Spring 2014 Genomics Exam #1
Genomic Sequences

There is no time limit on this test, though I don't want you to spend too much time on it. I have tried to design an exam that will take less time that exams in the past. You do not need to read any additional papers other than the ones I send to you. There are 4 pages, including this cover sheet, for this test. There are no Discovery Questions on this exam. You are not allowed discuss the test with anyone until all exams are turned in at 10:30 am on Wednesday February 12. **ELECTRONIC COPIES OF YOUR EXAM ANSWERS ARE DUE BY 10:30 am ON WEDNESDAY FEBRUARY 12**. You may use a calculator, a ruler, your notes, the book, and the internet. You may work on this exam in as many blocks of time as you want. Submit your electronic version before 10:30 am (eastern time zone).

The **answers to the questions must be typed in a Word file and emailed to me as an attachment**. Be sure to backup your test answers just in case (I suggest a thumb drive or other removable medium). You will need to capture screen images as a part of your answers which you may do without seeking permission since your test answers will not be in the public domain. Remember to explain your thoughts in *your* own words and use screen shots to support your answers. Screen shots without ***your*** words are worth very few points. Support your answers with data using screen shots liberally.

DO NOT READ or DOWNLOAD ANY PAPERS FOR THIS EXAM. You may search and read abstracts. RELY ON YOUR EXPERIENCE, AND YOUR SKILLS. Spell out your logic for each answer.

-3 pts if you do not follow this direction.
Please do not write or type your name on any page other than this cover page.
Staple all your pages (INCLUDING THE TEST PAGES) together when finished with the exam.

Name (please type): **ANSWER KEY**

Write out the full pledge and sign (electronic signature is ideal):

How long did this exam take you to complete?

20 pts

1) I want you to analyze some sequences. All of them can be found in the Word file called "Mystery_One.docx".

a) Translate **>First_Sequence**. How many amino acids does the encoded protein have? (5 pts)

