MSA AND PHYLOGENETIC TREES

Align any 10 sequences of your choice by using Clustal Omega and explain their relation using a phylogenetic tree.

INTRODUCTION:

MULTIPLE SEQUENCE ALIGNMENT: It is generally the alignment of three or more sequences of similar length. From the output, homology can be inferred and the evolutionary relationships between the sequences are studied.

There are two techniques involved:

1.**PROGRESSIVE ALIGNMENT METHOD**: This method builds the final MSA by combining pairwise sequence alignments based on a guide tree, starting from the closed pair and progressing to more distantly related sequences.

Eg: Cluster Omega, ClusterW, T-Coffee.

2.**ITERATIVE ALIGNMENT METHOD**: This works similar to the progressive method, however repeatedly realigns the initial sequences as new sequences are added to it.

Eg: PRR, MUSCLE, SAGA.

A phylogenetic tree (also phylogeny or evolutionary tree) is a branching diagram or a tree showing the evolutionary relationships among various biological species or other entities based upon similarities and differences in their physical or genetic characteristics.

There are Two types of phylogenetic trees:

1. Rooted Tree: A rooted phylogenetic tree (see two graphics at top) is a directed tree with a unique node —the root — corresponding to the (usually imputed) most recent common ancestor of all the entities at the leaves of the tree.

2. Unrooted Tree: Unrooted trees illustrate the relatedness of the leaf nodes without making assumptions about ancestry.

METHODOLOGY:

Clustal Omega implements progressive alignment techniques for MSA and phylogenetic tree construction.

Progressive alignment also known as hierarchical or tree method, this method was developed by Hogewag and Hesper in 1984. Initially a guide tree is constructed from the pairwise sequence alignments using a clustering method such as neighbour joining or UPGMA. The MSA construction begins with the closest pair (or seed pair). More distantly related sequences are gradually added to the alignment based on the guide tree.

Neighbour joining methods apply general data clustering techniques to sequence analysis using genetic distance as a clustering metric. The simple neighbour-joining method produces unrooted trees, but it does not assume a constant rate of evolution (i.e., a molecular clock) across lineages. It begins with an unresolved star-like tree. Each pair is evaluated for being joined and the sum of all branch length is calculated of the resultant tree. The pair that yields

the smallest sum is considered the closest neighbours and is thus joined . A new branch is inserted between them and the rest of the tree and the branch length is Recalculated. This process is repeated until only one terminal is present.

The collagen protein sequence is obtained from 10 different species:

Accession Number	Organism
NP_723046.1	Drosophila melanogaster
KRT86529.1	Oryctes borbonicus
NP_001255859.1	Caenorhabditis elegans
BAA04809.1	Homo sapiens
BAA04483.1	partial [Mus musculus]
CAL69616.1	Aphrocallistes vastus
EJW81814.1	Wuchereria bancrofti
AZZ86130.1	Bemisia tabaci
YP_008436674.1	Pandoravirus salinus
XP_003150425.1	Loa loa

