

HOMEWORK #11 KEY

1. All 3 clones overlap and are likely part of the same gene (see Appendix 1 for alignment). The order from 5'-3' is clone_z, clone_y, and clone_x. Appendix 2 shows how the 3 sequences can be compiled into a single sequence.
2. There is one ambiguity in that clone_y has an extra G residue at position 384 which is not found in clone_x. It is usually possible just to re-examine the original data of the two sequences and see if there are problems with one of the sequences in the region of question. It may also be possible to sequence more clones in this region or to re-sequence both of the clones. Sometimes examining the reading frames will help resolve the problem since the addition or loss of a nucleotide will result in a frame shift.
3. Analysis of the reading frames of the compiled sequence identified a single long open reading frame starting with a Met. Appendix 3 shows the translation. In addition, searching databases reveals that the size of this reading frame is similar to other homologues.
4. Searching the databases revealed that this sequence is a casein kinase-1 (see appendix 4), a particular type of protein kinase.

APPENDIX 1

Alignment (DIALIGN format):

```

clone_X 1 -----
clone_Y 1 -----
clone_Z 1 ttgtacctgt atatgtatta atactatatt gctgcagata aaattttata

clone_X 1 -----
clone_Y 1 -----
clone_Z 51 attcaaaatt taaaaaatgg aaattagagt ggcaaataaa tatgcttttag

clone_X 1 -----
clone_Y 1 -----
clone_Z 101 gaaaaaaatt agggagtggt tcctttggtg atatatatgt tgccaaggat

clone_X 1 -----
clone_Y 1 -----
clone_Z 151 attgtaacga tggaagaatt tgctgtaaaa ttagaatcaa cacgatcaaa

clone_X 1 -----
clone_Y 1 -----
clone_Z 201 acatcccaa ctattgtatg agtcaaaatt atataaaata ttagggaggag

clone_X 1 -----
clone_Y 1 -----
clone_Z 251 gaattgggtg acctaaagta tattggtacg gtatagaagg ggattttact

clone_X 1 -----
clone_Y 1 -----
clone_Z 301 atcatggttc ttgatttatt aggcccatcc cttgaagatt tatttacctt

clone_X 1 -----
clone_Y 1 -----
clone_Z 351 atgtaataga aaattttctt taaagactgt ccttatgaca gcagaTCAAA

clone_X 1 -----
clone_Y 6 TGTTAAATAG AATTGAATAT GTGCATTCAA AAAATTTTAT ACATAGAGAT
clone_Z 401 TGTTAAATAG AATTGAATAT GTGCATTCAA AAAATTTTAT ACATAGAGAT

clone_X 1 -----
clone_Y 56 ATTAAACCAG ATAACCTTTT AATAGGACGA GGGAAAAAAG TTAaccttaat
clone_Z 451 ATTAAACCAG ATAACCTTTT AATAGGACGA GGGAAAAAAG TTA-----

clone_X 1 -----
clone_Y 106 acatattatt gattttgggt tagcaaaaaa atatagagat tcaagatcac
clone_Z 494 -----

clone_X 1 -----
clone_Y 156 atactcatat tccatataaa gaagggaaaa atttaacagg aacagcaaga
clone_Z 494 -----

clone_X 1 -----
clone_Y 206 tatgcaagta taaatacaca cttaggaatt gaacaatctc gtagggatga
clone_Z 494 -----

```

```

clone_X 1 -----
clone_Y 256 tattgaagcc cttggatatg ttctcatgta ttttctaaga ggaagtttac
clone_Z 494 -----

clone_X 1 ----GCAAGG TCTAAAGGCT ATATCTAAGA AGGATAAATA TGATAAAATT
clone_Y 306 catgGCAAGG TCTAAAGGCT ATATCTAAGA AGGATAAATA TGATAAAATT
clone_Z 494 -----

clone_X 47 ATGGAAAAAA AAATATCAAC ATCAGTGG-A AGTTTTGTGT AGAAATGCAA
clone_Y 356 ATGGAAAAAA AAATATCAAC ATCAGTGGgA AGTTTTGTGT AGAAATGCAA
clone_Z 494 -----

clone_X 96 GCTTCGAATT CGTAACATAT TTAAATTATT GCCGATCATT AAGATTTGAA
clone_Y 406 GCTTCGAATT CGTAACATAT TTAAATTATT GCCGATCATT AAGATTTGAA
clone_Z 494 -----

clone_X 146 GATAGACCCG ATTATACATA TTTAAGAaga cttttaaagg acttgtttat
clone_Y 456 GATAGACCCG ATTATACATA TTTAAGA--- -----
clone_Z 494 -----

clone_X 196 aagagagggga tttacttatg actttttatt tgactggaca tgtgtatatg
clone_Y 483 -----
clone_Z 494 -----

clone_X 246 cctccgaaaa ggataagaag aagatgtag aaaataaaaa ccgttttgat
clone_Y 483 -----
clone_Z 494 -----

clone_X 296 caaacagcag atcaagaagg aagagatcaa cgaaataatt gatatgcctt
clone_Y 483 -----
clone_Z 494 -----

clone_X 346 atgcaacatt tttagttctt tgatgtaatt catcttaaca ggtaccatta
clone_Y 483 -----
clone_Z 494 -----

clone_X 396 tgtatatgca gcatataatc aaaa
clone_Y 483 -----
clone_Z 494 -----

```

Legend. The three sequences are align. Regions of identity are in capital letters. Clone_Z is the beginning. Then clone_Y begins at residue 396 in clone_Z. Then clone_X begins at residue 310 in clone_Y. There is a G at residue 384 in clone_Y which is not found in clone_X (arrow).

