

**MODUL PETUNJUK PRAKTIKUM
MK DASAR BIOINFORMATIKA**



Disusun oleh

Dr.rer.nat. Yunus Effendi, M.Sc.

**PROGRAM STUDI BIOLOGI (BIOTEKNOLOGI)
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS AL AZHAR INDONESIA
JAKARTA - 2018**

PENGANTAR

Bismillahirrahmanirrahim,

Buku modul praktikum Dasar Bioinformatika merupakan buku petunjuk pelaksanaan praktikum untuk MK Dasar Bioinformatika yang berisikan topik-topik praktikum untuk tingkat penguasaan kemampuan analisis data bioinformatika dasar. Topik yang diperkenalkan pada tingkat ini adalah pengenalan dan akses database yang ditujukan selain pengenalan laman database bioinformatika yang ada, juga untuk memberikan skil mahasiswa dalam me'retrieve' data dari database; pengenalan data molekuler (sekuens DNA, RNA dan asam amino); analisis kekerabatan organisme berbasis data molekuler menggunakan analisis multiple alignment analysis; analisis filogenetik yang diakhiri dengan desain pohon filogenetik. Mahasiswa juga akan dihadapkan pada proyek mandiri bioinformatika yang meliputi semua topik tersebut di atas.

Tentunya materi praktikum akan selalu di update sesuai dengan kebutuhan dan perkembangan dunia bioinformatika. Insya Allah akan selalu ada pengembangan dari modul praktikum untuk MK Dasar Bioinformatika kedepannya.

Penyusun

DAFTAR ISI

Pengantar	2
Daftar Isi	3
Modul 1 PENGENALAN WEBSITE BIOINFORMATIKA I - NCBI	4
Modul 2 PENGENALAN WEBSITE BIOINFORMATIKA II – UniPROT, PDB	8
Modul 3 PAIRWISE DAN BLAST	11
Modul 4 ANALISIS FILOGENETIK I – SEQUENCE ALIGNMENT	15
Modul 5 ANALISIS FILOGENETIK II – POHON FILOGENETIK	17
Modul 6 DNA BARCODING MENGGUNAKAN GEN 16srRNA PADA BAKTERI	20

MODUL 1

PENGENALAN WEBSITE BIOINFORMATIKA

Akses sekuens DNA, RNA maupun protein mempersyaratkan kemampuan mengakses database yang ada di Internet. Terdapat berbagai macam database baik yang bebas diakses tanpa bayar maupun yang berbayar. Database-database tersebut berfungsi sebagai tempat repository berbagai macam data molekuler baik DNA, RNA, Protein, data publikasi, taksonomi, dan banyak lagi informasi lainnya terkait data molekuler. Data molekuler yang jumlahnya sangat luar biasa banyak, untuk mengaksesnya diperlukan suatu skill tertentu agar efisiensi dan efektifitas pencarian data tersebut dapat dilakukan dengan lebih baik.

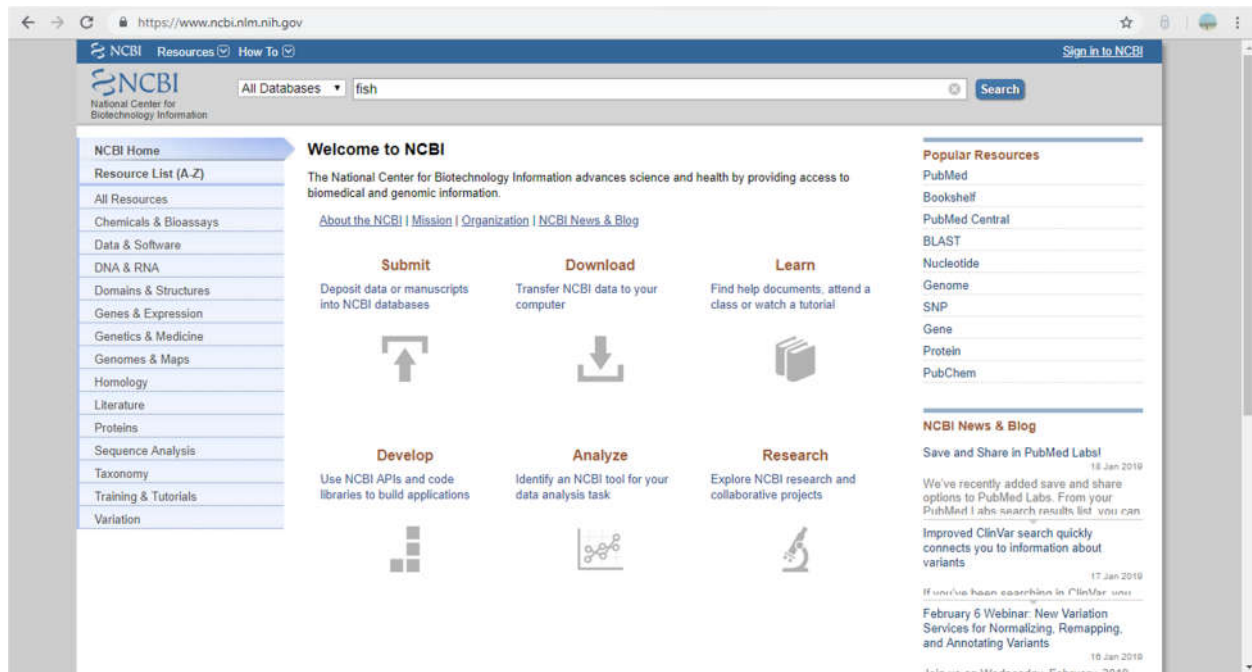
Tujuan:

Dalam modul praktikum ini, mahasiswa diharapkan mampu mengakses data molekuler berupa sekuens DNA ataupun RNA dari berbagai macam organisme (Manusia, hewan, tumbuhan, bakteri, archea maupun virus), mampu mengakses database repository publikasi ilmiah dan informasi turunannya dari gene database yang ada.