RESULT:

```
CLUSTAL O(1.2.4) multiple sequence alignment


EJW81814.1          ------------------------------------------------------------    0
AZZ86130.1          ------------------------------------------------------------    0
YP_008436674.1      ------------------------------------------------------------    0
CAL69616.1          ------------------------------------------------------------    0
BAA04483.1          ------------------------------------------------------------    0
XP_003150425.1      ------------------------------------------------------------    0
NP_001255859.1      ------------------------------------------------------------    0
BAA04809.1          MHPGLWLLLVTLCL---------------------------------------------TE   16
NP_723046.1         -MLPFWKRLLYAAVIAGALVGADAQFWKTAGTAGSIQDSVKHYNRNEPKFPIDDSYDIVD   59
KRT86529.1          ------------------------------------------------------------    0


EJW81814.1          ------------------------------------------------------------    0
AZZ86130.1          ------------------------------------------------------------    0
YP_008436674.1      ------------------------------------------------------------    0
CAL69616.1          ------------------------------------------------------------    0
BAA04483.1          ------------------------------------------------------------    0
XP_003150425.1      ------------------------------------------------------------    0
NP_001255859.1      ------------------------------------------------------------    0
BAA04809.1          ELAAAGEKSYGKPCGGQDCSGSCQCFPEKGARGRPGPIGIQGPTGPQGFTGSTGLSGLKG   76
NP_723046.1         SAGVARGDLPPKNCTAGYAGCVPKCIAEKGNRGLPGPLGPTGLKGEMGFPGMEGPSGDKG  119
KRT86529.1          ------------------------------------------------------------    0


EJW81814.1          ------------------------------------------------------------    0
AZZ86130.1          ------------------------------------------------------------    0
YP_008436674.1      ------------------------------------------------------------    0
CAL69616.1          ------------------------------------------------------------    0
BAA04483.1          ------------------------------------------------------------    0
XP_003150425.1      ------------------------------------------------------------    0
NP_001255859.1      ------------------------------------------------------------    0
BAA04809.1          ERGFPGLLGPYGPKGDKGPMGVPGFLGINGIPGHPGQPGPRGPPGLDGCNGTQGAVGFPG  136
NP_723046.1         QK---GDPGPYGQRGDKGERGSPGLHGQAGVPGVQGPAGNPGAPGINGKDGCDGQDGIPG  176
KRT86529.1          ------------------------------------------------------------    0


EJW81814.1          ------------------------------------------------------------    0
AZZ86130.1          ------------------------------------------------------------    0
YP_008436674.1      ------------------------------------------------------------    0
CAL69616.1          ------------------------------------------------------------    0
BAA04483.1          ------------------------------------------------------------    0
XP_003150425.1      ------------------------------------------------------------    0
NP_001255859.1      ------------------------------------------------------------    0
BAA04809.1          PDGYPGLLGPPGLPG---QKGSKGDPVLAPGSF-KGMKGDPGLPGLDGITGPQGAPGFPG  192
NP_723046.1         LEGLSGMPGPRGYAGQLGSKGEKGEPAKENGDYAKGEKGEPGWRGTAGLAGPQGFPGEKG  236
KRT86529.1          ------------------------------------------------------------    0
```

```
EJW81814.1       ----------------------------------------------------------   0
AZZ86130.1       ----------------------------------------------------------   0
YP_008436674.1   ----------------------------------------------------------   0
CAL69616.1       ----------------------------------------------------------   0
BAA04483.1       ----------------------------------------------------------   0
XP_003150425.1   ----------------------------------------------------------   0
NP_001255859.1   ----------------------------------------------------------   0
BAA04809.1       AVGPAGPPGLQGPPGPPGPLGPDGNMGLG--FQGEKGVKGDVGLPGPAGPPPSTGELEFM 250
NP_723046.1      ERGDSGPYGAKGPRGEHGLKGEKGASCYGPMKPGAPGIKGEKGEPASSFPVK--PTHTVM 294
KRT86529.1       ----------------------------------------------------------   0


EJW81814.1       ----------------------------------------------------------   0
AZZ86130.1       ----------------------------------------------------------   0
