1. Metagenomic sequencing of the gut content from *Tipula* sp.

Metagenomic sequencing of a sample taken from the gut of a *Tipula* sp. larvae revealed the following protein sequence:

MGREKAHISLVVIGHVDVGKSTTTGHLIYKCGGIDKRTIEKFEQEADNIGKASFKYAWVLDKLKAERERGITIDISLW
KFETNKFYFTIIDAPGHRDFIKNMITGTSQADAAILIVAANVGEFEAGISKDGQTREHALLAYTLGVRQMIVCVNKM
DDKSVNWSEARYTEIKNEMGNYLKKVGYNPEKIPVLPISGFNGDNMLEKSVNMPWYKGLTLFDTLDTLEVPKRPS
DKPLRLPIQDVFKIGGIGTVPVGRVETGTLTPGQIVTIAPASITTECKSVEMHHETLTQAVPGDNVGFNLKNVSVKDI
KRGFVCGDSKNDPPMEADTFQAQVIVMAHPGQISNGYTPVCDCHTSHIAVKFKEIQSKIDRRTNKVMEETPKFIKS
GDSAFVLMQPTKPMVVEKFQEYPPLGRFAVRDMRATVAVGVIKEVTKKSRDAKVTKAAAKAIKK

Based on the sequence similarity, determine which protein it is and from which group of organisms it comes. Show this on a phylogenetic tree constructed by any method which you learned that includes at least 20 of the closest homologues from the GenBank database. When you calculate the tree do not forget to calculate the branch support (bootstrap or posterior probability).

2. Metagenomic sequencing of the gut content from *Chinchilla* sp.

Metagenomic sequencing of a sample taken from the gut of *Chinchilla* sp. revealed the following protein sequence:

MSTIRKIFAREILDSRALPTVEVDLFTDKGMFRAAVPSGKSTGEHEAVELRDGGSRFLGKGVTHAVSNINTVIAPALI
GKSIMDQHALDKIMCDLDGDPKKSKLGTNAILPVSIAACRAAASERGIPLYRYIAELSGNSTLRLPVPSLNVINGGKH
AGNKLPFQEYMIAPCVASSFSEAIRMGAEVYGCLKEILKARFGLDATNVGDEGGFAPPVTDVMEPLELLVEAIEKAG
HTGRVKICIDPAASEFYESKEKMYDLGFKDTTHKKITSEEMQMIFEQMVEKYPIVSLEDTFDENDFDAFAKLTAVM
NKKKIQVMGDDLLVTCPERIRMAVEKKSCTSLLLKVNQIGTVSETIEAAKIALDAGWSVMVSHRSGETEDSFIADLA
VGLGCGQIKSGAPCRGERTAKYNQLIRIEEELGKEAIFGYEAWKK

Based on the sequence similarity, determine which protein it is and from which group of organisms it comes. Show this on a phylogenetic tree constructed by any method which you learned that includes at least 20 of the closest homologues from the GenBank database. When you calculate the tree do not forget to calculate the branch support (bootstrap or posterior probability).

3. Metagenomic sequencing of freshwater sediment sample

Metagenomic sequencing of a fresh-water sediment sample revealed the following nucleotide sequence:

TTCTGGTTGATCCTGCCAGAGGCCACTGCTATCGGAGTTCGATTAAGCCATGCGAGTCGAGAGGTGTAAGACC
TCGGCATATTGCTCAGTAACACGTGGACAACCTACCCTAAGGAGGGGGATAACCCCGGAAAACTGGGGATAA
TACCCCATAGACTATGGATGCTGGAATGCTTTGTAGTCGAAAGGTCCGCCGCCATAGGATGGGTCTGCGGCC
GATTAGGTTGTTGTTGGGGTAACGGCCCAACAAGCCTGTAATCGGTACGGGTTGTGGGAGCAAGAGCCCGGA
GATGGATTCTGAGACACGAATCCAGGCCCTACGGGGCGCAGCAGGCGCGAAAACTTTACAATGCGGGAAAC
CGTGATAAGGGAACTCCGAGTGCCCGTTAAATCGGGCTGTCCACCAGTTTAAACAACTGGTGAAGAAAGGGC
CGGGCAAGACCGGTGCCAGCCGCCGCGGTAATACCGGCGGCTCGAGTGGTGGCCACTATTACTGGGCTTAAA
GCGTTCGTAGCTGGTTTGTTAAGTCTCTGGGGAAATCTACCGGCTTAACCGGTAGGCGTCTCAGGGATACTGG
CAGACTAGGGACCGGGAGAGGTGAGAGGTACTCCAGGGGTAGGAGTGAAATCCTGTAATCCTTGGGGGACC
ACCTGTGGCGAAGGCGTCTCACCAGAACGGCTCCGACAGTGAGGGACGAAAGCTGGGGGAGCAAACCGGAT
TAGATACCCGGGTAGTCCCAGCTGTAAACGATGCGCGTTAGGTGTATCGGTGACCACGAGTCACCGAGGTGC
CGAAGAGAAATCGTGAAACGTGCCGCCTGGGAAGTACGGTCGCAAGGCTGAAACTTAAAGGAATTGGCGGG
GGAGCACCACAACGGGTGGAGCCTGCGGTTTAATCGGACTCAACGCCGGAAAACTCACCAGTTAGGACAGCT

GAATGATAGTCGGGCTGAAGACTCTACTTGACTAGCTGAGAGGAGGTGCATGGCCGTCGTCAGTTCGTACTG
TGAAGCATCCTGTTAAGTCAGGCAACGAGCGAGACCCACGCCAACAGTTGCTAGCTCGTCCTCAGGGATGGA
GAGGACACTGTTGGGACCGCCTCTGCTAAAGAGGAGGAAGGAATGGGCAACGGTAGGTCAGCATGCCCCGA
ATTAACTGGGCTACACGCGGGCTACAATGGGTGGGACAATGGGTATCGACACCGAAAGGTGAAGGCAATCTC
CTAAACCCATCCGTAGTTCGGATTGTGGGCTGCAACTCGCCCACATGAAGCTGGAATCCGTAGTAATCGTGTC
TCAATATGGCACGGTGAATATGTCCCTGCTCCTTGCACACACCGCCCGTCAAACCACCCGAGTGAGGTCTTGAT
GAGGCCGTAGTTTTTGCTGCGGTCGAATCTAGGTTTTGCAAGGGGGGTTAAGTCGTAACAAGGTAGCCGTAG
GGGAATCTGCGGCTGGATCACCTCCT

Based on the sequence similarity, determine which gene it is and from which group of organisms it comes. Show this on a phylogenetic tree constructed by any method that includes at least 20 of the closest homologues from the GenBank database. When you calculate the tree do not forget to calculate the branch support (bootstrap or posterior probability).