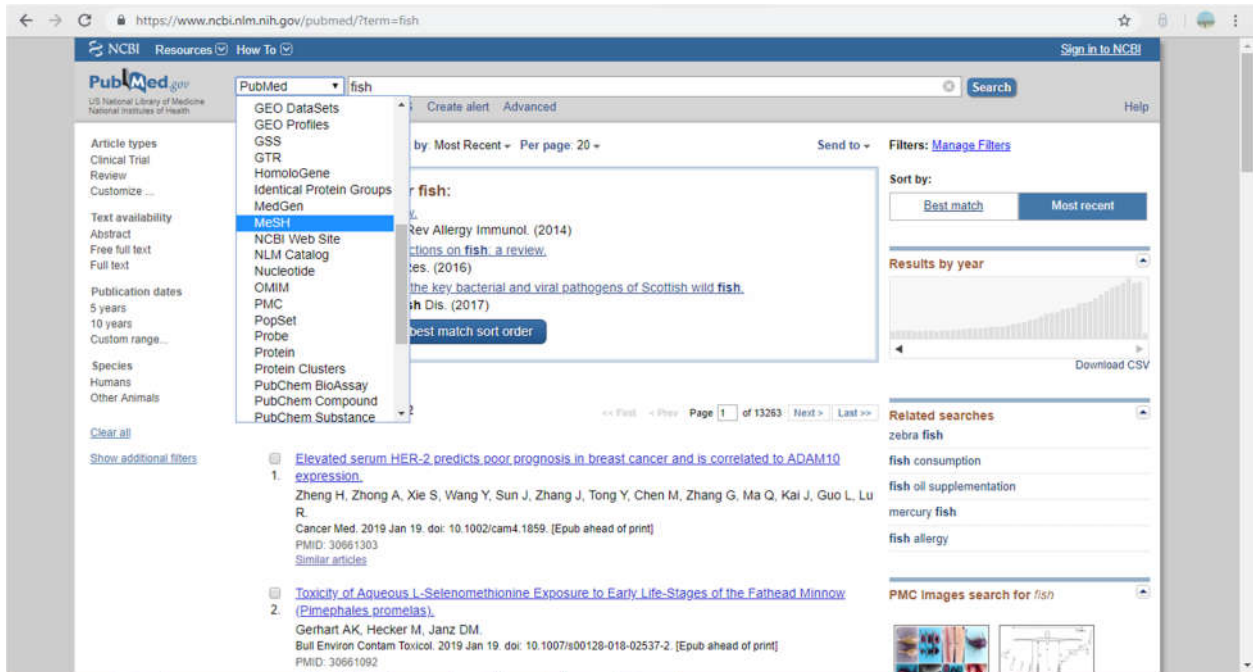
Prosedur:

PENGENALAN NCBI (Gene Bank)

- Buka alamat berikut, www.ncbi.nlm.nih.gov
- Pada kotak query ketik kata “*fish*”, lalu klik [Search]



1. Berapa jumlah DATA PubMed yang Anda peroleh?
2. Berapa jumlah DATA Nucleotide yang Anda peroleh?
3. Berapa jumlah DATA Taxonomy yang Anda peroleh?
 - Klik link hasil DATA [PubMed]
 - Cari box [Search Details]
4. Bagaimana entrez menginterpretasikan query “fish” yang Anda masukan untuk data PubMed ini? "fishes"[MeSH Terms] OR "fishes"[All Fields] OR "fish"[All Fields]
 - Pada tombol pull-down, ganti opsi “PubMed” menjadi opsi “MeSH” (*Medical Subject Heading*), lalu klik [Search]



5. Berapa jumlah DATA PubMed yang Anda peroleh sekarang?

6. Apakah sama atau berbeda dengan jumlah DATA awal?

- Lihat box [Search Details]

7. Bagaimana entrez menginterpretasikan query “fish” Anda sekarang? "fishes"[MeSH Terms] OR fish[Text Word]

- Kembali ke halaman muka NCBI
- Pada pull-down menu “search” pilih opsi “Nucleotide”
- Pada kotak search ketik kata “fish”, lalu klik [Search]

8. Berapa entry yang muncul?

- Buka pada tab baru entry yang pertama

9. Berasal dari organisme apakah entry nucleotide tersebut? Berapa pasang basa sekuen nukleotida pada entry tersebut? (*Print screen full browser, lingkari bagian informasi*)

- Kembali ke halaman muka NCBI
- Pada pull-down menu “Search” pilih opsi “Taxonomy” pada kotak query ketik “fish”, klik [Search]

10. Berapa jumlah DATA Taxonomy yang muncul?

11. Apa saja? Sebutkan!

- Buka pada tab baru, opsi “*protein*” pada kotak query ketik “*fish*”, klik [Search]

12. Berapakah jumlah DATA Protein yang muncul?

- Buka pada tab baru, opsi “*SNP*” pada kotak query ketik “*fish*”, klik [Search]

13. Berapakah jumlah DATA SNP yang diperoleh?

- Kembali ke halaman muka NCBI
- Pada pull-down menu “Search” pilih opsi “*Nucleotide*” pada kotak query ketik “*fish*”, klik [Search]
- Masih pada halaman nucleotide, klik pilihan [mRNA]. Selanjutnya pilih tab [advanced]. Pada halaman advanced search builder, klik pilihan “organism” di pull-down menu, dan ketik “*fish*” pada kolom pencarian. Selanjutnya klik [search]
- Cari box [Search Details]

14. Tulis kembali query yang tertulis pada kotak tersebut! ("Chondrichthyes"[Organism] OR "Dipnoi"[Organism] OR "Actinopterygii"[Organism] OR "Hyperotreti"[Organism] OR "Hyperoartia"[Organism] OR "Coelacanthimorpha"[Organism]) AND biomol_mrna[PROP]

- Masih pada halaman nucleotide, klik pilihan [mRNA]. Selanjutnya pilih tab [advanced]. Pada halaman advanced search builder, klik pilihan “organism” di pull-down menu, dan ketik “*fish*” pada kolom pencarian. Selanjutnya klik pilihan “gene name” di pull-down menu, dan ketik “GFP” pada kolom pencarian. Selanjutnya klik [search]

15. Berapa jumlah DATA yang muncul? Apa GenBank ID yang muncul pertama? (*Print screen full browser, lingkari bagian informasi*)

MODUL 2

PENGENALAN WEBSITE BIOINFORMATIKA II

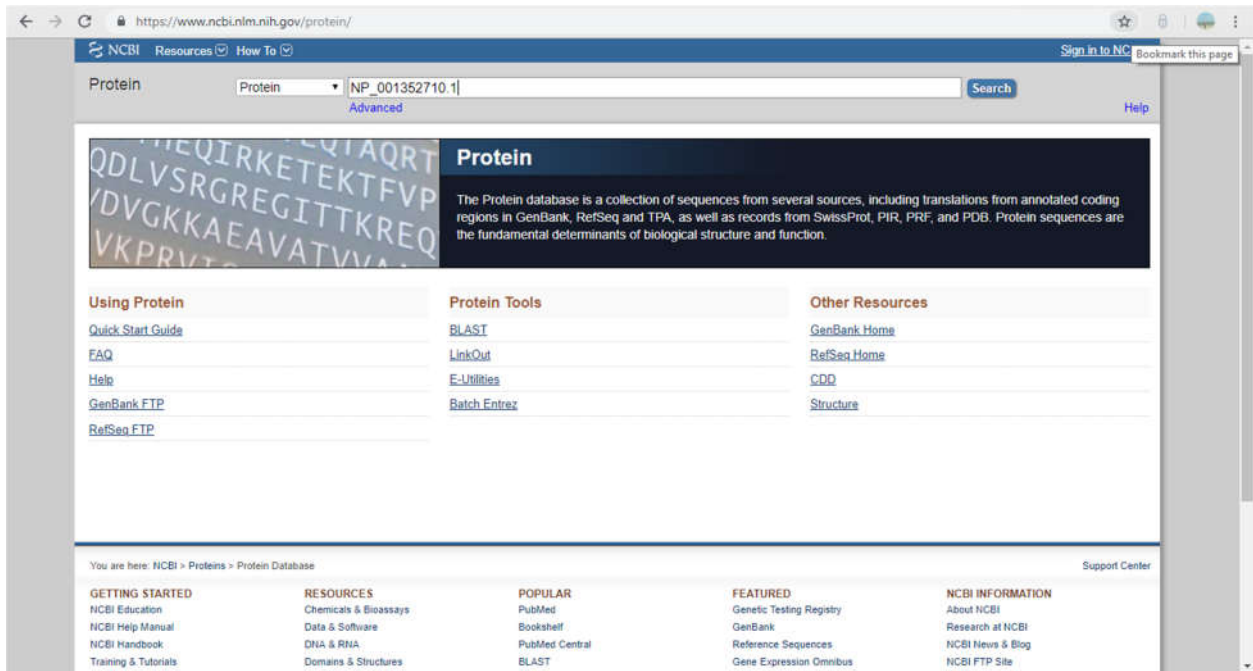
Dalam modul praktikum ini, Anda akan diperkenalkan dengan laman-laman bioinformatika khususnya yang menyimpan database protein. NCBI, PDB, dan UniProt merupakan 2 laman database protein yang cukup lengkap dan dapat diakses secara bebas. Selain itu terdapat fitur-fitur analisis bioinformatik pada kedua laman tersebut.