YP_008436674.1   ----------------------------------------------------------   0
CAL69616.1       ----------------------------------------------------------   0
BAA04483.1       ----------------------------------------------------------   0
XP_003150425.1   ----------------------------------------------------------   0
NP_001255859.1   ----------------------------------------------------------   0
BAA04809.1       GFPKGKKGSKGEPGPKGFPGISGPPGFPGLGTTGEKGEKGEKGIPGLPGPRGPMGSEGVQ 310
NP_723046.1      -GPRGDMGQKGEPGLVGRKGEPGPE-----GDTGLDGQKGEKGLPGGPGDRGR---QGNF 345
KRT86529.1       ----------------------------------------------------------   0


EJW81814.1       ----------------------------------------------------------   0
AZZ86130.1       ----------------------------------------------------------   0
YP_008436674.1   ----------------------------------------------------------   0
CAL69616.1       ----------------------------------------------------------   0
BAA04483.1       ----------------------------------------------------------   0
XP_003150425.1   ----------------------------------------------------------   0
NP_001255859.1   ----------------------------------------------------------   0
BAA04809.1       GPPGQQGKKGTLGFPGLNGFQGIEGQKGDI---------------------------- 340
NP_723046.1      GPPGSTGQKGDRGEPGLNGLPGNPGQKGEPGRAGATGKPGLLGPPGPPGGGRGTPGPPGP 405
KRT86529.1       ----------------------------------------------------------   0


EJW81814.1       ----------------------------------------------------------   0
AZZ86130.1       ----------------------------------------------------------   0
YP_008436674.1   ----------------------------------------------------------   0
CAL69616.1       ----------------------------------------------------------   0
BAA04483.1       ----------------------------------------------------------   0
XP_003150425.1   ----------------------------------------------------------   0
NP_001255859.1   ----------------------------------------------------------   0
BAA04809.1       ------------------GLPGPDVFIDIDGAVISGNPGDPGVPGLPGLKGDEGIQGLR 381
NP_723046.1      KGPRGYVGAPGPQGLNGVDGLPGPQGY-------NGQKGGAGLPGRPGNEGPPGKKGEK 457
KRT86529.1       ---------MGPKGTDGFDGAPGPRGP-------IGQPGGPGIPGLPGPEGAPGEKGLK  43
```

```
EJW81814.1         ---------------------------------------------------------    0
AZZ86130.1         ---------------------------------------------------------    0
YP_008436674.1     ---------------------------------------------------------    0
CAL69616.1         ---------------------------------------------------------    0
BAA04483.1         ---------------------------------------------------------    0
XP_003150425.1     ---------------------------------------------------------    0
NP_001255859.1     ---------------------------------------------------------    0
BAA04809.1         GPSGVPGLPALSGVPGALGPQGFPGLKGDQGNPGRTTIGAAG---LPGRDGLPGPPGPPG  438
NP_723046.1        GTAGLNGPKGSIGPIGHPGPPGPEGQKGDAGLPGYGIQGSKGDAGIPGYPGLKGSKGERG  517
KRT86529.1         GESGSVGLPGQQGIKGFSGPQGPSGPRGLTGEKGLTIPGARGLDGSPGRDGEKGQKGERG  103


EJW81814.1         ---------------------------------------------------------    0
AZZ86130.1         ---------------------------------------------------------    0
YP_008436674.1     ---------------------------------------------------------    0
CAL69616.1         ---------------------------------------------------------    0
BAA04483.1         ---------------------------------------------------------    0
XP_003150425.1     ---------------------------------------------------------    0
NP_001255859.1     ---------------------------------------------------------    0
BAA04809.1         PPSPEFETETLHNKESGF---------PGLRG---EQGPKGNLGLKGIKGDSGF-C-ACD  484
NP_723046.1        FKGN---AGAPGDSKLGRPGTPGAAGAPGQKGDAGRPGTPGQKGDMGIKGDVGGKCSSCR  574
KRT86529.1         YSGL---RGTPGDALGGIPGAPGPIGSKGDKGEPGRPGQHGTPGLSGEKGDIGGRCTDCF  160


EJW81814.1         ---------------------------------------------------------    0
AZZ86130.1         ---------------------------------------------------------    0
YP_008436674.1     ---------------------------------------------------------    0
CAL69616.1         ---------------------------------------------------------    0
BAA04483.1         --------------QRVPFARAS---QGLELERGAGLFVGQA---GTADPDKFQGMISE   39
XP_003150425.1     ---------------------------------------------------------    0
NP_001255859.1     ---------------------------------------------------------    0
