



**MODUL MATA KULIAH  
PENGANTAR BIOINFORMATIKA  
(IBT 431)**

**Disusun Oleh  
Seprianto, S.Pi, M.Si**

**PROGRAM STUDI BIOTEKNOLOGI  
FAKULTAS ILMU - ILMU KESEHATAN  
UNIVERSITAS ESA UNGGUL**

2017



gggul



Universitas  
**Esa Unggul**



Universitas  
**Esa Un**



gggul



Universitas  
**Esa Unggul**



Universitas  
**Esa Un**



aaqul



Universitas  
**Esa Unaaqul**



Universitas  
**Esa Un**

## KATA PENGANTAR

Dengan mengucapkan puji syukur kehadiran Allah SWT yang telah memberikan rahmatNya sehingga penyusunan Modul Matakuliah Pengantar Bioinformatika ini dapat terselesaikan dengan baik. Modul matakuliah ini disusun bagi mahasiswa program studi Bioteknologi, Fakultas Ilmu-ilmu Kesehatan, Universitas Esa Unggul yang mengikuti RPS Pengantar Bioinformatika agar dapat melaksanakan kegiatan perkuliahan dengan sebaik-baiknya.

Modul mata kuliah ini dapat disusun dengan bantuan dari berbagai pihak. Ucapan terima kasih kami sampaikan ke berbagai pihak yang telah memberikan kontribusi, baik secara langsung maupun tidak langsung dalam penyusunan Modul Mata Kuliah ini

Penulis berharap semoga Modul Mata Kuliah ini dapat bermanfaat bagi para pembaca dan dapat membantu khususnya bagi para mahasiswa yang menempuh matakuliah Bioinformatika ini. Penulis menyadari bahwa Modul Mata Kuliah ini masih jauh dari sempurna sehingga penulis sangat mengharapkan kritik dan saran dari pembaca yang sifatnya membangun demi terus meningkatkan kualitas dan kesempurnaan Modul ini.

## DAFTAR ISI

Kata Pengantar.....	2
Daftar Isi.....	3
Bab 1. Bioinformatika dan Perkembangan dalam Dunia Sains.....	4
Bab II. Ruang Lingkup Bioinformatika.....	14
Bab III. Bioinformatika DNA .....	24
Bab IV. Pengenalan Online Data Base.....	34
Bab V. Analisis BLAST.....	45
Bab VI. Desain Primer DNA.....	58
Bab VII. Analisis Pohon Filogenetik.....	72

## **BAB 1. BIOINFORMATIKA DAN PERKEMBANGAN DALAM DUNIA SAINS**

### **I. PENDAHULUAN**

#### **A. Pengantar**

Mata kuliah ini membahas pengantar bioinformatika yang meliputi peranan informasi sekuens DNA dan protein dalam memahami proses biologi, sumber daya (basis data) dan aplikasi-aplikasi yang digunakan secara luas di bidang bioinformatika, algoritme-algoritme yang digunakan untuk memecahkan permasalahan di bidang bioinformatika, khususnya yang terkait dengan sekuens DNA dan protein, seperti persoalan sequence alignment beserta struktur datanya, algoritme untuk *phylogenetic tree*, dan pengenalan penerapan *machine learning* pada bioinformatika. Setelah menyelesaikan mata kuliah ini, mahasiswa diharapkan dapat memahami dan mampu menerapkan algoritme bioinformatika serta membuat aplikasinya untuk memecahkan permasalahan dalam bidang bioinformatika, khususnya yang terkait dengan analisis sekuen DNA dan protein.

#### **B. Kompetensi Dasar**

Mahasiswa dapat mengetahui istilah bioinformatika dan manfaat bioinformatika dalam sains

#### **C. Kemampuan Akhir yang Diharapkan**

Mahasiswa diharapkan mampu :

- Dapat menguraikan pengertian bioinformatika
- Mengetahui perkembangan bioinformatika dalam dunia sains.

#### **D. Kegiatan Pembelajaran**

- Pembelajaran dilakukan dengan metoda presentasi dosen, diskusi dan presentasi kelompok
- Mahasiswa memahami penjelasan dosen selama 30 menit dan selanjutnya diajukan masalah ke setiap kelompok untuk didiskusikan dan setiap kelompok presentasi di depan kelas

## II. MATERI

### 1.1. Pengertian Bioinformatika

Bioinformatika didefinisikan sebagai cabang komputasi dari biologi molekuler yang merupakan teknologi pengumpulan, penyimpanan, analisa, interpretasi, penyebaran, dan aplikasi dari informasi biologi. Bioinformatika menggunakan program komputer maupun website untuk analisa data biologi dan penyimpanan sejumlah data biologiyang dihasilkan oleh proyek genom. Bioinformatika banyak berhubungan dengan sekuen nukleotida termasuk desain primer, struktur, fungsi, perbandingan seluruh genom dan gen, struktur tiga dimensi protein, dan manajemen data Melalui bioinformatik kita juga dapat melakukan berbagai desain eksperimen untuk mengetahui penyakit manusia dan pembuatan peta genom.

#### **Bioinformatika "klasik"**

Sebagian besar ahli Biologi mengistilahkan 'mereka sedang melakukan Bioinformatika' ketika mereka sedang menggunakan komputer untuk menyimpan, melihat atau mengambil data, menganalisa atau memprediksi komposisi atau struktur dari biomolekul. Ketika kemampuan komputer menjadi semakin tinggi maka proses yang dilakukan dalam bioinformatika dapat ditambah dengan melakukan simulasi. Bagian yang termasuk dalam biomolekul diantaranya adalah materi genetik dari manusia --asam nukleat-- dan produk dari gen manusia, yaitu protein. Hal-hal diataslah yang merupakan bahasan utama dari Bioinformatika "klasik", terutama berurusan dengan analisis sekuen (sequence analysis).

Definisi Bioinformatika menurut Fredj Tekaia dari Institut Pasteur adalah: "metode matematika, statistik dan komputasi yang bertujuan untuk menyelesaikan masalah-masalah biologi dengan menggunakan sekuen DNA dan asam amino dan informasi-informasi yang terkait dengannya." Dari sudut pandang Matematika, sebagian besar molekul biologi mempunyai sifat yang menarik, yaitu molekul-molekul tersebut adalah polymer; rantai-rantai yang tersusun rapi dari modul-modul molekul yang lebih sederhana, yang disebut monomer. Monomer dapat dianalogikan sebagai bagian dari bangunan, dimana meskipun bagian bagian tersebut berbeda warna dan bentuk, namun semua memiliki ketebalan yang sama dan cara yang sama untuk dihubungkan antara yang satu dengan yang lain.

Monomer yang dapat dikombinasi dalam satu rantai ada dalam satu kelas umum yang sama, namun tiap jenis monomer dalam kelas tersebut mempunyai karakteristik masing-masing yang terdefinisi dengan baik. Beberapa molekul-molekul monomer dapat digabungkan bersama membentuk sebuah entitas yang berukuran lebih besar, yang disebut macromolecule. Macromolecule dapat mempunyai informasi isi tertentu yang menarik dan sifat-sifat kimia tertentu.

Berdasarkan skema di atas, monomer-monomer tertentu dalam macromolecule dari DNA dapat diperlakukan secara komputasi sebagai huruf-huruf dari alfabet, yang diletakkan dalam sebuah aturan yang telah diprogram sebelumnya untuk membawa pesan atau melakukan kerja di dalam sel. Proses yang diterangkan di atas terjadi pada tingkat molekul di dalam sel. Salah satu cara untuk mempelajari proses tersebut selain dengan mengamati dalam laboratorium biologi yang sangat khusus adalah dengan menggunakan Bioinformatika sesuai dengan definisi "klasik" yang telah disebutkan di atas.

### **Bioinformatika "baru"**

Salah satu pencapaian besar dalam metode Bioinformatika adalah selesainya proyek pemetaan genom manusia (Human Genome Project). Selesainya proyek raksasa tersebut menyebabkan bentuk dan prioritas dari riset dan penerapan Bioinformatika berubah. Secara umum dapat dikatakan bahwa proyek tersebut membawa perubahan besar pada sistem hidup kita, sehingga sering disebutkan --terutama oleh ahli biologi bahwa kita saat ini berada di masa pascagenom. Selesainya proyek pemetaan genom manusia ini membawa beberapa perubahan bagi bioinformatika diantaranya: Setelah memiliki beberapa genom yang utuh maka kita dapat mencari perbedaan dan persamaan di antara gen-gen dari spesies yang berbeda. Dari studi perbandingan antara gen-gen tersebut dapat ditarik kesimpulan tertentu mengenai spesies-spesies dan secara umum mengenai evolusi. Jenis cabang ilmu ini sering disebut sebagai perbandingan genom (comparative genomics).

Sekarang ada teknologi yang didisain untuk mengukur jumlah relatif dari kopi/cetakan sebuah pesan genetik (level dari ekspresi genetik) pada beberapa tingkatan yang berbeda pada perkembangan atau penyakit atau pada jaringan yang berbeda. Teknologi tersebut, contohnya seperti DNA

microarrays akan semakin penting. Akibat yang lain, secara langsung, adalah cara dalam skala besar untuk mengidentifikasi fungsi-fungsi dan keterkaitan dari gen (contohnya metode yeast twohybrid) akan semakin tumbuh secara signifikan dan bersamanya akan mengikuti

Bioinformatika yang berkaitan langsung dengan kerja fungsi genom (functional genomics). Akan ada perubahan besar dalam penekanan dari gen itu sendiri ke hasil-hasil dari gen yang pada akhirnya akan menuntun ke usaha untuk mengkatalogkan semua aktivitas dan karakteristik interaksi antara semua hasil-hasil dari gen (pada manusia) yang disebut proteomics; usaha untuk mengkristalisasi dan memprediksikan struktur-struktur dari semua protein (pada manusia) yang disebut structural genomics. Apa yang disebut orang sebagai research informatics atau medical informatics, manajemen dari semua data eksperimen biomedik yang berkaitan dengan molekul atau pasien tertentu --mulai dari spektroskop massal, hingga ke efek samping klinis akan berubah dari semula hanya merupakan kepentingan bagi mereka yang bekerja di perusahaan obat-obatan dan bagian TI Rumah Sakit akan menjadi jalur utama dari biologi molekuler dan biologi sel, dan berubah jalur dari komersial dan klinikal ke arah akademis. Dari uraian di atas terlihat bahwa Bioinformatika sangat mempengaruhi kehidupan manusia, terutama untuk mencapai kehidupan yang lebih baik. Penggunaan komputer yang notabene merupakan salah satu keahlian utama dari orang yang bergerak dalam TI merupakan salah satu unsur utama dalam Bioinformatika, baik dalam Bioinformatika "klasik" maupun Bioinformatika "baru.

Sebelum era bioinformatika, terdapat hanya dua cara untuk melakukan percobaan biologi yaitu percobaan dalam organisme hidup (in vivo) atau pada lingkungan buatan (in vitro). Melalui bidang ilmu bioinformatika, kita dapat melakukan percobaan biologi secara in silico. Kata silico berasal dari kata lempengan (chip) silikon yang membentuk mikroprosesor komputer. Bioinformatika menjadi salah satu pilihan utama di masa mendatang sebab merupakan titik penting yang paling berkembang sekarang ini di bidang biologi seperti menguraikan genom manusia, teknik pengesahan di bidang biologi dan forensik berdasarkan informasi DNA, serta pengobatan

Saat ini bioinformatika sangat mudah dilakukan karena sudah bisa diakses melalui system jaringan World Wide Web secara gratis di internet.



Beberapa website yang penting antara lain adalah <http://www.ebi.ac.uk> dari European Bioinformatics Institute (EBI), <http://www.ncbi.nlm.nih.gov> dari National Center for Biotechnology Information (NCBI), serta <http://www.ddbj.nig.ac.jp> dari DNA Data Bank of Japan (DDBJ). Ketiga website tersebut saling berintegrasi dalam menggabungkan, memaparkan, maupun memperbaiki informasi tentang sekuen DNA atau protein dari suatu organism. Secara garis besar, orang menggunakan bioinformatika untuk menganalisa sekuen DNA melalui

Beberapa tahapan, antara lain yaitu:

- 1) Mencari sekuen DNA yang diinginkan secara tepat.
- 2) Membandingkan sekuen yang di dapat dengan yang tersedia menggunakan BLAST
- 3) Melakukan analisis sekuen DNA dengan ClustalW.
- 4) Membangun pohon filogenetik.

Pada website NCBI dapat di akses program BLAST (Basic Local Alignment Search Tool) yang merupakan program untuk menganalisa kesamaan yang didisain dalam mengeksplorasi semua database sekuen yang diminta, baik berupa DNA maupun protein. Program ini juga dapat digunakan untuk mendeteksi hubungan antara sekuen yang hanya berbagi daerah tertentu yang memiliki kesamaan/daerah conserve dengan ClustalW. Selain website NCBI, software BIOEDIT juga bisa digunakan untuk melihat daerah conserve dengan ClustalW (<http://www.mbio.ncsu.edu/bioedit/page2.html>).

## **1.2. Perkembangan Bioinformatika Dalam Dunia Sains**

Perkembangan Penetrasi Teknologi Informasi (TI) dalam berbagai disiplin ilmu telah melipatgandakan perkembangan ilmu bersangkutan. Berbagai kajian baru bermunculan, sejalan dengan perkembangan TI itu sendiri dan disiplin ilmu yang didukungnya. Aplikasi TI dalam bidang biologi molekul telah melahirkan bidang Bioinformatika. Kajian ini semakin penting, sebab perkembangannya telah mendorong kemajuan bioteknologi di satu sisi, dan pada sisi lain memberi efek domino pada bidang kedokteran, farmasi, lingkungan dan lainnya.

Kajian baru Bioinformatika ini tak lepas dari perkembangan biologi molekul modern yang ditandai dengan kemampuan manusia untuk

memahami genom, yaitu cetak biru informasi genetik yang menentukan sifat setiap makhluk hidup yang disandi dalam bentuk pita molekul DNA (asam deoksiribonukleat). Kemampuan untuk memahami dan memanipulasi kode genetik DNA ini sangat didukung oleh TI melalui perangkat perangkat keras maupun lunak. Hal ini bisa dilihat pada upaya Celera Genomics, perusahaan bioteknologi Amerika Serikat yang melakukan pembacaan sekuen genom manusia yang secara maksimal memanfaatkan TI sehingga bisa melakukan pekerjaannya dalam waktu yang singkat (hanya beberapa tahun), dibanding usaha konsorsium lembaga riset publik AS, Eropa, dan lain-lain, yang memakan waktu lebih dari 10 tahun.

Kelahiran Bioinformatika modern tak lepas dari perkembangan bioteknologi di era tahun 70-an, dimana seorang ilmuwan AS melakukan inovasi dalam mengembangkan teknologi DNA rekombinan. Berkat penemuan ini lahirlah perusahaan bioteknologi pertama di dunia, yaitu Genentech di AS, yang kemudian memproduksi protein hormon insulin dalam bakteri, yang dibutuhkan penderita diabetes. Selama ini insulin hanya bisa didapatkan dalam jumlah sangat terbatas dari organ pankreas sapi. Bioteknologi modern ditandai dengan kemampuan pada manipulasi DNA. Rantai/sekuen DNA yang mengkode protein disebut gen. Gen ditranskripsikan menjadi mRNA, kemudian mRNA ditranslasikan menjadi protein. Protein sebagai produk akhir bertugas menunjang seluruh proses kehidupan, antara lain sebagai katalis reaksi biokimia dalam tubuh (disebut enzim), berperan serta dalam sistem pertahanan tubuh melawan virus, parasit dan lain-lain (disebut antibodi), menyusun struktur tubuh dari ujung kaki (otot terbentuk dari protein actin, myosin, dan sebagainya) sampai ujung rambut (rambut tersusun dari protein keratin), dan lain-lain. Arus informasi, DNA -> RNA -> Protein, inilah yang disebut sentral dogma dalam biologi molekuler.