Tujuan:

Setelah melakukan praktikum ini, diharapkan mahasiswa mampu mengakses database protein minimal pada database NCBI, PDB, dan UniProt, mampu menemukan sekuens protein tertentu yang di'generate' dari sekuens DNA maupun RNA.

Prosedur:

- Buka NCBI protein database <https://www.ncbi.nlm.nih.gov/protein/> dan masukkan NP_001352710.1 pada kotak query.



16. Apakah NP_001352710.1 termasuk primary key atau secondary key?

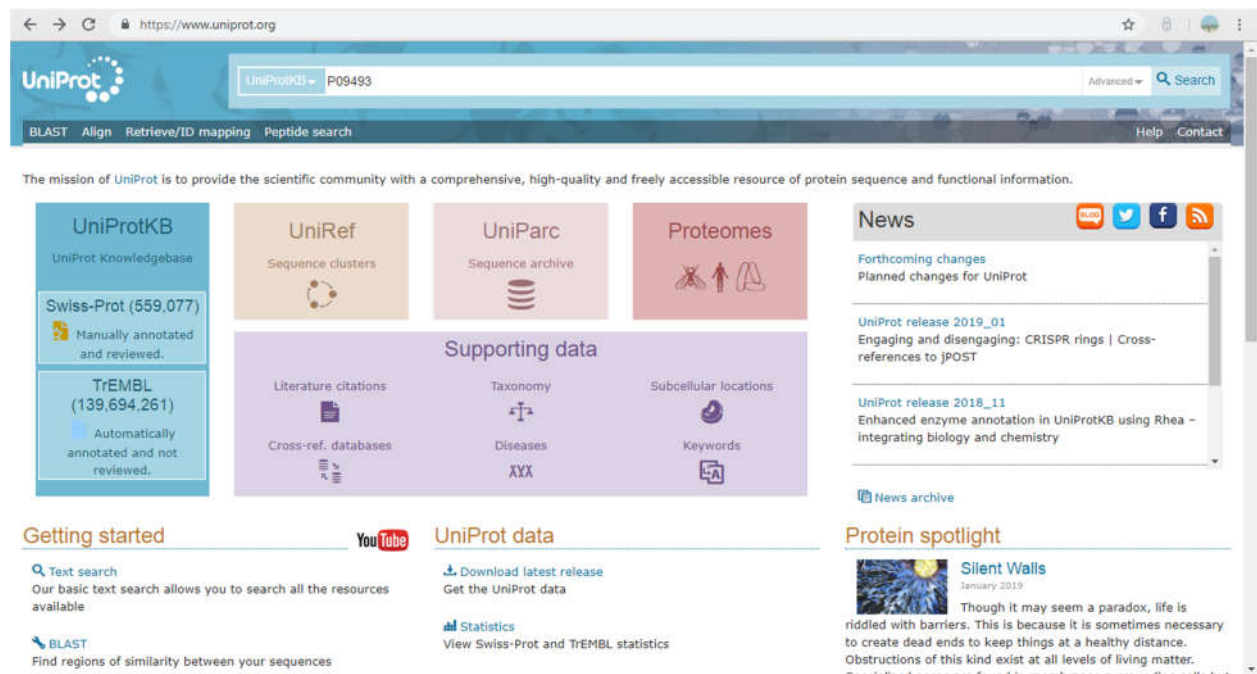
17. Tuliskan nama protein tersebut! tropomyosin alpha-1 chain isoform 18 [Homo sapiens]

18. Berasal dari mana protein tersebut?

19. Apakah fungsi protein tersebut?

UNIPROT

- Buka www.uniprot.org
- Search in: (UniprotKB) dengan query “P09493”



The screenshot shows the UniProt website interface. At the top, there is a search bar with the UniProt logo on the left and a search input field containing "UniProtKB P09493". Below the search bar, there are navigation links for "BLAST", "Align", "Retrieve/ID mapping", and "Peptide search". The main content area features a mission statement: "The mission of UniProt is to provide the scientific community with a comprehensive, high-quality and freely accessible resource of protein sequence and functional information." Below this, there are several colored boxes representing different databases: UniProtKB (559,077 manually annotated and reviewed), UniRef (Sequence clusters), UniParc (Sequence archive), and Proteomes. A "Supporting data" section includes links for Literature citations, Taxonomy, Subcellular locations, Cross-ref. databases, Diseases, and Keywords. On the right side, there is a "News" section with articles about UniProt releases and planned changes. At the bottom, there are sections for "Getting started" (Text search, BLAST), "UniProt data" (Download latest release, Statistics), and "Protein spotlight" (Silent Walls).