BAA04809.1         ----GGVPNTGPPGEPGPPGPWGLIGLPGLKGARGDRGSGGAQ------GPAGAPGLVGP  534
NP_723046.1        AGPKGDKGTSGLPGIPGKDGARGPPGERGYPGERGHDGINGQTGPPGEKGEDGRTGLP--  632
KRT86529.1         PGAKGVKGERGRDGIPGTPGQRGSSGLPGLKGEMGLDGLIGSAGPPGQKGKDGLPGRTGQ  220


EJW81814.1         ---------------------------------------------------------    0
AZZ86130.1         ---------------------------------------------------------    0
YP_008436674.1     ---------------------------------------------------------    0
CAL69616.1         ---------------------------------------------------------    0
BAA04483.1         LKV------RKTPRVSPVHCLDEEDDDEDRASGDFGSGFEESSKSHKEDTSLLPGLPQPP   93
XP_003150425.1     ---------------------------------------------------------    0
NP_001255859.1     ---------------------------------------------------------    0
BAA04809.1         LGPSGPKGKKGEPILSTIQGMPGDRGDSGS--QGFRGVIGEPGKDGVPGLPGLPGLPGD-  591
NP_723046.1        -GATGEPGKPALCDLSLIEPLKGDKGYPGA--------PGAKGVQGFKGAEGLPGIPGPK  683
KRT86529.1         PGPRGPPGSLAA--GSITQCPKGSVGKDGS--PGPKGDKGERGIQGVQGSEGLRGLPGER  276
```

```
EJW81814.1      --------------------------------MSVTHISLFLRIERAK-QHLKKRMTT        25
AZZ86130.1      ----------------------------------------------------------        0
YP_008436674.1  ----------------------------------------------------------        0
CAL69616.1      ----------------------------------------------------------        0
BAA04483.1      PVTS-----------------------PPLAGGSTTEDPRTEETEEDAAVDSIGAETL       128
XP_003150425.1  ----------------------------------------------------------        0
NP_001255859.1  ----------------------------------------------------------        0
BAA04809.1      GGQGFPGEKGLPGLPGEKGHP------GPPGLPGNGLPGLPGPRGLP-----GDKGKDGL      640
NP_723046.1     GEFGFKGEKGLSGAPGNDGTPGRAGRDGYPGIPGQSIKGEPGFHGRD-----GAKGDKGS      738
KRT86529.1      GSQ---GHPGFQGSDGRDGEPGRN---GTKGEKGDSIKGESGPPGYM-----GEPGEKGE      325


EJW81814.1      LEETRRNIWRDTLALRK--------------GQ------T-MKRDAHSQQLIM--SEYEQ       62
AZZ86130.1      ----------------------------------------------------------        0
YP_008436674.1  ---------------MAHYNKKHAAYKRDDDDWYSEDRSVAYKHDDDDEDEDDVQDYH        43
CAL69616.1      -----------------------------------------------MSTLVDPK         8
BAA04483.1      PGTGSSGAWDEAIQN-PG-RGLIKGGMKGQKGEPGAQGPPGPAGPQGPAGP-V--VQSPN      183
XP_003150425.1  ------MHLISPFLLVICAVFMIQACP----S------Q-------------------        23
NP_001255859.1  -------MWAIFVFIVPALGYNIRYPSYEPNRQYGPYSNDVIHYDDDDS----------G        43
BAA04809.1      --PGQQGLPGSKGITLPCIIP----GSYGPSGFPG---TPGFPGPKGSRGL-PGTPGQPG      690
NP_723046.1     --FGRSGEKGEP---GSCALDEIKMPAKGNKGEPG---QTGMPGPPGEDGS-PGERGYTG      789
KRT86529.1      --RGQKGIAGTD---ARCPQY-HEINNKGEKGAPGDKGDTGVSGRPGLKGD-KGLPGLDG      378


EJW81814.1      CESCVNLQ-------CPPGPRGAPGLDG-EPGADGFPGTPGKAGIDGIDIL----GHVES      110
AZZ86130.1      ----------------------------------------------------------        0
YP_008436674.1  KKKKD-AKHHDGKKAHKKHHDD----DD-DTDGDDYDKCPDK-----------------        79
CAL69616.1      NSARDGPQGPE----------GP-------------------KGARG-------------        26
BAA04483.1      SQPVPGAQGPPGPQ-GPPGKDGTPGRDG-EPGDPGEDGRPGDTGPQGFPGTPGDVGPKGE      241
XP_003150425.1  ----------------------------------------------------------        23
NP_001255859.1  SYERDSRSHEKEIV-YLPGPPGPPGPPGTGTGKDGVPGTNGKDGTPGDKGDTGDKGDKGD      102
BAA04809.1      SSGSKGEPGSPGLV-HLPELPGFPGPRG-EKGLPGFPGLPGKDGLPGMIGSPGLPGSKGA      748
NP_723046.1     LKGNTGPQGPPGVE-GPRGLNGPRGEKG-NQGAVGVPGNPGKDGLRGIPGRNGQPGPRGE      847
KRT86529.1      ARGIPGERGPSGPV-GPRGLPGPRGDKG-EQGPMGFPGEPGKSGERGTPGQIGQQGSKGE      436


EJW81814.1      NFPCVICPAGPPGSSGQQGEQGRPGVPGKTGP--------------------------      142
AZZ86130.1      ----------------------------------------------------------        0
YP_008436674.1  KCIKVCAVRGPVGPRGPKGPKGDCGKCGPCGPC-----------------------GPK      115
CAL69616.1      ----------NEGSSGLQGDKGDTGLPGGIGLPGPSGLGGNDGTVGDSGDQGEAGLKGDK       76
BAA04483.1      KGDPGIGPRGPPGPPGPPGPSFRQDKLTFIDMEGSG----FSGDI--------------      282
XP_003150425.1  ----------------------------------------------------------        23
NP_001255859.1  TGAPGVGVKGDPGAQGPPGEKGKDGDKGAPGAKGSKGDDGKKGDKGSSGEKGEKGSPGQK      162
BAA04809.1      TGDIFGAENGAPGEQGLQGLT-------------------------------------      769
NP_723046.1     PGISRPGPMGPPGLNGLQGEKGDRGPTGPIGFPGADGSVGYPGDRGDAGLPGVSGRPGIV      907