4. Origin of the glycolytic enzyme in *Monocercomonoides* sp.

The glycolytic enzyme fructose biphosphate aldolase from *Monocercomonoides* (accession number ABG56068) has a complicated past. Using phylogenetic methods try to investigate the evolutionary history of this enzyme. Include in the phylogenetic tree prokaryotic and eukaryotic sequences. When you calculate the tree do not forget to calculate the branch support (bootstrap or posterior probability).

5. Environmental sequence identification

The following small subunit rRNA sequence was obtained from a freshwater sample:

TCTGGTTGATCCTGCCAGCAGTCATATGCTTTGTTCAAGGGCTAAGCCATGCACGTCTCAGCGCAAACGGAGT
GACAGTGGATCTGTGAATGGCTCCTTACATCAGCAGTCATCTACGTGATAGAGTGTGCTCGGTCCACCTGCAA
GGACCCCATTGGACATCCACCAAAACCTTGTGGCTAATACACGTTCGACCCAGTCAGCCATGCAACACTCGGC
AGGGATCCTGTCTCCGGACAGTCCCTTCACCGGTGGTGGCGGATGTATGCCCAGCTGATACGAAGACCAGCG
GCCGCAAGGCCAGTGTGTTGGCATGGTTGACTCAGGCTGGCCCTCCGTGGCCGCAGTGCTGGTGGATTTCGT
GCATGCCTCGTGCATGCCCCACTTGATCGCAAGAGCTTCTGACCTATCAGCTTGACTGTGGTGTATCGGACCAC
AGTGGCCTTGACGGGTAACGGAGAATCAGGGTTCGATTCCGGAGAGGGAGCCTGAGAGACGGCTACCACTA
CCAAGGTGGGCAGCAGGCACGCAAATTGCCCCATGCAAAGACAGTCTGTGAGGCAGCGACGAACAGTAGCA
ACCCCGTCGGCCTTACGTGCCGATGGGGCTTGGAATGGACGCTATCCAAAGACAGCCGTGAGTATCAACCGG
AGGGCAAGTCTGGTGCCAGCAGCTGCGGTAATTCCAGCTCCGAGGGCGTATACTAACATTGCTGCTGTTAAAA
CACTTGTAGTCTGCCTACGGGCTGCAGGTCTGCTGGGTGGCCGGTTTGTTGTTTCTCTGGCCAGGGAAGGACC
TCGGTTCGACCCTGTGTTGGGCTGCAACGGCTGGACTCAACCCCCAGTGGTACGTCCCTGCGCCCACCTCTCA
GTCGATGGTGAGATCTGCTCCTGCCAAAAGTCTGCTTCACTGCAGGCCAAAGCGGTTTATGCCTCCCGCACTG
GCAACGGACACCAACAGGGGACCCAGCCTCGAGCTGGGTAGTCTACCTCTGGTCCACCACCGGAGCCCACCG
TCTTCGACACCCTGGAAAACTCAGTGTGCTCAAAGCATCCCCGCGACGGCTGAATGTCCATCCATGGAATGTC
AAGGCATCGACCAAGTGTGGCATTGGAGTTGTGCTGgCCTTGGGGCCCACTCTGGACAACCTGGTGGTGTGTT
CCTGCAGGATCAACAGGATCGTTGCCCTGCCTGGCCTTCGGGTCTCGTCAGGCTTCGTCCCCTGTCCCTGCAGC
TTGCACCCATCGATCGTAAGTGATGGGACTGTTCGGGGTGAAAGATACGGGAGCGCCAGAGGTGAAATTCTT
AGATCGCTGCCAGATCCACTGCAGCGAAGGCGTTCTGCAAGTGCACGTCCGTCGATCAAGAATGAGAGTTCG
GGGAGCAAAGATGATCAGACACCGTCGTAGTCCGGCCACTGTAAACGATGCCGGCCAGGCCTTGGCAGAGC
AAGAATCCTAGACTCTGTCAGGGCCACTCCTCCCACAACGAGAAATCCACAGCCTGTGGGTTCAGGGGGGAG
TACTGTCGCAAGGCTGAAACTTAAAGGAATTGACGGAATGGCACCACAAGGCGTGGAGTATGCGGCTTAATT

TGACTCAACGCGGGGAATGTTACCAGGTCAGGACGCAACTGGGATTGACAGATTGAGAGCTCTTTCTTGATCT
TGTGGACGGTGGTGCATGGCCGCTCCTGATTGGTGGAGTGATTTGTCTGGTTGATTCCGATAACGAGTGAGA
CATCTGCCTCCCACTAGCCTGAGGCTCGCATTTGGTAGGGTTCGGCTGCTCGGTGGCAGCCCCCTGGCAACAG
GGGGAGATGTACCGGTGCATGCTCCCGAGAGCCTCCAGTTCAGCTTCTCTGAGGTGCTGTGTCCGCCACAAAG
GGCATGCATGCTAGAGCCAACAGCAGGTCTGTGATGCTCCCAGATGTCCTGGGCCGCACGCGCACTACATTGT
CACAGTGAAGGTGTCGACATGCCCACTCCGGTGGGCCCTGGCCTGAAGAGGCTGGGAAATCCTGCAAGCCTG
TGACGTACTGGGGATAGATGGTTGCAACTGTCTGCCTTGAACGTGGAATGCCTAGTATGCCTGGGTCATCAGC
CCAGACCGATTGTGTCCCTGCCATTTGTACACACCGCCCGTCGTTGCTACCGATGGTGGCTGGATAGAGTCAT
CAGGAGGTTGTGGAGGCGGCCTCTCGGGGCAGTAGAGCATCCCAGCCGAAATTGGACGATGTCCAGCCACTA
GAGGAAGCAAAAGTCGTAACAAGGTTGCTGTAGGTGAACCTGCAGCAG

Determine from which group of organisms it comes. Show this on a phylogenetic tree constructed by any method that includes at least 20 of the closest homologues from the GenBank database. When you calculate the tree do not forget to calculate the branch support (bootstrap or posterior probability).