From 151 to 11133.
Translation 3660 a.a. MW=422892.6499999991

```
     1  M  S  A  H  V  L  W  Y  E  d  V  E  D  D  Y  E  R  E  D  V  Q  K  K  T  F  25
   151 ATGTCTGCTCACGTGCTTTGGTATGAGGAcGTGGAAGATGATTATGAACGGGAAGATGTTCAAAAGAAAACATTC 225

    26  T  K  W  I  N  A  Q  F  A  K  C  G  R  R  C  I  E  D  L  F  N  D  F  R  D  50
   226 ACGAAATGGATAAATGCACAGTTTGCTAAGTGTGGAAGACGTTGCATTGAAGATCTTTTTAATGATTTTCGAGAT 300

    51  G  R  K  L  L  E  L  L  E  C  L  T  G  Q  K  I  A  K  E  K  G  S  T  R  V  75
   301 GGACGAAAACTTCTGGAGCTCCTGGAATGCCTTACAGGCCAAAAAATTGCAAAAGAAAAGGGCTCCACAAGAGTT 375

    76  H  A  L  N  N  V  N  K  A  L  Q  I  L  Q  R  N  N  V  D  L  V  N  I  G  S  100
   376 CATGCTCTGAACAACGTCAACAAAGCACTGCAGATTTTGCAAAGAAATAATGTTGATTTGGTAAATATTGGGAGC 450

   101  S  D  I  V  D  G  N  H  K  L  T  L  G  L  I  W  N  I  I  L  H  W  Q  V  K  125
   451 TCAGACATTGTGGACGGCAATCATAAACTGACCCTTGGTTTGATCTGGAATATAATCCTCCACTGGCAGGTCAAA 525

   126  D  V  M  K  N  I  M  A  G  L  Q  Q  T  N  S  E  K  I  L  L  S  W  V  R  Q  150
   526 GATGTAATGAAAAACATTATGGCTGGACTGCAGCAGACAAACAGTGAGAAGATTCTGCTGAGCTGGGTCCGTCAA 600

   151  S  T  R  N  Y  P  Q  V  N  V  I  N  F  T  S  S  W  S  D  G  L  A  F  N  A  175
   601 TCAACTCGTAATTACCCACAGGTCAATGTTATCAATTTCACCAGTAGCTGGTCTGATGGATTGGCTTTCAATGCA 675

   176  L  L  H  S  H  R  P  D  L  F  D  W  N  A  V  A  S  Q  Q  S  P  V  Q  R  L  200
   676 CTCCTTCACAGTCACAGACCAGACCTGTTTGATTGGAATGCTGTTGCTTCTCAGCAGTCACCTGTGCAACGATTA 750

   201  D  H  A  F  N  I  A  R  Q  H  L  G  I  E  K  L  L  D  P  E  D  V  A  T  A  225
   751 GACCATGCATTTAACATAGCCAGGCAACACCTGGGCATAGAGAAGCTCCTTGATCCTGAAGACGTTGCAACTGCC 825

   226  C  P  D  K  K  S  I  L  M  Y  V  T  S  L  F  Q  V  L  P  Q  Q  V  T  M  E  250
   826 TGTCCAGATAAGAAGTCCATCTTAATGTATGTGACTTCCCTCTTCCAAGTTCTGCCACAGCAAGTCACTATGGAG 900

   251  A  I  R  E  V  E  M  L  P  R  H  S  R  V  T  T  E  E  H  I  Q  V  H  H  Q  275
   901 GCCATCAGGGAGGTGGAAATGCTGCCACGGCACTCAAGGGTCACTACAGAGGAGCACATACAAGTACATCATCAA 975

   276  Q  H  F  S  Q  E  I  T  V  N  I  P  Q  R  P  S  P  S  P  K  P  R  F  K  S  300
   976 CAGCATTTTTCACAAGAGATCACAGTCAATATACCCCAGAGACCTTCACCTTCTCCTAAACCACGGTTCAAAAGT 1050

   301  Y  A  Y  A  Q  T  A  Y  V  I  P  P  D  Q  K  R  R  Q  V  P  P  Q  F  L  E  325
  1051 TATGCATATGCACAAACTGCGTATGTCATACCCCCTGACCAAAAAAGGAGGCAGGTTCCTCCACAGTTTTTAGAA 1125

   326  T  V  E  K  R  T  Y  T  T  T  V  M  R  S  E  M  D  L  D  S  Y  Q  T  A  L  350
  1126 ACTGTTGAAAAAAGAACGTATACCACCACAGTGATGAGGTCCGAAATGGATCTTGACAGCTACCAAACAGCTTTA 1200

   351  E  E  V  L  T  W  L  L  S  A  E  D  A  L  Q  A  Q  G  D  I  S  S  D  V  E  375
  1201 GAAGAAGTACTCACGTGGCTTCTCTCTGCTGAAGATGCATTACAAGCACAAGGGGATATATCCAGTGATGTAGAA 1275

   376  V  V  K  E  Q  F  H  T  H  E  G  F  M  M  E  L  T  A  H  Q  G  R  V  G  N  400
  1276 GTTGTTAAAGAGCAGTTTCATACTCATGAGGGTTTCATGATGGAATTAACAGCTCACCAAGGCCGTGTTGGTAAT 1350

   401  V  L  Q  V  G  S  Q  L  L  A  M  G  K  L  S  D  D  E  E  N  E  I  Q  E  Q  425
  1351 GTTTTGCAAGTTGGAAGTCAACTTCTAGCAATGGGAAAGCTTTCTGATGATGAAGAAAATGAAATACAAGAACAA 1425

   426  M  N  L  L  N  S  R  W  E  S  L  R  V  A  S  M  E  K  Q  S  N  L  H  K  I  450
  1426 ATGAATCTACTCAATTCTCGATGGGAAAGTCTCAGGGTTGCAAGTATGGAAAAACAAAGCAATTTACATAAAATC 1500

   451  L  M  D  L  Q  N  Q  Q  L  A  Q  L  A  D  W  L  T  K  T  E  E  R  T  K  K  475
  1501 CTAATGGATCTGCAGAATCAGCAACTAGCCCAGTTAGCTGACTGGCTAACAAAAACAGAGGAAAGGACAAAGAAG 1575

   476  I  D  S  E  P  L  G  P  D  L  E  D  L  K  R  Q  V  E  E  H  K  A  F  Q  D  500
  1576 ATAGATTCAGAGCCCTTAGGTCCGGATCTAGAGGACCTAAAGCGTCAAGTAGAAGAACATAAGGCATTTCAAGAT 1650

   501  D  L  E  Q  E  Q  V  K  V  N  S  L  T  H  M  V  V  V  V  D  E  N  S  G  D  525
  1651 GATTTGGAACAGGAACAAGTGAAGGTGAACTCTCTAACCCATATGGTGGTGGTGGTGGATGAAAACAGCGGAGAC 1725

   526  K  A  T  A  A  L  E  E  Q  L  Q  H  F  G  S  R  W  A  A  I  C  R  W  T  E  550
  1726 AAAGCGACTGCTGCACTGGAAGAGCAGCTTCAGCACTTTGGAAGCCGGTGGGCAGCCATCTGCAGATGGACAGAA 1800
```

```
 551  D  R  W  V  L  L  Q  D  I  L  R  K  W  Q  H  F  A  E  E  Q  C  L  F  D  A  575
1801 GATAGATGGGTCCTTCTGCAAGATATTCTTAGAAAATGGCAGCACTTTGCAGAAGAGCAGTGCCTTTTTGATGCA 1875

 576  W  L  T  E  K  E  G  S  L  S  K  I  Q  T  S  D  F  K  D  E  N  E  M  L  T  600
1876 TGGCTTACTGAGAAAGAAGGATCGCTTAGCAAAATCCAGACAAGTGACTTTAAAGATGAAAATGAGATGCTGACT 1950

 601  S  L  R  K  L  A  I  L  K  G  D  I  E  M  K  K  Q  M  M  S  K  L  K  S  L  625
1951 AGTCTACGCAAATTAGCTATCCTAAAAGGAGATATAGAAATGAAAAAGCAAATGATGAGTAAACTGAAGTCGCTC 2025

 626  S  R  D  L  L  V  A  V  K  N  K  A  V  A  Q  K  L  E  S  R  L  E  N  F  A  650
2026 AGCAGAGACCTCCTGGTAGCTGTGAAAAATAAAGCAGTGGCTCAGAAGCTGGAAAGCCGGCTTGAAAATTTTGCC 2100

 651  Q  R  W  D  S  L  V  Q  K  L  E  S  D  S  K  Q  V  S  Q  A  V  T  T  T  Q  675
2101 CAGCGCTGGGATAGCCTTGTGCAGAAGTTGGAGAGCGATTCTAAGCAGGTTTCGCAGGCTGTCACTACCACTCAG 2175

 676  T  S  L  T  Q  T  T  V  M  E  T  V  T  M  V  T  T  R  E  Q  I  L  V  K  H  700
2176 ACATCACTAACACAGACAACTGTAATGGAAACTGTCACTATGGTGACCACAAGAGAACAAATCCTGGTTAAGCAT 2250

 701  A  K  E  E  L  P  P  P  P  P  H  K  K  R  Q  L  L  V  D  S  E  I  R  K  R  725
2251 GCTAAAGAGGAGCTCCCACCACCACCTCCCCACAAAAAGAGGCAACTCCTTGTGGATTCGGAAATTAGGAAGAGG 2325

 726  F  D  S  D  T  T  E  L  H  S  W  M  T  R  S  E  A  V  L  Q  S  P  E  F  A  750
2326 TTTGACTCGGATACAACAGAACTCCATAGCTGGATGACACGTTCAGAAGCAGTACTTCAGAGTCCTGAGTTTGCA 2400

 751  I  Y  R  K  E  G  N  L  S  D  L  R  E  R  V  N  A  I  Q  R  E  K  P  E  K  775
2401 ATATATCGAAAAGAAGGAAATCTCTCAGATCTCAGAGAAAGAGTCAATGCCATCCAGCGAGAAAAGCCTGAGAAG 2475

 776  Y  R  K  L  Q  D  A  S  R  S  A  E  A  L  V  E  Q  M  V  N  E  G  L  N  A  800
2476 TACAGAAAGCTGCAAGATGCCAGCAGATCAGCTGAGGCCCTGGTGGAACAGATGGTGAATGAGGGTCTGAATGCT 2550

 801  D  N  I  R  Q  A  S  E  Q  L  K  S  R  W  I  E  F  C  Q  L  L  S  E  R  L  825
2551 GACAACATTAGACAAGCCTCAGAGCAGCTGAAGAGTCGCTGGATTGAGTTCTGTCAGTTACTGAGTGAAAGACTG 2625

 826  V  W  L  E  Y  Q  N  S  I  I  D  F  Y  S  Q  L  Q  R  L  E  Q  T  A  I  T  850
2626 GTATGGTTGGAGTACCAAAATAGCATCATTGACTTCTATTCCCAGCTGCAGCGACTGGAGCAGACAGCAATTACA 2700

 851  A  E  N  W  L  K  A  Q  P  T  P  A  T  D  P  A  T  V  K  I  Q  L  E  K  C  875
2701 GCAGAAAACTGGCTGAAAGCACAACCCACACCAGCAACAGATCCTGCTACAGTAAAAATTCAGTTGGAAAAATGC 2775

 876  K  D  E  I  I  R  M  S  T  L  Q  P  Q  I  E  R  L  K  A  Q  S  Q  A  L  K  900
2776 AAGGACGAAATCATCCGAATGTCAACTCTTCAGCCTCAAATTGAACGGCTTAAGGCTCAGAGTCAAGCTCTGAAA 2850

 901  E  K  E  Q  C  P  V  F  L  D  A  D  L  A  A  F  T  S  H  F  K  Q  I  L  A  925
2851 GAGAAAGAACAATGCCCAGTGTTTCTGGATGCTGACCTTGCTGCTTTTACCAGCCACTTCAAACAAATACTTGCT 2925

 926  D  M  H  T  R  E  K  Q  L  Q  T  I  F  D  S  L  P  P  A  R  Y  K  D  T  V  950
2926 GACATGCACACCAGAGAAAAGCAACTACAGACCATTTTTGACAGTTTGCCTCCTGCACGCTATAAAGACACAGTG 3000

 951  T  T  I  L  S  W  I  Q  Q  S  E  T  K  V  S  I  P  P  V  A  V  A  E  Y  E  975
3001 ACTACTATACTTTCATGGATCCAGCAGTCAGAAACTAAAGTCTCCATACCTCCAGTTGCAGTGGCTGAATATGAA 3075

 976  I  M  E  Q  R  L  G  E  L  K  A  L  Q  S  S  L  Q  E  Q  Q  K  G  L  K  Y  1000
3076 ATCATGGAACAGAGACTCGGGGAGCTCAAGGCTCTACAAAGTTCTCTGCAAGAGCAGCAAAAAGGCCTGAAATAT 3150

1001  L  N  T  T  V  E  D  L  S  R  K  A  P  A  E  V  S  Q  K  Y  R  S  E  V  E  1025
3151 CTCAACACAACTGTTGAAGACTTGTCTAGGAAAGCCCCTGCAGAAGTCAGCCAGAAATACCGATCAGAGGTTGAG 3225

1026  L  I  V  G  R  W  K  K  L  S  S  Q  L  V  E  H  C  Q  K  L  E  D  L  M  T  1050
3226 TTGATCGTTGGCCGCTGGAAGAAGCTGTCATCACAGTTGGTGGAACATTGCCAGAAACTGGAGGATCTTATGACT 3300

1051  K  L  Q  R  F  Q  N  D  T  K  T  L  K  K  W  M  A  E  V  D  V  F  L  K  E  1075
3301 AAACTCCAACGATTCCAGAATGACACAAAAACATTGAAGAAGTGGATGGCTGAAGTAGATGTCTTTCTGAAGGAG 3375

1076  E  W  P  A  L  G  D  S  E  A  L  E  K  Q  L  E  Q  C  T  A  L  V  N  D  I  1100
3376 GAATGGCCTGCTCTTGGTGATTCAGAAGCTCTGGAAAAGCAACTTGAGCAGTGTACAGCTTTAGTAAATGATATC 3450

1101  Q  T  I  Q  P  S  L  N  S  V  N  E  I  G  K  K  M  K  R  E  A  E  P  E  F  1125
3451 CAGACTATCCAGCCGAGTTTGAACAGTGTTAATGAGATTGGGAAGAAAATGAAGAGGGAAGCAGAGCCAGAATTT 3525

1126  A  S  R  I  A  T  E  L  K  D  L  N  A  Q  W  E  H  I  C  Q  Q  A  H  A  K  1150
3526 GCTTCCAGAATAGCAACAGAACTAAAGGATCTCAATGCTCAATGGGAACATATTTGCCAACAGGCACATGCTAAG 3600

1151  K  A  A  L  K  G  G  L  D  K  T  V  S  L  R  K  D  L  S  E  M  H  E  W  I  1175
3601 AAGGCAGCATTAAAAGGTGGTTTGGATAAGACTGTGAGCCTCAGAAAGGATTTGTCAGAGATGCATGAATGGATA 3675

1176  T  Q  A  E  E  E  Y  L  E  R  D  F  E  Y  K  T  P  E  E  L  Q  K  A  V  E  1200
```

```
3676 ACACAAGCTGAGGAAGAATATCTGGAAAGAGATTTTGAGTATAAAACACCCGAAGAATTACAGAAAGCTGTTGAA 3750

1201  E  L  K  R  A  K  E  D  A  M  Q  K  E  V  K  V  K  L  I  T  D  S  V  N  N  1225