KRT86529.1      VGRPGIGPPGPPGSMGLPGEKGIVGFPGKPGVPGSSGIPGPSGEKGDRGEPGWPGLPGAH      496
```

```
EJW81814.1        ----------PGLPGKPGLPGRVGPPGLPGKQGEEGDAGIKGPPGDDAIGGVGIKGPPG    191
AZZ86130.1        ----------------------------------------------------------    0
YP_008436674.1    GDCGKCGPCGPR---------GPRGPKGKDGKDGCDGERGPKGAKGDC--GKCGPCGPCG    164
CAL69616.1        GQQGDQGQVGQPGNQGDGGDVGTTGQKGEVGAPGRDGTDGLTGSTGDT--GDSGIQGPSG    134
BAA04483.1        -----ESLRGPRGFPGPPGPPGVPGLPGEPGRFGINGS----YAPGPA--GLPGVPGKEG    331
XP_003150425.1    ----------------------------------------------------------    23
NP_001255859.1    G-------------------------DKGEKGEKGSNGSPGSKGDK--GEKGADGNPG    193
BAA04809.1        ---------GHKGFLGDSGLPGLKGVHGKPGLLGPKGERGSPGTPGQV--GQPGTPGSSG    818
NP_723046.1       GEKGDVGPIGPAGVAGPPGVPGIDGVRGRDGAKGEPGSPGLVGMPGNK--GDRGAPGNDG    965
KRT86529.1        GEKGEPGPIGQPGVMGPLGVAGERGSKGEPGMKGEQGRNGLAGFPGSK--GEPGVPGPEG    554


EJW81814.1        PQGPRGPKGPPGPNGIPS-----------------------------------------    209
AZZ86130.1        ----------------------------------------------------------    0
YP_008436674.1    PKGKDGKDGCDGAR--------------------GPKGEKGDKGP--------------    189
CAL69616.1        ANGATGPVGPPGNDGTPGTNGQAGSQGSRGDTGPAGAVGPSGDQGPTGNNGQQGPPGTQG    194
BAA04483.1        PPGFPGPPGPPGPPGKEGPPGVAGQKGSVGDVGIPGPKGSKGDLGPIGMPGKSG------    385
XP_003150425.1    ----------------------------------------------------------    23
NP_001255859.1    PKGDKGADGAKGADGTPGSKGDKGSDGAKGADGNPGSKGDTGDKGEKGSDGAKGD-----    248
BAA04809.1        PYGIKGKSGLPGAPGFPGI------------SGHPGKKGTRGKKGPPGSIVKKGLPGLKG    866
NP_723046.1       PKGFAGVTGAPGKRGPAGI------------P---GVSGAKGDKGATGLTGNDGPVGGRG    1010
KRT86529.1        PKGFTGPRGHPGTPGSPGM------------DGMPGIRGEKGDRGSPGLPGH------VG    596


EJW81814.1        ----------NHMGPVGPIGETGPAGPPGKRGEMGPPGSIGLPGEPGLPAGHCESSCSI    258
AZZ86130.1        ----------------------------------------------------------    0
YP_008436674.1    ----------------RGPKGEDGCHGPRGPKGEKGAPGPCG---KCGPCGPCG------    224
CAL69616.1        LQGSQGNQGL------------------------QGQIGNPGPPGPQGPAGSEGPR-    226
BAA04483.1        ------L---AGSPGPVGPPGPPGPPGPPGPGFAAGFDDM------------------    416
XP_003150425.1    ----------------------------------------------------------    23
NP_001255859.1    -KGANGTPGSKGDKGEKGADGAKGADGTPGSKGDKGADGAKGADG-------------    292
BAA04809.1        LPGNPGLVGLKGSPGSPGVAGLPALSGPKGEKGSVG---FVGFPGIPGLPGISGTRGLK    922
NP_723046.1       PPGAPGLMGIKGDQGLAGAPGQQGLDGMPGEKGNQGFP---GLDGPPGLPGDASEKGQK-    1066
KRT86529.1        PPGPMGIVGMPGEKGDAGPTGPHGLQGPPGERGQKGEPGIFGLPGEKGEPGETSEKGQK-    655


EJW81814.1        QEVVAPSLLELEV---------------------------------------------    271
AZZ86130.1        ----------------------------------------------------------    0
YP_008436674.1    ----------------------------------------------------------    224
CAL69616.1        ---GEPGSLGLSGDTGGSGVGGGDGEKGEKGSTGIG---GSEGPPGEQGQIGDSGATGSD    280
BAA04483.1        ------------------------------EGSGIPLWTT----AR---SSD    431
XP_003150425.1    ----------------------------------------------------------    23
NP_001255859.1    ------------------TPGQKGDKGEKGLDGPKGDKGADGTPGAKGDKGDKGADGAK    333
BAA04809.1        ---GIPGSTGKMGPSGRAGTPGEKGDRGNPGPVGIPSPRRPMSNLWLK---GDKGSQGSA    976
NP_723046.1       ---GEPGPSGLRGDTGPAGTPGWPGEKGLPGLAVH----GRAGPPGEK---GDQGRSGID    1116
KRT86529.1        ---GEPGASGLRGPSGHPGRHGIPGRKGEQGLPGE----GRMGPPGIK---GDRGSRGRD    705
```

```
EJW81814.1       ------------------------------------------------------------   271
AZZ86130.1       ------------------------------------------------------------   0
YP_008436674.1   ------------------------------------------------------------   224
CAL69616.1       ------------GGAGDSGLNGLQGETGKRGTD----------------------------   433
BAA04483.1       YGRPGHKGEIGFPGRPGRPGTNGLKGEKGEPGDASLGFSMRGLPGPPGPPGPPGPPGMPI   657
XP_003150425.1   ------------------------------------------------------------   23
NP_001255859.1   -----------------GQDGQPGQKGDK--GDKGQGEKGQDGQPGQKGQDGQPG       591
BAA04809.1       LGSPGLIGPKGFPGFPGLHGLNGLPGTKGTHGTP--GPSITGVPGPAGLPGPKGEKGYPG   1212
NP_723046.1      IGAPGLIGERGLPGLAGEPGLVGLPGPIGPAGSK--GER-----GLAGSPGQPGQDGFP-   1365
KRT86529.1       LGRPGEKGEKGLAGLVGIPGMQGPPGLVGLKGEK--GDR-----GISGVPGAPGMDGFP-   954


EJW81814.1       ------------------------------------------------------------   271
AZZ86130.1       ----------------MSFSF------KIKRPCFLLSIILSLQLYQVSVKADGPPGMPG    37