APPENDIX 2

>clone_Z

TTGTACCTGTATATGTATTAATACTATATTGCTGCAGATAAAAATTTTATAATTCAAAATT
TAAAAAATGGAAATTAGAGTGGCAAATAAATATGCTTTAGGAAAAAAATTAGGGAGTGGT
TCCTTTGGTGATATATATGTTGCCAAGGATATTGTAACGATGGAAGAATTTGCTGTAAAA
TTAGAATCAACACGATCAAAACATCCCCAACTATTGTATGAGTCAAAATTATATAAAATA
TTAGGAGGAGGAATTGGTGTACCTAAAGTATATTGGTACGGTATAGAAGGGGATTTTACT
ATCATGGTTCTTGATTTATTAGGCCCATCCCTTGAAGATTTATTTACCTTATGTAATAGA
AAATTTTCTTTAAAGACTGTCCTTATGACAGCAGATCAAATGTTAAATAGAATTGAATAT
GTGCATTCAAAAAATTTTATACATAGAGATATTAACCAGATAAATTTTAAATAGGACGA
GGGAAAAAAGTT

>clone_Y

TCAAATGTTAAATAGAATTGAATATGTGCATTCAAAAAATTTTATACATAGAGATATTTAA
ACCAGATAACTTTTTAATAGGACGAGGGAAAAAAGTTCCCTTAATACATATTATTGATTT
TGGTTTAGCAAAAAAATATAGAGATTCAAGATCACATACTCATATTCCATATAAAGAAGG
GAAAAATTTAACAGGAACAGCAAGATATGCAAGTATAAATACACACTTAGGAATTGAACA
ATCTCGTAGGGATGATATTGAAGCCCTTGGATATGTTCTCATGATTTTTCTAAGAGGAAG
TTTACCATCAAGGTCTAAAGGCTATATCTAAGAAGGATAAATATGATAAAATTATGGA
AAAAAATATCAACATCAGTGGGAAGTTTTGTGTAGAAATGCAAGCTTCGAATTCGTAA
CATATTTAAATTATTGCCGATCATTAAGATTTGAAGATAGACCCGATTATACATATTTAA
GA

>clone_X

CAAGGTCTAAAGGCTATATCTAAGAAGGATAAATATGATAAAATTATGGAAAAAAT
ATCAACATCAGTGGAAAGTTTTGTGTAGAAATGCAAGCTTCGAATTCGTAACATATTTAAA
TTATTGCCGATCATTAAGATTTGAAGATAGACCCGATTATACATATTTAAGAAGACTTTT
AAAGGACTTGTTTATAAGAGAGGGATTTACTTATGACTTTTTATTTGACTGGACATGTGT
ATATGCCTCCGAAAAGGATAAGAAGAAGATGTTAGAAAATAAAAACCGTTTTGATCAAAC
AGCAGATCAAGAAGGAAGAGATCAACGAAATAATTGATATGCCTTATGCAACATTTTTAG
TTCTTTGATGTAATTCATCTTAACAGGTACCATTATGTATATGCAGCATATAATCAAAA

Legend. Boxes and lines indicate the overlapping residues in the 3 clones. The sequences can be compiled into a single sequence by removing the overlapping residues. The extra G in clone_Y is underlined and not included in the compiled sequence.