3751 GAACTGAAGAGAGCAAAGGAGGATGCCATGCAGAAAGAAGTGAAAGTGAAACTTATTACTGATTCCGTGAATAAT 3825

1226  F  I  A  K  A  P  P  A  A  N  E  A  L  K  K  E  L  D  V  L  I  T  S  Y  Q  1250
3826 TTTATAGCAAAGGCTCCACCTGCAGCTAATGAGGCTTTGAAAAAGGAGCTTGATGTTCTAATTACCAGCTACCAG 3900

1251  R  L  C  S  R  L  N  G  K  C  K  T  L  E  E  V  W  A  C  W  H  E  L  L  S  1275
3901 AGACTCTGCAGCAGACTGAATGGAAAGTGCAAAACTTTGGAGGAAGTATGGGCATGTTGGCATGAATTATTGTCA 3975

1276  Y  L  D  A  E  N  K  W  L  N  E  V  E  L  K  L  K  A  T  E  N  I  Q  G  G  1300
3976 TATTTGGATGCAGAAAACAAATGGTTAAATGAGGTGGAATTGAAACTGAAGGCAACTGAAAATATCCAGGGAGGT 4050

1301  A  E  E  I  S  E  S  L  D  S  L  E  R  L  M  R  H  P  E  D  N  R  N  Q  I  1325
4051 GCAGAAGAGATTTCTGAGTCTTTAGATTCTTTGGAACGTTTAATGAGACATCCAGAAGATAATCGCAATCAGATT 4125

1326  R  E  L  A  Q  T  L  T  D  G  G  I  L  D  E  L  I  N  E  K  L  E  K  F  N  1350
4126 CGGGAATTAGCTCAGACTTTAACTGATGGTGGAATCTTGGATGAACTGATCAATGAGAAACTTGAGAAGTTCAAT 4200

1351  T  R  W  E  E  L  Q  Q  E  A  V  R  R  Q  K  S  L  E  Q  S  I  Q  S  A  Q  1375
4201 ACTCGATGGGAAGAACTGCAGCAGGAGGCTGTGAGAAGACAAAAGAGTCTTGAACAGAGTATTCAATCTGCCCAG 4275

1376  E  T  D  K  T  L  R  L  I  Q  E  S  L  A  A  I  D  K  Q  L  T  A  Y  T  A  1400
4276 GAGACTGACAAAACCCTCCGCTTAATTCAAGAGTCTCTTGCTGCTATAGACAAACAGCTGACAGCCTACACTGCA 4350

1401  D  R  V  D  A  A  Q  V  P  Q  E  A  Q  K  I  Q  S  E  L  T  S  H  E  I  S  1425
4351 GACAGAGTTGATGCAGCACAAGTGCCTCAGGAAGCACAGAAAATACAATCTGAATTAACAAGCCATGAGATTAGT 4425

1426  L  E  E  M  K  K  R  N  R  G  K  E  S  A  K  R  V  L  S  Q  I  D  V  A  Q  1450
4426 TTGGAAGAAATGAAGAAACGAAACCGAGGCAAGGAATCTGCAAAAAGAGTTCTTTCCCAAATTGATGTGGCACAG 4500

1451  K  K  L  Q  D  V  S  M  K  F  R  L  F  Q  K  P  A  N  F  E  Q  R  L  Q  E  1475
4501 AAAAAGCTGCAGGATGTCTCCATGAAGTTTCGCTTGTTTCAGAAACCAGCTAATTTTGAACAGCGTCTACAAGAA 4575

1476  C  K  R  I  L  D  E  V  K  L  Q  V  P  K  L  E  T  K  S  V  E  Q  E  V  V  1500
4576 TGCAAAAGAATTCTAGATGAAGTGAAGTTGCAAGTGCCCAAGTTGGAGACGAAGAGTGTTGAGCAGGAAGTAGTG 4650

1501  Q  S  H  L  D  H  C  M  K  L  Y  K  S  L  S  E  V  K  S  E  V  E  T  V  I  1525
4651 CAGTCACATTTGGACCACTGCATGAAATTATATAAAAGCCTGAGTGAGGTGAAGTCTGAAGTGGAAACAGTGATA 4725

1526  K  T  G  R  Q  I  V  Q  K  Q  Q  T  E  N  P  K  E  L  D  E  R  L  T  A  L  1550
4726 AAAACTGGGAGGCAGATTGTTCAAAAGCAGCAGACAGAGAACCCAAAAGAACTGGATGAAAGGCTTACAGCTTTG 4800

1551  K  L  Q  Y  N  E  L  G  A  K  V  T  E  K  K  Q  E  L  E  K  C  L  K  L  S  1575
4801 AAGTTGCAGTATAATGAATTGGGTGCGAAGGTGACAGAAAAAAAGCAGGAGTTAGAGAAATGCTTGAAATTGTCC 4875

1576  R  K  L  R  K  E  I  N  S  L  T  E  W  L  A  A  T  D  V  E  L  T  K  R  S  1600
4876 CGGAAGCTACGAAAAGAAATTAATTCGCTGACAGAATGGCTTGCAGCAACAGATGTGGAATTGACAAAGAGATCA 4950

1601  A  V  Q  G  M  P  S  N  L  D  A  E  I  A  W  G  K  A  T  R  K  E  I  E  K  1625
4951 GCTGTGCAAGGGATGCCTAGCAATTTGGATGCTGAAATTGCCTGGGGCAAGGCAACACGGAAAGAGATTGAGAAA 5025

1626  R  Q  V  Q  L  K  N  I  C  D  L  G  E  N  L  K  T  V  L  K  G  K  E  S  L  1650
5026 CGCCAAGTCCAGCTTAAGAATATCTGTGATTTAGGAGAGAATTTGAAGACAGTACTGAAAGGAAAGGAAAGTCTA 5100

1651  V  E  D  K  L  S  L  L  N  S  N  W  I  A  V  T  S  R  A  E  E  W  L  N  L  1675
5101 GTGGAAGATAAACTCAGTCTCCTGAACAGTAATTGGATAGCAGTAACTTCACGTGCCGAGGAATGGTTAAATCTG 5175

1676  L  M  E  Y  Q  K  H  M  E  A  F  D  Q  K  V  A  N  V  T  T  W  I  Y  R  A  1700
5176 TTAATGGAATATCAAAAGCACATGGAGGCTTTTGATCAGAAAGTAGCTAATGTCACGACTTGGATATATCGTGCT 5250

1701  E  I  L  L  D  E  S  D  K  Q  K  P  Q  Q  K  E  E  T  L  K  R  L  K  A  E  1725
5251 GAAATACTGTTGGATGAATCTGATAAGCAAAAGCCCCAGCAAAAAGAGGAAACTCTTAAGCGCTTAAAGGCTGAG 5325

1726  L  N  D  M  H  P  K  V  D  S  V  R  D  Q  A  V  D  L  M  T  N  R  G  D  H  1750
5326 CTGAATGATATGCATCCAAAGGTGGACTCTGTGCGTGACCAAGCAGTAGACTTGATGACAAACCGTGGTGATCAC 5400

1751  C  R  K  V  I  E  P  K  L  S  E  L  N  H  R  F  A  A  I  S  Q  R  I  K  S  1775
5401 TGCAGGAAAGTAATAGAGCCTAAACTATCTGAGCTCAACCATCGATTTGCTGCCATATCACAAAGAATTAAGAGT 5475

1776  G  K  P  F  I  P  L  K  E  L  E  Q  F  D  F  D  I  Q  K  L  L  E  P  L  E  1800
5476 GGAAAGCCCTTCATTCCTTTGAAGGAATTGGAGCAATTTGACTTCGATATACAAAAATTGCTTGAACCACTGGAG 5550

1801  V  E  I  Q  Q  G  V  N  L  K  E  E  D  F  N  K  D  M  S  E  D  D  E  S  T  1825
5551 GTTGAAATTCAGCAGGGGGTGAATCTGAAAGAGGAGGACTTCAATAAAGATATGAGTGAAGATGATGAGAGCACA 5625
```

```
1826  V  K  E  L  L  Q  R  G  D  T  L  Q  K  R  I  T  D  E  R  K  R  E  E  I  K  1850
5626 GTGAAAGAATTGCTGCAAAGGGGCGACACGCTTCAGAAAAGAATCACAGATGAGAGAAAACGGGAGGAAATAAAG 5700

1851  I  K  Q  Q  L  L  Q  T  K  H  N  A  L  K  D  L  R  S  Q  R  R  K  K  A  L  1875
5701 ATAAAACAACAGCTGTTGCAGACTAAACATAATGCTCTCAAGGACTTGAGGTCTCAAAGAAGAAAAAAGGCTTTA 5775

1876  E  I  S  H  Q  W  Y  Q  Y  K  R  Q  A  D  D  L  M  T  W  L  D  D  I  E  K  1900
5776 GAGATTTCTCATCAATGGTATCAGTACAAGAGGCAGGCTGATGACCTAATGACATGGCTGGATGACATTGAGAAA 5850

1901  K  L  A  S  L  P  D  H  K  D  E  Q  K  L  K  E  I  G  G  E  L  E  K  K  K  1925
5851 AAGTTAGCCAGTCTACCAGACCACAAAGATGAGCAGAAGCTAAAGGAGATTGGTGGAGAGTTGGAGAAGAAGAAG 5925

1926  E  D  L  N  A  V  N  R  Q  A  E  R  L  S  K  D  G  A  A  K  A  V  E  P  T  1950
5926 GAAGATCTGAATGCGGTGAACAGACAGGCTGAACGCCTGTCTAAGGATGGGGCTGCAAAAGCAGTGGAGCCAACC 6000

1951  L  V  Q  L  S  K  R  W  R  D  F  E  S  K  F  A  Q  F  R  R  L  N  Y  A  Q  1975
6001 CTGGTACAGCTCAGCAAGCGCTGGCGAGATTTTGAGAGCAAATTTGCTCAGTTTCGAAGACTCAACTATGCACAA 6075

1976  I  Q  T  V  L  E  D  T  T  F  V  M  T  E  S  M  T  V  E  T  T  Y  V  P  S  2000
6076 ATTCAAACAGTTCTAGAAGATACAACTTTTGTGATGACTGAAAGTATGACTGTGGAAACCACTTACGTGCCTTCT 6150

2001  T  Y  L  A  E  I  L  Q  L  L  Q  A  L  S  E  V  E  E  R  L  N  S  P  V  L  2025
6151 ACATACCTGGCAGAGATCCTTCAGCTTCTGCAAGCCTTGTCTGAAGTAGAAGAACGCCTTAATTCTCCTGTTCTG 6225

2026  Q  A  K  D  C  E  D  L  L  K  Q  E  E  C  L  K  N  I  K  D  C  L  G  R  L  2050
6226 CAGGCCAAGGACTGTGAGGATCTCTTGAAACAAGAAGAATGTCTCAAGAACATTAAAGATTGCCTGGGGAGACTC 6300

2051  Q  G  H  I  D  I  I  H  S  K  K  T  P  A  L  Q  S  A  T  P  R  E  T  A  N  2075
6301 CAGGGTCATATAGACATTATTCACAGCAAGAAGACACCGGCTTTGCAAAGTGCTACACCACGGGAAACAGCAAAT 6375

2076  I  Q  D  K  L  T  Q  L  N  S  Q  W  E  K  V  N  K  M  Y  R  D  R  Q  A  R  2100
6376 ATACAAGACAAGCTGACTCAGCTTAATTCCCAATGGGAGAAAGTTAACAAGATGTACCGGGACCGGCAGGCACGC 6450

2101  F  D  K  S  K  E  K  W  R  L  F  H  C  E  M  K  S  F  N  E  W  L  T  E  T  2125
6451 TTTGACAAATCCAAGGAAAAGTGGCGGCTTTTTCATTGCGAAATGAAGAGTTTTAATGAGTGGCTAACTGAAACT 6525

2126  E  E  K  L  S  R  A  Q  I  E  A  G  D  V  G  H  V  K  T  K  Q  F  L  Q  E  2150
6526 GAAGAGAAACTGTCAAGAGCACAGATAGAGGCTGGAGACGTGGGTCATGTGAAAACCAAGCAATTTCTTCAGGAG 6600

2151  L  Q  D  G  I  G  R  Q  Q  T  V  V  K  T  L  N  V  T  G  E  E  I  I  E  Q  2175
6601 CTTCAGGATGGCATTGGGCGACAGCAAACTGTTGTCAAAACACTGAATGTAACTGGCGAAGAAATTATTGAGCAG 6675

2176  S  S  A  A  D  A  N  V  L  K  E  Q  L  G  N  L  N  T  R  W  Q  E  I  C  R  2200
6676 TCATCAGCAGCAGATGCTAACGTGCTGAAGGAGCAACTGGGAAATCTGAATACCCGGTGGCAGGAGATCTGCAGA 6750

2201  Q  L  V  E  K  R  K  R  I  E  E  E  K  N  I  L  S  E  F  Q  E  D  L  N  K  2225
6751 CAGCTGGTAGAGAAAAGAAAGAGGATAGAGGAAGAAAAGAATATTTTATCAGAATTTCAAGAAGACTTGAACAAG 6825

2226  L  I  L  W  L  E  E  T  E  N  V  I  A  I  P  L  E  P  G  N  E  D  Q  L  R  2250
6826 CTGATTTTATGGTTAGAGGAAACAGAGAACGTCATTGCTATTCCCCTTGAACCAGGGAATGAAGACCAGCTAAGA 6900

2251  D  C  L  G  K  V  K  L  R  V  E  E  L  L  P  H  K  G  I  L  K  R  L  N  E  2275
6901 GACTGCCTTGGCAAAGTAAAGTTAAGAGTTGAAGAGCTGCTGCCACACAAGGGAATACTGAAACGATTAAATGAA 6975

2276  T  G  G  T  T  L  G  S  A  S  L  N  P  E  R  K  H  K  L  E  S  T  L  K  E  2300
6976 ACTGGAGGAACAACGCTTGGAAGTGCATCACTGAACCCAGAAAGAAAACATAAGCTTGAGAGTACACTGAAGGAG 7050

2301  A  S  R  R  L  L  K  V  S  R  D  L  P  E  K  Q  K  E  I  E  I  L  L  K  D  2325
7051 GCTAGCCGTCGCTTGTTAAAGGTGTCCAGAGATCTACCAGAGAAGCAAAAAGAAATAGAGATTCTGCTAAAGGAT 7125

2326  F  I  E  L  N  Q  Q  I  N  Q  L  T  L  W  I  T  P  V  K  N  Q  L  E  L  Y  2350
7126 TTCATCGAACTTAATCAGCAAATAAATCAACTGACACTCTGGATAACACCTGTCAAAAACCAGCTAGAGCTTTAT 7200

2351  N  Q  V  G  Q  P  G  A  F  D  I  K  E  T  E  A  A  V  Q  A  K  Q  P  N  V  2375
7201 AACCAAGTGGGTCAACCAGGAGCTTTTGATATTAAGGAAACCGAAGCAGCAGTGCAGGCTAAACAGCCGAATGTG 7275

2376  E  E  V  L  S  K  G  C  H  L  Y  K  E  K  P  A  T  H  P  V  K  K  K  L  E  2400
7276 GAAGAGGTTTTGTCTAAAGGGTGTCATTTATATAAGGAAAAACCAGCCACTCATCCAGTAAAGAAAAAACTAGAA 7350

2401  D  L  N  A  D  W  K  A  I  N  H  L  I  L  Q  L  K  E  K  P  T  F  G  E  P  2425
7351 GACTTGAATGCTGACTGGAAGGCAATAAACCACTTAATTCTACAACTGAAGGAGAAGCCAACATTTGGAGAGCCT 7425

2426  A  L  T  S  P  G  V  L  T  S  G  Q  T  V  A  V  D  T  Q  A  R  V  T  K  E  2450
7426 GCCCTTACCTCTCCAGGTGTCTTAACTTCTGGTCAAACTGTTGCTGTGGATACACAAGCCAGGGTAACCAAGGAA 7500