20. Tuliskan nama protein tersebut!

21. Berasal dari mana protein tersebut?

22. Apa nama gen dari protein tersebut?

23. Apakah fungsi protein tersebut pada level molekular (molecular function)?

24. Apakah fungsi protein tersebut pada level proses biologi (biological process)?

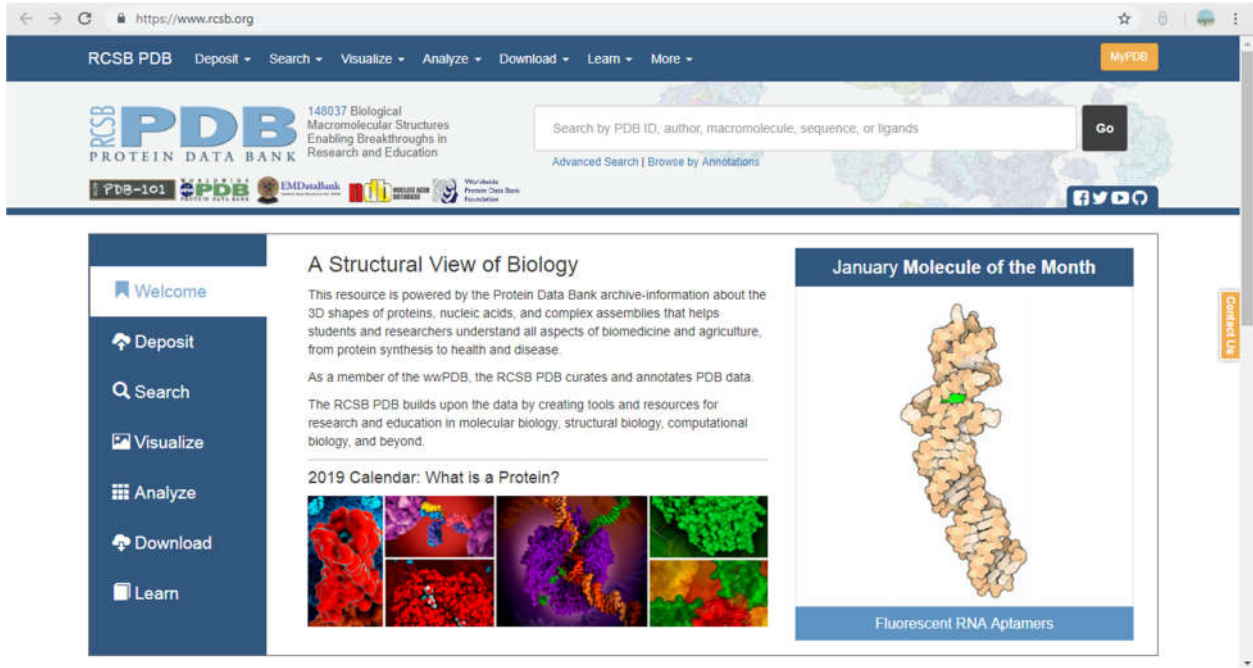
25. Di organela mana dapat dijumpai protein tersebut di dalam sel?

26. Apa penyakit yang disebabkan oleh mutasi gen tersebut?

27. Bagaimana struktur 3D protein model tersebut? (*Print screen & lingkari bagian informasi*)

RCSB PDB

- Buka <https://www.rcsb.org/>
- Pada search engine, ketik “tropomyosin”, klik tombol “Go”



28. Ada berapa jumlah data “structures” yang muncul?
29. Klik tab “Citations”, Sebutkan judul jurnal pada urutan pertama! (*Print screen full browser, lingkari bagian informasi*)
30. Apakah ID protein dari jurnal di atas?
 - Klik nomor ID protein artikel tersebut.
31. Dari organisme apakah protein tersebut?
32. Berapa total berat molekulnya? Dan ada berapa rantai unik pada struktur protein tersebut?
33. Metode apa yang digunakan untuk mendapat gambar molekul tersebut? (*Print screen full browser, lingkari bagian informasi*)
34. Berdasarkan praktikum yang telah Anda kerjakan, apa fungsi dari:
 - NCBI
 - UNIPROT
 - PDB
35. Buat analisis perbandingan mengenai ketiga website tersebut! (Buat Tabel)

MODUL 3

PAIRWISE DAN BLAST

BLAST (Basic Local Alignment Search Tool) merupakan algoritma dan program bioinformatika yang digunakan untuk membandingkan informasi sekuens primer biologi seperti sekuens asam amino untuk protein atau sekuens nukleotida (DNA dan RNA). Fitur BLAST merupakan salah satu fitur penting dalam gene database terbesar, NCBI, yang sangat berguna sebagai mesin pencarian sekuens nukleotida maupun asam amino. Dengan menggunakan BLAST, peneliti mampu membandingkan urutan protein atau urutan nukleotida yang tidak diketahui identitasnya (disebut query) dengan pustaka atau database sekuens yang ada di database NCBI dan mengidentifikasi sekuens serupa atau yang memiliki similaritas dengan sekuens yang ada di database dan sudah diidentifikasi.

Gene1

```
TCATCACTTAGACCTCACCCCTGTGGAGCCACACCCTAGGGTTGGCCAATCTACTCCCAGGAGCAGGGAGGGCAGGAG
CCAGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGACACAACCTGTGTTCACTAGCA
ACCTCAAACAGACACCATGGTGCACCTGACTCCTGAGGAGAAGTCTGCCGTTACTGCCCTGTGGGGCAAGGTGAACG
TGGATGAAGTTGGTGGTGGAGCCCTGGGCAGTTGGTATCAAGTTACAAGACAGGCTTAAGGAGACCAGTAGAAAC
TGGGCATGTGGAGACAGAGAAGACTCTTGGGTTTCTGATAGGCACTGACTCTCTGCCTATTGGGCTATTTTCCCA
CCCTTAGGCTGCTGGTGGTCTACCCCTGGACCCAGAGGTTCTTTGAGTCCTTTGGGGATCTGTCCACTCCTGATGCT
GTTATGGGCAACCCTAAGGTGAAGGCTCATGGCAAGAAAGTCTCGGTGCCTTTAGTGATGGCCTGGCTCACCTGGA
CAACCTCAAGGGCACCTTTGCCACACTGAGTGAGCTGCACTGTGACAAGCTGCACGTGGATCCTGAGAACTTCAGGG
TGAGTCTATGGGACCCTTGATGTTTTCTTTCCCTTCTTTTCTATGGTTAAGTTCATGTCATAGGAAGGGGATAAGT
AACAGGGTACAGTTTAGAATGGGAAACAGACAAATGATTGCATCAGTGTGGAACCTCAGGATCATTTTAGTTTCTT
TTATTTGCTGTTCAACAATTGTTTTCTTTGTTAATTCTTGCTTCTTTTTTTTTCTTCTGCAATTTTTACT
ATTATACTTAATGCCTTAACGTTGTGTATAACAAAAGGAAATATCTCTGAGATACATTAAGTAACCTAAAAAAATG
TTTACGCAGTCTGCCTAGTACATACTATTGGAATATATGTGTGCTTATTTGCATATTCATAATCTACCTACTTA
TTTTCTTTATTTTTAATTGATACATAATCATTATACATATTTATGGGTTAAAGTGTAATGTTTTAATATGTGTACA
CATATTGACCAAATCAGGGTAATTTTGCATTTGTAATTTAAAAAATGCTTTCTTTAATATACTTTTTTTGTTT
ATCTATTTCTACTACTTTCCCTAATCTCTTTCTTTTCCAGGGCAATAATGATACAATGTATCATGCCTCTTTGCACCA
TTCTAAAGAATAACAGTGATAATTTCTGGGTTAAGGCAATAGCAATATTTCTGCATATAAATATTTCTGCATATAAA
TTGTAACCTGATGTAAGAGGTTTCATATTGCTAATAGCAGCTACAATCCAGCTACCATTCTTCTTTATTTTATGGTT
GGGATAAGGCTGGATTATTCTGAGTCCAAGCTAGGCCCTTTGCTAATCATGTTTCATACCTCTTGTCTTCTCCAC
AGCTCCTGGGCAACGTGCTGGTCTGTGTGCTGGCCCATCACTTTGGCAAAGAATCACCCACCCAGTGCAAGGCTGCC
TATCAGAAAGTGGTGGCTGGTGTGGCTAATGCCCTGGCCACAAGTATCACTAAGCTCGCTTTCTGTCTGTTCAATT
TCTATTAAGGTTCTTTGTTCCCTAAGTCCAACCTAACTGGGGGATATTATGAAGGGCCTTGAGCATCTGGAT
TCTGCCTAATAAAAAACATTTATTTTTCATTGCAA
```