YP_008436674.1   ------------------------------------------------------------   224
CAL69616.1       ----------------------------------------------------GTPGQSG   440
BAA04483.1       YD------------------------------------------SNAFVESGRPGLPG   673
XP_003150425.1   -----------------------------------------TGPRGPDGVDGKDG    37
NP_001255859.1   ----------------------------------------------------QKG   594
BAA04809.1       IGIGAPGKPGLRGQKGDRGFPGLQGPAGLPGAPGISLPS-----LIAGQPGDPGRPGLDG   1267
NP_723046.1      ----------------------------------------------------GAPG   1369
KRT86529.1       -----------------------------------------------------GLKG   958


EJW81814.1       ------------------------------------------------------------   271
AZZ86130.1       LPGLPGRQGQKGEPGIPGLHGPPGPPGPPGLPD------------------------   70
YP_008436674.1   ------------------------------------------------------------   224
CAL69616.1       REGDPGSPGSSGPPGPPGIIGRPGAPGFKMENDGTEMEETEISRGSAGSRGSRGAPGFEG   500
BAA04483.1       QQGVQGPSGPKGDKGEV---GPPGPPGQFP------IDLFHLEAEMKGDKGDRGDAG---   721
XP_003150425.1   EMGPPGPVGPRGETGPIGRQGDPGPKGRPG------VQGPMGPPGKQGEQGAVGPEGK--   89
NP_001255859.1   DK----------------------------------------------------------G   597
BAA04809.1       ERGRPGPAGPPGPPGPSSNQGDTGDPGFPG------IPGFSGLPGELGLKGMRGEPGFMG   1321
NP_723046.1      LKGDTGPQGFKGERGLNGFEGQKGDKGDRG------LQGPSGLPGLVGQKGDTGYPGLNG   1423
KRT86529.1       DEGLAGFVGEKGERGPPGLIGAPGQKGERG------ITGLQGERGVQGERGERGFQGVVG   1012


EJW81814.1       ------------------------------------------------------------   271
AZZ86130.1       --------------G---SPGKSGLAGLPGAPG----------MPGPPGMPGHPGMPGVK   103
YP_008436674.1   ------------------------------------------PRGPKGEPGCHGPRGPK   241
CAL69616.1       DHGNKGDRGNVGATGDKGYNGPQGAPGATGSQGGV----GAIGPPGPAGSPG---PPGPG   553
BAA04483.1       -----------------QKGE---RGEPGAPGGGFFSSS---VPGPPGPPGYPGIPGPK   757
XP_003150425.1   ----------PKG--------------------------------------------    92
NP_001255859.1   EKGDIGQPGQKGDKGDKGQDGQ---PGQKGQDG----------QPGQKG---QDGQPGQK   641
BAA04809.1       TPGKVGPPGDPGFPGMKGKAGARGSSGLQGDPGQTPTAEAVQVPPGPLGLPGIDGIPGLT   1381
NP_723046.1      NDGPVGAPGERGFTGPKGRDGRDGTPGLPGQKGEP----GMLPPPGPKGEPGQPGRNGPK   1479
KRT86529.1       LEGPIGPPGQKGEKGSPGFN------GLPGAQG----------EKGSKGERGRDGIRGLA   1056
```

```
EJW81814.1       --------------------------------------------------------   271
AZZ86130.1       GDMGFPGLAGPKGDPGDCCISEGQK----TSAREKRGL----------------YLRPYG   143
YP_008436674.1   GEKGEKGDCGKCGPCGPCGPKGPKGKDGKDGCDGARGPKGEKGDKGPRGAKGEDGCPGAR   301
CAL69616.1       GPPGR-----------------------D---AENGPPGQPGPAGSPGRS            577
BAA04483.1       GESIR-------GPPGPPGPQGPPG----------IGYE---GRQGPPGPPGPPGPPSFP   797
XP_003150425.1   --------------------------------------------------------     92
NP_001255859.1   GDKGE------KGEKGEKGEKG---------QKG-------------------------   660
BAA04809.1       GDPGAQGPVGLQGSKGLP---------GIPGKDGPSGL------PGPPGALGDPGLPGLQ   1426
NP_723046.1      GEPGRPGERGLIGIQGERGEKGERGLIGETGNVGR---------PGPKGDRGEPGERGYE   1530
KRT86529.1       GERGEQGLAGLIGQKGERGRDGNQGLIGERGEKGVVGLD---GKPGYPGPTGPKGERGID   1113


EJW81814.1       --------------------------------------------------------   271
AZZ86130.1       GAQ---GMKGDKGDQGPIGEPGHCDVRVIKSVNLDDEAKK--INLETSRNGSREAPAITC   198
YP_008436674.1   GPK---GPKGDQGECGPRGPRGPKGEDAVVDECLIEKLVYKAVHKALDKDDTDDEDHKKK   358
CAL69616.1       ---CSDGQDGQSGPQGPPGPPGPPGTVLNNGAGG----------K-------------Y   610
BAA04483.1       GPHRQ-----TVSVPGPPGPPGPPGPPGA--------------MG-------------A   824
XP_003150425.1   --------------------------------------------------------     92
NP_001255859.1   --------------------------------------------------------   660
BAA04809.1       GPP---GFEGAPGQQGPFGMPGMPGQSMRVGYTLVKHSQSEQVPP-------------C   1469
NP_723046.1      GAIGLIGQKG------EPGAPAPAALDYLTGILITRHSQSETVPA-------------C   1570
KRT86529.1       GVIGIKGEEGDRGVPGDRGEPCVPQPDYLTGTLLVRHSQSPEVPR-------------C   1159


EJW81814.1       --------------------------------------------------------   271
AZZ86130.1       RELK------------------------------------STHPNLKSSLYWID--   216
YP_008436674.1   KDDKKKKHHKYQAEDDDDDDDDEDYKKKKDQDDEE-------------------EDDD   397
CAL69616.1       SSSHMEFHNG-------------------------------LPVVRSCADWFQDE     634
BAA04483.1       SAGQVRIWATYQTMLDKIREVPEGWLIFVAEREELYVRVRNGFRKVLLEARTALPRG---   881
XP_003150425.1   --------------------------------------------------------     92