Sekuen DNA satu organisme, yaitu pada sejenis virus yang memiliki kurang lebih 5.000 nukleotida/molekul DNA atau sekitar 11 gen, berhasil dibaca secara menyeluruh pada tahun 1977. Sekuen seluruh DNA manusia terdiri dari 3 milyar nukleotida yang menyusun 100.000 gen dapat dipetakan dalam waktu 3 tahun. Saat ini terdapat milyaran data nukleotida yang tersimpan dalam database DNA, GenBank di AS yang didirikan tahun 1982. Di Indonesia, ada Lembaga Biologi Molekul Eijkman yang terletak di Jakarta.

Di sini kita bisa membaca sekuen sekitar 500 nukleotida hanya dengan membayar \$15. Trend yang sama juga nampak pada database lain seperti database sekuen asam amino penyusun protein, database struktur 3D protein, dan sebagainya. Inovasi teknologi DNA chip yang dipelopori oleh perusahaan bioteknologi AS, Affymetrix di Silicon Valley telah mendorong munculnya database baru mengenai RNA. Desakan kebutuhan untuk mengumpulkan, menyimpan dan menganalisa data-data biologis dari database DNA, RNA maupun protein inilah yang semakin memacu perkembangan kajian bioinformatika

Berikut ini adalah beberapa perkembangan bioinformatika dalam dunia sains meliputi:

#### **A. Bioinformatika dalam Bidang Klinis**

Bioinformatika dalam bidang klinis sering disebut sebagai informatika klinis (clinical informatics). Aplikasi dari informatika klinis ini berbentuk manajemen data-data klinis dari pasien melalui Electrical Medical Record (EMR) yang dikembangkan oleh Clement J. McDonald dari Indiana University School of Medicine pada tahun 1972. McDonald pertama kali mengaplikasikan EMR pada 33 orang pasien penyakit gula (diabetes). Sekarang EMR ini telah diaplikasikan pada berbagai penyakit. Data yang disimpan meliputi data analisa diagnosa laboratorium, hasil konsultasi dan saran, foto rontgen, ukuran detak jantung, dan lain lain. Dengan data ini dokter akan bisa menentukan obat yang sesuai dengan kondisi pasien tertentu dan lebih jauh lagi, dengan dibacanya genom manusia, akan memungkinkan untuk mengetahui penyakit genetik seseorang, sehingga penanganan terhadap pasien menjadi lebih akurat.

#### **B. Bioinformatika untuk Identifikasi Agent Penyakit Baru**

Bioinformatika juga menyediakan tool yang sangat penting untuk identifikasi agent penyakit yang belum dikenal penyebabnya. Banyak sekali penyakit baru yang muncul dalam dekade ini, dan diantaranya yang masih hangat adalah SARS (Severe Acute Respiratory Syndrome). Pada awalnya, penyakit ini diperkirakan disebabkan oleh virus influenza karena gejalanya mirip dengan gejala pengidap influenza. Akan tetapi ternyata dugaan ini salah karena virus influenza tidak terisolasi dari pasien. Perkiraan lain penyakit ini

disebabkan oleh bakteri Candida karena bakteri ini terisolasi dari beberapa pasien. Tapi perkiraan ini juga salah. Akhirnya ditemukan bahwa dari sebagian besar pasien SARS terisolasi virus Corona jika dilihat dari morfologinya. Sekuen genom virus ini kemudian dibaca dan dari hasil analisa dikonfirmasi bahwa penyebab SARS adalah virus Corona yang telah berubah (mutasi) dari virus Corona yang ada selama ini.

Kedua pada proses mencari kemiripan sekuen (homology alignment) virus yang didapatkan dengan virus lainnya. Dari hasil analisa virus SARS diketahui bahwa genom virus Corona penyebab SARS berbeda dengan virus Corona lainnya. Perbedaan ini diketahui dengan menggunakan homology alignment dari sekuen virus SARS. Selanjutnya, Bioinformatika juga berfungsi untuk analisa posisi sejauh mana suatu virus berbeda dengan virus lainnya

### **C. Bioinformatika untuk Diagnosa Penyakit Baru**

Untuk menangani penyakit baru diperlukan diagnosa yang akurat sehingga dapat dibedakan dengan penyakit lain. Diagnosa yang akurat ini sangat diperlukan untuk pemberian obat dan perawatan yang tepat bagi pasien. Ada beberapa cara untuk mendiagnosa suatu penyakit, antara lain: isolasi agent penyebab penyakit tersebut dan analisa morfologinya, deteksi antibodi yang dihasilkan dari infeksi dengan teknik enzyme-linked immunosorbent assay (ELISA), dan deteksi gen dari agent pembawa penyakit tersebut dengan Polymerase Chain Reaction (PCR).

Teknik yang banyak dan lazim dipakai saat ini adalah teknik PCR. Teknik ini sederhana, praktis dan cepat. Yang penting dalam teknik PCR adalah disain primer untuk amplifikasi DNA, yang memerlukan data sekuen dari genom agent yang bersangkutan dan software seperti yang telah diuraikan di atas. Disinilah Bioinformatika memainkan peranannya. Untuk agent yang mempunyai genom RNA, harus dilakukan reverse transcription (proses sintesa DNA dari RNA) terlebih dahulu dengan menggunakan enzim reverse transcriptase. Setelah DNA diperoleh baru dilakukan PCR. Reverse transcription dan PCR ini bisa dilakukan sekaligus dan biasanya dinamakan RT-PCR.

Teknik PCR ini bersifat kualitatif, oleh sebab itu sejak beberapa tahun yang lalu dikembangkan teknik lain, yaitu Real Time PCR yang bersifat kuantitatif. Dari hasil Real Time PCR ini bisa ditentukan kuantitas suatu agent

di dalam tubuh seseorang, sehingga bisa dievaluasi tingkat emergensinya. Pada Real Time PCR ini selain primer diperlukan probe yang harus didisain sesuai dengan sekuen agent yang bersangkutan. Di sini juga diperlukan software atau program Bioinformatika.

#### **D. Bioinformatika untuk Penemuan Obat**

Cara untuk menemukan obat biasanya dilakukan dengan menemukan zat/senyawa yang dapat menekan perkembangbiakan suatu agent penyebab penyakit. Karena perkembangbiakan agent tersebut dipengaruhi oleh banyak faktor, maka faktor-faktor inilah yang dijadikan target. Diantaranya adalah enzim-enzim yang diperlukan untuk perkembangbiakan suatu agent. Mula-mula yang harus dilakukan adalah analisa struktur dan fungsi enzim-enzim tersebut. Kemudian mencari atau mensintesa zat/senyawa yang dapat menekan fungsi dari enzim-enzim tersebut.

Analisa struktur dan fungsi enzim ini dilakukan dengan cara mengganti asam amino tertentu dan menguji efeknya. Analisa penggantian asam amino ini dahulu dilakukan secara random sehingga memerlukan waktu yang lama. Setelah Bioinformatika berkembang, data-data protein yang sudah dianalisa bebas diakses oleh siapapun, baik data sekuen asam amino-nya seperti yang ada di SWISS-PROT (<http://www.ebi.ac.uk/swissprot/>) maupun struktur 3D-nya yang tersedia di Protein Data Bank (PDB) (<http://www.rcsb.org/pdb/>). Dengan database yang tersedia ini, enzim yang baru ditemukan dapat dibandingkan sekuen asam amino-nya, sehingga bisa diperkirakan asam amino yang berperan untuk aktivitas (active site) dan kestabilan enzim tersebut. Setelah asam amino yang berperan sebagai active site dan kestabilan enzim tersebut ditemukan, kemudian dicari atau disintesa senyawa yang dapat berinteraksi dengan asam amino tersebut. Dengan data yang ada di PDB, maka dapat dilihat struktur 3D suatu enzim termasuk active site-nya, sehingga bisa diperkirakan bentuk senyawa yang akan berinteraksi dengan active site tersebut. Dengan demikian, kita cukup mensintesa senyawa yang diperkirakan akan berinteraksi, sehingga obat terhadap suatu penyakit akan jauh lebih cepat ditemukan. Cara ini dinamakan “docking” dan telah banyak digunakan oleh perusahaan farmasi untuk penemuan obat baru.

Meskipun dengan Bioinformatika ini dapat diperkirakan senyawa yang berinteraksi dan menekan fungsi suatu enzim, namun hasilnya harus

dikonfirmasi dahulu melalui eksperimen di laboratorium. Akan tetapi dengan Bioinformatika, semua proses ini bisa dilakukan lebih cepat sehingga lebih efisien baik dari segi waktu maupun finansial.

### III. EVALUASI

#### A. Latihan

- 1) Jelaskan pengertian Bioinformatika ?
- 2) Jelaskan perbedaan antara Bioinformatika klasik dan Bioinformatika baru ?
- 3) Jelaskan Perkembangan Bioinformatika dalam dunia sains?
- 4) Jelaskan kelemahan Bioinformatika dalam masing – masing bidang ?

#### B. Tugas

1. Buatlah makalah tentang perkembangan bioinformatika dalam dunia sains !
2. Presentasikan makalah tersebut oleh masing – masing kelompok !

#### C. Penilaian Tugas

2. Tugas dibuat di blog mahasiswa
3. Blog di link ke web hybrid learning.
4. Blog tersebut harus mencantumkan logo dan nama Universitas Esa Unggul
5. Diselesaikan sebelum batas akhir penyerahan tugas (Tanggal .

### IV. DAFTAR PUSTAKA

Aprijani, DA. Elfaizi, MA. 2004. Bioinformatika : Perkembangan, Disiplin Ilmu dan Penerapannya di Indonesia, <http://www.gnu.org/copyleft/fdl.html>. per 6 Juli 2017

[UTAMA2003] Utama, Andi (2003), *Peranan Bioinformatika dalam Dunia Kedokteran*, <http://ikc.vlsm.org/populer/andi-bioinformatika.php>

[BIOINFORMATICS2004] Bioinformatics.org: The Open-Access Institute, <http://bioinformatics.org> per 20 Juni 2011

## **BAB II. RUANG LINGKUP BIOINFORMATIKA**

### **I. Pendahuluan**

#### **A. Pengantar**

Ruang lingkup Bioinformatika Dari pengertian Bioinformatika baik yang klasik maupun baru, terlihat banyak terdapat cabang-cabang disiplin ilmu yang terkait dengan Bioinformatika –terutama karena Bioinformatika itu sendiri merupakan suatu bidang interdisipliner-- . Hal tersebut menimbulkan banyak pilihan bagi orang yang ingin mendalami Bioinformatika. Sehari-harinya bioinformatika dikerjakan dengan menggunakan program pencari sekuen (sequence search) seperti BLAST, program analisa sekuen (sequence analysis) seperti EMBOSS dan paket Staden, program prediksi struktur seperti THREADER atau PHD atau program imaging/modelling seperti RasMol dan WHATIF.

#### **B. Kompetensi Dasar**

Mahasiswa dapat mamahami dan menjelaskan ruang lingkup bioinformatika serta perkembangannya di Indonesia.

#### **C. Kemampuan Akhir yang Diharapkan**

Mahasiswa diharapkan mampu :

- Memahami dan menjelaskan berbagai cabang ilmu bioinformatika
- Memahami teknologi dan penerapan bioinformatika
- Menjelaskan perkembangan bioinformatika di Indonesia.

#### **D. Kegiatan Pembelajaran**

- Pembelajaran dilakukan dengan metoda contextual learning dan project based learning
- Mahasiswa mencari bahan pustaka, membuat bahan presentasi dan mempresentasikan hasil literasinya

### **II. MATERI**

#### **2.1 Cabang –cabang Ilmu yang Terkait dengan Bioinformatika**

Pengertian Bioinformatika baik yang klasik maupun baru, terlihat banyak terdapat cabang-cabang disiplin ilmu yang terkait dengan

Bioinformatika –terutama karena Bioinformatika itu sendiri merupakan suatu bidang interdisipliner-- . Hal tersebut menimbulkan banyak pilihan bagi orang yang ingin mendalami Bioinformatika. Dibawah ini akan disebutkan beberapa bidang yang terkait dengan Bioinformatika.

### **2.1.1. Biophysics**

Biologi molekul sendiri merupakan pengembangan yang lahir dari biophysics. Biophysics adalah sebuah bidang interdisipliner yang mengaplikasikan teknik-teknik dari ilmu Fisika untuk memahami struktur dan fungsi biologi (British Biophysical Society). Sesuai dengan definisi di atas, bidang ini merupakan suatu bidang yang luas. Namun secara langsung disiplin ilmu ini terkait dengan Bioinformatika karena penggunaan teknik-teknik dari ilmu Fisika untuk memahami struktur membutuhkan penggunaan TI.

### **2.1.2. Computational Biology**

Computational biology merupakan bagian dari Bioinformatika (dalam arti yang paling luas) yang paling dekat dengan bidang Biologi umum klasik. Fokus dari computational biology adalah gerak evolusi, populasi, dan biologi teoritis daripada biomedis dalam molekul dan sel. Tak dapat dielakkan bahwa Biologi Molekul cukup penting dalam computational biology, namun itu bukanlah inti dari disiplin ilmu ini. Pada penerapan computational biology, model-model statistika untuk fenomena biologi lebih disukai dipakai dibandingkan dengan model sebenarnya. Dalam beberapa hal cara tersebut cukup baik mengingat pada kasus tertentu eksperimen langsung pada fenomena biologi cukup sulit. Tidak semua dari computational biology merupakan Bioinformatika, seperti contohnya Model Matematika bukan merupakan Bioinformatika, bahkan meskipun dikaitkan dengan masalah biologi.

### **2.1.3. Medical Informatics**

Menurut Aamir Zakaria Pengertian dari medical informatics adalah "sebuah disiplin ilmu yang baru yang didefinisikan sebagai pembelajaran, penemuan, dan implementasi dari struktur dan algoritma untuk meningkatkan komunikasi, pengertian dan manajemen informasi medis." Medical informatics lebih memperhatikan struktur dan algoritma untuk pengolahan data medis, dibandingkan dengan data itu sendiri. Disiplin ilmu ini, untuk alasan praktis,



kemungkinan besar berkaitan dengan data-data yang didapatkan pada level biologi yang lebih "rumit" --yaitu informasi dari sistem-sistem superselular, tepat pada level populasi—di mana sebagian besar dari Bioinformatika lebih memperhatikan informasi dari sistem dan struktur biomolekul dan selular.

#### **2.1.4. Cheminformatics**

Cheminformatics adalah kombinasi dari sintesis kimia, penyaringan biologis, dan pendekatan data-mining yang digunakan untuk penemuan dan pengembangan obat (Cambridge Healthtech Institute's Sixth Annual Cheminformatics conference). Pengertian disiplin ilmu yang disebutkan di atas lebih merupakan identifikasi dari salah satu aktivitas yang paling populer dibandingkan dengan berbagai bidang studi yang mungkin ada di bawah bidang ini. Salah satu contoh penemuan obat yang paling sukses sepanjang sejarah adalah penisilin, dapat menggambarkan cara untuk menemukan dan mengembangkan obat-obatan hingga sekarang --meskipun terlihat aneh--. Cara untuk menemukan dan mengembangkan obat adalah hasil dari kesempatan, observasi, dan banyak proses kimia yang intensif dan lambat. Sampai beberapa waktu yang lalu, disain obat dianggap harus selalu menggunakan kerja yang intensif, proses uji dan gagal (trial-error process).

