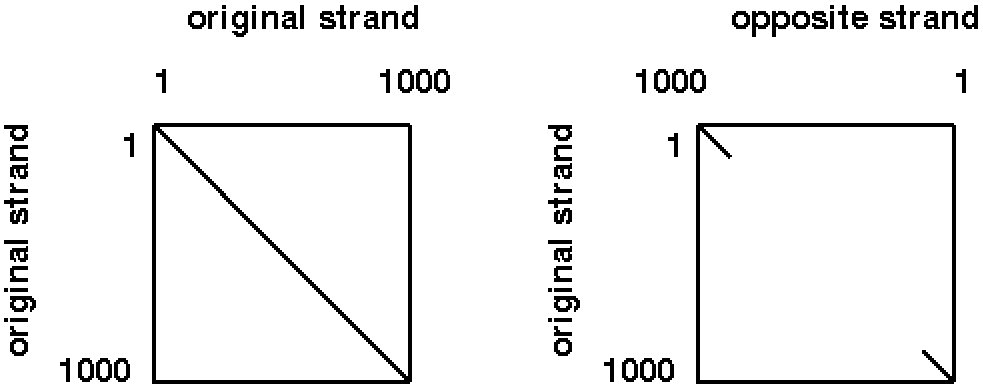
Answer any combination of questions totalling to exactly 100 points. If you answer questions totalling more than 100 points, answers will be discarded at random until the total points equal 100. This exam is worth 20% of the course grade.

Hand in this question sheets along with your exam book. All questions must be answered in the exam book. The exam sheets will be shreded after the exam.

1. (20 points) Define the following:

a) similar
b) homologous
c) analogous
d) orthologous
e) paralogous

2. (10 points) A sequence was compared with its opposite strand, showing short diagonals at each end. Explain this observation.



3. (20 points) There are many advantages to searching a protein sequence against a protein database, rather than searching a DNA sequence against a DNA database.

a) List two ways in which the protein search is faster
b) List two ways in which the protein search is more sensitive

4. (5 points) It might seem trivial to generate the opposite strand of a sequence, so simple, in fact that you might be able to do it by a simple search and replace:

original sequence	AATCGTTTGCCCCCCTA
replace A with 1	11TCGTTTGCCCCCCT1
replace G with 2	11TC2TTT2CCCCCCT1
replace T with A	11AC2AAA2CCCCCCA1
replace C with G	11AG2AAA2GGGGGGA1
replace 1 with T	TTAG2AAA2GGGGGGAT
replace 2 with C	TTAGCAAACGGGGGGAT

What is the problem with this approach?

5. (15 points) A number of alternative genetic codes have been discovered. Examples are found in mitochondria, plastids, bacteria and archea. In all of the alternative genetic codes seen so far, most of the codons code for the same amino acids as in the Standard Genetic Code, with a few codons differing. For example in some cases, a stop codon codes for an amino acid, or a codon for an amino acid is used as a stop codon. In other cases, one or two codons are reassigned to a different amino acid.

Type of search	NCBI	FASTA
a)DNA vs. DNA database	blastn	fasta3 ssearch3 (slow, full Smith-Waterman alignment)
b) protein vs. protein database	blastp	fasta3 ssearch3 (slow, full Smith-Waterman alignment)
c) protein vs. translated DNA database	tblastn	tfasta3
d) translated DNA vs. translated DNA database	tblastx	tfastx3, tfasty3
e) translated DNA vs. protein database	blastx	fastx3, fasty3 (especially well-suited for cDNAs, which often contain frameshift errors)

Yeast mitochondria use a non-starndard genetic code. Suppose you had the seqeunces for a yeast mitochondrial gene, and its corresponding protein, and wished to find homolgues in other species.

How would the difference in genetic codes affect each of the types of searches listed above?

6. (10 points) The CLUSTAL/TCOFFEE family of programs begin creating a multiple alignment by building a guide tree based on the distances between sequences. Next, sequences are added to the alignment one at a time, and the alignment is adjusted to include each new sequence. What is the main shortcoming of this strategy?

7. (15 points)
For a given restriction endonuclease, how many possible restriction fragments can be generated

a) in a **complete** digest of a **linear** molecule with **n** sites

b) in a **complete** digest of a **circular** molecule with **n** sites

c) in a **partial*** digest of a **circular** molecule with **n** sites

*A partial digest means that if there are n sites, not all sites necessarily cut, resulting in a population of fragments with 1 cut, 2 cuts ... n cuts.

8. (5 points) In pairwise DNA sequence alignments, matches are scored as +1, mismatches as -1, and gaps as -2. Why are gaps more strongly weighted than mismatches?

9. (5 points) Describe what is meant by the term "E-value", for BLAST and FASTA database searches.

10. (10 points) Part of a multiple sequence alignment for PR10 defense genes from several species of pea is shown below. There are 5 paralogues of PR10, numbered 1 through 5, found in different pea species *Pisum sativum* (Ps), *Pisum humile* (Ph), *Pisum elatius* (Pe) and *Pisum fulvum* (Pf). (Note: blanks in the alignment indicate that the clone was truncated at the 5' or 3' end. In the case of Ypr10.Ps.2 blanks are also seen within the intron because that sequence comes from a cDNA clone. Thus, no intron sequence is available.). Which regions of the alignment would be best for designing gene-specific PCR primers:

- that would only work for a particular copy of the gene?
- that would be most likely to amplify any PR10 genes from more distant species from which PR10 had never been cloned before?