```

```
2451  T  T  S  F  T  P  E  M  P  S  S  V  L  L  E  V  P  A  L  A  D  F  N  K  A  2475
7501 ACCACCAGCTTCACACCAGAAATGCCATCTTCTGTGCTTTTGGAGGTTCCAGCCTTAGCTGACTTCAATAAGGCA 7575

2476  W  A  E  L  T  D  W  L  S  R  L  D  R  E  I  K  A  Q  R  V  T  V  G  D  L  2500
7576 TGGGCAGAACTCACTGACTGGCTTTCTCGACTGGATCGAGAGATAAAAGCTCAGAGAGTGACAGTAGGTGATCTT 7650

2501  D  D  I  N  D  M  I  I  K  Q  K  A  N  M  Q  D  L  E  Q  R  R  P  Q  L  D  2525
7651 GATGATATCAACGACATGATCATCAAACAAAAGGCTAACATGCAAGATCTGGAGCAAAGACGTCCCCAGCTGGAT 7725

2526  E  L  I  T  A  A  Q  N  L  K  N  K  T  S  N  Q  E  A  R  T  I  I  T  D  R  2550
7726 GAACTAATAACTGCAGCACAAAATCTCAAAAACAAGACGAGCAATCAAGAGGCCAGAACAATAATTACTGACCGC 7800

2551  I  E  K  I  Q  S  Q  W  D  D  V  H  G  Y  L  Q  N  R  R  Q  Q  L  H  E  M  2575
7801 ATTGAAAAGATACAGAGCCAGTGGGATGATGTGCACGGATACCTCCAAAACCGAAGACAACAGCTTCATGAGATG 7875

2576  Q  K  D  S  T  Q  W  L  E  A  K  Q  E  A  E  Q  V  L  E  Q  A  K  A  K  L  2600
7876 CAAAAGGATTCAACACAGTGGCTAGAAGCTAAACAAGAAGCTGAACAGGTTCTTGAACAAGCAAAAGCAAAGCTT 7950

2601  E  S  W  K  E  I  S  Y  T  V  E  A  L  K  K  Q  N  S  E  L  K  Q  F  S  K  2625
7951 GAGTCATGGAAAGAAATTTCCTATACTGTGGAAGCTCTGAAAAAGCAGAACTCTGAGCTTAAGCAATTTTCAAAA 8025

2626  E  I  R  Q  W  Q  M  N  I  E  G  V  N  D  V  A  L  K  P  V  R  D  Y  S  A  2650
8026 GAGATACGACAGTGGCAAATGAATATAGAAGGGGTGAATGACGTGGCACTTAAGCCTGTCCGCGATTATTCAGCA 8100

2651  D  D  T  R  K  V  E  L  M  T  D  N  I  N  A  T  W  A  T  I  N  K  R  V  S  2675
8101 GATGACACCAGAAAAGTAGAACTGATGACAGATAACATTAATGCGACATGGGCTACAATCAATAAGAGGGTTAGT 8175

2676  E  R  E  A  A  L  E  S  A  L  L  M  L  Q  E  F  Y  L  D  L  E  K  F  L  A  2700
8176 GAACGTGAAGCCGCACTGGAATCAGCTCTACTGATGTTGCAGGAATTCTACCTGGATCTTGAAAAGTTCCTTGCT 8250

2701  W  L  T  E  A  E  T  T  A  N  V  L  Q  D  A  T  H  K  E  K  T  L  E  D  P  2725
8251 TGGCTTACAGAAGCTGAAACAACTGCTAATGTCCTGCAGGATGCTACACACAAGGAAAAGACACTAGAGGATCCC 8325

2726  Q  M  V  R  E  L  M  K  Q  W  Q  D  L  Q  A  E  I  D  A  H  T  D  I  F  H  2750
8326 CAGATGGTTCGGGAGCTCATGAAGCAGTGGCAGGATCTACAGGCAGAAATTGATGCACATACTGACATCTTCCAC 8400

2751  N  L  D  E  N  G  Q  K  I  L  R  S  L  E  G  S  E  D  A  V  L  L  Q  R  R  2775
8401 AACCTGGATGAAAACGGGCAGAAAATCCTGAGATCCCTGGAAGGCTCAGAGGATGCTGTCCTGTTGCAGAGACGT 8475

2776  L  D  N  M  N  F  R  W  S  E  L  R  K  K  S  L  N  I  R  S  H  L  E  A  S  2800
8476 CTGGATAACATGAACTTCAGATGGAGTGAGCTTAGGAAGAAATCTCTAAACATTAGATCTCATTTGGAAGCCAGC 8550

2801  T  D  Q  W  K  R  L  H  L  S  L  Q  E  L  L  A  W  L  Q  L  K  E  D  E  L  2825
8551 ACAGACCAGTGGAAGCGTTTACATCTCTCTCTTCAGGAACTTTTGGCATGGCTGCAATTGAAGGAGGATGAATTA 8625

2826  K  Q  Q  A  P  I  G  G  D  I  P  T  V  Q  K  Q  N  D  V  H  R  T  F  K  R  2850
8626 AAACAGCAAGCACCCATTGGTGGAGATATTCCCACTGTGCAGAAGCAGAATGATGTTCATAGGACTTTCAAGAGG 8700

2851  E  L  K  T  K  E  P  V  I  M  N  A  L  E  T  V  R  L  F  L  A  D  Q  P  V  2875
8701 GAGCTGAAAACAAAAGAACCTGTTATCATGAATGCACTTGAGACTGTGCGACTCTTCCTGGCAGATCAACCAGTA 8775

2876  E  G  L  E  K  V  Y  P  E  P  R  D  L  S  P  E  E  R  A  Q  N  V  T  K  V  2900
8776 GAGGGACTGGAAAAGGTCTATCCAGAACCAAGAGACCTATCACCTGAGGAGAGGGCCCAGAATGTCACTAAAGTT 8850

2901  L  R  R  Q  A  D  D  V  R  T  E  W  D  K  L  N  L  R  S  A  D  W  Q  K  K  2925
8851 CTCCGAAGGCAAGCAGATGATGTCAGAACTGAGTGGGATAAGCTAAATCTACGTTCTGCTGATTGGCAAAAGAAG 8925

2926  I  D  D  A  L  E  R  L  Q  G  L  Q  E  A  M  D  E  L  D  L  K  L  R  Q  A  2950
8926 ATAGATGATGCTCTTGAAAGACTGCAGGGTCTTCAGGAGGCAATGGATGAACTAGACCTGAAACTGCGCCAGGCT 9000

2951  E  A  F  K  G  S  W  Q  P  V  G  D  L  L  I  D  S  L  Q  D  H  L  E  K  V  2975
9001 GAAGCATTCAAGGGATCCTGGCAGCCAGTGGGGGATCTGCTGATAGACTCTCTGCAGGATCACTTAGAAAAAGTC 9075

2976  K  V  Y  R  A  E  M  V  P  L  K  E  K  V  H  Q  V  N  E  L  A  H  R  F  A  3000
9076 AAGGTTTATCGAGCAGAAATGGTGCCCCTTAAAGAGAAGGTGCATCAAGTCAATGAGCTGGCTCACCGGTTCGCT 9150

3001  P  P  D  I  Q  L  S  P  Y  T  L  S  C  L  E  D  L  N  T  R  W  K  V  L  Q  3025
9151 CCCCCTGATATTCAGCTCTCCCCATACACTCTCAGCTGTCTGGAGGACCTGAACACAAGGTGGAAGGTGCTACAG 9225

3026  V  A  I  D  E  R  I  R  Q  L  H  E  A  H  R  D  F  G  P  T  S  Q  H  F  L  3050
9226 GTGGCCATTGATGAGCGCATCAGGCAACTGCATGAAGCTCACAGGGATTTTGGCCCTACTTCCCAGCATTTTCTT 9300

3051  T  T  S  V  Q  G  P  W  E  R  A  I  S  P  N  K  V  P  Y  Y  I  N  H  E  T  3075
9301 ACCACTTCTGTCCAAGGCCCCTGGGAGAGGGCAATCTCGCCAAACAAAGTGCCCTATTACATCAACCATGAGACG 9375

3076  Q  T  T  C  W  D  H  P  K  M  T  E  L  Y  Q  S  L  A  D  L  N  N  V  R  F  3100
```

```
 9376 CAGACAACCTGCTGGGATCATCCCAAAATGACCGAGCTCTACCAGTCTTTAGCGGACCTGAACAATGTCAGATTC 9450

 3101  S  A  Y  R  T  A  M  K  L  R  R  L  Q  K  A  L  C  L  D  L  L  N  L  S  A  3125
 9451 TCAGCATACAGAACTGCCATGAAGCTCCGCAGGCTGCAGAAAGCTCTCTGCTTGGATCTCCTGAATCTGTCTGCT 9525

 3126  A  C  D  A  L  D  Q  H  N  L  K  Q  N  D  Q  P  M  D  I  L  Q  I  I  N  C  3150
 9526 GCATGCGATGCCTTGGACCAGCACAACCTCAAGCAAAATGACCAGCCGATGGATATTCTGCAGATCATTAACTGC 9600

 3151  L  T  T  I  Y  D  R  L  E  Q  E  H  N  N  L  V  N  V  P  L  C  V  D  M  C  3175
 9601 TTGACCACTATTTATGATCGACTGGAACAGGAGCACAATAATCTGGTCAATGTTCCTCTCTGCGTAGACATGTGC 9675

 3176  L  N  W  L  L  N  V  Y  D  T  G  R  T  G  R  I  R  V  L  S  F  K  T  G  V  3200
 9676 CTCAACTGGCTGCTGAATGTCTATGACACGGGTCGAACAGGAAGGATCCGTGTCTTATCTTTCAAAACTGGTGTT 9750

 3201  V  S  L  C  K  A  H  L  E  D  K  Y  R  Y  L  F  K  Q  V  A  S  S  T  G  F  3225
 9751 GTATCCCTTTGTAAAGCACATCTGGAAGATAAGTATAGATACCTGTTCAAGCAGGTGGCGAGCTCCACTGGCTTC 9825

 3226  C  D  Q  R  R  L  G  L  L  L  H  D  S  I  Q  I  P  R  Q  L  G  E  V  A  S  3250
 9826 TGTGACCAGCGCCGGCTGGGACTGCTGCTGCACGACTCCATCCAGATCCCACGGCAGCTGGGGGAGGTCGCTTCG 9900

 3251  F  G  G  S  N  I  E  P  S  V  R  S  C  F  Q  F  A  N  N  K  P  E  I  E  A  3275
 9901 TTTGGGGGCAGCAACATCGAGCCGAGTGTCAGAAGCTGCTTCCAGTTTGCCAATAACAAGCCTGAGATCGAAGCA 9975

 3276  A  L  F  L  D  W  M  R  L  E  P  Q  S  M  V  W  L  P  V  L  H  R  V  A  A  3300
 9976 GCCTTGTTCCTGGACTGGATGAGGCTGGAACCACAATCCATGGTGTGGCTGCCCGTGCTGCACAGGGTGGCTGCT 10050

 3301  A  E  T  A  K  H  Q  A  K  C  N  I  C  K  E  C  P  I  I  G  F  R  Y  R  S  3325
10051 GCCGAAACTGCCAAACACCAAGCAAAGTGTAACATCTGCAAGGAGTGCCCCATTATTGGATTCAGGTACAGAAGC 10125

 3326  L  K  H  F  N  Y  D  I  C  Q  S  C  F  F  S  G  R  V  A  K  G  H  K  M  H  3350
10126 TTAAAGCACTTTAACTATGACATCTGCCAAAGTTGCTTCTTCTCTGGCCGTGTTGCAAAAGGTCACAAAATGCAC 10200

 3351  Y  P  M  V  E  Y  C  T  P  T  T  S  G  E  D  V  R  D  F  A  K  V  L  K  N  3375
10201 TATCCCATGGTGGAGTACTGCACACCGACAACTTCTGGAGAAGATGTCCGTGACTTTGCCAAGGTACTAAAAAAC 10275

 3376  K  F  R  T  K  R  Y  F  A  K  H  P  R  M  G  Y  L  P  V  Q  T  V  L  E  G  3400
10276 AAATTTCGAACAAAAAGATATTTTGCAAAGCACCCACGAATGGGCTACCTGCCTGTGCAAACTGTCTTGGAGGGA 10350

 3401  D  N  M  E  T  P  V  T  L  I  N  F  W  P  V  D  S  A  L  A  E  M  E  N  S  3425
10351 GACAACATGGAAACTCCTGTTACTCTGATCAACTTCTGGCCAGTAGATTCTGCGCTAGCAGAAATGGAGAACAGC 10425

 3426  N  G  S  Y  L  N  D  S  I  S  P  N  E  S  I  D  D  E  H  L  L  I  Q  H  Y  3450
10426 AATGGTTCTTACCTAAATGACAGTATTTCACCTAATGAGAGCATCGATGATGAACACTTGTTAATCCAGCACTAC 10500

 3451  C  Q  S  L  N  Q  E  S  P  L  S  Q  P  R  S  P  A  Q  I  L  I  S  L  E  S  3475
10501 TGCCAAAGTCTGAACCAGGAATCCCCCCTGAGCCAGCCCCGAAGCCCTGCCCAGATCTTGATTTCTCTGGAGAGT 10575

 3476  E  E  R  G  E  L  E  R  I  L  A  D  L  E  E  E  N  R  N  L  Q  A  E  Y  D  3500
10576 GAAGAAAGAGGTGAACTGGAGAGAATTCTTGCAGATCTGGAAGAAGAGAATCGAAACTTGCAAGCGGAGTATGAC 10650

 3501  R  L  K  Q  Q  H  D  H  K  G  L  S  P  L  P  S  P  P  E  M  M  P  V  S  P  3525
10651 CGTTTGAAGCAACAGCATGATCACAAAGGATTATCTCCACTGCCATCCCCACCAGAGATGATGCCAGTTTCTCCA 10725

 3526  Q  S  P  R  D  A  E  L  I  A  E  A  K  L  L  R  Q  H  K  G  R  L  E  A  R  3550
10726 CAGAGTCCTCGCGATGCTGAACTCATTGCAGAAGCCAAACTGCTTCGCCAGCACAAGGGCCGCCTGGAGGCCAGG 10800

 3551  M  Q  I  L  E  D  H  N  K  Q  L  E  S  Q  L  H  R  L  R  Q  L  L  E  Q  P  3575
10801 ATGCAGATTCTGGAGGATCACAACAAACAGCTGGAGTCACAGCTGCACAGGCTGAGGCAGCTGCTGGAGCAGCCA 10875

 3576  Q  A  D  A  K  V  N  G  T  T  L  S  S  P  S  T  S  L  Q  R  S  D  S  S  Q  3600
10876 CAGGCAGATGCCAAGGTGAATGGTACAACACTATCATCTCCTTCTACCTCTTTGCAGAGGTCAGACAGCAGTCAG 10950

 3601  P  M  L  L  R  V  V  G  S  Q  T  S  E  T  M  G  E  D  D  L  L  S  P  P  Q  3625
10951 CCAATGCTTCTTCGTGTAGTTGGCAGCCAGACTTCAGAAACCATGGGCGAGGACGACCTGCTCAGCCCTCCCCAG 11025

 3626  D  T  S  T  G  L  E  E  V  M  E  Q  L  N  N  S  F  P  S  S  R  G  R  N  A  3650
11026 GACACAAGCACAGGTTTGGAGGAAGTGATGGAGCAGCTTAACAACTCCTTCCCCAGTTCCAGAGGAAGAAATGCC 11100

 3651  P  G  K  P  V  R  E  A  T  M  *  3661
11101 CCTGGAAAGCCAGTGAGAGAGGCCACAATGTAG 11133
```