NP_001255859.1   --------------------------------------------------------   660
BAA04809.1       PIGMSQLWVGYSLLFVEGQ-------------EKAHNQDLGFAG------SCLPRF---   1506
NP_723046.1      SAGHTELWTGYSLLYVDGN-------------DYAHNQDLGSPG------SCVPRF---   1607
KRT86529.1       EPGHIKLWEGYSLLYIEGD-------------EKAHSQDLGQAG------SCIRKF---   1196


EJW81814.1       --------------------------------------------------------   271
AZZ86130.1       -PNGQDIQDAVLVFCDMELGATCLAASQTVSGSIPAPDQHSSPIWLTDVRD---------   266
YP_008436674.1   EDDDQDYKKKHHKYADKKNKK--------H-------KDWSDTFWSDAKRKAYHH-----   437
CAL69616.1       YSTGNRFASKKFAYCNPNGGSTY--DTMAC-------ACERTEIDAKTSVVCS------C   679
BAA04483.1       --TGNEVAALQPPLVQLHEGSPY--TRREY-------SYSTARPWRADDILANPPRLPDR   930
XP_003150425.1   --------------------------------------------------------     92
NP_001255859.1   --------------------------------------------------------   660
BAA04809.1       -------STMPFIYCNINEVCHY--ARRND----------KSYWLST-----------T   1535
NP_723046.1      -------STLPVLSCGQNNVCNY--ASRND----------KTFWLTT-----------N   1636
KRT86529.1       -------STMPFLFCDFNNVCNY--ASRND----------RSYWLST-----------S   1225
```
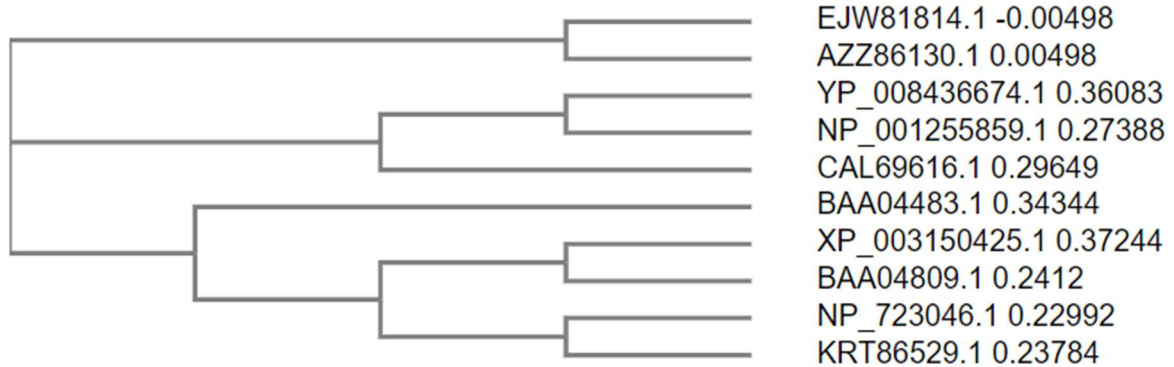
```
EJW81814.1       ------------------------------------------------------------   271
AZZ86130.1       ----------EIQVSYGITRKQLV------RLQQISD------------------------   287
YP_008436674.1   ------------------------------------------------------------   437
CAL69616.1       MNSTITGSDEDEVIQYNPSVSNQ-----VRYLGLLSDVMKIKVP--------------CG   720
BAA04483.1       QPYPGVPHHHSSYVHLPPARPTLSLAHTHQDFQPVLHLVALNTPLSGGMRGIRGADFQCF   990
XP_003150425.1   ------------------------------------------------------------   92
NP_001255859.1   ------------------------------------------------------------   660
BAA04809.1       APIPMMPVSQTQIPQYI------S------------------------------RCSVCE   1559
NP_723046.1      AAIPMMPVENIEIRQYI------S------------------------------RCVVCE   1660
KRT86529.1       QSIPMMPVEETDIVRYI------S------------------------------RCVVCE   1249


EJW81814.1       ------------------------------------------------------------   271
AZZ86130.1       ----------------------------------NASQNITVRASDTEENTYFIRENM--   311
YP_008436674.1   ------------------------------------------------------------   437
CAL69616.1       FG--KIE----------------------------LPGYNVEVSECRFDLETSVQE--I--   747
BAA04483.1       QQARAVGLSGTFRAFLSSRLQDLYSIVRRADRGSVPIVNLKDEVLSPSWDSLFSGSQG--   1048
XP_003150425.1   ------------------------------------------------------------   92
NP_001255859.1   ------------------------------------------------------------   660
BAA04809.1       APSQAIA------------------------VHSQDITIPQCPLGWRSLWIGYSFLM   1592
NP_723046.1      APANVIA------------------------VHSQTIEVPDCPNGWEGLWIGYSFLM   1693
KRT86529.1       APANVIA------------------------VHSQSLSMPDCPTGWTGLWIGYSFIM   1282


EJW81814.1       ------------------------------------------------------------   271
AZZ86130.1       ---------GLSLLGYNDAIISPKNSKSLRYETETEQTYTEHDEKITTDYLLSYT----T   358
YP_008436674.1   ------------------------------------------------------------   437
CAL69616.1       ---------------LLQ--------------------------------PEFLQT   756
BAA04483.1       -----------QLQPGARIFSFDGRDVLRHPAWPQKSVWHGSDPSGRRLMESYCETWRT   1096
XP_003150425.1   ------------------------------------------------------------   92
NP_001255859.1   ------------------------------------------------------------   660
BAA04809.1       HTAAGAEGGGQSLVSPGSCLEDFRATPFIECSGAR----------GTCHYFANKYSFWLT   1642
NP_723046.1      HTAVGNGGGGQALQSPGSCLEDFRATPFIECNGAK----------GTCHFYETMTSFWMY   1743
KRT86529.1       HTAAGAEGGGQSLSSPGSCLEDFRASPFIECNGAR----------GSCHYFANKLSFWLA   1332


EJW81814.1       --------------------------------------------------   271
AZZ86130.1       DRPER-LPIMDVDLSEVLKASTSLTELTVTVGSACFY---------   394
YP_008436674.1   --------------------------------------------------   437
CAL69616.1       KLP-----IASFPSSYSLNA----------RNEVCFGVRMSQK---   784
BAA04483.1       ETTGATGQASSLLSGRLLEQKAASC--HNSYIVLCIENSFMTSFSK   1140
XP_003150425.1   --------------------------------------------------   92
NP_001255859.1   --------------------------------------------------   660
BAA04809.1       TVEER-QQFGELPVSETLKAGQLHT--RVSRCQVCMKSL-------   1678
NP_723046.1      NLESS-QPFE-RPQQQTIKAGERQS--HVSRCQVCMKNSS------   1779
KRT86529.1       TIEIE-QQFR-TPQRQTLKAGNVRD--RVSRCQVCIKNT-------   1367
```