Gene2

ATTAATTAATTCTGAGAGCTGCTGAGTTGTGTTTACTGAGAGATTGTGTATCTGCGAGAGAAGTCTGTAGCAAGTAG
CTAGACTGTGCTTGACCTAGGAACATATACAGTAGATTGCTAAAATGTCTCACTTGGGGAATTTTAGACTAAACAGT
AGAGCATGTATAAAAATACTCTAGTCAAGTGCTGCTTTTGAAACAAATGATAAAACCACACTCCCATAGATGAGTGT
CATGATTTTCATGGAGGAAGTTAATATTCATCCTCTAAGTATACCCAGACTAGGGCCATTCTGATATAAAACATTAG
GACTTAAGAAAAGATTAATAGACTGGAGTAAAGGAAATGGACCTCTGTCTCTCTCGCTGTCTCTTTTTTGAGGACTTT
GTGGTCAGTGGGGCTGGAATAAAGTAGAATAGACCTGCACCTGCTGTGGCATCCATTCACAGAGTAGAAGCAAGCT
CACAATAGTGAAGATGTCAGTAAGCTTGAATAGTTTTTCAGGAACCTTGAATGCTGATTTAGATTTGAAACTGAGGC
TCTGACCATAACCAAATTTGCACTATTTATTGCTTCTTGAAACTTATTTGCTGGTATGCCTGGGCTTTTGATGGTC
TTAGTATAGCTTGCAGCCTTGTCCCTGCAGGGTATTATGGGTAATAGAAAGAAAAGTCTGCGTTACACTCTAGTCAT
ACTAAGTAACTACCATTGGAAAAGCAACCCCTGCCTTGAAGCCAGGATGATGGTATCTGCAGCAGTTGCCAACACAA
GAGAAGGATCCATAGTTCATCATTTAAAAAGAAAACAAAATAGAAAAAGGAAAACTATTTCTGAGCATAAGAAAT
GTAGGGTAAGTCTTTAAGAAGGTGACAATTTCTGCCAATCAGGATTTCAAAGCTCTTGCTTTGACAATTTTGGTCTT
TCAGAATACTATAAATAACCTATATTATAATTTTATAAAAGTCTGTGCATTTTCTTTGACCCAGGATATTTGCAA
AGACATATTCAAACTTCCGCAGAACACTTTATTTTACATATACATGCCTCTTATATCAGGGATGTGAAAACAGGGTCT
TGAAAAGTGTCTAAATCTAAAACAATGCTAATGCAGGTTTAAATTTAATAAAAATAAAATCCAAAATCTAACAGCCAA
GTCAAATCTGTATGTTTTAACATTTAAAATATTTTAAAGACGCTTTTTCCAGGATTCACATGTGAAATCTTTTCT
CAGGGATACACGTGTGCCTAGATCCTCATTGCTTTAGTTTTTACAGAGGAATGAATATAAAAAGAAAATACTTAAA
TTTTATCCCTCTTACCTCTATAATCATAATAGGCATAATTTTTTAACTAGGCTCCAGATAGCCATAGAAGAACCA
AACACTTTCTGCGTGTGTGAGAATAATCAGAGTGAGATTTTTTACAAGTACCTGATGAGGGTTGAGACAGGTAGAA
AAAGTGAGAGATCTCTATTTATTTAGCAATAATAGAGAAAGCATTTAAGAGAATAAAGCAATGAAATAAGAAATTT
GTAAATTTCTTCTGATAACTAGAAATAGAGGATCCAGTTTCTTTGGTTAACCTAAATTTTATTTTATTATTGT
TTTATTTTATTTTATTTTATTTTGTGTAATCGTAGTTTCAGAGTGTTAGAGCTGAAAGGAAGAAGTAGGAGAAACAT
GCAAAGTAAAAGTATAACACTTTTCTTACTAAACCGACATGGGTTTCCAGGTAGGGGCAGGATTCAGGATGACTGAC
AGGGCCCTTAGGGAACACTGAGACCCTACGCTGACCTATAAATGCTTGTACCTTTGCTGTTTTAATTACATCTTT
TAATAGCAGGAAGCAGAACTCTGCACTTCAAAAGTTTTTCTCACCTGAGGAGTTAATTTAGTACAAGGGGAAAAAG
TACAGGGGGATGGGAGAAAGGCGATCACGTTGGGAAGCTATAGAGAAAGAAGAGTAAATTTTAGTAAAGGAGGTTTA
AACAACAAAATATAAAGAGAAATAGGAACTTGAATCAAGGAAATGATTTTAAAACGCAGTATTCTTAGTGGACTAG
AGGAAAAAATAATCTGAGCCAAGTAGAAGACCTTTCCCTCCTACCCCTACTTTCTAAGTCACAGAGGCTTTTTG
TTCCCCCAGACACTCTTGCAATTAGTCCAGGCAGAAACAGTTAGATGTCCCCAGTTAACCTCCTATTTGACACCAC
TGATTACCCATTGATAGTCACACTTTGGGTTGTAAGTACTTTTTATTTATTTGATTTTTGACTGCATTAAGAGG
TCTCTAGTTTTTTACCTCTGTTTCCAAAACCTAATAAGTAACTAATGCACAGAGCACATTGATTTGTATTTATTC
TATTTTTAGACATAATTTATTAGCATGCATGAGCAAATTAAGAAAAACAACAATAATGAATGCATATATATGTATA
TGTATGTGTATATATACATATATATATATATTTTTCTTACCAGAAGGTTTTAATCCAATCAGGAGAAGATATGC
TTAGAAGTGAAGTAGAGTTTTTATCCATTCTGTCTGTAAAGTATTTGATATTTCTGGAGACGCAGGAAGAGATCCA
TCTACATATCCCAAAGCTGAATTATGGTAGACAAAACCTTTCCACTTTTAGTGCATCAATTTCTTATTTGTGTAATA
AGAAAATTGGGAAAACGATCTTCAATATGCTTACCAAGCTGTGATTCCAATATTACGTAATAACACTTGCAAAGGA
GGATGTTTTTAGTAGCAATTTGACTGATGGTATGGGGCCAAGAGATATATCTTAGAGGGAGGGCTGAGGGTTGAA
GTCCAACCTCTAAGCCAGTGCCAGAAGAGCCAAGGACAGGTACGGCTGTCATCACTTAGACCTCACCTGTGGAGCC
ACACCCTAGGGTTGGCCAATCTACTCCAGGAGCAGGGAGGGCAGGAGCCAGGGCTGGGCATAAAAGTCAGGGCAGA
GCCATCTATTGCTTACATTTGCTTCTGACACAACCTGTGTTCACTAGCAACCTCAAACAGACACCATGGTGCACCTGA
CTCCTGAGGAGAAGTCTGCCGTTACTGCCCTGTGGGGCAAGGTGAACGTGGATGAAGTTGGTGGTGAGGCCCTGGGC
AGGTTGGTATCAAGGTTACAAGACAGGTTTAAAGGAGACCAATAGAAACTGGGCATGTGGAGACAGAGAAGACTCTG
GGTTTCTGATAGGCACTGACTCTCTCTGCCTATTGGTCTATTTTCCACCCTTAGGCTGCTGGTGGTCTACCCTGG
ACCCAGAGGTTCTTTGAGTCCTTTGGGGATCTGTCCACTCCTGATGCTGTTATGGGCAACCCTAAGGTGAAGGCTCA
TGGCAAGAAAGTCTCGGTGCCTTTAGTGATGGCTGGCTCACCTGGACAACCTCAAGGGCACCTTTGCCACACTGA
GTGAGCTGCACTGTGACAAGCTGCACGTGGATCCTGAGAAGTTCAGGGTGAGTCTATGGGACCCTTGATGTTTTCTT