Kemungkinan penggunaan TI untuk merencanakan secara cerdas dan dengan mengotomatiskan proses-proses yang terkait dengan sintesis kimiawi dari komponen-komponen pengobatan merupakan suatu prospek yang sangat menarik bagi ahli kimia dan ahli biokimia. Penghargaan untuk menghasilkan obat yang dapat dipasarkan secara lebih cepat sangatlah besar, sehingga target inilah yang merupakan inti dari cheminformatics. Ruang lingkup akademis dari cheminformatics ini sangat luas. Contoh bidang minatnya antara lain: Synthesis Planning, Reaction and Structure Retrieval, 3-D Structure Retrieval, Modelling, Computational Chemistry, Visualisation Tools and Utilities.

#### **2.1.5. Genomics**

Genomics adalah bidang ilmu yang ada sebelum selesainya sekuen genom. Genomics adalah setiap usaha untuk menganalisa atau membandingkan seluruh komplemen genetik dari satu spesies atau lebih.

Secara logis tentu saja mungkin untuk membandingkan genom-genom dengan membandingkan kurang lebih suatu himpunan bagian dari gen di dalam genom yang representatif.

#### **2.1.6. Mathematical Biology**

Mathematical biology lebih mudah dibedakan dengan Bioinformatika daripada computational biology dengan Bioinformatika. Mathematical biology juga menangani masalah-masalah biologi, namun metode yang digunakan untuk menangani masalah tersebut tidak perlu secara numerik dan tidak perlu diimplementasikan dalam software maupun hardware. Bahkan metode yang dipakai tidak perlu "menyelesaikan" masalah apapun; dalam mathematical biology bisa dianggap beralasan untuk mempublikasikan sebuah hasil yang hanya menyatakan bahwa suatu masalah biologi berada pada kelas umum tertentu. Secara umum mathematical biology melingkupi semua ketertarikan teoritis yang tidak perlu merupakan sesuatu yang beralgoritma, dan tidak perlu dalam bentuk molekul, dan tidak perlu berguna dalam menganalisis data yang terkumpul.

#### **2.1.7. Proteomics**

Istilah proteomics pertama kali digunakan untuk menggambarkan himpunan dari protein-protein yang tersusun (encoded) oleh genom. Ilmu yang mempelajari proteome, yang disebut proteomics, pada saat ini tidak hanya memperhatikan semua protein di dalam sel yang diberikan, tetapi juga himpunan dari semua bentuk isoform dan modifikasi dari semua protein, interaksi diantaranya, deskripsi struktural dari protein –protein dan kompleks-kompleks orde tingkat tinggi dari protein, dan mengenai masalah tersebut hampir semua pasca genom. Michael J. Dunn Pemimpin Redaksi dari Proteomics mendefinisikan kata "proteome" sebagai: "The PROTEin complement of the genOME". Dan mendefinisikan proteomics berkaitan dengan: "studi kuantitatif dan kualitatif dari ekspresi gen di level dari protein-protein fungsional itu sendiri". Yaitu: "sebuah antarmuka antara biokimia protein dengan biologi molekul". Mengkarakterisasi sebanyak puluhan ribu protein-protein yang dinyatakan dalam sebuah tipe sel yang diberikan pada waktu tertentu --apakah untuk mengukur berat molekul atau nilai-nilai isoelektrik protein-protein tersebut-- melibatkan tempat penyimpanan dan

perbandingan dari data yang memiliki jumlah yang sangat besar, tak terhindarkan lagi akan memerlukan Bioinformatika.

#### **2.1.8. Pharmacogenomics**

Pharmacogenomics adalah aplikasi dari pendekatan genomik dan teknologi pada identifikasi dari target-target obat. Contohnya meliputi menjangkau semua genom untuk penerima yang potensial dengan menggunakan cara Bioinformatika, atau dengan menyelidiki bentuk pola dari ekspresi gen di dalam baik patogen maupun induk selama terjadinya infeksi, atau maupun dengan memeriksa karakteristik pola-pola ekspresi yang ditemukan dalam tumor atau contoh dari pasien untuk kepentingan diagnosa (kemungkinan untuk mengejar target potensial terapi kanker). Istilah pharmacogenomics digunakan lebih untuk urusan yang lebih "trivial" -- tetapi dapat diargumentasikan lebih berguna-- dari aplikasi pendekatan Bioinformatika pada pengkatalogan dan pemrosesan informasi yang berkaitan dengan ilmu Farmasi dan Genetika, untuk contohnya adalah pengumpulan informasi pasien dalam database.

#### **2.1.9. Pharmacogenetics**

Tiap individu mempunyai respon yang berbeda-beda terhadap berbagai pengaruh obat; sebagian ada yang positif, sebagian ada yang sedikit perubahan yang tampak pada kondisi mereka dan ada juga yang mendapatkan efek samping atau reaksi alergi. Sebagian dari reaksi-reaksi ini diketahui mempunyai dasar genetik. Pharmacogenetics adalah bagian dari pharmacogenomics yang menggunakan metode genomik/Bioinformatika untuk mengidentifikasi hubungan-hubungan genomik, contohnya SNP (Single Nucleotide Polymorphisms), karakteristik dari profil respons pasien tertentu dan menggunakan informasi-informasi tersebut untuk memberitahu administrasi dan pengembangan terapi pengobatan. Secara menakjubkan pendekatan tersebut telah digunakan untuk "menghidupkan kembali" obat-obatan yang sebelumnya dianggap tidak efektif, namun ternyata diketahui manjur pada sekelompok pasien tertentu. Disiplin ilmu ini juga dapat digunakan untuk mengoptimalkan dosis kemoterapi pada pasien-pasien tertentu.

Gambaran dari sebagian bidang-bidang yang terkait dengan Bioinformatika di atas memperlihatkan bahwa Bioinformatika mempunyai

ruang lingkup yang sangat luas dan mempunyai peran yang sangat besar dalam bidangnya. Bahkan pada bidang pelayanan kesehatan Bioinformatika menimbulkan disiplin ilmu baru yang menyebabkan peningkatan pelayanan kesehatan

## **2.2. Teknologi dan Penerapan Bioinformatika**

Teknologi Bioinformatika secara umum pada saat ini banyak pekerjaan Bioinformatika berkaitan dengan teknologi database. Penggunaan database ini meliputi baik tempat penyimpanan database "umum" seperti GenBank atau PDB maupun database "pribadi", seperti yang digunakan oleh grup riset yang terlibat dalam proyek pemetaan gen atau database yang dimiliki oleh perusahaan-perusahaan bioteknologi. Konsumen dari data Bioinformatika menggunakan platform jenis komputer dalam kisaran: mulai dari mesin UNIX yang lebih canggih dan kuat yang dimiliki oleh pengembang dan kolektor hingga ke mesin Mac yang lebih bersahabat yang sering ditemukan menempati laboratorium ahli biologi yang tidak suka komputer.

Database dari sekuen data yang ada dapat digunakan untuk mengidentifikasi homolog pada molekul baru yang telah dikuatkan dan disekuenkan di laboratorium. Dari satu nenek moyang mempunyai sifat-sifat yang sama, atau homology, dapat menjadi indikator yang sangat kuat di dalam Bioinformatika. Setelah informasi dari database diperoleh, langkah berikutnya adalah menganalisa data. Pencarian database umumnya berdasarkan pada hasil alignment / pensejajaran sekuen, baik sekuen DNA maupun protein. Kegunaan dari pencarian ini adalah ketika mendapatkan suatu sekuen DNA/protein yang belum diketahui fungsinya maka dengan membandingkannya dengan yang ada dalam database bisa diperkirakan fungsi daripadanya. Salah satu perangkat lunak pencari database yang paling berhasil dan bisa dikatakan menjadi standar sekarang adalah BLAST (Basic Local Alignment Search Tool) yang merupakan program pencarian kesamaan yang didisain untuk mengeksplorasi semua database sekuen yang diminta, baik itu berupa DNA atau protein. Program BLAST juga dapat digunakan untuk mendeteksi hubungan di antara sekuen yang hanya berbagi daerah tertentu yang memiliki kesamaan.

Di bawah ini diberikan contoh beberapa alamat situs yang berguna untuk bidang biologi molekuler dan genetika

Tabel 1. Beberapa situs yang digunakan dalam bidang bioinformatika

Deskripsi	Alamat
National Center for Biotechnology Information	<a href="http://www.ncbi.nlm.nih.gov/">http://www.ncbi.nlm.nih.gov/</a>
GenBank (NIH Genetic Sequence Database)	<a href="http://www.ncbi.nlm.nih.gov/Web/Genbank/index/html">http://www.ncbi.nlm.nih.gov/Web/Genbank/index/html</a>
European Molecular Biology Laboratory Nucleotide Sequence	<a href="http://www.ebi.ac.uk/ebi_docs/embl_db.html">http://www.ebi.ac.uk/ebi_docs/embl_db.html</a>
Protein Information Resource	<a href="http://www.nbrf.georgetown.edu/pir">http://www.nbrf.georgetown.edu/pir</a>
Protein Data Bank	<a href="http://www.pdb.bnl.gov/">http://www.pdb.bnl.gov/</a>
Restriction Enzyme Database	<a href="http://www.neb.com/rebase/rebase.html">http://www.neb.com/rebase/rebase.html</a>
National Center for Genome Research (NCGR)	<a href="http://www.ncgr.org/gpi/">http://www.ncgr.org/gpi/</a>
GeneMark	<a href="http://www.dixie.biology.gatech.edu/GeneMark/eukhmm.cgi">http://www.dixie.biology.gatech.edu/GeneMark/eukhmm.cgi</a>
Biotechnology Industry Organization (BIO)	<a href="http://www.bio.org">http://www.bio.org</a>

### 2.3. Kondisi dan Penerapan Bioinformatika di Indonesia

Di Indonesia, Bioinformatika masih belum dikenal oleh masyarakat luas. Hal ini dapat dimaklumi karena penggunaan komputer sebagai alat bantu belum merupakan budaya. Bahkan di kalangan peneliti sendiri, barangkali hanya para peneliti biologi molekuler yang sedikit banyak mengikuti perkembangannya karena keharusan menggunakan perangkat-perangkat Bioinformatika untuk analisa data. Sementara di kalangan TI masih kurang mendapat perhatian.

Ketersediaan database dasar (DNA, protein) yang bersifat terbuka/gratis merupakan peluang besar untuk menggali informasi berharga daripadanya. Database genom manusia sudah disepakati akan bersifat terbuka untuk seluruh kalangan, sehingga dapat digali/diketahui kandidat-kandidat gen yang memiliki potensi kedokteran/farmasi. Dari sinilah Indonesia dapat ikut berperan mengembangkan Bioinformatika. Kerjasama antara peneliti bioteknologi yang memahami makna biologis data tersebut dengan praktisi TI seperti programmer, dan sebagainya akan sangat berperan dalam kemajuan Bioinformatika Indonesia nantinya.

Sebagai kajian yang masih baru, Indonesia seharusnya berperan aktif dalam mengembangkan Bioinformatika ini. Paling tidak, sebagai tempat tinggal lebih dari 300 suku bangsa yang berbeda akan menjadi sumber genom, karena besarnya variasi genetiknya. Belum lagi variasi species flora maupun fauna yang berlimpah. Memang ada sejumlah pakar yang telah mengikuti perkembangan Bioinformatika ini, misalnya para peneliti dalam Lembaga Biologi Molekul Eijkman. Mereka cukup berperan aktif dalam memanfaatkan kajian Bioinformatika. Bahkan, lembaga ini telah memberikan beberapa sumbangan cukup berarti, antara lain:

### **2.3.1. Deteksi Kelainan Janin**

Lembaga Biologi Molekul Eijkman bekerja sama dengan Bagian Obstetri dan

Ginekologi Fakultas Kedokteran Universitas Indonesia dan Rumah Sakit Cipto Mangunkusumo sejak November 2001 mengembangkan klinik genetik untuk mendeteksi secara dini sejumlah penyakit genetik yang menimbulkan gangguan pertumbuhan fisik maupun retardasi mental seperti antara lain, talasemia dan sindroma down. Kelainan ini bisa diperiksa sejak janin masih berusia beberapa minggu. Talasemia adalah penyakit keturunan di mana tubuh kekurangan salah satu zat pembentuk hemoglobin (Hb) sehingga mengalami anemia berat dan perlu transfusi darah seumur hidup. Sedangkan sindroma down adalah kelebihan jumlah untaian di kromosom 21 sehingga anak tumbuh dengan retardasi mental, kelainan jantung, pendengaran dan penglihatan buruk, otot lemah serta kecenderungan menderita kanker sel darah putih (leukemia).

Dengan mengetahui sejak dini, pasangan yang hendak menikah, atau pasangan

yang salah satunya membawa kelainan kromosom, atau pasangan yang mempunyai anak yang menderita kelainan kromosom, atau penderita kelainan kromosom yang sedang hamil, atau ibu yang hamil di usia tua bisa memeriksakan diri dan janin untuk

memastikan apakah janin yang dikandung akan menderita kelainan kromosom atau tidak, sehingga mempunyai kesempatan untuk mempertimbangkan apakah kehamilan akan diteruskan atau tidak setelah

mendapat konseling genetik tentang berbagai kemungkinan yang akan terjadi.

Di bidang talasemia, Eijkman telah memiliki katalog 20 mutasi yang mendasari talasemia beta di Indonesia, 10 di antaranya sering terjadi. Lembaga ini juga mempunyai informasi cukup mengenai spektrum mutasi di berbagai suku bangsa yang sangat bervariasi. Talasemia merupakan penyakit genetik terbanyak di dunia termasuk di Indonesia.

### **2.3.2. Pengembangan Vaksin Hepatitis B Rekombinan**

Lembaga Biologi Molekul Eijkman bekerja sama dengan PT Bio Farma (BUMN Departemen Kesehatan yang memproduksi vaksin) sejak tahun 1999 mengembangkan vaksin Hepatitis B rekombinan, yaitu vaksin yang dibuat lewat rekayasa genetika. Selain itu Lembaga Eijkman juga bekerja sama dengan PT Diagnosia Dipobiotek untuk mengembangkan kit diagnostik.

### **2.3.3. Meringankan Kelumpuhan dengan Rekayasa RNA**

Kasus kelumpuhan distrofi (Duchenne Muscular Dystrophy) yang menurun kini dapat dikurangi tingkat keparahannya dengan terapi gen. Kelumpuhan ini akibat ketidaknormalan gen distrofin pada kromosom X sehingga hanya diderita anak laki-laki. Diperkirakan satu dari 3.500 pria di dunia mengalami kelainan ini. Dengan memperbaiki susunan ekson atau bagian penyusun RNA gen tersebut pada hewan percobaan tikus, terbukti mengurangi tingkat kelumpuhan saat pertumbuhannya menjadi dewasa.