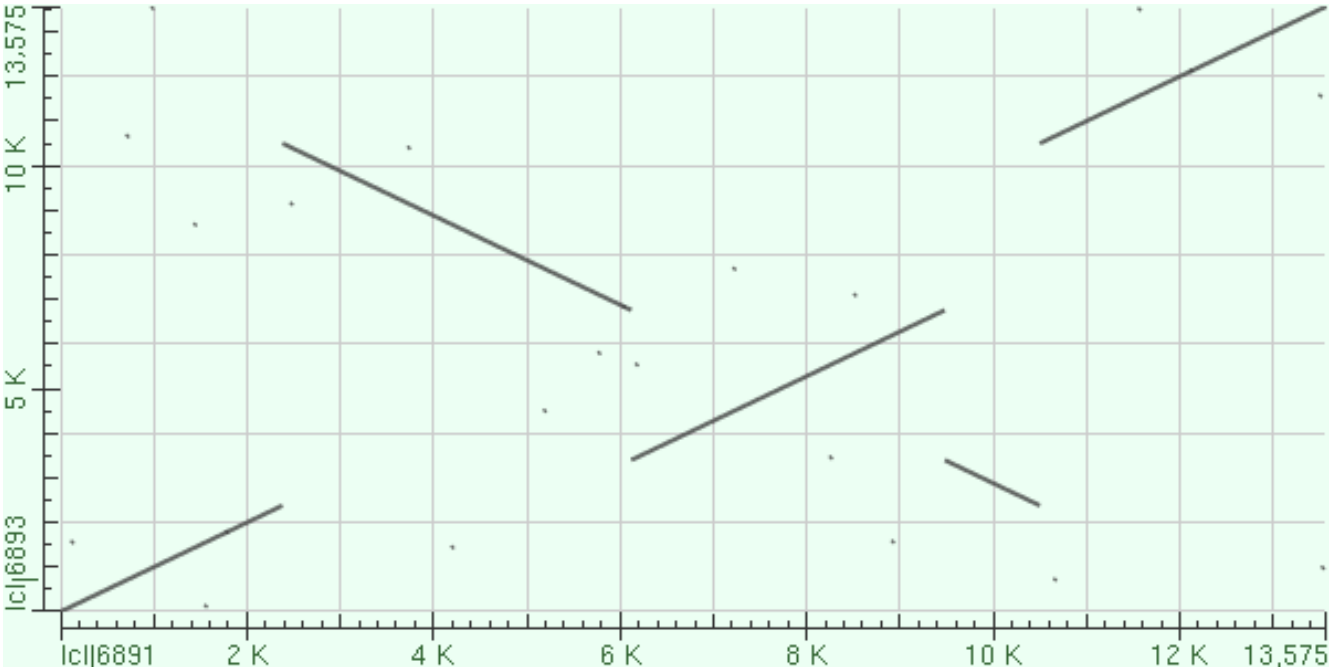
b) From what species was this sequence taken? Support your answer with evidence.