6. Diplomonads from a human stool sample

Using diplomonad specific primers, the following sequence was obtained from a human stool sample:

GCGCCCGGCCATCCGGTCGATCCTGCCGGAGCGCGACGCTCTCCCCAAGGACGAAGCCATGCATGCCCGCTC
ACCCGGGACGCGGCGGACGGCTCAGGACAACGGTTGCACCCCCCGCGGCGGTCCCTGCTAGCCGGACACCG
CTGGCAACCCGGCGCCAAGACGTGCGCGCAAGGGCGGGCGCCCGCGGGCGAGCAGCGTGACGCAGCGACG
GCCCGCCCGGGCTTCCGGGGCATCACCCGGTCGGCGCGGTCGCGGCGCGCCGAGGGCCCGACGCCTGGCGG
AGAATCAGGGTTCGACTCCGGAGAGCGGGCCTGCGAGACGGCCCGCACATCCAAGGACGGCAGCAGGCGCG
GAACTTGCCCAATGCGCGGCGCGCGAGGCAGCGACGGGGAGCGCGCGAGCGAGGCGGGCCCACAGCCCCC
GCCGCGGAGCCGAGGGCAAGGTCTGGTGCCAGCAGCCGCGGTAATTCCAGCTCGGCGAGCGTCGCGCGGCG
CTGCTGCAGTTGAAACGCCCGTAGTTGGCCCCCCGCCGCCACGAGGAAACGGGAGCGCTCCAGGCAGGCCCG
TTGGACCCGCCGCGTGGGACCGCGCAGCGGGCGCGGCGCGCCGCGGCAGCCCCGAGGAGAGCGGGCGGGG
GCACCGGTACCGGCCGGGGACGGGTGAAACAGGATGATCCCGCCGAGACCGCCGGCCGCGCAGGCGCCTGC
CAAGACCGCCTCTGTCAATCAAGGGCGAAGGCCGGGGGCTAGAAGGCGATCAGACACCACCGTATTCCCGGC
CGTAAACGGTGCCGCCCCGCGGCCGGCGCGCGCGTCCCGCCGGCCGCCCAGGGAAACCGGGAGGCTCCGGG
CTCTGGGGGGGAGTATGGCCGCAAGGCTGAAACTTGAAGGCATTGACGGAGGGGTACCACCAGACGTGGAG
TCTGCGGCTCAATCTGACTCAACGCGCGCACCTCACCAGGCCCGGACGCGCGGAGGACCGACAGCCGGGCGC
GCTTTCGCGATCGCGCGGGCGGTGGTGCATGGCCGCTCCAGCCCGTGGCGCGAGCCGTCTGCTCCATTGCG
ACAACGAGCGAGACCCCGGCCGCGGGCGCCGCGGGACGGCCCGCGCGAGCGGGAGGACGGCGGGGCGAT
AGCAGGTCTGTGATGCCCTCAGACGCCCTGGGCCGCACGCGCGCTACACTGGCGGGGCCAGCCGGCGCCCGC
GAGGACGCGCGGAGCCCCCGCCGTGGCCGGGACCGCGGGCTGGAACGCCCCCGCGCACCAGGAATGTCTTG
TAGGCGCCCGCCCCCACCGCGCGCCGGATGCGTCCCTGCCCCTTGTACACACCGCCCGTCGCTCCTACCGACT
GGGCGCGGCGGCGAGCGCCCCGGACGCGCGAAGGGCCGCGAGCCCCCGCGCCTGGAGGAAGGAGAAGTCG
TAACAAGGTATCCGTAGGTGAACCTGCGGATGGATCCCTC

Based on the sequence similarity, determine which gene it is and try to identify if the gene comes from a parasite or not. Show this on a phylogenetic tree constructed by any method that includes at least 20 of the closest homologues from the GenBank database. When you calculate the tree do not forget to calculate the branch support (bootstrap or posterior probability).

7. Antarctica environmental sample

During sequencing of an environmental sample off the coast of Antarctica, the following sequence was obtained:

GAGTTTGATCCTGGCTCAGGATGAACGCTAGCGGCAGGCTTAACACATGCAAGTCGAGGGGTAACATTGTAG
CTTGCTACAGATGACGACCGGCGCACGGGTGCGTAACGCGTATACAATTTACCTATTACTAAGAGATAGCCCA
GAGAAATTTGGATTAATATTTTATAGCATTATCGATTGGCATCAATTGGTAATTAAAGATTACGGTAATAGATG
AGTATGCGTCCTATTAGCTTGATGGTAAGGTAACGGCTTACCATGGCTACGATAGGTAGGGGTCCTGAGAGG
GAGATCCACCACACTGGTACTGAGACACGGACCAGACTCCTACGGGAGGCAGCAGTGAGGAATATTGGACAA
TGGGAGCAATCCTGATCCAGCCATGCCGCGTGCAGGAAGACTGCCCTATGGGTTGTAAACTGCTTTTATACAG
GAAGAAAACGGTTCACGTGTGAACTGTTGACGGTACTGTAGGAATAAGGATCGGCTAACTCCGTGCCAGCAG
CCGCGGTAATACGGAGGATCCAGGCGTTATCCGGAATCATTGGGTTTAAAGGGTCCGTAGGCGGGACAATCA
GTCAGTGGTGAAAGTTTGCGGCTCAACCGTAAAATTGCCATTGATACTGTTGTTCTTGAGTGCTTGTGAAGTG
GTTAGAATGAGTAGTGTAGAAATGAAATGCATAGATATTACTCAGAATACCGATTGCGAAGGCAGATCACTA
ACAATTCACTGACGCTGATGGACGAAAGCGTAGTAGCGAACAGGATTAGATACCCTGGTAGTCTACGCCGTA
AACGATGGTTACTAGCTGTTCGGACTAATTGCGGTCTGAGTGGCTAAGCGAAAGTGATAAGTAACCCACCTG
GGGAGTACGTTCGCAAGAATGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGTT
TAATATGATACGCGAGGAACCTTACCAGGGCTTAAATGCAGTTTGACAGGGGTGGAAACATCTTTTTCTTCGG
ACAAATTGCAAGGTGCTGCATGGTTGTCGTCAGCTCGTGCCGTGAGGTGTCAGGTTAAGTCCTTATAACGAGC
GCAACCCCTTTGTTTAGTTACCAGCATGTAAAGATGGGGACTCTAGACATACTGCCAGTGTAAACTGTGAGGA
AGGTGGGGATGACGTCAAATCATCACGGCCCTTACGTCCTGGGCTACACACGTGCTACAATGGTAGGGACAG
AGAGCAGCCACTGGGTGACCAGGAGCGAATCTACAAACCCTATCACAGTTCGGATCGCAGTCTGCAACTCGA
CTGCGTGAAGCTGGAATCGCTAGTAATATACAGCCATGATGCGGTGAATACGTTCCCGGGCCTTGTACACACC
GCCCGTCAAGCCATGGAAGCTGGGGGTACCTGAAGTCGGTCGCCGCAAGGAGCCGCCTAGGGTAAAACTAG
TAACTGGGGCTAAGTCGTAACAAGGTAACC

Based on the sequence similarity, determine which gene it is and from which group of organisms it comes. Show this on a phylogenetic tree constructed by any method that includes at least 20 of the closest homologues from the GenBank database. When you calculate the tree do not forget to calculate the branch support (bootstrap or posterior probability).

8. Sequence from the algae *Micromonas* sp.

From the genome sequencing project of the alga *Micromonas* sp. the following sequence was obtained. Based on the homology, find out which protein it encodes. Based on phylogeny, try to estimate in which cell compartment this protein works.

TTCGGCACGAGGCCTCGTGCCGAATTCGGCACGAGGATCGACCTCGGCACGACCAACTCGTGCGTCGCCGTG
ATGGAGGGCCGCGTGGCGAAGGTGATTGAGAACTCGGAGGGCGCGCGTACCACGCCGTCGATGGTGGCGTT
CACGGAGAAAGGCGAGCGGTTGGTGGGGCAGCCCGCGAAGCGCCAGGCGGTGACGAACCCCACGAACACG
CTGTACGCGTGCAAGCGCTTGATCGGCCGCAGGTTCGACGACGCGCACACGCGAAAAGAAAGCAAGATGGTT
CCGTACGAGATCGTGTCCGCTTCGAACGGCGACGCGTGGGTGGAGGCGGGTGGTAAGCAGTACTCGCCGTC
GCAGGTGGGGGCGTTCGTGTTGCAGAAAATGAAGGAAACCGCGGAGAGCTACCTCGGGCACGGCGTGACGC
AAGCCGTCGTCACGGTACCGGCGTATTTCAACGACGCGCAGCGGCAAGCCACGAAGGACGCCGGCCGAATC
GCCGGCCTCGACGTTTTGAGAATCATAAACGAACCCACCGCGGCGGCGCTTTCGTATGGGGTGGATAAAAAA
GAAGgCCTGGTGGCGGTGTACGACCTCGGTGGAGGTACCTTCGACGTGTCCATTTTGGAAATCAGCGGCGGC
GTGTTTGAAGTCAAAGCCACGAACGGCGACACGTTTCTGGGGGGGCGAGGATTTCGACACCGCGCTTCTGGAC
AACTTTCTGCAGACGTTCAAAAAAGAATCCGGCGTCGACTTGCTCGGGGATAAACTCGCGATTCAGCGACTAC
GTGAAGCCGCTGAAAAAGCCAAGATCGAACTGTCATCCGCAGCGTCGACGGATATTAATTTGCCCTTCATCAC
CGCGGACGCCACGGGACCCAAACACATGGCCATGAGCCTTTCCCGAGCGCAACTCGAGAAACTGGTGGCGCC
GCTGTTGGAGCGCACGAAAGATCCGTGCCGCGCTTGTATCAAGGACGCCGGAATCGCCGCGTCCGAAATAAA

CGAAGTGTTGTTGGTCGGGGGGTATGTCACGCATGCCCAAAGTCAACGAGCTCGTGAAAGAGCTCTTCGCGGG
GAAGGAACCCAACAAAGGAGTGAACCCCGATGAGGTGGTCGCGATGGGCGCCGCGATTCAAGGCGGCGTGC
TGCGCGGGGACGTGAAGGACATCTTACTGTTGGACGTCACGCCCctgtccttgGGTatcGAGACcCtgGGCGGGGT
TTTCACGCGTTTGATCAACCGGAACACCACGATTCCCACGAAAAAGTCCC

9. Environmental sequencing from a freshwater pond

The following small subunit rRNA sequence was obtained from a freshwater sample:

CCTGGTTGATCCTACCAGCGAACGTATGCTTCTCTCAAGGGCTAAGCCATGCACGCGTGAGAGCAAGTTGCGT
GCCTTTGCCAAACCTGTCTGCGAATGGCTCATTGGACCAGCGATAATCTCAGCAGAGAGTGgaTccTTcTcgTgA
GCGGGGAGTTCGTGCCTTGCCTGGACTTCCAGACCGTCGGAGTTCTGTCTCTGCGGGGGAAGGGAGGAAGG
CCTTACTCCGAGGCCGGTTCGGCATACCCTCTCTTTCTGAAACAGTTGGCTAATACATGTGCGAGGGCACTCCT
GTCCATTCCCTCGGGAATGGGACGGGATGCACTGGGCTGACCCAAAGAGGGaACTTTGCCGCCCCGGCGGGG
GCAATTTTCCTGGGGTGGAACCAAGCCTTCTGTGGAAGACCGAGGGGGTGGGTGCTGCTCTGGGTGTCTCAC
CCAGGGTGGCACCGCCCACGAGAACGGATTCTGCTGCGTATGGTGCGCTAGTGGCGGCCCTGCGAAGCCCCA
CTACTCTGCTTGAGTTTCTGCCCTATCAGCCTGCACGGTGGTCCAGGGGACCACGGCGGCGATGACGGGTAAC
GCGGAATTAGGGTTCGATTCCGGAGAGGGAGCCTGAAATACGGCTACTACGACCAAGGTTGGCAGCAGGCG
CGCAAATTGCTCAATATCGATGTGCATTGGTGAAGCAGCGACGAGGAGTGGTACcTTTTATCGATCACTTGAT
CGGTACTGGTGGAATGAATGACGACCCAAGGAATGCAGTGAGCATCCACTTGTGGGCCAGTCTGGTGCCAGC
AGCCGCGGTAATTCCAGCACAAGTGGCCTACTTTGACGTTGTTGCCGTTAAAAAGCTCGTAGTTGGTTGGCAT
CCGAGAGAGCCCCGGAAGCTCCCCAGGCCCTGTTTGGGCGTGGTGCGGGCGGAAACGGGCGATCTTGCGGT
GCCCTGCTCAACTCGGGCTGCTCTCGACCTCTGCCCCTTCCATCAGGGACATCGGCTCTGCTGTTCCCTCGGGG
ACAGTCCGGAGGCGGGCAGACGAGGGTGGTTGGGTCTGAGCTTCTCAGTGGGAACGATTTTCCGATGATGG
GAGACGGGCCAGGCCCTCACCGCGCAGCTCTTCTGTTGCGAGGTGCTCCGGACCGCCGCCTCTCGCAAACCGC
AACGCGGTTATTTGGCCTACATCCTGAAGCAAGGCACTGCAGAATGGCCCGATGCACCCCTTCTGGCACCGGG
CCCCCCCTCGGGGGGGGACCGTACCCGGGGTGCAAAAAGCCAAGGAAACGCCCTCTGGTTTTGTACCATGAGC
AAAGCTAAGCGCTCCAAGCAAGTTGTCCTTCCTACCGATCGGATGTTCAAGCGTGGCATAACAAGGGAcCGGT
cCGAaTgGACCCGATGTTCTGGCATTCTTTACTGGCCCGCTTTGGACGAACTTCGCTTCCCTTGGACCATCTCGG
ATGACCCCCCCAGATCTGGGCCGGTGCCCTGCTGAAGTCCCCGGGACCGGGGTTGAACTCTGGTCCCTCTTCG
GCAGGCCCATCCCTCTGGGGGAATCAGGCGCCGCCCTGGGCCAGGGAACGGAGGTGTTGACCGAGGTCGTT
TCTGTGCTCAGGGGACTGGGCGGCGAGCGGTGAAACGCTTTGACCCGTCTCAGGACCACCAAAGGCGAAAG
CACTGAGCAGGACTCGACCTGCCAATCAAGAGCGAGTGTGCGGTGATCGAAGAGGATCAGATACCCTCGTAG
TCCGCGCTGTAAACCGTGCCTACCGaGGATgCGGGCTGCGTTCCTTCGGAATGTCAGCCCGTGTCTTTCCCCCG
CAACGAGAAATCCAAGTGTATGGGTTCAGGGGGGGAGTACGGTTGCAAGACTGAAACTTAAAGGAATTGACG
GAATGGCACCACAAGGCGTGGAGTTTGCGGCTTAATTTGACTCAACACGGGGAACGTTACCTGGATAGCACG
TGGAAAGGATATGACAGGCTGAAGCACTCTCTTcGATAACTCGGAAAGTGGTGCATGGTCGTTCCTGGTTGGT
GAAGTGATTTGTCTGGTTGATTCCGATAACGAGCGAGGTCTTCACCTTCCAGTAGCCGGCGTTCTCGCTTGCGT
AGGGCCATTGTGTGGTCCCCTCGCGCTCAGTGcTggcccgcaagggggacccagggcTgggaTgccgaggTcGCCaTGGCCT
GAGAGTGACGCACCGGTTGCTTCTCAGTGGGTTCAGGCACTTCGTCCTAGACTAAGTGTCTACAACCGGAAGC
CGACAGCAGGTCTGTGATGCCCTTAGATGCCCAGGGCTGCACGCGCACGACAATGCCCCCGTGAGCGTGCaG
TTATTCCCCTGTGGCCGTCGTGATGCCCTGGAACAGGGGGCTCGCCCTCTGGACTGGTGGCTTCACGATTCCG
CAGAGCTATTCAGGATACCGTGGTTCGATGATGGACTGGGAAATCCCCGAAGCTGCACTCTCCGgacgTGACT
GGGGGATCGGTGATTGCCACTGTTCTCCTTCAACGAGGAATGCCCAGTAAGTGCAGGTCATCAGCCTGCGCTGA
CTGCGTCCCTGCCATTTGTACACACCGCCCGTCGTTGCTACCGATGGCGTTGGGATAGAGGCCTGAGGAAACT
AGTAGCGAGACGCGCCTCCGGGTGCCCTCGTGAACGAAGAATTCGGCCAATATCCCATTGCTGGAGGAAGCA
AAAGTCGTAACAAGGTTGCTGTAGGTGAACCTGCTGCAG

Determine from which group of organisms it comes. Show this on a phylogenetic tree constructed by any method that includes at least 20 of the closest homologues from the GenBank database. When you calculate the tree do not forget to calculate the branch support (bootstrap or posterior probability).