Gen distrofin pada kasus kelumpuhan paling sering disebabkan oleh delesi atau hilangnya beberapa ekson pada gen tersebut. Normalnya pada gen atau DNA distrofin terdapat 78 ekson. Diperkirakan 65 persen pasien penderita DMD mengalami delesi dalam jumlah besar dalam gen distrofinnya. Kasus kelumpuhan ini dimulai pada otot prosima seperti pangkal paha dan betis. Dengan bertambahnya usia kelumpuhan akan meluas pada bagian otot lainnya hingga ke leher. Karena itu dalam kasus kelumpuhan yang berlanjut dapat berakibat kematian.

Teknologi rekayasa RNA seperti proses penyambungan (slicing) ekson dalam satu rangkaian terbukti dapat mengoreksi mutasi DMD. Bila bagian ekson yang masih ada disambung atau disusun ulang, terjadi perubahan

asam amino yang membentuk protein. Molekul RNA mampu mengenali molekul RNA lainnya dan melekat dengannya.

### III. EVALUASI BELAJAR

#### a. Latihan

- Jelaskan tentang ruang lingkup Bioinformatika
- Jelaskan Bidang – bidang ilmu yang terkait dengan bidang bioinformatika
- Menurut pengetahuan saudara, bagaimana perkembangan dan penerapan bioinformatika di Indonesia
- Tuliskan minimal 5 situs yang sering digunakan dalam analisis data dalam Bioinformatika

#### b. Tugas

- Buatlah makalah tentang Ruang lingkup bioinformatika serta perkembangan bioinformatika di Indonesia

#### Penilaian Tugas

1. Tugas dibuat di blog mahasiswa
2. Blog di link ke web hybrid learning.
3. Blog tersebut harus mencantumkan logo dan nama Universitas Esa Unggul
4. Diselesaikan sebelum batas akhir penyerahan tugas (Tanggal ...)

### IV. DAFTAR PUSTAKA

Aprijani, DA. Elfaizi, MA. 2004. Bioinformatika : Perkembangan, Disiplin Ilmu dan Penerapannya di Indonesia, <http://www.gnu.org/copyleft/fdl.html>. per 6 Juli 2017

[UTAMA2003] Utama, Andi (2003), *Peranan Bioinformatika dalam Dunia Kedokteran*, <http://ikc.vlsm.org/populer/andi-bioinformatika.php>

[BIOINFORMATICS2004] Bioinformatics.org: The Open-Access Institute, <http://bioinformatics.org> per 20 Juni 2017



## **BAB III. BIOINFORMATIKA DNA**

### **I. PENDAHULUAN**

#### **A. Pengantar**

Bioinformatika ilmu yang mempelajari penerapan teknik komputasional untuk mengelola dan menganalisis informasi biologis. Analisis bioinformatika tidak terlepas penggunaan sekuens DNA asam amino serta informasi yang berkaitan dengan analisis. Contohnya topik utama yang meliputi basis data untuk mengelola informasi biologis, penyejajaran sekuens (sequens alignment), prediksi struktur untuk meramalkan bentuk struktur protein maupun struktur sekunder RNA, analisis filogenetik dan analisis ekspresi gen. Semua makhluk hidup memiliki materi genetik untuk mempertahankan kelangsungan struktur, sifat, fungsi dan aktivitas-aktivitas kimia dalam selnya. DNA merupakan salah satu jenis asam nukleat yang berperan sebagai materi genetik yang menurunkan sifat tertentu dari satu generasi ke generasi turunannya. Materi ini mengarahkan pembentukan protein dan RNA tertentu yang penting dalam sel makhluk hidup. DNA juga mengatur pertumbuhan dan pembelahan sel, termasuk informasi untuk diferensiasi sel sehingga terbentuk tumbuhan, hewan, manusia dan mikroorganisme lainnya. Begitu pentingnya DNA ini sehingga disebut sebagai molekul utama kehidupan. DNA berfungsi untuk menyimpan informasi genetik secara lengkap yang diperlukan untuk mencirikan struktur semua protein dan RNA tiap-tiap spesies organisme, untuk membuat program pada saat yang tepat dan menempatkan biosintesis sel dan jaringan secara teratur, untuk menentukan aktivitas organisme sepanjang siklus hidupnya, dan untuk menentukan kekhususan organisme tertentu.

#### **B. Kompetensi Dasar**

Memiliki kemampuan dalam memahami struktur DNA sebagai materi genetik analisis paternitas dan maternitas terhadap individu serta DNA 16S dan 18S sebagai identifikasi organisme.

#### **C. Kemampuan Akhir yang Diharapkan**

Mahasiswa diharapkan :

- a. Mampu mendefinisikan komponen struktur DNA dan RNA
- b. Mampu melakukan Identifikasi organisme secara molekuler

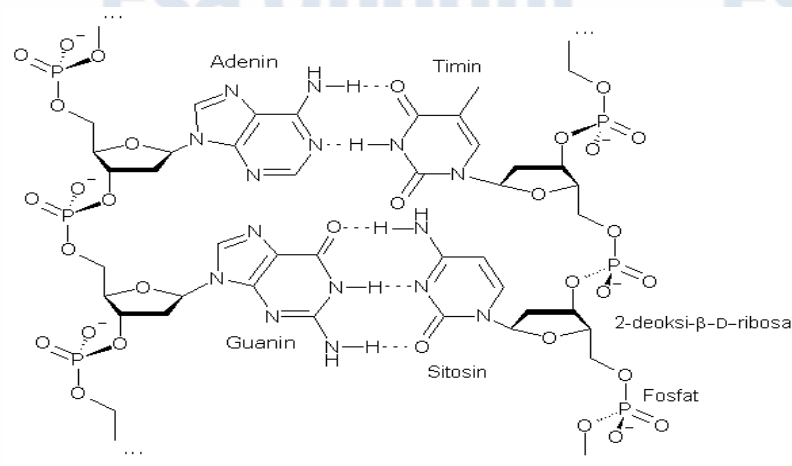
#### D. Kegiatan Pembelajaran

1. Pembelajaran dilakukan dengan metoda contextual learning *Small grup discussion* dan Presentasi
2. Mahasiswa mencari bahan pustaka, membuat bahan presentasi dan mempresentasikan hasil literasinya

## II. MATERI

### A. DNA Sebagai Material Genetik

Rantai DNA adalah sebuah polimer panjang, tak bercabang, yang hanya tersusun dari empat macam subunit. Subunit – subunit ini adalah deoksiribonukleotid yang mengandung basa – basa adenine (A), sitonin (C), guanin (G), dan timin (T). Nukleotid – nukleotid itu saling dirangkaikan dengan ikatan ikatan fosfodiester kovalen yang menghubungkan karbon 5' pada sebuah gugus deoksiribosa dengan karbon 3' pada gugus berikutnya. Keempat macam basa tadi tersambung ke rantai gula –fosfat yang hampir seperti empat macam manik- manik yang menjadi sebuah kalung (Gambar 1).



Gambar 2. Sruktur DNA

Terdapat tiga hal penting pada struktur DNA dari hasil analisa kimia

1. DNA mengandung sejumlah basa purin dan pirimidin yang sama dalam heliks ganda.
2. Terdapat kesetimbangan antara jumlah adenin dengan timin, dan

guanin dengan sitosin (A=T dan G=C).

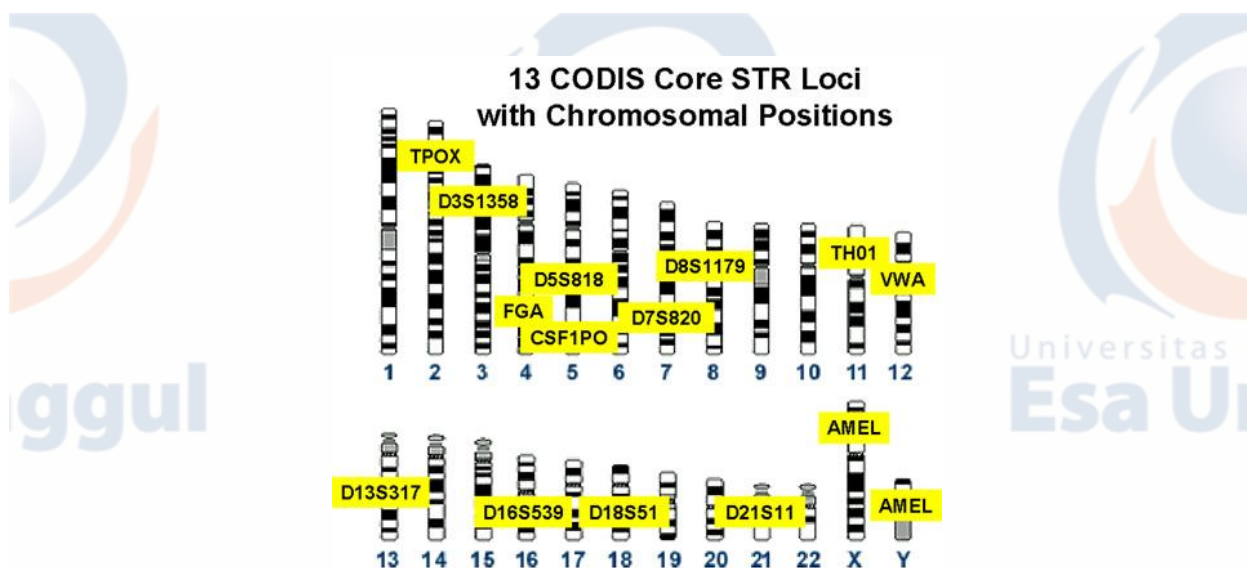
3. Perbandingan antara (A+T) dengan (G+C) dapat bervariasi, tetapi mempunyai nilai yang tetap pada masing-masing spesies

DNA tersusun atas 3 komponen utama yaitu gula deoksiribosa, basa nitrogen, dan fosfat. DNA yang menyusun kromosom ini merupakan nukleotida rangkap yang tersusun heliks ganda, dimana basa nitrogen dan kedua benang polinukleotida saling berpasangan dalam pasangan yang tetap melalui ikatan hidrogen dan antara nukleotida yang satu dengan nukleotida yang lain dihubungkan dengan ikatan fosfat. Ikatan-ikatan fosfat itu sangat kuat dan dikenal sebagai ikatan ester kovalen, atau ikatan fosfodiester. Residu fosfat ( $PO_4^{4-}$ ) sepanjang rantai ini bersifat asam, sehingga diberi nama asam nukleat.

Bioinformatika DNA merupakan kajian bioinformatika yang menjadi sangat penting dalam analisis suatu data sekuensing. Aplikasi ini merupakan alat komputasi dan analisa untuk menangkap dan menginterpretasikan data-data biologi molekuler. Biologi molekuler sendiri juga merupakan bidang interdisipliner, mempelajari kehidupan dalam level molekuler. Mula-mula bidang kajian ini muncul atas inisiatif para ahli biologi molekuler dan ahli statistik, berdasarkan pola pikir bahwa semua gejala yang ada di alam ini bisa dibuat secara artificial melalui simulasi dari data-data yang ada. Pada bidang Bioinformatika, data-data atau tindak-tanduk gejala genetika menjadi inti pembentukan simulasi. Pada saat ini, Bioinformatika ini mempunyai peranan yang sangat penting, diantaranya adalah untuk manajemen data-data biologi molekuler, terutama sekuen DNA dan informasi genetika. Perangkat utama Bioinformatika adalah software dan didukung oleh kesediaan internet

#### **B. Identifikasi Manusia berdasarkan *short Tandem Repeats* (STR) DNA**

Genom manusia terdiri dari untaian unit DNA berulang dalam berbagai ukuran yang terpola. Rasio DNA dengan pengulangan unit pendek (2 -6 bp) di sebut dengan *Short Tandem Repeats* (STR). Seorang individu mewarisi satu salinan STR masing – masing dari orang tuanya. Pengulangan unit STR DNA menjadi marka polimorfisme yang memiliki variasi yang sangat tinggi dalam kelompok individu, sehingga marka STR sangat efektif digunakan untuk identifikasi manusia.



Gambar 3 Marka STR 13 CODIS loki inti pada kromosom manusia. Kit Identifiler AmpFISTR menambahkan dua marka pada kromosom 2 dan 19

Semangkin kecil ukuran alel STR maka marka STR tersebut menjadi lebih baik untuk aplikasi forensik. Mengingat didalam forensik DNA seringkali dalam keadaan terdegradasi. Selain itu alel STR menjadi lebih mudah dipisahkan dari lokasi kromosomal lainnya untuk menghindari terpilihnya loki yang berdekatan yang dapat mengganggu pola distribusi acak populasi yang sangat penting untuk di analisis statistik. Alel STR juga memiliki tingkat mutasi yang lebih rendah, sehingga data yang diperoleh juga semakin stabil dan dapat diprediksi. Berdasarkan karakternya yang unit tersebut, maka STR DNA menjadi alat dengan keakuratan yang tinggi di dalam upaya identifikasi individu pada kasus – kasus forensik. STR DNA dapat digunakan untuk identifikasi korban kecelakaan, pelaku kriminal, penelusuran jejak maupun orang hilang

Marka STR DNA yang digunakan adalah *Combined DNA index system* (CODIS) pada 16 loki yaitu CF1PO, FGA, TH01, TPOX, VWA, D3S1358, D5S818, D7S820, D8S1179, D16S539, D18S51, D21S11, D19S433, dan D2S1338 serta amelogenin untuk jenis kelamin. Codis ini dikeluarkan oleh laboratorium FBI dan telah menjadi standar internasional untuk identifikasi Individu.

Pemeriksaan biologis berbasis STR maka profi DNA individu dapat diinterpretasikan secara lengkap dan jelas. STR adalah loki DNA yang tersusun atas pengulangan 2 – 6 basa. Dalam genom manusia dapat ditemukan pengulangan basa yang bervariasi jumlah dan jenisnya. Identifikasi DNA dengan penanda loki STR merupakan salahsatu prosedur tes DNA yng sangat sensitif karena penanda STR memiliki tingkat variasi yang tinggi baik antar loki STR maupun antar individu. Proses pemeriksaan STR terhadap barang bukti biologis menggunakan metode PCR dan *capillary electrophoresis* (CE). CE dapat digunakan untuk memisahkan ion – ion spesies melalui muatan listriknya dan indentifikasi STR dari satu loki digunakan perangkat *AmpF/STR identifier*.

Hasil dari proses identifikasi terhadap barang bukti biologis disebut dengan elektroforegram. Berupa print out yang terdiri dari 16 loki berupa sinyal elektroforegram. Enam belas loki tersebut adalah D8s1179, D21S11, D7S820, CSF1PO, D3S1358, TH01, D13S317, D16S539, D2S1338, D19S433, vWA, TPOX, D18S51, D5S818, FGA dan sebuah loki menentukan jenis kelamin. XX untuk wanita dan XY untuk pria. Alel sinyal yang digambarkan pada setiap lokus merupakan deskripsi dari profil DNA individu bersangkutan. Setiap lokus memiliki sepasang alel yang diwarisi dari ayah dan ibu biologisnya.

### **C. Tes Paternitas dan Matesnitas DNA dengan software Bioinformatika**

Setiap anak akan menerima satu alel kromosom dari ayah dan satu alel kromosom dari ibu. Dengan perkembangan teknologi, pemeriksaan DNA dapat digunakan untuk mengintifikasi dan membedakan individu yang satu dengan individu yang lain.