`/organism="Gallus gallus"`

c) I happen to know that **>Second_Sequence** evolved from **>First_Sequence**. Analyze **>Second_Sequence**. List and describe all the mutations you detect. Support each mutation you describe with the evidence you used. Your analysis should provide the most parsimonious answer.

1) SNP in 10th codon to change amino acid

2) double inversion: first one at 2,200 and 10,500; second one at 6,100 and 9,400.



d) Search NCBI to find a tabular presentation of some of the known SNPs in the original gene for First Sequence. Be sure you are looking at the correct species (see part b above). Your answer must include a screen show showing multiple SNPs and the effects they would have on the encoded protein.

http://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgi?locusId=396236

something like this:

										intron	-/AAC					
	114077213		rs13925555	0.500							-/AAA					
	114111422		rs15393454	0.500							(>6bp)					
	114125248		rs15393513	0.500							(>6bp)					
	114125252		rs15393515	0.500							(>6bp)					
	114131790	348	rs15393546	0.500						frame shift	-	Met [M]	3	66		
										contig reference	T	Ile [I]	3	66		
	114144336	632	rs15393606	0.500						missense	C	Thr [T]	2	161		
										contig reference	T	Ile [I]	2	161		
	114144352	648	rs316293821	N.D.						synonymous	T	Ser [S]	3	166		
										contig reference	C	Ser [S]	3	166		

20 pts

2) Over the last few years, we have sequenced more and more Neaderthal DNA. These questions pertain to those findings.

a) **Neander Fig1a** shows the analysis of DNA from 6 different fossil bone samples. Which fossil would you use for additional studies and why? Support your answer with data.

VI-80 has the least amount of modern DNA contamination and the most Neanderthal DNA.

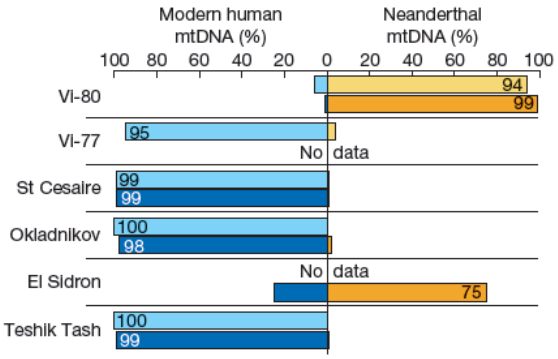


Figure 1 | Ratio of Neanderthal to modern human mtDNA in six hominid fossils. For each fossil, primer pairs that amplify a long (119 base pairs; upper lighter bars) and short (63 base pairs; lower darker bars) product were used to amplify segments of the mtDNA hypervariable region. The products were sequenced and determined to be either of Neanderthal (yellow) or modern human (blue) type.

b) From an abstract: "Neanderthals are the extinct hominid group most closely related to contemporary humans, so their genome offers a unique opportunity to identify genetic changes specific to anatomically fully modern humans. Direct high-throughput sequencing of a DNA extract from this fossil has thus far yielded over one million base pairs of hominoid nuclear DNA sequences. Comparison with the human and chimpanzee genomes reveals that modern human and Neanderthal DNA sequences diverged on average about 500,000 years ago." What can you conclude from **Neander Fig1b**?