10. Plastid origin

From the plastid genome of *Arabidopsis thaliana* (accession number NC_000932) choose one gene and download its prokaryotic and eukaryotic homologues (at least 30 sequences). Try to find out the origin of the primary plastid based on phylogeny. When you calculate the tree do not forget to calculate the branch support (bootstrap or posterior probability).

11. Environmental sequencing of a freshwater sediment sample

Metagenomic sequencing of a fresh-water sediment sample revealed the following protein sequence:

MEGEEVQALVIDNGSGMCKAGFAGDDAPRAVFPSIVGRPRHTQVMAGMGNKDAYVGDEAQSKRGILTLKYPIE
HGIVTNWDDMEKIWHHTFYNELRVAPEEHPVLLTEAPLNPKANREKMTQLMFETFNTPAMYVAIQAVLSLYASG
RTTGIVMDTGDGVTHTVPIYEGYALPHAILRLDLAGRDLTDYMMKILTERGYSFTTTAEREIVRDIKEKLAYVALDFE
AEMKTAAESSSLEKSYELPDGQVITIGNERFRCPEALFQPSFLGMEAAGIHETCYNSIMKCDVDIRKDLYGNIVLSGG
TSMFPGIADRMQKEMTTLAPTTMKIKIIAPPERKYSVWIGGSILASLSTFQQMWISKEEYDESGPAIVHRKCF

Based on the sequence similarity, determine which protein it is and from which group of organisms it comes. Show this on a phylogenetic tree constructed by any method that includes at least 20 of the closest homologues from the GenBank database. When you calculate the tree do not forget to calculate the branch support (bootstrap or posterior probability).

12. Termite gut metagenomic sequencing

Metagenomic sequencing of a sample taken from the gut of *Zootermopsis angusticolis* revealed the following nucleotide sequence:

TATGGAGAGTTTGATCCTGGCTCAGAGTGAACGCTGGCGGCGTGCCTAACACATGCAAGTCGTGCGGGATTC
TGCAGAGTATCGCAGAGTCTAGCGGCAAACGGGTGAGTAACACGTAGGGAACCTCCCTCAAAATGGGGAAT
ATTTCTGAGAAATCGGAGTCAATACCGCATAAGACCACAGTTTGGCATCAGACAGGGGTTAAAGCAGTGATG
CGTTTTGAGATGGCCCTGCGCCCTATCAGCTAGTTGGCGGGGTAACGGCTCGCCAAGGCTATGACGGGTATC
CGGCCTTAAAGGGTGAACGGACACACTGGAACTGAGACACGGTCCAGACTCCTAAGGGAGGCAGCAGTGGG
GAATTTTGGACAATGGGGGAAACCCTGATCCAGCGACGCCGCGTGGAGGATGAAGGCCTTCGGGTTGTAAAC
TCCTTTTAGAGGGGAAGAAATAAATGACGTTACCCTCAGAAAAAGCCACGGCTAACTACGTGCCAGCAGCCG
CGGTAATACGTAGGTGGCGAGCGTTACTCGGAATTACTAGGCGTAAAGCGTATGTAGGCTGTCGTATAAGTC
TTTAATAAAATTCCCCGGCTCAACCGGGAAAAGCTTAAGGATACTGTTCGACTTGAGTGTGGGAGAGGGAGA
CGGAATTCCTGGTGTAGCGGTGAAATGCGTAGATATCAGGAGGAACACCGATGGCGAAAGCAGTCTCCTGGA
CCAATACTGACGCTGATGTACGAAAGTTAGGGTAGCAAACAGGATTAGATACCCTGGTAGTCCTAACCGTAAA
CTATGTTCACTAGATGTGGGGGGGTATCGACCCCTTCCGCGTCGTCGCTAACGCATTAAGTGAACCGCCTGGGG
AGTACGGCCGCAAGGTTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGTGGAGTATGTGGTTTAA
TTCGACGCAACGCGAAAAACCTTACCTGGGCTTGAACTGCTGATGGTAAAAACCAGAAACGGCGATGACCCG
CAAGGGAGTCAGCAGAGGTGTTGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAA
CGAGCGCAACCCTTATCCCATGTTACCTGTAGCGATACAGGGTCTCTGAGGAAACTGCCATTGATAAAATGGA
GGAAGGTGGGGACGACGTCAAGTCATCATGTCCTTTATGTCCAGGGCTACACACGTACTACAATGGCCGGTA
CAGAGGGTAGCAATATCGTAAGATGGAGCAAATCCCTGAAAGCCGGCCCCAGTTCAGATTGTGGGCTGCAAT
TCGTCCACATGAAGTAGGAATCGCTAGTAATCGCAGATCAGCAGGCTGCGGTGAATACGTTCCCGGGCCTTGT
ACACACCGCCCGTCACACCACGAAAGTCAGTCATAGCAGAAGTCGTCGGGCTAACCGAAAGGAGGCAGTCGC
CTAAGTTATGGTTGGTGATTGGGGTGAAGTCGTAACAAGGTAGCCGTACGAGAACGTGCGGCTGGATCACCT
CCTTT

Based on the sequence similarity, determine which gene it is and from which group of organisms it comes. Show this on a phylogenetic tree constructed by any method that includes at least 20 of the closest homologues from the GenBank database. When you calculate the tree do not forget to calculate the branch support (bootstrap or posterior probability).