TCCCCTTCTTTCTATGGTTAAGTTCATGTCATAGGAAGGGGATAAGTAACAGGGTACAGTTTAGAATGGGAAACAG
ACGAATGATTGCATCAGTGTGGAAGTCTCAGGATCGTTTTAGTTTCTTTTATTTGCTGTTCATAACAATTGTTTTCT
TTTGTTAATTCTTGCTTTCTTTTTTTTTCTTCTCCGCAATTTTTACTATTATACTTAATGCCTTAACATTGTGTAT
AACAAAAGGAAATATCTCTGAGATACATTAAGTAACCTAAAAAAAACCTTACACAGTCTGCCTAGTACATTACTAT
TTGGAATATATGTGTGCTTATTTGCATATTCATAATCTCCCTACTTTATTTTCTTTTATTTTAAATTGATACATAAT
CATTATACATATTTATGGGTTAAAGTGAATGTTTTAATATGTGTACACATATTGACCAAATCAGGGTAATTTTGCA
TTTGTAATTTTAAAAAATGCTTTCTTCTTTAATATACTTTTTGTTTATCTTATTTCTAATACTTTCCCTAATCTC
TTTCTTTCCAGGGCAATAATGATACAATGTATCATGCCTCTTTGCACCATTCTAAAGAATAACAGTGATAATTTCTGG
GTTAAGGCAATAGCAATATTTCTGCATATAAATATTTCTGCATATAAATTGTAAGTACTGATGTAAGAGGTTTCATATTG
CTAATAGCAGCTACAATCCAGCTACCATTCTGCTTTTATTTATGGTTGGGATAAGGCTGGATTATTCTGAGTCCAA
GCTAGGCCCTTTTGCTAATCATGTTTCATACCTCTTATCTTCTCCACAGCTCCTGGGCAACGTGCTGGTCTGTGTG
CTGGCCCATCACTTTGGCAAAGAATTCACCCACCAGTGCAGGCTGCCTATCAGAAAAGTGGTGGCTGGTGTGGCTAA
TGCCCTGGCCACAAGTATCACTAAGCTCGCTTTCTGCTGTCCAATTTCTATTAAGGTTCCCTTTGTTCCCTAAGT
CCAATACTAAACTGGGGGATATTATGAAGGGCCTTGAGCATCTGGATTCTGCCTAATAAAAAACATTTATTTTCAT
TGCAATGATGTATTTAAATTATTTCTGAATATTTACTAAAAAGGGAATGTGGGAGGTGAGTGCATTTAAACATAA
AGAAATGAAGAGCTAGTTCAAACCTTGGGAAAATACACTATATCTTAAACTCCATGAAAGAAGGTGAGGCTGCAAA
AGCTAATGCACATTGGCAACAGCCCCTGATGCCTATGCCTTATTCATCCCTCAGAAAAGGATTCAAGTAGAGGCTTG
ATTTGGAGGTTAAAGTTTTGCTATGCTGTATTTACATTACTTATTGTTTTAGCTGTCCTCATGAATGCTTTTTAC
TACCCATTTGCTTATCCTGCATCTCTCAGCCTTGACTCCACTCAGTTCTCTTGCTTAGAGATACCACCTTTCCCTG
AAGTGTCTTCCATGTTTTACGGCGAGATGGTTTCTCCTCGCCTGGCCACTCAGCCTTAGTTGTCTCTGTTGTCTT
ATAGAGGTCTACTTGAAGAAGGAAAAACAGGGGTCATGGTTTGACTGTCCTGTGAGC

Prosedur:

1.
 - a. Gen apakah yang dikode oleh kedua urutan di atas?
 - b. Berasal dari organisme apa saja?
 - c. Berapa panjang gen tersebut?
2. Perhatikan hasil BLAST yang anda peroleh. Pada hasil tersebut terlihat beberapa parameter hasil BLAST seperti di bawah ini. Apakah maksud dari parameter parameter tersebut? Jelaskan!
 - a. Description
 - b. Max Score
 - c. Total Score
 - d. Query Score
 - e. E value
 - f. Ident
 - g. Accesion
3. Lakukan pencarian menggunakan fitur BLAST-X, untuk masing-masing gen tersebut di atas!
 - a. Protein apa yang dikode oleh kedua gen tersebut di atas? (nama dan GenBank Accession)
 - b. Berapa panjang protein dari masing-masing gen? dan apa fungsi masing-masing?
4. Untuk membandingkan kedua urutan gen tersebut, lakukanlah analisis pairwise sequence alignment!.

- Buka website EBI (<http://www.ebi.ac.uk>) kemudian pilih services → DNA&RNA → see all tools
 - Lakukan analisis dengan menggunakan program EMBOSS Needle → nucleotide alignment
5. Tentukan nilai *gaps*, *identity* dan *similarity* dari *analisis pairwise alignment* tersebut!
 6. Kembali ke halaman see all tools, kemudian lakukan analisis dengan menggunakan program EMBOSS Water → nucleotide alignment
 - Tentukan nilai *gaps*, *identity* dan *similarity* dari *analisis pairwise alignment* tersebut!
 7. Apa pengertian dari nilai-nilai yang disebutkan pada no. 5 dan 6?
 - a. *Gaps*
 - b. *Identity*
 - c. *Similarity*
 8. Pada hasil pairwise alignment, apa arti simbol :
 - a. garis vertikal (|)
 - b. titik (.)
 - c. garis horizontal (-)
 9. Buatlah pembahasan mengenai kedua analisis pairwise alignment yang Anda lakukan! (poin: **perbedaan kedua program alignment, kegunaan, cara kerja, interpretasi hasil**)