PHYLOGENETIC TREE:

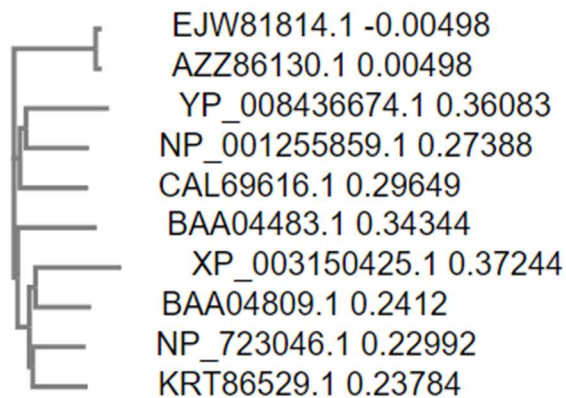In a cladogram, the branch length is not proportional to the number of evolutionary changes/mutations.

Hence, it doesn't indicate the evolutionary distance between taxa.

Branch length: ◉ Cladogram   ○ Real

EJW81814.1 -0.00498
AZZ86130.1 0.00498
YP_008436674.1 0.36083
NP_001255859.1 0.27388
CAL69616.1 0.29649
BAA04483.1 0.34344
XP_003150425.1 0.37244
BAA04809.1 0.2412
NP_723046.1 0.22992
KRT86529.1 0.23784

A phylogram is an estimate of phylogeny. The branch lengths are proportional to the amount of evolutionary change.

Branch length: ○ Cladogram   ◉ Real

EJW81814.1 -0.00498
AZZ86130.1 0.00498
YP_008436674.1 0.36083
NP_001255859.1 0.27388
CAL69616.1 0.29649
BAA04483.1 0.34344
XP_003150425.1 0.37244
BAA04809.1 0.2412
NP_723046.1 0.22992
KRT86529.1 0.23784

CONCLUSION: Wuchereria bancrofti and Bemisia tabaci , Pandoravirus salinus

 And Caenorhabditis elegans,Loa loa and Mus musculus ,Drosophila melanogaster and Oryctes borbonicus are genetically closest pairs.

Pandoravirus salinus,Caenorhabiditids elegans and Aphrocallisters are evolved from same ancestor.

Mus musculus,Loa loa ,Homosapiens,Drosophila melaogas,Oryctes borbonicus are evoloved from same ancestor.