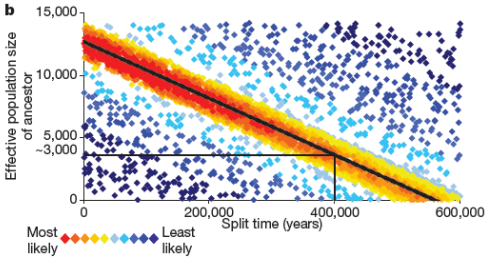


Figure 6 | Estimate of the effective population size of the ancestor of humans and Neanderthals. **a**, Schematic illustration of the model used to estimate ancestral effective population size. By split time, we mean the time, in the past, after which there was no more interbreeding between two groups. By divergence, we mean the time, in the past, at which two genetic regions separated and began to accumulate substitutions independently. Effective population size is the number of individuals needed under ideal conditions to produce the amount of observed genetic diversity within a population. **b**, The likelihood estimates of population split times and ancestral population sizes. The likelihoods are grouped by colour. The red–yellow points are statistically equivalent based on the likelihood ratio test approximation. The black line is the line of best fit to red–yellow points (see Supplementary Methods). This graph is scaled assuming a human–chimpanzee average sequence divergence time of 6,500,000 years.

c) From the figure legend of **Neander Fig1c:** "a, Individual maps; the marginal probability of Neanderthal ancestry for one European-American, one east-Asian and one sub-Saharan-African phased genome across chromosome 9. b, Population maps; estimated the proportion of Neanderthal ancestry in European individuals (red) and east-Asian individuals (green), averaged across all individuals from each population in non-overlapping 100-kb windows on chromosome 9. The black bar denotes the coordinates of the centromere. The plot is limited to segments of the chromosome that pass quality filters. CEU, residents of Utah, US, with northern and western European ancestry; CHB, Han Chinese in Beijing, China; LWK, African Luhya in Webuye, Kenya." Interpret panels a and b from this figure. Support your answer with data from the

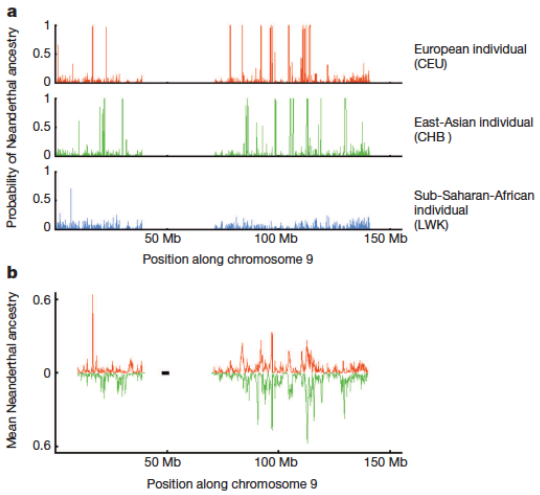


figure.

panel a:

It appears East Asian and Europeans still carry some Neanderthal DNA on chromosome 9. It looks like they are not identical segments, but maybe the resolution is not precise enough to be sure. The Luhya person does not appear to carry any Neanderthal DNA on chromosome 9.

panel b:

Similar results for populations of Asians and Europeans with both sharing similar parts of chromosome 9 with Neanderthals.

d) Consider **Neander Fig1d** and summarize the results. Support your answer with data from the figure.

Functionally important DNA from he Neanderthal DNA seems to be less abundant on the X chromosomes of Asians and Europeans than on the autosomes. Suggests some decreased fertility or other form of fitness.

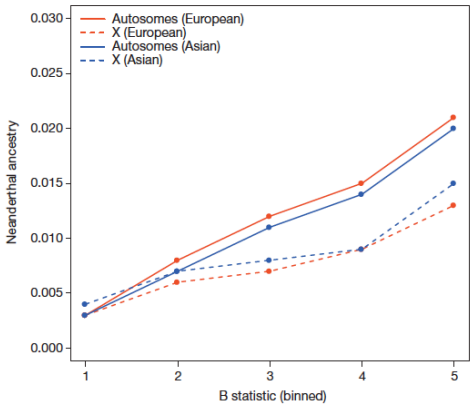


Figure 2 | Functionally important regions are deficient in Neanderthal ancestry. The median of the proportion of Neanderthal ancestry (estimated as the average over the marginal probability of Neanderthal ancestry assigned to each individual allele at a SNP) within quintiles of a B statistic that measures proximity to functionally important regions (1-low, 5-high). We show results on the autosomes and the X chromosome, and in European and east-Asian populations.

15 pts

3) This question focuses on the biggest news in biology for several years. I suspect this will win a Nobel prize not long from now.

a) Look up "STAP cells" and briefly summarize what they are. Provide me with a link to your source(s).

These are induced pluripotent stem cells but they are not produced by giving them DNA encoding 4 transcription factors like iPSCs. Treated with acid to stress them. They had GFP and were derived from white blood cells that had rearranged DNA to verify they were somatic.
http://www.cnn.com/2014/01/29/health/stem-cell-discovery/

b) What genomic evidence did the investigators use to convince skeptics that they had converted mature, differentiated cells into stem cells? Provide citation for your source(s).

"gene rearrangement analysis" for T cell receptor or antibodies (T cells is correct, but the abstract did not differentiate which one).
http://www.ncbi.nlm.nih.gov/pubmed/?term=Stimulus-triggered+fate+conversion+of+somatic+cells+into+pluripotency

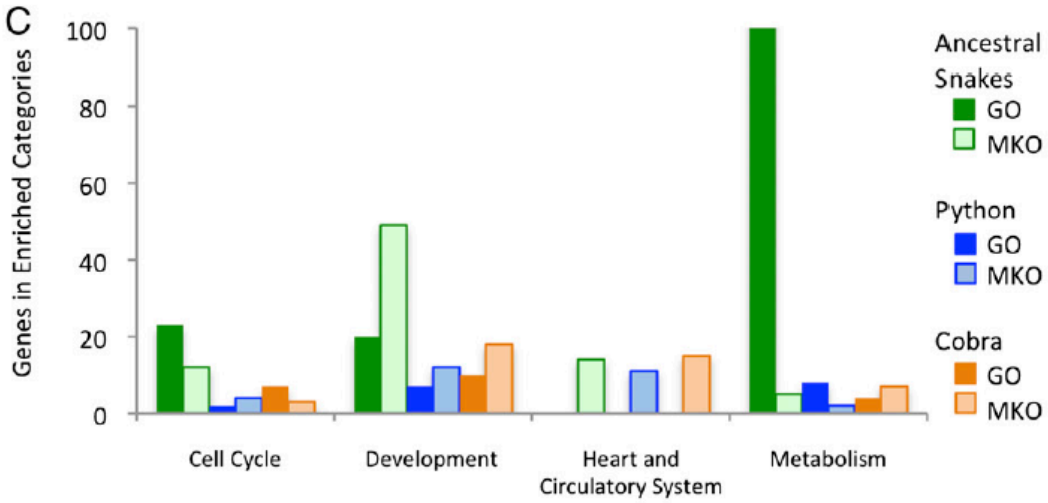
c) What sort of changes would you expect to see at the genome level if you compared the original cells to the STAP cells? Do not talk about protein or mRNA differences – focus on the genome.

Epigenetic changes
http://www.utsandiego.com/news/2014/Jan/29/new-stem-cell-stap-vacanti/

25 pts

4) The first two snake genomes (Burmese Python and King Cobra) were sequenced in December, 2013. Answer the following questions.

a) Look at **Snake Fig1** (MKO = mouse knock out phenotype). Summarize what you see about snakes in general as well as each species of snake.



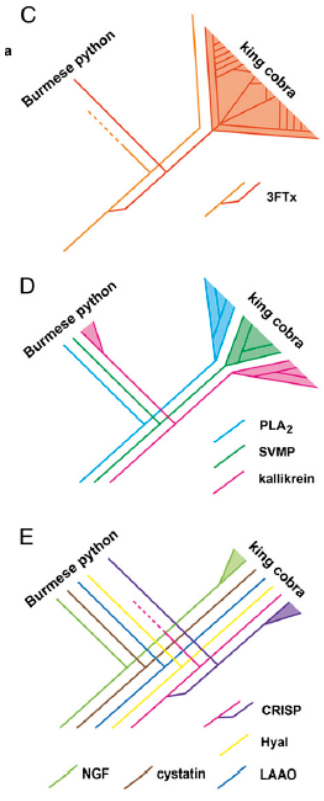
Snakes have gene enrichment in all four categories but metabolism is by far the most enriched.
Cobra has slightly more than python in all but metabolism.

b) **Snake Fig2** focuses on snake venom that evolved in king cobras. (C) … the three-finger toxin gene family (D) other pathogenic toxin families of venom-expressed genes, and (E) ancillary toxin families after the split of the Burmese python from the advanced snakes. Compare and contrast evolution of the three toxin gene families.

C three-finger toxin – lots of adaptive radiation in cobra, but not in python.

D other venom genes – adaptive radiation in 3 genes for cobra, not as much as three-finger toxin, and only one paralog in python.

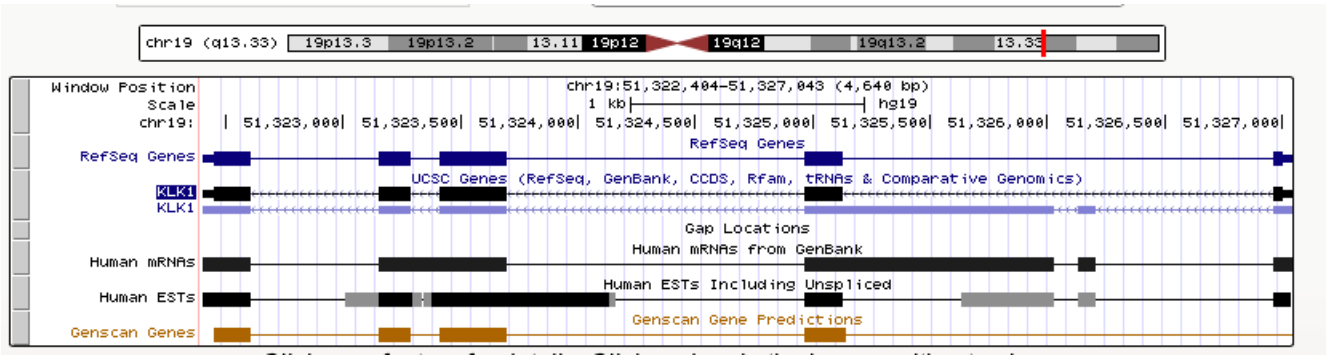
E ancillary toxins – not much adaptive radiation in cobras or pythons in this category. So cobra mainly diversified three-finger toxins.



c) Where are Bermese Pythons found in nature outside their normal range? Are they having an impact on their new location? Support your answer with data including a peer-reviewed journal citation.
Florida Everglades: killing all vertebrates and mammals.
Our own Mike Dorcas:
http://www.pnas.org/content/109/7/2418.abstract?sid=93e23f28-3d72-46e1-bba1-308342442f7d

d) What would you do next to determine if the Burmese Python is evolving in its new habitat? Design the experiment you would like to perform if expense and human resources were not limiting.
Lots of options; genome, transcriptome, epigenetic changes in methylome….

e) Find the Homo sapiens kallikrein 1 gene in a human genome browser. What compounds bind to human kallikrein 1? Support your answer with screen shots.