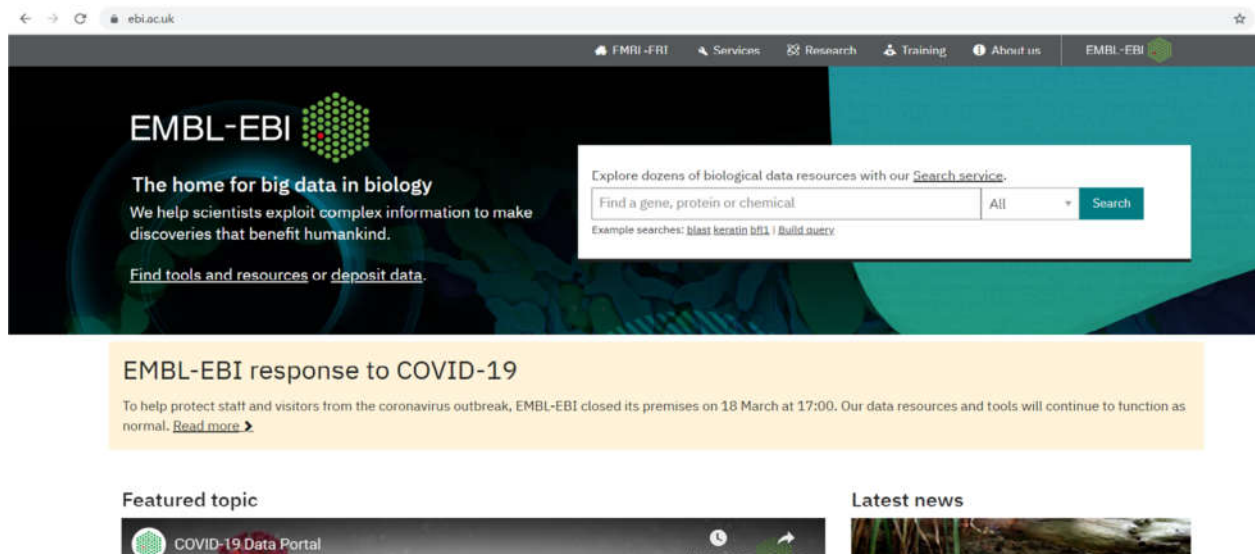
MODUL 4

ANALISIS FILOGENETIK I – SEQUENCE ALIGNMENT

Filogenetik banyak diterjemahkan sebagai kajian mengenai hubungan di antara kelompok-kelompok organisme yang dikaitkan dengan proses evolusi. Berdasarkan teori evolusi, tiap makhluk hidup yang ada memiliki hubungan kekerabatan. Organisme yang kompleks merupakan hasil evolusi dari organisme yang lebih sederhana. Proses perubahan secara bertahap ini terjadi dalam kurun waktu yang sangat lama. Kemampuan organisme untuk bertahan hidup dalam kondisi lingkungan yang ada saat ini, pada tempat dan lokasi tertentu yang menjadi penentu keberhasilan organisme tersebut untuk bertahan dan melanjutkan kelangsungan eksistensi kelompoknya dengan terus berkembang biak dan beradaptasi. Sedangkan organisme yang tidak mampu bertahan dengan perubahan kondisi lingkungan, akan terseleksi dan pada akhirnya akan menuju kepunahan. Untuk bertahan dan beradaptasi, ada penyesuaian-penyesuaian yang dilakukan oleh organisme dalam hal sifat morfologi, fisiologi maupun karakter lainnya yang harus terpenuhi. Karakter-karakter fenotip tersebut, pada dasarnya dikendalikan oleh karakter genetic yang dikode oleh gen-gen yang terletak di dalam pita DNA.

Karakter genetic tersebut, umumnya bersifat cukup stabil, tidak mudah mengalami perubahan, karena karakter genetic yang merupakan material genetic tersebut adalah blueprint dari tiap organisme yang harus diturunkan kepada keturunannya agar tetap eksis. Namun material genetic tersebut masih mampu mengalami perubahan yang bersifat diturunkan juga, melalui mekanisme mutasi. Mutasi-mutasi yang bermakna dan tidak bersifat letal yang nantinya akan menjadi faktor penentu keberlangsungan hidup dari suatu kelompok organisme tertentu. Analisis filogenetik menjadi salah satu alat yang dapat digunakan untuk mengetahui dan merekonstruksi hubungan kekerabatan diantara organisme-organisme berbeda spesies ataupun spesies yang sama, dengan cara membandingkan dan menghitung perubahan yang terjadi pada sekuens-sekuens gen-gen atau fragmen-fragmen DNA tertentu yang bersifat terkonservasi. Umumnya gen 16srRNA digunakan sebagai gen penanda untuk mengetahui hubungan kekerabatan pada kelompok organisme prokariot, sedangkan gen 18srRNA sering digunakan untuk mengetahui kekerabatan organisme eukariot. Selain itu terdapat gen-gen tertentu yang umumnya digunakan sebagai gen penanda untuk menentukan hubungan filogenetik diantara

organisme, misalnya sekuens Internal Transcribed Spacers (ITS) untuk tanaman, COI untuk hewan, dan lain sebagainya.



Tujuan:

Dalam modul praktikum ini mahasiswa diharapkan mampu melakukan analisis alignment beberapa gen atau sekuens DNA/RNA yang diakses secara mandiri dari database (NCBI) dengan menggunakan beberapa software multiple sequence alignment (MSA), diantaranya ClustalX, T-Coffe, dan MUSCLE, dan menginterpretasikan hasil tersebut

Prosedur:

1. Akses database NCBI (www.ncbi.nlm.nih.gov)
2. Temukan sekuens gen 16srRNA (usahakan dapat full sequence) dari 10 bakteri/mikroba.
3. Salin dan simpan sekuens yang sudah diperoleh dalam format FASTA. Pastikan masing-masing sekuens sudah diberi indentifikasi (ID gene cantumkan juga) dari masing-masing spesies.
4. Akses laman ebi.ac.uk/tools/msa
5. Lakukan analisis multiple sequence alignment dari file fasta yang sudah disimpan untuk software: CLUSTAL omega, MUSCLE, T-COFFE, KALIGN.
6. Bandingkan hasil MSA tersebut dan intepretasikan
7. Buat laporan praktikum ini

MODUL 5

ANALISIS FILOGENETIK II – POHON FILOGENETIK

Analisis Multiple sequence alignment yang ditujukan untuk membandingkan dan melihat kesamaan atau ketidaksamaan dari sekuens yang dibandingkan, yang umumnya berasal dari organisme yang berbeda baik spesies maupun strain, perlu dideskripsikan dalam bentuk tampilan yang menggambarkan hubungan kekerabatan diantara sekuens tersebut. Pohon kekerabatan ini lazim disebut dengan dendogram. Melalui dendogram kita akan lebih mudah mengetahui hubungan kekerabatan diantara spesies/organisme-organisme yang dibandingkan. Model dendrogram yang dihasilkan dari suatu pohon filogenetik, akan sangat ditentukan selain oleh karakter dari sekuens yang dibandingkan, juga dipengaruhi oleh program yang digunakan untuk melakukan MSA.

Terdapat tiga jenis kelompok filogenetik yang dikenal yakni:

1. Kelompok monofiletik yaitu mengandung leluhur dan semua keturunannya
2. Kelompok parafiletik yaitu berisi leluhur tetapi hanya beberapa keturunannya
3. Kelompok polifiletik yaitu berisi segala macam organisme tanpa nenek moyang yang sama baru-baru ini.

Tujuan:

Dalam modul ini, mahasiswa diharapkan mampu men'generate' pohon filogenetik yang dibangun menggunakan beberapa model MSA dan mengintepretasikannya.

Prosedur:

Terdapat 3 proyek yang harus diselesaikan oleh mahasiswa dalam modul ini. Lakukan analisis seperti yang sudah Anda lakukan pada modul Analisis filogenetik I. Lanjutkan analisis dengan membuat pohon filogenetiknya dan analisis hasilnya.