Notes on PCR primers: DNA synthesizers can create degenerate primers. For example, if some genes have A at a position, and other genes have G at that position, the primer pool will have two sets of primers with either A or G at that position. Thus, in the oligo specified as AYCCTCGTA, Y stands for purine. So two primers would be produced, AGCCTCGTA, and AACCTCGTA. In practical application, degeneracies can only be at a small number of positions in a primer.

```
                     100       110       120       130       140       150       160       170       180       190
                                                                                                         <--intron
                                                    TNTTGAAGGAAANGGTGGTGCTGGAACCATCAAGAAACTCACTTTCGTTGAAGgtcagtat-
 ┌ Ypr10.Ps.1 ACTCCAAAGGTTATTGATGCCATCAAAAGTATCGAAAT.G..........C....................A............................
1│ Ypr10.Ph.1                                                    ....................A...A........................
 │ Ypr10.Pe.1                                                    ....................A............................
 └ Ypr10.Pf.1                                                    ....................A..........................c.
 ┌ Ypr10.Ps.2 ACTCCAAAGGTTATTGATGCCATCAAAAGTATCGAAAT.G..........C..A.....................................
2│ Ypr10.Ph.2                                                    ..A..............................................
 └ Ypr10.Pf.2 ACTCCAAAGGTTATTGATGCCATCAAAAGTATCGAAAT.G..........C.................A..A..........................c.
 ┌ Ypr10.Ps.3 ACTCCAAAGGTTATTGATGCCATCAAAAGTATTGAAAT.G..........C.....CC.C.......................................t
 │ Ypr10.Ph.3                                       .G..........C.....CC.C.......................................t
3│ Ypr10.Pe.2                                                 ..A..............A.............A................t
 │ Ypr10.Pe.3                                       .G..........C.....CC.C.......................................t
 └ Ypr10.Pf.3                                       .G..........C.....CC.C..G....................................t
4  Ypr10.Ps.4 GTCCCAAAGGTGATCAAGGAAGCACAAGGAGTCGAAAT.A.C........T..A...C.A..............G..AT.CA.TC......
5  Ypr10.Ps.5 GTCCCAAAAGTTGTTGATTCAATCAAGACTGTTGAAATCC..........T......C.A..C..TG.......G........T.......

                     200       210       220       230       240       250       260       270       280       290
              -------------------------------------intron---------------------------------------------->
              a-aat--atnc-t--t-tt-ac--ga-atat-c-t-t-anta-ta-tannatt-tt-a--a-t-tgnaat---t---t--tntgt-gcagATGGTGAAAC
 ┌ Ypr10.Ps.1 .....tt..a.a.ga.......tt.......g...c.c.a........aa................c..............a..................
1│ Ypr10.Ph.1 .....tt..a.a.g..g.....tt.......g...c.c.a........aa................c..............a..................
 │ Ypr10.Pe.1 .....tt..a.a.g..g.....tt.......g...c.c.a........aa................c..............a..................
 └ Ypr10.Pf.1 .....tt..a.a.g..g.....tt.......g...c.c.a........aa................c..............a................C.
 ┌ Ypr10.Ps.2                                                                                           ..........
2│ Ypr10.Ph.2 .....tt..a.a.g..g.....tt.......g...c.c.a........aa................c..............a..................
 └ Ypr10.Pf.2 .....tt..a.a.g..g.....tt.......g...c.c.a........aa................c..............a..................
 ┌ Ypr10.Ps.3 .g.......t.........c.t....t......a.....t..c..g..tt...g..t.tg.t.g..a...gaa.caa.tg.g...t.....C........
 │ Ypr10.Ph.3 .g.......t.........g......t......a.....t..c..g..tt...g....tg.c.g..a...gaa.caa.gg.g...t..............
3│ Ypr10.Pe.2 .g.......t.........g......t......a.....t..c..g..tt...g..t.tg.c.g..a...gaa.caa.gg.g...t..............
 │ Ypr10.Pe.3 .g.......t.........g......t......a.....t..c..g..tt...g..t.tg.c.g..a...gaa.caa.gg.g...t..............
 └ Ypr10.Pf.3 .g.g.....t.........g.t....t......a.....t..c..g..tt...g..t.tg.t.g..a...gaa.caa.gg.g...t..............
4  Ypr10.Ps.4                                                                                           ....AA....
5  Ypr10.Ps.5                                                                                           GA..AC.G..

                     300       310       320       330       340       350       360       370       380       390
              CAAGNATGTGTTGCACAAAGTGGAGTTAGTAGATGNTGCTAACTTGGCTTACAACTATAGCATAGTTGGNGGTGTTGGANTTCCAGACACAGTTGAGAAG
 ┌ Ypr10.Ps.1 ...AC..............................T.................................T.........T....................