Tes paternitas adalah tes DNA unguK menentukan apakah seorang pria aadalah ayah biologis dari seorang anak. Tes paterniats membandingkan pola DNA anak dengan terduga ayah untuk memeriksa bukti pewarisan DNA yang menunjukkan kepastian adanya hubungan biologis dengan menggunakan DNA inti

Tes maternitas adalah tes DNA untuk menentukan apakah seorang wanita adalah ibu biologis dari seorang anak. Seperti pada tes paternitas, tes ini membandingkan pola DNA akan dengan terduga ibu untuk menentukan

kecocokan DNA anak yang diwariskan dari terduga ibu menggunakan DNA mitokondria. Umumnya tes maternitas dilakukan untuk kasus, seperti kasus dugaan tertukarnya bayi, kasus bayi tabung, kasus anak angkat dan lain sebagainya.

Setiap orang memiliki DNA yang unik. DNA adalah materi genetik yang membawa informasi yang dapat diturunkan. Di dalam sel manusia DNA dapat ditemukan di dalam inti sel dan di dalam mitokondria. Di dalam inti sel, DNA membentuk satu kesatuan untaian yang disebut kromosom. Setiap sel manusia yang normal memiliki 46 kromosom yang terdiri dari 22 pasang kromosom somatik dan 1 pasang kromosom sex (XX atau XY). DNA Mitokondria → Tes mtDNA penurunan maternal

Dalam tes paternitas dan tes maternitas beban bioinformatika sangat berperan dalam memberikan hasil yang akurat. Dalam analisis sekuens genetik yang dihasilkan, data diolah berdasarkan software bioinformatika. Dengan analisis DNA, tes paternitas dapat dilakukan sebelum anak dilahirkan (prenatal). Tes DNA dapat dilakukan dengan sampel dari jaringan janin (Chorionic Villi Sample, CVS) umumnya pada umur kehamilan 10-13 minggu, atau dengan cara amniosentesis pada umur kehamilan 14-24 minggu. Untuk pengambilan jaringan janin ini harus dilakukan oleh ahli kebidanan/kandungan. Ibu yang ingin melakukan tes DNA prenatal harus berkonsultasi dengan ahli kebidanan/kandungan.

Tes DNA adalah 100% akurat bila dikerjakan dengan benar. Tes DNA ini memberikan hasil lebih dari 99.99% probabilitas paternitas bila DNA terduga ayah dan DNA anak cocok (matched). Apabila DNA terduga ayah dan anak tidak cocok (mismatched) maka terduga ayah yang di tes 100% bukanlah merupakan ayah biologis anak tersebut. Konfirmasi dilakukan dengan mengulang tes terhadap terduga ayah

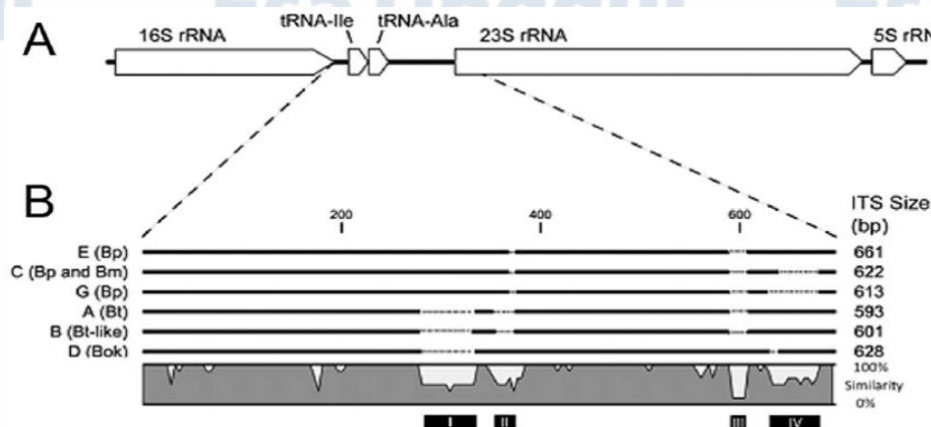
#### **D. Identifikasi organisme berdasarkan *specific gene identified***

##### **1. Gene 16S rRNA**

Studi keanekaragaman genetik pada prinsipnya bertujuan untuk mengkaji komposisi genetik individu di dalam atau antar populasi dan untuk mengetahui faktor-faktor yang menyebabkan terjadinya modulasi atau dinamika keanekaragaman genetik dari populasi tersebut. Secara umum

keanekaragaman genetik pada suatu populasi dapat terjadi karena gen mengalami mutasi, rekombinasi, dan perpindahan sekelompok populasi dari suatu tempat ke tempat yang lain.

Metode identifikasi bakteri yang banyak direkomendasikan adalah analisa genotip bakteri melalui pembacaan sekuen basa nitrogen pada nukleotida penyusun fragmen gen 16S rDNA bakteri. Dinilai dari beberapa pertimbangan, metode ini dinilai lebih baik dibandingkan dengan metode analisa fenotipik. Pertimbangan yang pertama adalah gen 16S rDNA dari hampir seluruh spesies bakteri telah ditentukan urutan basa nitrogennya sehingga dapat dijadikan pedoman jika ditemukan spesies Baru. Pertimbangan yang kedua adalah urutan basa nitrogen gen 16S rDNA memiliki keragaman intraspesifik yang lebih rendah dibandingkan gen pengkode protein yang lain, serta sifat dari fragmen 16S rDNA yang lestari Struktur dan komponen penyusun ribosom pada organisme prokariotik dan eukariotik. Diantara komponen-komponen penyusun ribosom tersebut, rRNA-lah yang menarik perhatian Woese, karena memiliki area konservatif, yakni area yang tidak mengalami perubahan banyak selama proses evolusi. Area konservatif bermanfaat menunjukkan kekerabatan dan perjalanan evolusi untuk mengkonstruksi pohon kehidupan (Gambar 3)



Gambar 4. Daerah lestari pada gen 16S rRNA

Gen 16S rDNA merupakan komponen penting dalam sel dan sangat menguntungkan di dalam analisis filogenik, karena terdiri daerah-daerah yang dikonversi sehingga mutasi akan terbatas, studi sistematika bakteri pada tingkat bakteri, genus, spesies, ataupun subspecies. Pada organisme

prokariotik juga mitokondria dan kloroplas, terdapat tiga jenis rRNA yakni 5S, 16S dan 23S. Diantara tiga jenis rRNA tersebut, 16S rRNA-lah yang dinilai memiliki efisiensi terbaik dalam menentukan kekerabatan dengan metode molekuler. 16S rRNA (terdiri dari 1.5 kb, kilo base) memiliki ukuran lebih besar dibanding 5S rRNA (121 kb) hingga banyak informasi yang bisa dikaji dari molekul tersebut, namun lebih kecil dari molekul 23S rRNA hingga analisis dapat dilakukan lebih cepat serta membutuhkan biaya yang lebih kecil

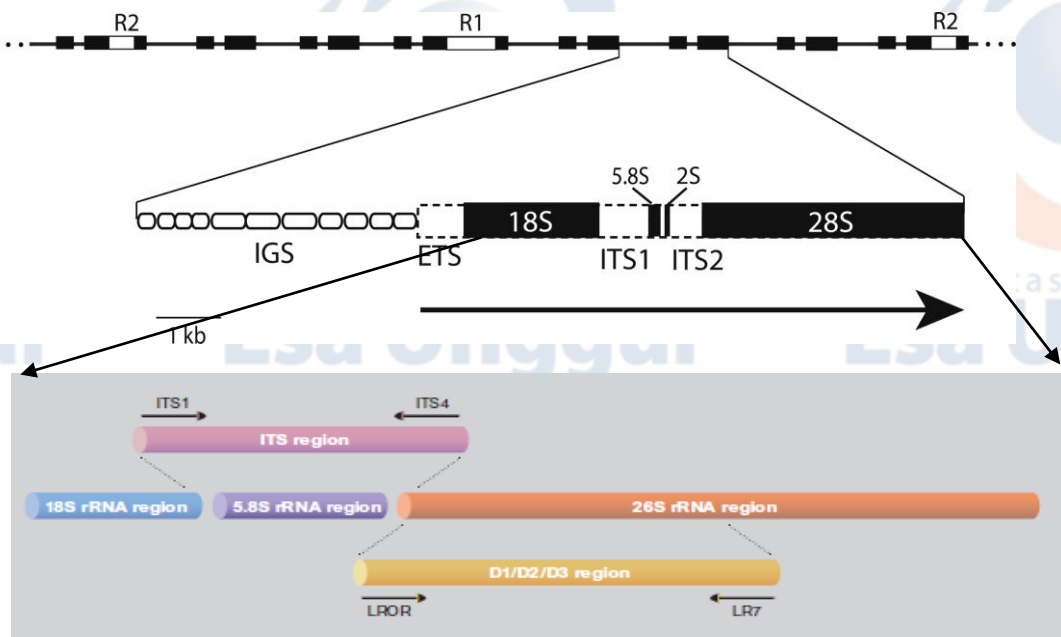
Brock dan Madigan (1991) menyatakan bahwa 16S rDNA sangat mudah penanganannya dari pada 23S rDNA, maka 16S rDNA lebih sering digunakan untuk melihat perkembangan filogenetik prokariota yang memiliki bagian sekuen konservatif. Sabdono (2001) menambahkan bahwa sekuen nukleotida 16S rDNA tidak hanya memudahkan identifikasi bakteri dari sampel lingkungan. Gen 16S rRNA merupakan kerangka penyusun ribosom yang peranannya sangat penting di dalam sintesis protein dan dimiliki oleh semua sel sehingga semua organisme dapat dibandingkan secara setara. Database sekuen gen penyandi 16S-rRNA dari berbagai organisme yang sudah dapat dikulturkan maupun yang tidak dapat dikulturkan telah dibuat databasenya. Tujuan utama pendataan ini adalah untuk mengukur secara tepat hubungan evolusioner di antara organisme-organisme, disamping dapat menjadi katalog untuk keragaman mikroba.

## **2. ITS (*Internal Transcribed Spacer*)**

DNA ribosomal (rDNA) adalah daerah penyandi genom untuk komponen RNA ribosom. Pada eukariot deret rDNA terletak pada nukleus/inti dan mitokondria. Daerah rDNA dipisahkan antara satu dengan yang lainnya oleh suatu pembatas yang disebut spacer Subunit rDNA baik yang besar maupun yang kecil dipisahkan oleh ETS (external transcribed spacer) dan IGS (intergenic spacer). Kedua pembatas tersebut kadang-kadang disebut NTS (nontranscribed spacer). Jamur termasuk organisme eukariotik, pada rDNA jamur terdapat daerah konservatif yaitu gen penyandi rRNA 18S, 5.8S dan 28S yang di antaranya terdapat daerah ITS (*Internal Transcribed Spacer*)



ITS adalah suatu urutan RNA dari proses transkripsi utama yang berada antar prekursor ribosomal subunit dan dihilangkan pada proses splicing ketika RNA precursor tanda molekul yang struktural diproses ke dalam suatu ribosom Organisme eukaryotik mempunyai dua daerah ITS; ITS-1 terletak di antara 18S gen dan 5.8S gen, dan ITS-2 terletak di antara 5.8S dan 28S gen. Ketiga gen ribosom tersebut mempunyai tingkat konservasi yang sangat tinggi (Gambar 4) . Daerah ITS biasanya mengalami perubahan atau mutasi sehingga dapat berbeda atau bervariasi di antara spesies. Sekuen rDNA subunit kecil 18S berkembang relatif lambat dan digunakan untuk studi hubungan kekerabatan pada tingkat spesies suatu organisme sedangkan daerah ITS dan IGS pada unit pengulangan rRNA berkembang lebih cepat dan memungkinkan terjadinya variasi diantara spesies dan populasi



Gambar 5. daerah ITS (*Internal Transcribed Spacer*)

Dua daerah ITS memisahkan gen subunit DNA ribosom. Daerah ITS1 memisahkan gen Subunit DNA ribosom. Daerah ITS1 memisahkan gen subunit ribosom kecil (18S) dengan subunit 5,8S; sedangkan daerah ITS2 memisahkan subunit 5,8S dengan gen subunit ribosom besar (28S). Daerah ITS dipilih karena mudah untuk di amplifikasi walaupun dari sejumlah kecil

DNA dan karena tingginya variasi bahkan antar spesies yang berkerabat dekat. Selain itu daerah ITS juga sangat terkonservasi dan konsisten

### **III. EVALUASI BELAJAR**

#### **a. Latihan**

1. Jelaskan Tentang Bioinformatika DNA ?
2. Jelaskan tentang STR (*short Tandem Repeats*) beserta fungsinya ?
3. Jelaskan tentang tes Paternitas dan maternitas?
4. Jelaskan identifikasi spesies berdasarkan DNA ribosomal

#### **b. Tugas**

Menganalisis hasil sekuens gen 16S dan ITS menggunakan software Bioinformatika

#### **Penilaian Tugas**

1. Tugas dibuat di blog mahasiswa
2. Blog di link ke web hybrid learning.
3. Blog tersebut harus mencantumkan logo dan nama Universitas Esa Unggul
4. Diselesaikan sebelum batas akhir penyerahan tugas

### **IV. Daftar Pustaka**

Brock, T. D. and M. T. Madigan. 1991. *Biology of Microorganisms*. Praticce Hall, Englewood Cliffs, New Jersey

Claveri, JM & Notredame, C. 2007. *Bioinformatics for Dummies*. 2nd Edition. Wiley Publishing. Indiana Canada.

Fatchyah.2015. *Prinsip Dasar Bioinformatika*. Universitas Brawijaya Malang

Jonathan Pevsner. 2015. *Bioinformatics and Functional Genomics*

Sabdono, A. A. 2001. *Identifikasi dan Analisis Genetik Bakteri Karang Pendegradasi Senyawa Herbisida 2,4 – Diklorofenoksi Asetat di Laut Jawa*. Yogyakarta. UGM Press.

Yuwono, T. 2005. *Biologi Molekular*. Erlangga, Jakarta. 269 hal.

## **BAB IV. PENGENALAN ONLINE DATA BASE**

### **I. PENDAHULUAN**

#### **A. Pengantar**

Teknologi Bioinformatika secara umum pada saat ini banyak pekerjaan Bioinformatika berkaitan dengan teknologi database. Penggunaan database ini meliputi baik tempat penyimpanan database "umum" seperti GenBank atau PDB maupun database "pribadi", seperti yang digunakan oleh grup riset yang terlibat dalam proyek pemetaan gen atau database yang dimiliki oleh perusahaan-perusahaan bioteknologi. Konsumen dari data Bioinformatika menggunakan platform jenis komputer dalam kisaran: mulai dari mesin UNIX yang lebih canggih dan kuat yang dimiliki oleh pengembang dan kolektor hingga ke mesin Mac yang lebih bersahabat yang sering ditemukan menempati laboratorium ahli biologi yang tidak suka komputer. Saat ini, perkembangan ilmu biologi sangat dipengaruhi oleh perkembangan ilmu bioinformatika. Tidaklah dapat dimungkiri kalau bioinformatika telah mempercepat kemajuan ilmu biologi. Lebih jauh lagi, kalau dilihat dari bidang yang lebih spesifik, kemajuan suatu bidang sangat dipengaruhi oleh kemajuan bioinformatika. Semakin maju bioinformatika di suatu bidang (ditandai dengan banyaknya software yang tersedia), semakin maju pulalah bidang tersebut.

#### **E. Kompetensi Dasar**

Memiliki kemampuan dalam memahami penyimpanan database secara online dan beberapa tempat penyimpanan data base diGenBank..