13. Metagenomic sequencing of the gut content of *Nasutitermes* sp.

Metagenomic sequencing of a sample taken from the gut of *Nasutitermes* sp. revealed the following protein sequence:

RPAGCWAALTLYDQTGARVLIAEKADIRRSGCLAAGVNALNAFIGPGKTPQDYVDYARNDADGIVRGDLLLTMAE
GLNRVTERLETLGLTILKDTRGQYVTRGWRNVKINGENIKPILAQAVLKRENITVLNRVNVFEYVVQENRILGAYAIA
VDEPILYVISAKAVLCATGGASGLYKPNNPGGSRHKLWYPPFNTGAGLSMGIRAGAEMTSLEMRFIALRCKDTIAP
TGTLAQGVGAKQINSLGEIYEDKYGLTTSQRLYGTVTENLSGRGPCYLKTAGITKEQDEDLVKAYLNMAPSQSLRWL
ESGLPPSVKDVEIVGTEPYVTGGHTGAGYWVDTHRETTVRGLFAVGDAAGGCPQKYVTGALVEGEIAAKRAVGY
MSEKIPPLARGGAETVKRKLESYLSHNSGEIYFSVDYLEEEMQTSMDKHAGGIGASYRFSEASLKSADKKIRELEARS
NHICAADMQELVYILELKERLILCKSLIVHLSTRKETRWRSFAERVDFPNKSDTWECYVNSRSENGKICAISRPLVME
GEFYEH

Based on the sequence similarity, determine which protein it is and from which group of organisms it comes. Show this on a phylogenetic tree constructed by any method that includes at least 20 of the closest homologues from the GenBank database. When you calculate the tree do not forget to calculate the branch support (bootstrap or posterior probability).

14. Metagenomic sequencing of the gut content of *Zootermopsis* sp.

Metagenomic sequencing of a sample from the gut of *Zootermopsis* sp. revealed the following protein sequence:

MSEDINDILNEFTNKDYEYGFSSNIHTEFIPKGLDETIVRTISAKKKEPKWLLDFRLKAFRYWLTLGMPQWAQLHIPPI
DYQDMIYYAAPKKKGGPKSLDEVDPELLKTFDRLGIPLHEQKMFSGMAIDAVMDSVSVKTTYKDTLIEKGIIFCSFSE
AVQEHPDLVKQYLGSLVSYRDNFFAALNSAVFSDGSFVFIPKGVRCPMELSTYFRINAMNTGQFERTLIVAEEDTTV
SYLEGCTAPVRDENQLHAAIVEIIVLDRAEVKYSTVQNWYPADKDGKGGIYNFVTKRGICKGDSSKISWTQVETGSA
ITWKYPGCILAGDNSVGEFYSVAVTNHFQQADTGTKMIHIGKNTKSRIVSKGISAGKSQNSYRGLVRVGTQAENAR
NHSQCDSLLLSDTSGAHTFPYMDIQNETAIVEHEATTSKVSEDQLFYCNQRGISTEEAVGLIVNGYAKEVLNKLPME
FAVEAQKLLQISLEGSVG

Based on the sequence similarity, determine which protein it is and from which group of organisms it comes. Show this on a phylogenetic tree constructed by any method that includes at least 20 of the closest homologues from the GenBank database. When you calculate the tree do not forget to calculate the branch support (bootstrap or posterior probability).

15. Metagenomic sequencing of the human gut

Metagenomic sequencing of the human gut revealed the following protein sequence:

ETTNMRKPYVILIGSASGIGKSTIAAELAKQLNIKHLIESDFIRAVVRGIIGKEYAPALHNSSYEAYKSLRNKSKYDNYDE
LVSAGFDEHASYVIPALEKVIQRAITDYDDIIIEGVHLVPGLIDIEQFYEDANIYFFILSSDEEAHKERFVKRAIQIHRGGK
QLEFFTENRIIHNHLISQAEKFNATIVKTENINNTLSKLLKTIKQTCKTVCLTNSVDELEEVVDIIIKQNNSSITKIVYKLG
GFKDSLVKTTNISDSDEATKFIKSInenkdkkedlnklYALSKYRKFTICAPDDDSLNNIIEELTKRGFVYNE

Based on the sequence similarity, determine which protein it is and from which group of organisms it comes. Show this on a phylogenetic tree constructed by any method that includes at least 20 of the closest homologues from the GenBank database. When you calculate the tree do not forget to calculate the branch support (bootstrap or posterior probability).

16. Origin of the euglenid plastid

From the plastid genome of *Euglena gracilis* (accession number NC_001603) choose one gene and download its eukaryotic homologues (at least 30 sequences). Try to find out the origin of the secondary plastid of euglenids based on phylogeny. When you calculate the tree do not forget to calculate the branch support (bootstrap or posterior probability).

17. Metagenomic sequencing of the human gut

Metagenomic sequencing of the human gut revealed the following protein sequence:

CIFITVRTKRPSRYTARDAQFTTKVDAWQPLSRGFLAAFGKSGLDSLTARERGFEPDLRKIAFLTFENSFAPLGGLAA
VVRILPVCLNRAGEDVIVLTPLYANIPKVKEALENGTLEPAILKQEFQASSYRGAVSCYRHVGAPVPTYFIAVEGRFTA
ALNPYSYQDSNEILDDSLAFCAAAPFVLNKLGVDRNVLFHAHDWETAPIALTSKIAVLEGLLGNARTVLTLHNSFDS
GIDVDRRRKFFGEKIPGDTVLKCVIPLLDGPLTTVSAPFANELRDDPLQRTVFTNHLQGEFFANPPIGVENGMFGRP
HLRYTYTALSHARQGVYDKLLVEKRRFRAEMLRTVGRMAGREGVIGGLTEAVVDKSAERCINSDNDADVNTSVNT
NADTDANVKNRPILFMSGRLDLMQKGFDVIFNAVQCFPVGKVALIFCPSSAEGARHSKELAFFRDIAAERPGDIVI
WPFRITEEDYVSAVLGSSFLLMPS

Based on the sequence similarity, determine which protein it is and from which group of organisms it comes. Show this on a phylogenetic tree constructed by any method that includes at least 20 of the closest homologues from the GenBank database. When you calculate the tree do not forget to calculate the branch support (bootstrap or posterior probability).

18. Origin of the chromatophore in *Paulinela chromatophora*

The amoeba *Paulinela chromatophora* contains a plastid-like organelle called chromatophore. Using phylogeny try to find out the origin of this organelle using one of the genes contained in its genome (GenBank accession NC_011087). When you calculate the tree do not forget to calculate the branch support (bootstrap or posterior probability).