APPENDIX 3

```
ttgtacctgtatatgtattaataactatattgctgcagataaaaatTTTataattcaaaatt
taaaaaatggaaattagagatggcaataaatatgcttttaggaaaaaaattagggagtggt
  M E I R V A N K Y A L G K K L G S G
tcctttggatgatataatggtgccaaggatattgtaacgatggaagaatttgctgtaaaa
  S F G D I Y V A K D I V T M E E F A V K
ttagaatcaacacgatcaaaacatccccaactattgtatgagtcaaaattatataaaata
  L E S T R S K H P Q L L Y E S K L Y K I
ttaggaggaggaattgggtgtacctaaagtatattgggtacggtatagaaggggattttact
  L G G G I G V P K V Y W Y G I E G D F T
atcatggttcttgatatttagggccatcccttgaagatTTTattttaccttatgtaataga
  I M V L D L L G P S L E D L F T L C N R
aaatTTTctTTaaagactgtccttatgacagcagatcaaatgtTaaatagaattgaatat
  K F S L K T V L M T A D Q M L N R I E Y
gtgcattcaaaaaatTTTatacatagagatattaaaccagataaactTTTtaataggacga
  V H S K N F I H R D I K P D N F L I G R
gggaaaaaagttaccttaatacatatttattgattttggtttagcaaaaaaatatagagat
  G K K V T L I H I I D F G L A K K Y R D
tcaagatcacatactcatattccatataaagaagggaaaaatTTaacaggaacagcaaga
  S R S H T H I P Y K E G K N L T G T A R
tatgcaagtataaatacacacttaggaattgaacaatctcgtagggatgatattgaagcc
  Y A S I N T H L G I E Q S R R D D I E A
cttgatattgcttcatgtatttttctaagaggaagtttaccatggcaaggtctaaaggct
  L G Y V L M Y F L R G S L P W Q G L K A
atatctaagaaggataaataatgataaaattatggaaaaaaatcaacatcagtgga
  I S K K D K Y D K I M E K K I S T S V E
gTTTTgtgtagaatgcaagcttcgaattcgtaacatattTaaattattgccgatcatta
  V L C R N A S F E F V T Y L N Y C R S L
agatttgaagatagaccggattatacatattTaaagaagactTTTaaaggacttgtttata
  R F E D R P D Y T Y L R R L L K D L F I
agagagggatttacttatgactTTTTatttgactggacatgtgtatatgcctccgaaaag
  R E G F T Y D F L F D W T C V Y A S E K
gataagaagaagatgTtagaaaataaaaaaccgTTTTgatcaaacagcagatcaagaagga
  D K K K M L E N K N R F D Q T A D Q E G
agagatcaacgaaataattgatatgccttatgcaacattTTTtagttctttgatgtaattc
  R D Q R N N *
atcttaacaggtaccattatgtatatgcagcatataatcaaaa
```

>protein_xyz

```
MEIRVANKYALGKKLGSFSGDIYVAKDIVTMEEFVAVKLESTRSKHPQLLYESKLYKILG
GGIGVPKVYWGIEGDFITIMVLDLLGPSLEDLFTLCNRKFSLKTVLMTADQMLNRIEYVH
SKNFIHRDIKPDNFLIGRKKVTLIHIIDFGLAKKYRDSRSHTHIPYKEGKNLTGTARYA
SINTHLGIEQSRDDIEALGYVLMYFLRGS LPWQGLKAI SKKDKYDKIMEKKI STSVEVL
CRNASFEFVTYLYNCRSLRFEDRPDYTYLRLLKDLFIREGFTYDFLFDWTCVYASEKDK
KKMLENKNRFDQTADQEGRDQRNN
```

Legend. Translation of the compiled nucleotide sequence showing the only long open reading frame beginning with a Met and ending with a Stop. No long open reading frames were found in the other five possible reading frames (not shown).

APPENDIX 4

Results of blastp search using the translated sequence indicates that sequence is a casein kinase type 1:

Sequences producing significant alignments:	Score (bits)	E Value
gi 23508567 ref NP_701236.1 casein kinase 1 [Plasmodium fa...	607	e-172
gi 6016422 sp O15726 KC1_PLAF4 CASEIN KINASE I >gi 2388708 ...	605	e-172
gi 23482049 gb EAA18147.1 casein kinase i [Plasmodium yoel...	603	e-171
gi 13122442 dbj BAB32922.1 casein kinasel epsilon-2 [Rattu...	437	e-121
gi 20804657 dbj BAB92346.1 putative casein kinase I [Oryza...	436	e-121
gi 9650968 dbj BAB03473.1 casein kinase 1 epsilon-3 [Rattu...	432	e-120
gi 15218569 ref NP_177415.1 putative casein kinase I; prot...	431	e-120
gi 7949025 ref NP_038795.2 casein kinase 1 epsilon; CKI, e...	430	e-119
gi 4503093 ref NP_001885.1 casein kinase 1 epsilon; CKI-ep...	429	e-119
gi 20544149 ref NP_620690.1 casein kinase 1, delta isoform...	429	e-119
gi 20149530 ref NP_001884.2 casein kinase 1, delta isoform...	429	e-119
gi 13928846 ref NP_113805.1 casein kinase 1 epsilon [Rattu...	428	e-119
gi 7644318 gb AAF65549.1 AF242536_1 casein kinase I epsilon...	428	e-119
gi 7512331 pir G01876 casein kinase I delta - human >gi 88...	428	e-119
gi 20070679 gb AAH26127.1 casein kinase 1, epsilon [Mus mu...	428	e-119
gi 15236114 ref NP_194340.1 Col-0 casein kinase I-like pro...	427	e-118
gi 20544147 ref NP_082150.1 casein kinase 1, delta isoform...	427	e-118
gi 20544145 ref NP_620693.1 casein kinase 1, delta isoform...	427	e-118