Project 1

1. Find protein sequence of Homoserine O-succinyltransferase from organism:
 - *Escherichia coli*

- *Pseudomonas aeruginosa*
 - *Xanthomonas campestris* pv. *campestris*
 - *Stenotrophomonas maltophilia*
 - *Burkholderia thailandensis*
 - *Salmonella typhimurium*
 - *Vibrio cholerae* serotype
 - *Bacillus cereus*
 - *Bradyrhizobium japonicum*
 - *Nitrosomonas europaea*
 - *Ralstonia solanacearum*
 - *Thiobacillus denitrificans*
 - *Acinetobacter baumannii*
 - *Ralstonia solanacearum*
 - *Marinobacter lipolyticus*
2. Keep the protein sequences in fasta format and save them in one file (.txt/.doc)
 3. Do multiple alignment using ClustalW by accessing one of these website :

EBI web server

Swiss Institute of Bioinformatics

4. Generate dendogramme

Project 2

Do the same work with otospiralin protein from organism below:

- *Homo sapiens*
- *Macaca mulatta*
- *Pan troglodytes*
- *Bos Taurus*

- *Canis familiaris*
- *Rattus norvegicus*
- *Cavia porcellus*
- *Mus musculus*
- *Monodelphis domestica*
- *Ornithorhynchus anatinus*
- *Xenopus laevis*
- *Xenopus tropicalis*

Project 3

Do the same work by using nucleotide sequences of otospiralin gene from organism:

- *Homo sapiens*
- *Macaca mulatta*
- *Pan troglodytes*
- *Bos Taurus*
- *Canis familiaris*
- *Rattus norvegicus*
- *Cavia porcellus*
- *Mus musculus*
- *Monodelphis domestica*
- *Ornithorhynchus anatinus*
- *Xenopus laevis*
- *Xenopus tropicalis*

MODUL 6

ANALISIS FILOGENETIK III - DNA BARCODING MENGGUNAKAN GEN 16srRNA PADA BAKTERI

Berdasarkan data statistik yang dihimpun WHO tahun 2011, penyakit yang disebabkan infeksi bakteri masih berada dalam peringkat lima besar penyebab kematian di dunia. Salah satu penyebab hal tersebut terjadi adalah penggunaan metode diagnostik infeksi bakteri yang masih berdasarkan *culture-based*, pengamatan fenotipik dan analisis biokimia. Metode tersebut membutuhkan waktu yang lama untuk mengetahui jenis bakteri apa yang menginfeksi pasien serta hasil yang ditunjukkan kurang akurat. Kemudian dikembangkan metode diagnostik molekuler yang cepat dan akurat untuk mengidentifikasi infeksi bakteri pada manusia menggunakan gen 16S rRNA. Gen 16S rRNA merupakan penanda molekuler untuk mengidentifikasi spesies bakteri yang telah memenuhi sifat sebagai penanda molekuler. Sifat-sifat yang harus dipenuhi oleh suatu gen apabila akan digunakan sebagai penanda molekuler antara lain : (i). memiliki sekuens yang dapat ditemukan pada hampir semua bakteri, (ii). fungsi dari gen penanda tersebut tidak mengalami perubahan sepanjang waktu, dan (iii). memiliki ukuran yang besar yang menyimpan informasi molekuler.

A. Database RDP dan NCBI

Anda adalah seorang mahasiswa Bioteknologi yang sedang mengerjakan proyek akhir salah satu matakuliah yaitu identifikasi bakteri dari pasien sebuah rumah sakit. Metode identifikasi yang dilakukan adalah menggunakan sekuens 16S rRNA. Dibawah ini dilampirkan sebagian sekuens *full-length* atau parsial 16S rRNA tiap isolat yang diperoleh untuk menyelesaikan tugas tersebut.

Prosedur:

Bukalah file isolat.txt – (Sequence Isolat). Mahasiswa dengan NIM ganjil mengcopy isolate nomer 1 -17. Mahasiswa dengan NIM genap mengcopy isolate nomer 18 -34

A.1. Lakukan analisis sekuens tersebut dengan menggunakan database RDP. Langkahl angka yang harus dilakukan sebagai berikut :

- Buka situs database RDP (<http://rdp.cme.msu.edu/>) kemudian pilih *sequencematch*

3. Jelaskan secara singkat berdasarkan hasil dari database RDP mengenai kekerabatan antarisolat!

Isolat	Filum	Kelas	Subkelas	Ordo	Subordo	Family	Genus	Prediksi spesies	
1									
2									
3									
4									
5									
...									
17									

A.2. Selanjutnya, lakukanlah analisis sekuens diatas dengan menggunakan BLAST-n NCBI. Analisis BLAST-n menggunakan database *16S ribosomal RNA sequence (bacteria and Archaea)*.

- Buka situs NCBI pada <http://blast.ncbi.nlm.nih.gov/> kemudian masukkan masing-masing sekuens bakteri pada kotak *query* dan pilih *BLAST*.
- Pilih database *16S ribosomal RNA sequence (bacteria and Archaea)*.
- Pilih organisme pertama yang muncul pada hasil *blast*.
- Perhatikan detail informasi “Organism” atau pilih “Taxonomy” pada *sidebar* “Related Information”.
- Jawablah pertanyaan-pertanyaan dibawah ini

Pertanyaan :

4. Jelaskan secara singkat berdasarkan hasil **Blast-n NCBI** mengenai informasi klasifikasi dari masing-masing isolat! (**gunakan tabel agar lebih rapi, mencakup *Class, Ordo, Family, Genus dan Spesies***).

Isolat	Filum	Kelas	Ordo	Family	Genus	Prediksi spesies
1						
2						
3						
4						
5						
...						
17						

5. Bandingkan hasil antara BLAST-n NCBI dengan hasil RDP (**sertakan kandidat terbaik dari hasil BLAST (nilai kesamaan tertinggi dari hasil BLAST) yang teridentifikasi**

untuk masing-masing isolat). Database manakah yang dapat mengidentifikasi bakteri secara lebih detail? Mengapa? (gunakan tabel untuk membandingkan agar lebih rapi)

Isolat	Spesies teridentifikasi	
	RDP	BLAST N
1		
2		
3		
.		
.		
17		

- **Buatlah file FASTA** dengan mengganti nama isolat *unknown* dengan nama bakteri sesuai yang Anda dapatkan. File ini akan Anda gunakan untuk analisis filogenetik.

6. Akses website www.ebi.ac.uk. Pilih service dan klik DNA & RNA (pada bagian Browse by Type). Pada lama yang sama lakukan hal sebagai berikut:

- analisis multiple alignment menggunakan MUSCLE, dan buatlah pohon filogenetiknya (screen shoot dan copy paste)
- analisis multiple alignment menggunakan Clustal W dan buatlah pohon filogenetiknya (screen shoot dan copy paste)
- analisis multiple alignment menggunakan T-Coffee, dan buatlah pohon filogenetiknya (screen shoot dan copy paste)

Bandingkan ketiga hasil alignment software di atas, berikan komentar Anda tentang ketiga software tersebut dan hasil pohon filogenetiknya