#### **F. Kemampuan Akhir yang Diharapkan**

Mahasiswa diharapkan :

- a. Mampu melakukan pencarian database secara online
- b. Mampu mengenal database sesuai dengan fungsinya
- c. Pemahaman tentang NCBI, ENSEMBL , EMBL, SWISS-PROT
- d. Pengenalan tentang GenBank

## G. Kegiatan Pembelajaran

3. Pembelajaran dilakukan dengan metoda contextual learning *Small grup discussion* dan Presentasi
4. Mahasiswa mencari bahan pustaka, membuat bahan presentasi dan mempresentasikan hasil literasinya

## II. MATERI

### A. Pengenalan Situs NCBI

NCBI merupakan server yang memuat data base tentang informasi kesehatan dan bioteknologi. Data base terus menerus di update sesuai dengan penemuan-penemuan terkini yang menyangkut DNA, Protein, Senyawa aktif dan taksonomi. Disamping data base, ncbi juga menyediakan berbagai macam software untuk analisis DNA, protein 3D, pencarian primer, pencarian conserve domain dan lain sebagainya. NCBI merupakan salah satu bank data gen, protein dan literature khususnya dibidang kesehatan yang terlengkap dan diacu oleh para peneliti di dunia.

NCBI (<http://www.ncbi.nlm.nih.gov>) memiliki database dan software (analysis tools) yang sering digunakan untuk analisis adalah sebagai berikut:

#### A.1 DNA-RNA TOOLS:

##### GenBank

Bank Gen ini bagian dari database nukleotida internasional yang bekerja sama dengan DataBank of Japan (DDBJ), the European Molecular Biology Laboratory (EMBL), and GenBank at NCBI.

##### BioSystems

Database yang berisi tentang korelasi biologi dari gen, protein dan small molecule berdasarkan literature. Kita dapat menggunakan untuk menganalisis fungsi protein dan interaksi protein yang kita teliti pada level sel. System ini terhubung dengan beberapa database biosystem seperti reactome dan KEGG. Misalnya kita ingin mengetahui keberadaan/fungsi protein p53 di dalam sel, maka tinggal memasukkan p53 kedalam keyword box, maka akan diperoleh hasil seperti dibawah yang menunjukkan p53 berfungsi sebagai DNA-damage response yang terdapat di database reactome

### Database of Expressed Sequence Tags (dbEST)

- GenBank yang berisi short single-pass reads of cDNA (transcript) sequences/EST.

### Database of Genome Survey Sequences (dbGSS)

- GenBank yang berisi short single-pass reads of genomic DNA/GSS.

### BLAST (*Basic Local Alignment Search Tool*)

Software yang dapat digunakan untuk menentukan homologi suatu urutan DNA atau asam amino dengan data yang ada di NCBI. BLAST memiliki beberapa pilihan menu sesuai dengan analisis yang akan dikerjakan seperti pada table di bawah ini

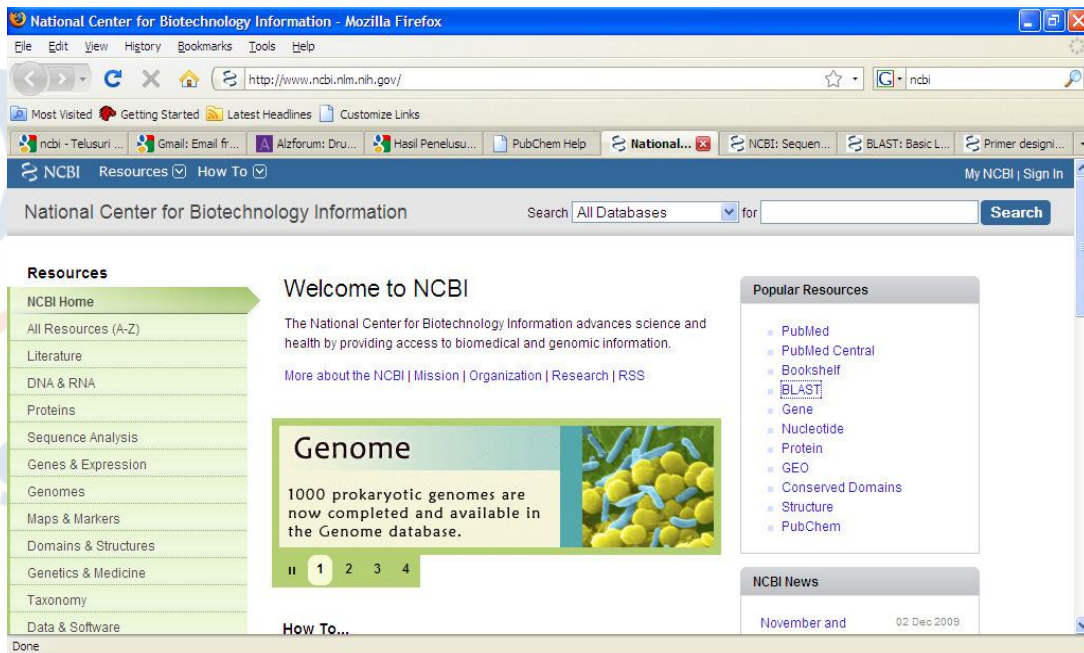
Program	Program Description
<a href="#">nucleotide blast</a>	Search a <b>nucleotide</b> database using a <b>nucleotide</b> query <i>Algorithms:</i> blastn, megablast, discontinuous megablast
<a href="#">protein blast</a>	Search <b>protein</b> database using a <b>protein</b> query <i>Algorithms:</i> blastp, psi-blast, phi-blast
<a href="#">blastx</a>	Search <b>protein</b> database using a <b>translated nucleotide</b> query
<a href="#">tblastn</a>	Search <b>translated nucleotide</b> database using a <b>protein</b> query
<a href="#">tblastx</a>	Search <b>translated nucleotide</b> database using a <b>translated nucleotide</b> query

### Allignment analysis

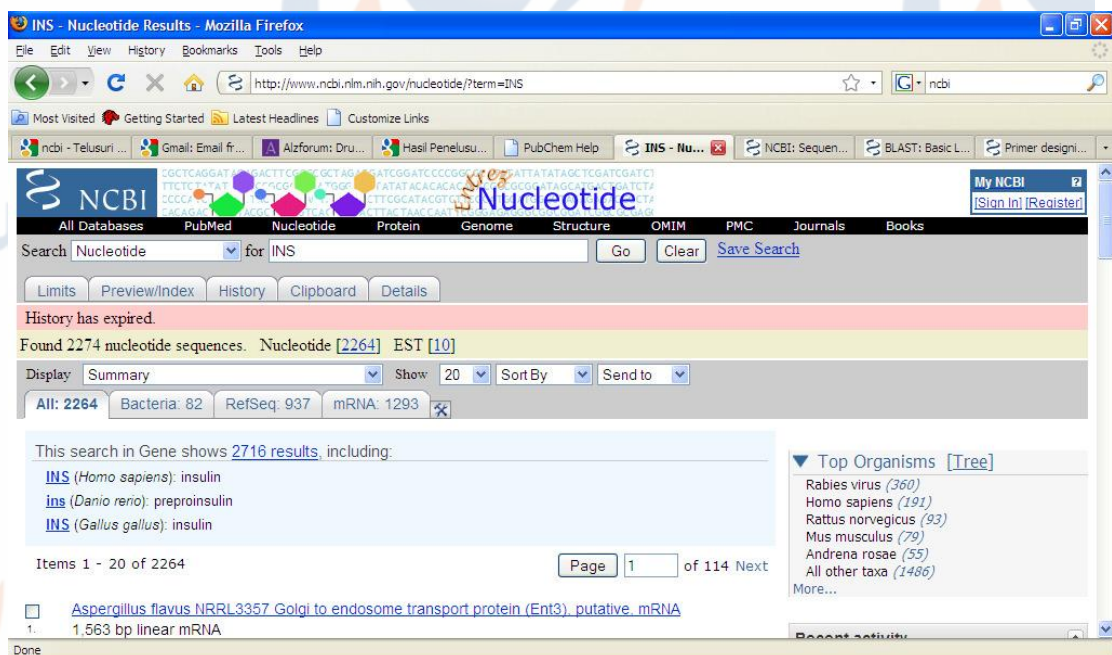
Sekuen yang diperoleh dari hasil penelitian di laboratorium dapat dianalisis dengan data serupa yang telah dipublikasikan sebelumnya di gen bank. Salah satu bentuk analisis yang dapat dilakukan misalnya adalah analisis penyejajaran. Analisis penyejajaran dapat digunakan untuk membandingkan dua sekuen atau lebih. Program yang digunakan untuk analisis penyejajaran yaitu program BLAST (Basic Local Allignment Search Tools). Program ini dapat diakses melalui website National Center for Biotechnology Information at The National Library of Medicine in Washington, DC (<http://www.ncbi.nlm.nih.gov/BLAST>)

Berikut merupakan langkah-langkah untuk mencari dan mendapatkan data dari genbank, misalnya untuk mencari sekuen insulin (INS)

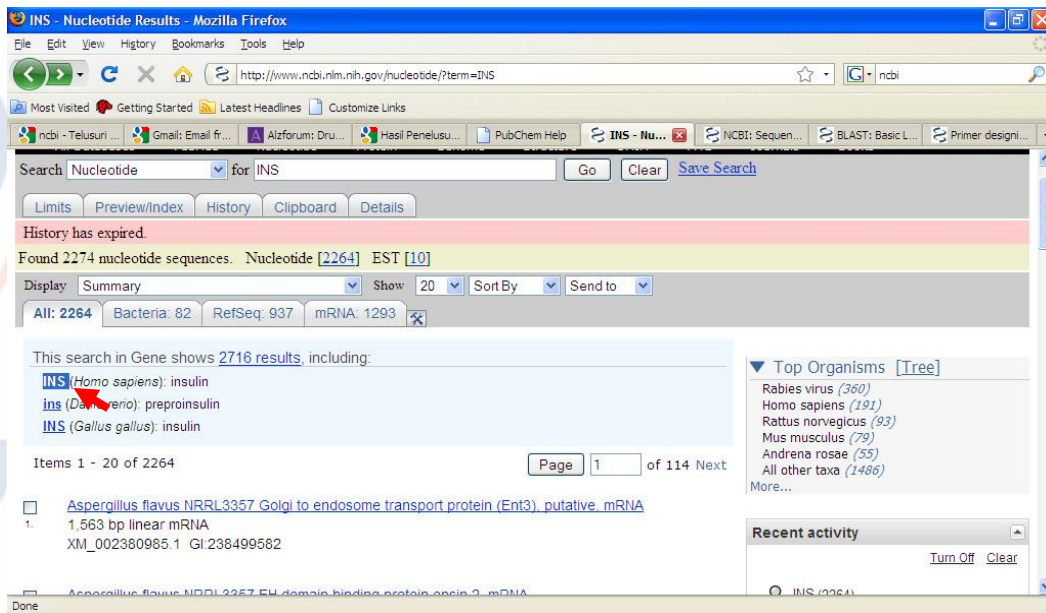
1. Ketikkan <http://www.ncbi.nlm.nih.gov> pada location bar pencarian



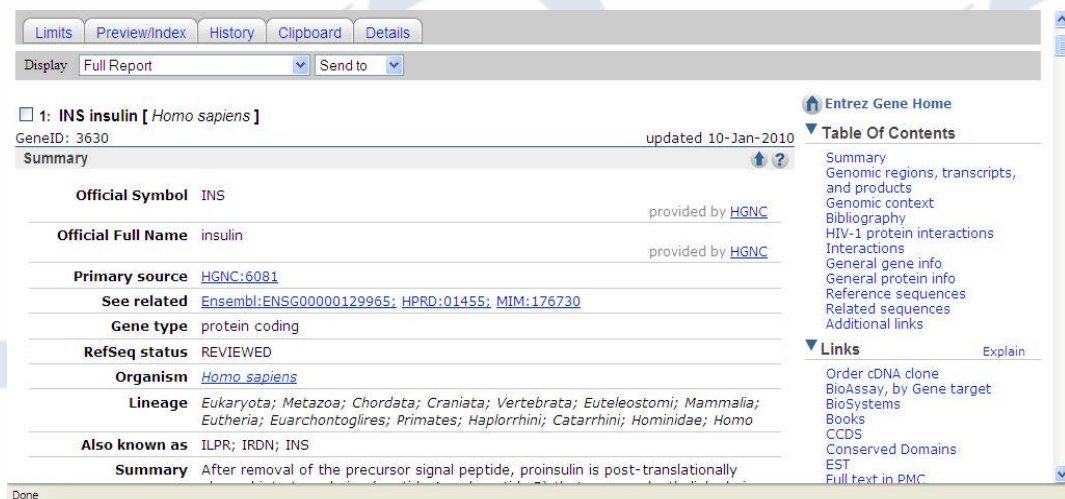
2. Pilih preferensi pencarian yang digunakan (pada contoh ini dipilih nucleotide) dan ketikkan juga molekul yang ingin dicari sebagai kata kunci pencarian (pada contoh ini diketikkan INS) dan diketik GO



3. Muncul berbagai pilihan sekuens yang berkaitan dengan INS dan dipilih (dengan cara mengklik kode) sekuens sesuai kebutuhan. Sekuens dengan kode awal NM menunjukkan sekuens nukleotida sedangkan NP menunjukkan sekuens protein.



4. Berdasarkan pilihan tersebut maka akan diperoleh tampilan sebagai berikut



5. Pada tampilan tersebut discroll ke bawah maka akan diperoleh tampilan berikut. Terdapat beberapa kode yaitu NM dan NP. NM menunjukkan kode untuk memperoleh informasi mengenai nukleotida, sedangkan NP menunjukkan kode untuk memperoleh informasi mengenai protein.

*Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo*

**Also known as** ILPR; IRDN; INS

**Summary** After removal of the precursor signal peptide, proinsulin is post-translationally cleaved into two chains (peptide A and peptide B) that are covalently linked via two disulfide bonds. Binding of this mature form of insulin to the insulin receptor (INSR) stimulates glucose uptake. A variety of mutant alleles with changes in the coding region have been identified. There is a read-through gene, INS-IGF2, which overlaps with this gene at the 5' region and with the IGF2 gene at the 3' region. [provided by RefSeq]

**Genomic regions, transcripts, and products**

(minus strand) Go to [reference sequence details](#) [Try our new Sequence Viewer](#)

**Genomic context**

chromosome: 11; Location: 11p15.5 [See INS in MapViewer](#)

- Books
- CCDS
- Conserved Domains
- EST
- Full text in PMC
- GEO Profiles
- Genome
- HomoloGene
- Map Viewer
- Nucleotide**
- OMIM
- Probe
- Protein
- PubChem Compound
- PubChem Substance
- PubMed
- PubMed (GeneRIF)
- PubMed (OMIM)
- SNP
- SNP: GeneView
- SNP: Genotype
- SNP: VarView
- Taxonomy
- UniSTS
- UniGene
- LinkOut
- HGNC
- Ensembl
- HPRD
- Evidence Viewer
- ModelMaker

6. Diklik link FASTA untuk memperoleh sekuen nukleotida dari INS dalam bentuk FASTA

INS insulin [Homo sapiens] - Gene Result - Mozilla Firefox

http://www.ncbi.nlm.nih.gov/nucleotide/?Db=gene&Cmd=retrieve&opt=full\_report&list\_uids=3630&log\$=data

**Genomic regions, transcripts, and products**

(minus strand) Go to [reference sequence details](#) [Try our new Sequence Viewer](#)

**Genomic context**

chromosome: 11; Location: 11p15.5 [See INS in MapViewer](#)

**Links**

- mRNA LINKS
- FASTA
- GENBANK

**Bibliography**

**Related Articles in PubMed**

**PubMed links**

javascript:window.location="/nuccore/109148525?report=genbank"

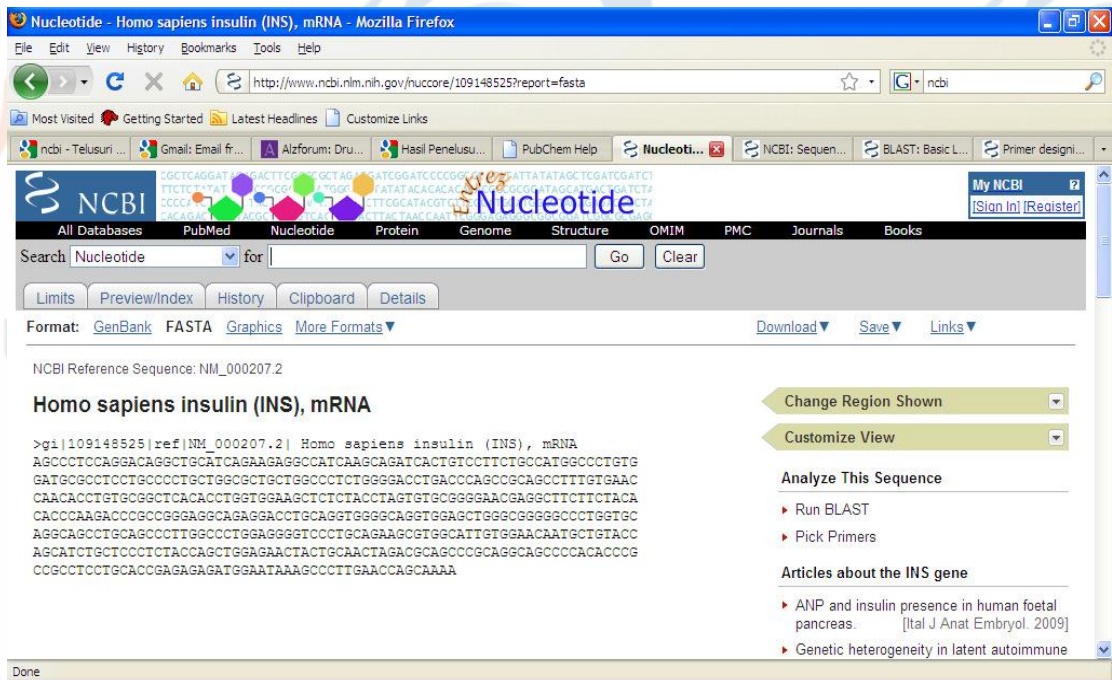
- Protein
- PubChem Compound
- PubChem Substance
- PubMed
- PubMed (GeneRIF)
- PubMed (OMIM)
- SNP
- SNP: GeneView
- SNP: Genotype
- SNP: VarView
- Taxonomy
- UniSTS
- UniGene
- LinkOut
- HGNC
- Ensembl
- HPRD
- Evidence Viewer
- ModelMaker
- AceView
- GeneTests for MIM: 176730
- UCSC
- MGC
- HUGE Navigator
- KEGG
- Reactome

**Entrez Gene Info**

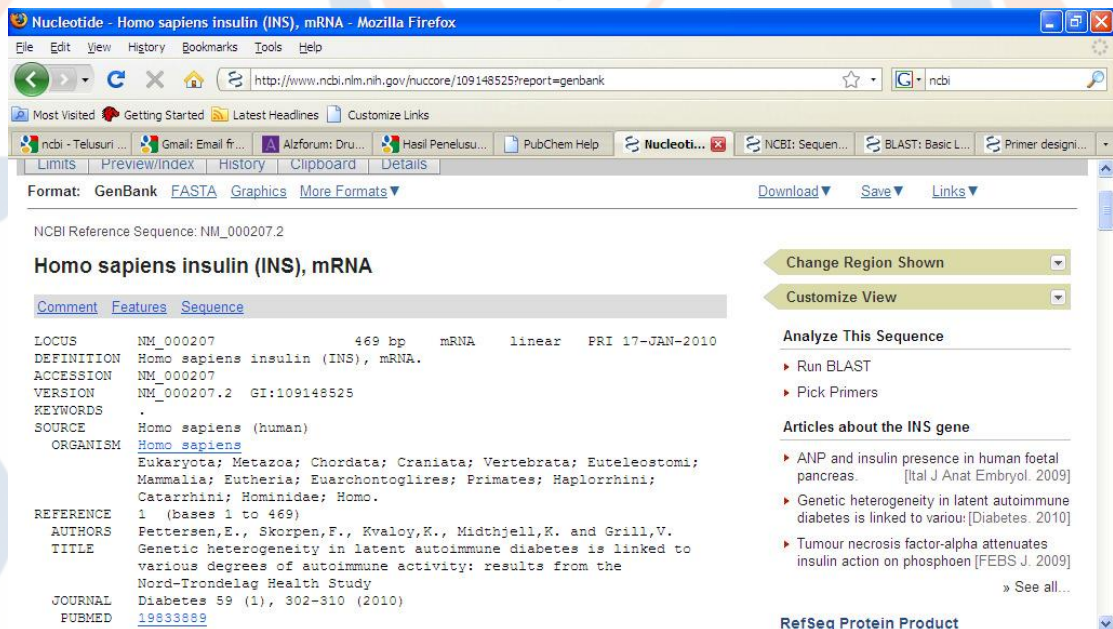
General info



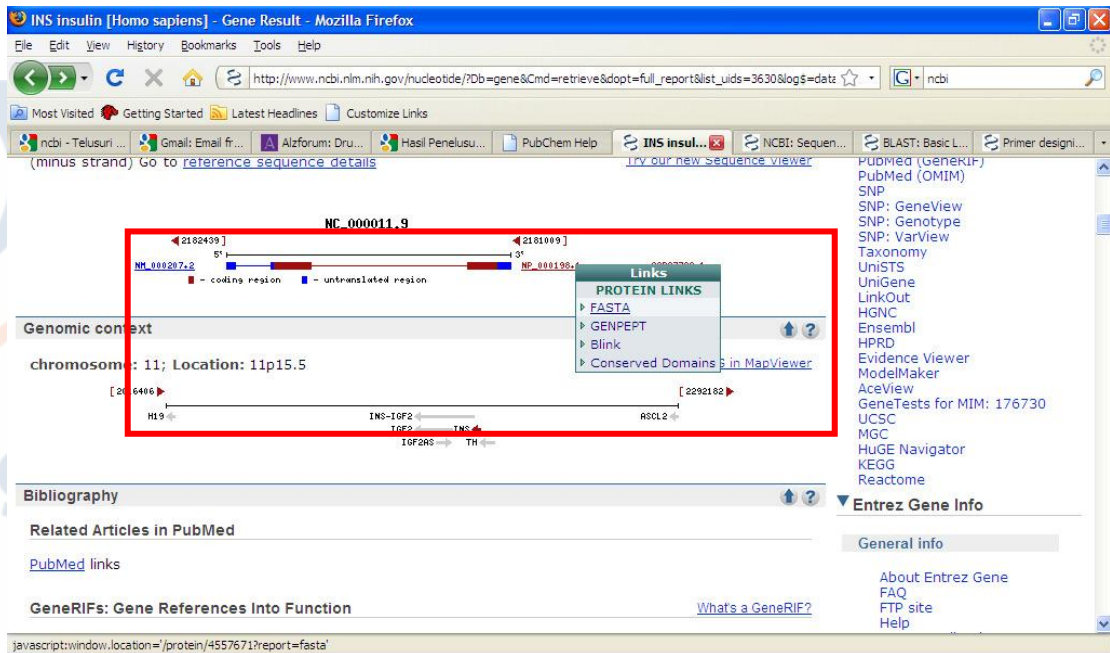
7. Format FASTA yang diperoleh adalah sebagai berikut



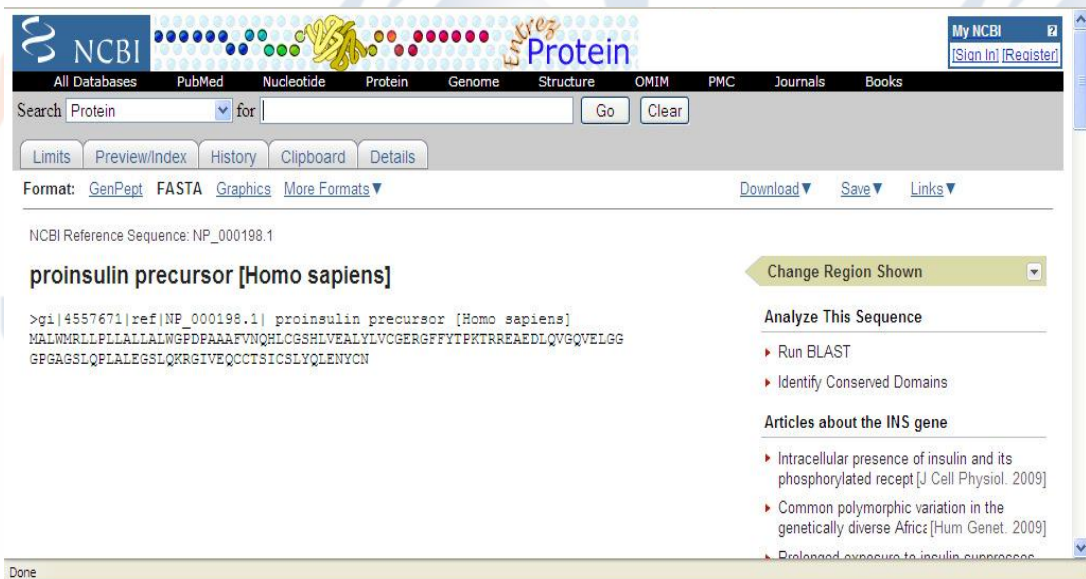
8. Apabila diklik kode NM dan Gen Bank, maka akan diperoleh informasi mengenai sekuens nukleotida, sebagai berikut



9. Kembali pada tampilan berikut, untuk memperoleh data sekuen protein dalam bentuk FASTA maka diklik bagian NP



## 10. Diperoleh sekuen asam amino dalam format FASTA



## Aplikasi BLAST

BLAST (Basic Local Alignment Search Tool) merupakan suatu program untuk pencarian kemiripan sekuen (sequence similarity) dan merupakan alat dalam identifikasi gen dan karakter genetik. Blast dapat melakukan pencarian sekuen melalui perbandingan dengan database DNA dalam waktu singkat (kurang dari 15 detik).

Ada 5 program utama dalam BLAST, yaitu :

- a) nucleotide blast (blastn) : membandingkan suatu sekuen nukleotida meragukan (query sequence) yang kita miliki dengan database sekuen nukleotida.
- b) protein blast (blastp) : membandingkan suatu sekuen asam amino yang kita miliki dengan database sekuen protein.
- c) blastx : membandingkan produk translasi konsep 6-frame sebuah sekuen nukleotida (translated nucleotide) yang kita miliki dengan database sekuen protein.
- d) tblastn : membandingkan suatu sekuen protein yang kita miliki dengan database sekuen nukleotida yang secara dinamis ditranslasi pada semua pembacaan 6 frame.
- e) tblastx : membandingkan suatu translasi 6 frame dari nukleotida

## B. Pangkalan Data Base lain

Pangkalan data primer untuk sekuens asam nukleat saat ini adalah

- [GenBank](#) (Amerika Serikat),
- [EMBL](#) (the European Molecular Biology Laboratory, Eropa), dan
- [DDBJ](#) (DNA Data Bank of Japan, [Jepang](#)).

Ketiga pangkalan data tersebut bekerja sama dan bertukar data secara harian untuk menjaga keluasan cakupan masing-masing pangkalan data.

Sumber utama data sekuens asam nukleat adalah submisi (pengumpulan) langsung dari

- peneliti individual,
- proyek sekuensing [genom](#), dan
- pendaftaran [paten](#).
- Selain berisi sekuens asam nukleat, entri dalam pangkalan data sekuens asam nukleat pada umumnya mengandung informasi tentang
  - jenis asam nukleat ([DNA](#) atau [RNA](#)),
  - nama [organisme](#) sumber asam nukleat tersebut, dan segala sesuatu yang berkaitan dengan sekuens asam nukleat tersebut

1. **OMIM**, (*Online Mendelian Inheritance in Man—woman*), adalah insiklopedia gen-gen manusia dan penyakit genetik, merupakan penghubung

untuk *entry* gen pada GenBank dan literatur ilmiah pada PubMed, berisi informasi berbagai gen manusia komplit dan paling baru.

**2. PDB (Protein Data Bank)** berisi semua publisitas yang ada secara eksperimen telah dideterminasi (oleh *x-ray crystallography* dan *NMR*) sebagai model structural proteins dan asam nukleat. Tidak berisi model homologi atau tipe model teoritis lainnya. **PDB (Protein Data Bank, Bank Data Protein)** ialah pangkalan data tunggal yang menyimpan model struktur tiga dimensi [protein](#) dan [asam nukleat](#) hasil penentuan eksperimental (dengan [kristalografi sinar-X](#), [spektroskopi NMR](#), dan [mikroskopi elektron](#)). PDB menyimpan data struktur sebagai [koordinat tiga dimensi](#) yang menggambarkan posisi [atom-atom](#) dalam protein atau pun asam nukleat.

**3. PubMed** adalah diskripsi pada Wikipedia sebagai “suatu kebebasan mengakses sititasi database MEDLINE dan abstrak artikel riset biomedik.

- Subjek utama adalah riset di bidang kedokteran, dan PubMed juga mempublikasi bidang yang terkait dengan bidang kedokteran, seperti kebidanan dan disipiin kesehatan lainnya.
- Hal ini secara menyeluruh meliputi keilmuan yang berhubungan dengan ilmu seperti biokemia dan biologi sel.

Situs ini ditawarkan oleh *the United States National Library of Medicine* di *the National Institutes of Health* sebagai bagian dari *the Entrez information retrieval system*.”

#### **4. UniProt Knowledgebase (Swiss-Prot and TrEMBL),**

dioperasikan oleh **SIB (Swiss Institute of Bioinformatics)** dan **EBI (European Bioinformatics Institute)**, berisi sebagian besar publikasi yang ada berupa **sekuens protein** (bukan DNA atau RNA).

- Sekuens dalam *Swiss-Prot* dijelaskan secara manual dan menyediakan atau menghubungkan pengguna dengan semua informasi publisitas yang berisi sekuens tersebut.
- Sequences pada TrEMBL dikoleksi dan dijabarkan secara otomatis dari sekuens database, dan akan membuat jalannya menuju Swiss-

Prot, tetapi tidak hanya setelah mereka secara manual menjabarkan *Swiss-Prot standards*.