1│ Ypr10.Ph.1 ...AC..............................T.................................T.........T.C..................
 │ Ypr10.Pe.1 ...AC..............................T.................................T.........T....................
 └ Ypr10.Pf.1 ....C..............................T.................................T.........T....................
 ┌ Ypr10.Ps.2 ...AC..............................T.................................T.........T....................
2│ Ypr10.Ph.2 ....C....C.........................T.................................T.....C...T....................
 └ Ypr10.Pf.2 ....C....C........................CT.................................T.........T....................
 ┌ Ypr10.Ps.3 ....T.......A......................A........G...AA.......C...........A.........C....G...............
 │ Ypr10.Ph.3 ....T.......A......................A........G..AAA.......C...........A.........C....G...............
3│ Ypr10.Pe.2 ....T.......A......................A........G..AAA.......C...........A.........C....G...............
 │ Ypr10.Pe.3 ....T.......A................C.....A........G..AAA.......C...........A.........C....G...............
 └ Ypr10.Pf.3 ....T.......A......................A........G...AA.......C...........A.........C....G...............
4  Ypr10.Ps.4 ...CT.....C.A......C.A..CGC...T....AA..A.....T.G.........C...T....A..A..ACCA..GC.A.AT..A.GTT.A.....A
5  Ypr10.Ps.5 .TT.T.C..............T..AGCCA.T....A...A..G..T.AA..T..T..C..T........A.....C..TA.AT......T..........

                     400       410       420       430       440       450       460       470       480       490
              ATCTCATTNGAGGCTAAACTGTCTGCAGGACCAAATGGAGGATCCATTGCAAAGCTGAGTGTGAAATATTACACAAAAGGTGAT---GCTGCTCCTANTG
 ┌ Ypr10.Ps.1 ........C...........................................................C.T..........................G..
1│ Ypr10.Ph.1 ........C.............................................T..........................................C..
 │ Ypr10.Pe.1 ........T............................................A..............C.T..........................C..
 └ Ypr10.Pf.1 ........C..........................................................................................C..
 ┌ Ypr10.Ps.2 ........T...........................................................................GCT..........C..
2│ Ypr10.Ph.2 ........C..........................................................................................C..
 └ Ypr10.Pf.2 ........T..........................................................................................C..
 ┌ Ypr10.Ps.3 .....G..T.........T....................................................T..C...............AT.....G..
 │ Ypr10.Ph.3 .....T..T.........T....................................................T..C...............AT.....G..
3│ Ypr10.Pe.2 .....G..C.........T.......................................................................AT.....G..
 │ Ypr10.Pe.3 .....G..C.........T.......................................................................AT.....G..
 └ Ypr10.Pf.3 .....T..T..................A.G....G...........................C........T..C...............AT.....G..
4  Ypr10.Ps.4 G.TG....C...A.A.TTA.T.TG..T..TT.TG.C..T........C.TT...A.ATC..........C....C..............A....TATC..
5  Ypr10.Ps.5 ..A.....T.....C...T...T..A...T..............TG...GT...A...T...T......C.T........A.........AAG....T..

                     500       510       520       530       540       550       560       570       580
              AAGAGNAACTCAAGANTGNCAAAGCTAAGGGNGATGNNNTTNTCAANGCNNTNGANNNTTNCNNTNTGGCNNATCCTNNTTACNANTNAN
 ┌ Ypr10.Ps.1 .....C.........C..A............G....GTC..T....G..TC.T..GGG..A.TG.T....TC.....GA....A.C.A.A
1│ Ypr10.Ph.1 .....C.........C..A............G...
 │ Ypr10.Pe.1 .....C.........C..A............G...
 └ Ypr10.Pf.1 .....C.........C..A............G...
 ┌ Ypr10.Ps.2 .....C.........G..A............T....GTC..T....G..TC.T..GCG..A.TG.T....TC.....GA....A.C.A.A
2│ Ypr10.Ph.2 .....C.........A..A............G....GTC..T....G..TC.T..GGG..A.TG.T....TC.....GA....A.C.A.A
 └ Ypr10.Pf.2 .....C.........C..A............G....GTC..T....A..TC.T..GGG..G.TG.T....TC.....GA....A.C.A.A
 ┌ Ypr10.Ps.3 .....G..A......A..G......C..A..T..A.GTA..T....G..TC.T..AGG..A.TG.G....TA.....GA....A.C.A.A
 │ Ypr10.Ph.3 .....G..A..
3│ Ypr10.Pe.2 .....G..A......A..G......C..A..T..A.GTA..T....G..TC.T..AGG..A.TG.G....TA.....GA....A.C.A.A
 │ Ypr10.Pe.3 .....G..A..
 └ Ypr10.Pf.3 .....G..A..
4  Ypr10.Ps.4 .T.CAGTT.GTG.TGAAACA..G..C..A..AAC..GAC..A....G..CA.A..AGG..A.GT.T....AA.....GG....T.A.T.GT
5  Ypr10.Ps.5 ..A..G..G.TG..GAA.G............T....CTC..T....G..CA.T..GGC..A.GT.T....CA.....AA....A.C.G.TC
```