Comparative Toxicogenomics Database (CTD)

The following chemicals interact with this gene

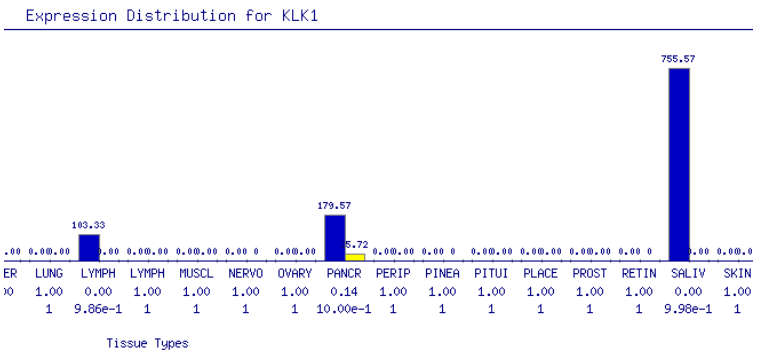
- D012964 Sodium
- D013311 Streptozocin
- D013749 Tetrachlorodibenzodioxin
- C472791 3-(4'-hydroxy-3'-adamantylbiphenyl-4-yl)acrylic acid
- C006500 4-aminobenzamidine
- C019498 4-nitroaniline
- D000535 Aluminum
- D000584 Amiloride
- D000643 Ammonium Chloride
- D001920 Bradykinin

f) What is kallikrein 1's normal function and what human tissues transcribe this mRNA? Support your answer with screen shots. How could this function be related to snake venom?
Protease
Molecular Function:
GO:0003824 catalytic activity
GO:0004252 serine-type endopeptidase activity
GO:0004293 tissue kallikrein activity
GO:0008233 peptidase activity
GO:0016787 hydrolase activity

***GeneHub-GEPIS* -- From Gene Integration to Expression Profiling**

Most in saliva, then pancreas then lymph system.

In blood, regulates BP.



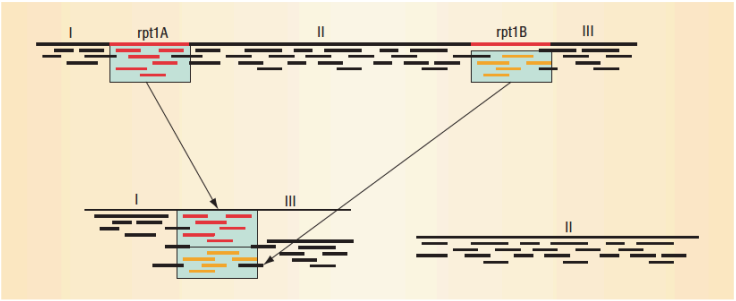
g) How could the cobra's kallikrein 1 gene evolve to be venomous in king cobra if snakes use kallikrein 1 the same way humans do?

They had gene duplication, then the paralogs changed expression pattern to be in venom and they use this to break down proteins in prey. Could also cause BP crash in prey.

20 pts

5) The final question is a series of mildly related parts.

a) Look at **Seq Fig5a** and evaluate the diagram. This is a hypothetical case of one chromosome with three distinct regions (I, II and III). The short segments under the long contiguous piece of DNA are sequencing reads from a 454 machine. On top is the assembly from algorithm #1 and below is the assembly from algorithm #2. Summarize what algorithm #2 did by mistake. Explain how this mistake could take place.

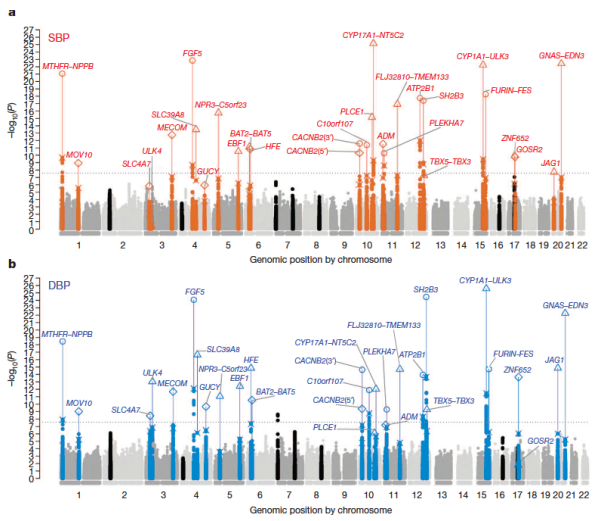


Regions of similarities and #2 combined them into one locus and put II on a separate scaffold. #2 did not take into consideration the number of reads as indicating a collapsed region.

b) Some authors stated, "hemizygosity of a 600-kilobase (kb) region on the short arm of chromosome 16 causes a highly penetrant form of obesity that is often associated with hyperphagia and intellectual disabilities. The corresponding reciprocal duplication is associated with being underweight. The reciprocal impact of these 16p11.2 copy-number variants indicates that severe obesity and being underweight could have mirror aetiologies, possibly through contrasting effects on energy balance." Do your best to determine how many distinct genes are in this region of the human genome. Explain to me how you determined your answer. Support your answer with screen shots.

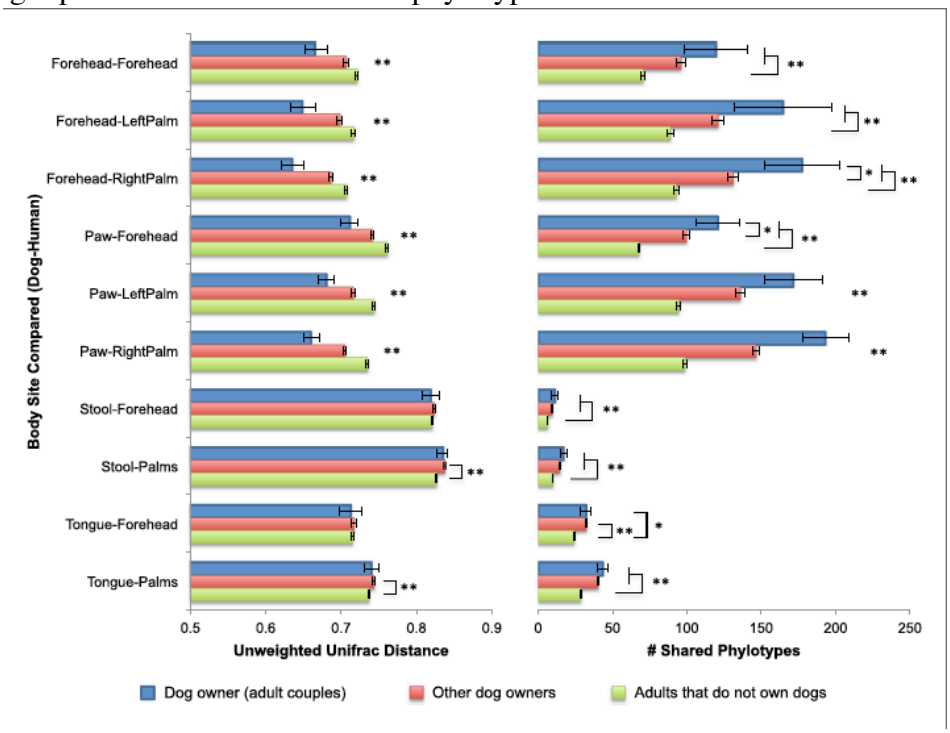
Gene names appear more than once but it looks like about 100 genes??
Genome Browser packed view.
Different correct answers. It is harder than you think it would be to know how many genes.

c) Look at **Seq Fig5b** which shows GWAS data plotted on negative $\log_{10}$ p-value on the Y-axes. They were looking for genes that affect systolic blood pressure (SBP) and diastolic blood pressure blood pressure (DBP). The horizontal dotted line is $p = 2.5 \times 10^{-8}$. Pick one gene to study and tell me why you would choose it. What is the function of the encoded protein for the gene you chose (use GO terms)?



Many options. Choose a gene high up on its column.

d) Look at **Seq Fig5c**. What can you conclude about the microbiome on humans and their pet dogs? The left panel shows the average unweighted UniFrac distance between adult dog-owners and their dogs (blue), between dog-owners and other (not their own) dogs (red), and between adults who do not own dogs and dogs (green). The right panel shows the number of phylotypes shared for the same categories. Comparisons are labeled on the y-axis such that the first body site listed corresponds to the dog and the second site corresponds to the human. Mean ± 95% CI shown. The presence of asterisks lacking brackets indicates that all pairwise comparisons within that group are significant. *p<0.05, **p<0.001 after Bonferroni correction (Wilcoxon test).



Do owners have more bacterial phylotypes than non-dog owners in all sites. The diversity between humans is greatest when compared to non-dog owners in most cases, though not all.