19. Metagenomic sequencing of a water sample

Metagenomic sequencing of a water sample revealed the following protein sequence:

MREIVHLQAGQCGNQIGAKFWEVMSAEHGITSTGTYAGDNDLQLERINVYFNEGMGGRYVPRAVLTDLEPGTM
DAIRAGGFGSLFRPDNFVFGQSGAGNNWAKGHYTEGAELVDSIMDVVRKESESCDMLQGFQITHSMGGGTGAG
MGTLLVSKIREEYPDRIMSTYSVVPSPKVSDTVVEPYNATLSVHQLVENADQCFALDNEALYDICFRTLKLTNPSYGD
LNHLIAAAITGTTCSLRFPGQLNCDLRKLAVNMVPFPRLHFFLVGFAPLTSRGSQDYRVLTVPELTQQAFDAKNMM
CAADPRHGRYLTCAMMFRGNMSSKEVDDQMLQIVNKNSSYFVEWIPNNLKASICDIPPKGMKMSSVFVGNSTSI
QEAWKRVAEQFTVMFRRKAFLHWYTGEGMDEMEFTEAESNLNDLVSEYQQYQDATADEEELSLGEDPDYE

Based on the sequence similarity, determine which protein it is and from which group of organisms it comes. Show this on a phylogenetic tree constructed by any method that includes at least 20 of the closest homologues from the GenBank database. When you calculate the tree do not forget to calculate the branch support (bootstrap or posterior probability).

20. Metagenomic sequencing of a brackish water sample

Metagenomic sequencing of a water sample revealed the following nucleotide sequence:

TCGAGTTTTTCAGCAAGATCTGGTTGATCCTGCCAGTAGTCATATGCTTGTCTCAAAGATTAAGCCATGCATGT
CTAAGTATAAACCACTTTATACGGTGAAACTGCGAATGGCTCATTACATCAGTTATAGTTTATTTGATCGTACCT
TACTACATGGATAACCGTAGTAATTCTAGAGCTAATACATGCAAGCAAACCCGACTTCTGGAAGGGTTGTATT
TATTAGATACAAAACCAATGCGGGGGCAACCTCAGTGGTTGGTGATTCATAATAACTGGTCGAATCGCATGGC
CTTTGTGCTGGCGATGGTTCATTCAAATTTCTGCCCTATCAGCTTTCGATGGTAGGATAGAGGCCTACCATGGC
GTTTACGGGTAACGGAGAATTAGGGTTCGATTCCGGAGAGGGAGCCTGAGAAACGGCTACCACATCCAAGG
AAGGCAGCAGGCGCGCAAATTACCCAATCCTGACTCAGGGAGGTAGTGACAATAAATAACAATACAGGGCTC
ATTGAGTCTTGTAATTGGAATGAGTACAATTTAAAATCCTTAACGAGGATCCATTGGAGGGCAAGTCTGGTGC
CAGCAGCCGCGGTAATTCCAGCTCCAATAGCGTATATTAAAGTTGTTGCAGTTAAAAAGCTCGTAGTTGAACC
TGAGAGATTGTCCGCCTGGCGGCCTTTCTGAGGTCTTGCATGGGTGTGTTTACAATTTCTTCTGGCCGGGGTTC
TGCGTCATTTGCCCTTCACTGGGTGGGCGCAGGAGGTCGGCCCTTTTACTTTGAACAAATTAGAGTGTTTAAA
GCAGGCCACTGCCTTTGAATACATTAGCATGGAATAATGGAATAGGACTTGGGTTCTATTTTGTTGGTTTCTAG
GACCGAAGTAATGATTAATAGGGATAGTTGGGGGCATTCATATTTCATTGTCAGAGGTGAAATTCTTGGATTT
ATGAAAGATGAACTTCTGCGAAAGCATTTGCCAAGGATGTTTTCATTAATCAAGAACGAAAGTTAGGGGATCG
AAGACGATCAGATACCGTCGTAGTCTTAACCATAAACTATGCCGACTGGGGATTGGAGGACGTTGTTTACGAC
TCCTTCAGCACCCTATGAGAAATCAAAGTCTTTGGGTTCCGGGGGGAGTATGGTCGCAAGGCTGAAACTTAAA
GGAATTGACGGAGGGGCACCACCAGGAGTGGAGCCTGCGGCTTAATTTGACTCAACACGGGGAAACTTACCA
GGTCCAGACATAGGAAGGATTGACAGATTGAGAGCTCTTTCTTGATTCTATGGGTGGTGGTGCATGGCCGTTC
TTAGTTGGTGGAGTGATTTGTCTGGTTAATTCCGATAACGAACGAGACCTCAGCCTGCTAAATAGGTGACGCG
AACCTTTGGGTTCGTGGACCACTTCTTAGAGGGACTATTCGTGTCTAACGAATGGAAGTTTGAGGCAATAACA
GGTCTGTGATGCCCTTAGATGTTCTGGGCCGCACGCGCGCTACACTGACGAAGGCAACGAGTGTTTACCTTGG
CCGGAAGGTCTGGGGAATCTGTGCAAACTTCGTCGTGATGGGGATAGATTCTTGCAATTATTAATCTTCAACG
AGGAATTCCTAGTAAGCGTGAGTCATCAGCTCGCGTTGATTACGTCCCTGCCCCTTGTACACACCGCCCGTCGC
TCCTACCGATTGGATGATCCGGTGAAATTCTCGGACGTGGCTGAGGGAGCGGGTTCGCCTGCACTCGCGGCC
GCGGAAGTTGCTTAAACCTTATCATCTAGAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCA
GAAGGATCAATCTTCTAGA

Based on the sequence similarity, determine which gene it is and from which group of organisms it comes. Show this on a phylogenetic tree constructed by any method that includes at least 20 of the closest homologues from the GenBank database. When you calculate the tree do not forget to calculate the branch support (bootstrap or posterior probability).