### **Beberapa situs informasi data base DNA, RNA dan protein**

- NCBI: [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)
- EMBL: [www.ebi.ac.uk](http://www.ebi.ac.uk)
- DDBJ: [www.ddbj.nig.ac.jp](http://www.ddbj.nig.ac.jp)
- SWISS-PROT: [www.expasy.ch/sprot/sprot\\_details.html](http://www.expasy.ch/sprot/sprot_details.html)
- ENSEMBL: [www.ensembl.org](http://www.ensembl.org)
- Univeristy California Santa Cruz: [genome.cse.ucsc.edu](http://genome.cse.ucsc.edu) MGD the Jackson Lab: [www.informatics.jax.org](http://www.informatics.jax.org)

### **III. EVALUASI BELAJAR**

#### **a. Latihan**

1. Jelaskan tentang Tool Penyimpan data base secara online ?
2. Jelaskan Tentang NCBI dan Tools yang ada didalamnya ?
3. Jelaskan pangkalan data base lainnya yang berkenaan tentang DNA dan protein

#### **b. Tugas**

Maklaah tentang penelusuran gen tertentu dengan menggunakan pangkalan data base NCBI

#### **Penilaian Tugas**

1. Tugas dibuat di blog mahasiswa
2. Blog di link ke web hybrid learning.
3. Blog tersebut harus mencantumkan logo dan nama Universitas Esa Unggul
4. Diselesaikan sebelum batas akhir penyerahan tugas (Tanggal ...)

### **IV. Daftar Pustaka**

Brock, T. D. and M. T. Madigan. 1991. *Biology of Microorganisms*. Praticce Hall, Englewood Cliffs, New Jersey

Claveri, JM & Notredame, C. 2007. *Bioinformatics for Dummies*. 2nd Edition. Wiley Publishing. Indiana Canada.

Jonathan Pevsner. 2015. *Bioinformatics and Functional Genomics*

## BAB V. ANALISIS BLAST

### 1. Pendahuluan

#### A. Pengantar

Analisis operasional database berupa sekuens DNA dan protein sangat perlu dalam menghasilkan data yang sesuai dengan hasil penelitian. Penggunaan pensejajaran berganda ini bertujuan untuk mensejajarkan dan mencocokkan hasil sekuensing yang diperoleh dari sampel penelitian dengan data di Genbank

#### Kompetensi Dasar

Mahasiswa dapat memahami dan melakukan analisis Sekuens DNA dan Protein berdasarkan metode BLAST (Basic Local Alignment Search Tool).

#### Kemampuan Akhir yang Diharapkan

Mahasiswa diharapkan :

1. Mahasiswa mampu memahami dan menggunakan program Blastn untuk identifikasi sekuen nukleotida melalui database Genbank
2. Mahasiswa dapat melakukan analisis homologi suatu protein dari organisme tertentu dengan organisme lain menggunakan database Genbank menggunakan Blastp.

#### Kegiatan Pembelajaran

1. Pembelajaran dilakukan dengan metoda contextual learning *Small grup discussion* dan Presentasi
2. Mahasiswa mencari bahan pustaka, membuat bahan presentasi dan mempresentasikan hasil literasinya

#### B. MATERI

**BLAST** (*Basic Local Alignment Search Tool*) merupakan perkakas bioinformatika yang berkaitan erat dengan penggunaan basis data sekuens biologis. Penelusuran BLAST (*BLAST search*) pada basis data sekuens memungkinkan ilmuwan untuk mencari sekuens asam nukleat maupun protein yang mirip dengan sekuens tertentu yang dimilikinya. Hal ini berguna

misalnya untuk menemukan gen sejenis pada beberapa organisme atau untuk memeriksa keabsahan hasil sekuensing maupun untuk memeriksa fungsi gen hasil sekuensing. Algoritma yang mendasari kerja BLAST adalah penyejajaran sekuens

## 1. Analisis Data Sekuen DNA dan Protein dengan BLAST

### a. Blastn

Blastn dapat digunakan untuk mengidentifikasi suatu sekuens nukleotida meragukan (*query sequence*) yang kita miliki dengan database nukleotida, sehingga output yang didapat berupa identitas nukleotida tersebut, antara lain nama gen dan spesies penghasil dari sekuens lengkapnya.

1. Buka situs [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)

The screenshot shows the NCBI homepage. At the top left is the NCBI logo and the text 'National Center for Biotechnology Information'. To the right is a search bar with a dropdown menu set to 'All Databases', a search button, and a clear button. Below the search bar is a 'Resources' sidebar with a list of categories. The main content area includes a 'Welcome to NCBI' message, a 'PubMed Central' banner, and a 'How To...' section. On the right, there are 'Popular Resources' and 'NCBI News' sections.

2. Pilih tool "BLAST", akan muncul tampilan pilihan program BLAST. Untuk mencari gen suatu sekuens nukleotida dari database nukleotida pilih "nucleotide blast" (blastn).

**BLAST** Basic Local Alignment Search Tool

Home Recent Results Saved Strategies Help

NCBI/BLAST Home

BLAST finds regions of similarity between biological sequences. [more...](#)

**New** Aligning Multiple Protein Sequences? Try the **COBALT** Multiple Alignment Tool. [Go](#)

**BLAST Assembled RefSeq Genomes**

Choose a species genome to search, or [list all genomic BLAST databases](#).

- [Human](#)
- [Mouse](#)
- [Rat](#)
- [Arabidopsis thaliana](#)
- [Oryza sativa](#)
- [Bos taurus](#)
- [Danio rerio](#)
- [Drosophila melanogaster](#)
- [Gallus gallus](#)
- [Pan troglodytes](#)
- [Microbes](#)
- [Apis mellifera](#)

**Basic BLAST**

Choose a BLAST program to run.

<a href="#">nucleotide blast</a>	Search a <b>nucleotide</b> database using a <b>nucleotide</b> query <i>Algorithms:</i> blastn, megablast, discontinuous megablast
<a href="#">protein blast</a>	Search <b>protein</b> database using a <b>protein</b> query <i>Algorithms:</i> blastp, psi-blast, phi-blast
<a href="#">blastx</a>	Search <b>protein</b> database using a <b>translated nucleotide</b> query
<a href="#">tblastn</a>	Search <b>translated nucleotide</b> database using a <b>protein</b> query
<a href="#">tblastx</a>	Search <b>translated nucleotide</b> database using a <b>translated nucleotide</b> query

3. Setelah tampilan muncul, entri sekuen nukleotida (query) yang akan dicari; pilih setting pencarian dari database "others" (jika belum diketahui spesiesnya); pilih program "megablast"; klik "BLAST" untuk memulai proses searching. Pada latihan/ccontoh digunakan sekuen nukleotida DNA berikut ini :

```

ATGTTCCCTGAAAAGTTCCTTTGGGGTGTGGCACAATCGGGTTTTTCAG
TTTGAATGGGGGATAAACTCAGGAGGAATATTGACACTAACACTGAT
TGGTGGCACTGGGTAAGGGATAAGACAAATATAGAGAAAGGCCT
CGTTAGTGGAGATCTTCCCGAGGAGGGGATTAACAATTACGAGCTTTA
TGAGAAGGACCATGAGATTGCAAGAAAGCTGGGTCTTAATGCTTACAG
AATAGGCATAGAGTGGAGCAGAATATTTCCCATGGCCAACGACAT
TTATTGATGTTGATTATAGCTATAATGAATCATATAACCTTATAGAAGAT
GTAAAGATCACCAAGGACACTTTGGAGGAGTTAGATGAGATCGCCAA
CAAGAGGGAGGTGGCCTACTATAGGTCAGTCATAAACAGCCTGAGGA
GCAAGGGGTTTAAGGTTATAGTTAATCTAAATCACTTCACCCTTCCATA
TTGGTTGCATGATCCCATTGAGGCTAGGGAGAGGGCGTTAACTAATAA
GAGGAACGGCTGGGTAAAC

```





```
> gb|&F043283.1|&F043283 Pyrococcus woesei beta-galactosidase gene, complete cds
Length=1533
```

```
Score = 998 bits (540), Expect = 0.0
Identities = 540/540 (100%), Gaps = 0/540 (0%)
Strand=Plus/Plus
```

```
Query 1 ATGTTCCCTGAAAAGTTCTTTGGGGTGTGGCACAAATCGGGTTTTAGTTTGAAATGGGG 60
      |||
Sbjct 1 ATGTTCCCTGAAAAGTTCTTTGGGGTGTGGCACAAATCGGGTTTTAGTTTGAAATGGGG 60

Query 61 GATAAACTCAGGAGGAATATTG&CACTAAC&CTGATTGGTGGCACTGGGTAAAGGATAAAG 120
      |||
Sbjct 61 GATAAACTCAGGAGGAATATTG&CACTAAC&CTGATTGGTGGCACTGGGTAAAGGATAAAG 120

Query 121 ACAAATATAGAGAAAAGGCCTCGTTAGTGGAGATCTTCCCGAGGAGGGGATTAAACAATTAC 180
      |||
Sbjct 121 ACAAATATAGAGAAAAGGCCTCGTTAGTGGAGATCTTCCCGAGGAGGGGATTAAACAATTAC 180

Query 181 GAGCTTTATGAGAAAGGACCATGAGATTGCAAGAAAAGCTGGGTCTTAAATGCTTACAGAATA 240
      |||
Sbjct 181 GAGCTTTATGAGAAAGGACCATGAGATTGCAAGAAAAGCTGGGTCTTAAATGCTTACAGAATA 240

Query 241 GGCATAGAGTGGAGCAGAATATCCC&TGGCC&ACG&CATTTATTGATGTTGATTATAGC 300
      |||
Sbjct 241 GGCATAGAGTGGAGCAGAATATCCC&TGGCC&ACG&CATTTATTGATGTTGATTATAGC 300

Query 301 TATAATGAATCATAT&AACCTTATAGA&GATGTAA&GATC&C&A&AGG&AC&CTTTGGAGGAG 360
      |||
Sbjct 301 TATAATGAATCATAT&AACCTTATAGA&GATGTAA&GATC&C&A&AGG&AC&CTTTGGAGGAG 360

Query 361 TTAGATGAGATCGCC&AAC&AGAGGGAGGTGGCCT&ACTATAGGTC&AGTC&ATA&AAC&AGC&CTG 420
      |||
Sbjct 361 TTAGATGAGATCGCC&AAC&AGAGGGAGGTGGCCT&ACTATAGGTC&AGTC&ATA&AAC&AGC&CTG 420

Query 421 AGG&GC&A&GGGGTTT&A&GGTTATAGTT&A&TCT&AA&T&CA&CTT&C&ACC&CTT&CC&AT&ATT&GGTTG 480
      |||
Sbjct 421 AGG&GC&A&GGGGTTT&A&GGTTATAGTT&A&TCT&AA&T&CA&CTT&C&ACC&CTT&CC&AT&ATT&GGTTG 480

Query 481 CATGATCCC&ATT&G&AGG&CT&AGGG&AG&GG&CGTT&A&CT&A&T&A&AG&G&A&AC&GG&CT&GG&TT&A&C 540
      |||
Sbjct 481 CATGATCCC&ATT&G&AGG&CT&AGGG&AG&GG&CGTT&A&CT&A&T&A&AG&G&A&AC&GG&CT&GG&TT&A&C 540
```

- Hasil blast umumnya akan menghasilkan lebih dari satu sekuen yang bersesuaian. Pilih hasil dengan skor paling tinggi dan query coverage mendekati 100%.

Description	Max score	Total score	Query coverage	E value	Max ident
Pyrococcus furiosus DSM 3638, complete genome	998	998	100%	0.0	100%
Pyrococcus woesei beta-galactosidase gene, complete cds	998	998	100%	0.0	100%
Pyrococcus furiosus beta-mannosidase (bmnA) gene, complete cds	998	998	100%	0.0	100%
Uncultured bacterium clone pWTS&A beta-galactosidase gene, complete cds	292	292	99%	1e-75	77%

- Klik "Accession" gen terpilih (hasil blastn) untuk keterangan lebih lanjut, (nucleotide origin dan CDS-nya).

## Pyrococcus woesei beta-galactosidase gene, complete cds

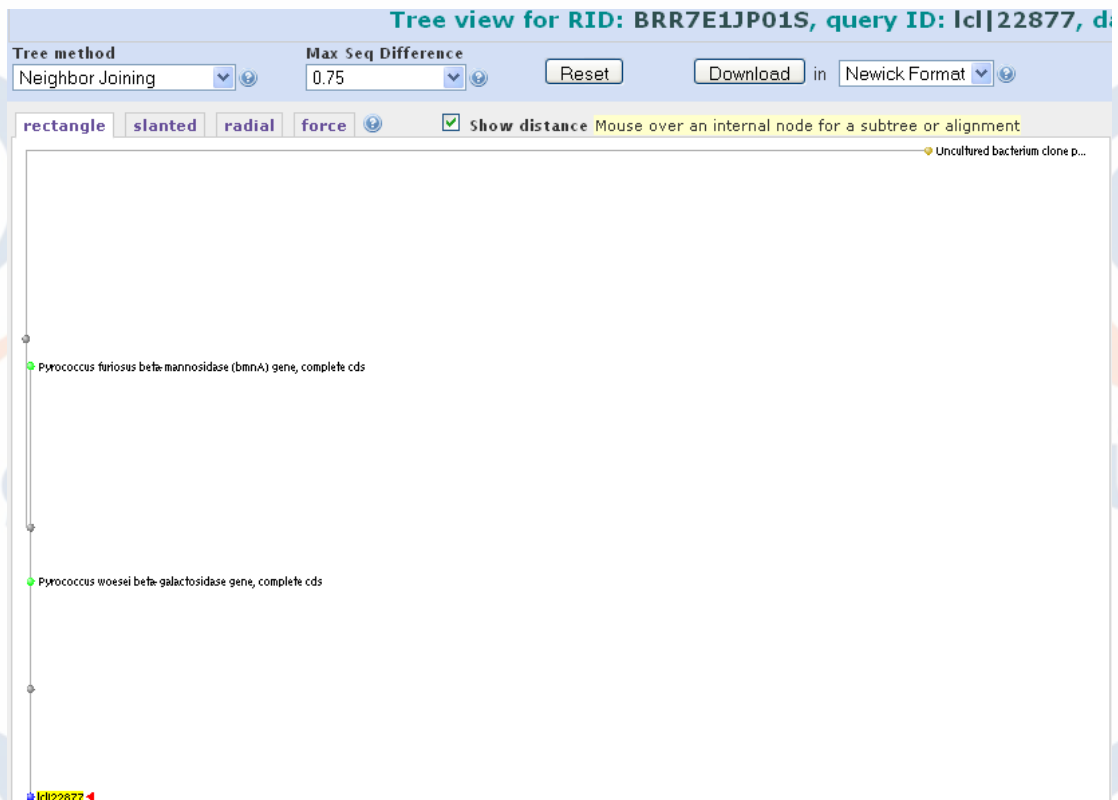
GenBank: AF043283.1

[FASTA](#) [Graphics](#)

[Features](#) [Sequence](#)

LOCUS AF043283 1533 bp DNA linear BCT 25-MAY-2001  
DEFINITION Pyrococcus woesei beta-galactosidase gene, complete cds.  
ACCESSION AF043283

7. Klik "Distance tree of results" Apabila ingin mengetahui phylogenetic tree antar sekuen yang didapatkan. Sebelum melakukan analisis ini, harus dipilih database sekuen yang akan dibandingkan.



## b. Blastp

Blastp dapat digunakan untuk mencari protein homolog dari protein yang kita miliki.

1. Buka situs [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)
2. Pilih tool "BLAST". Untuk mencari protein homolog dari query asam amino gunakan "protein blast" (blastp)
3. Setelah tampilan muncul, entri sekuen protein (*query*) yang akan dicari; pilih seting pencarian dari database (jika membatasi hanya ingin mencari pada spesies tertentu, ketik nama organisme); pilih program "blastp"; klik "BLAST" untuk memulai proses *searching*.

Pada latihan / contoh digunakan query sekuen protein berikut ini :

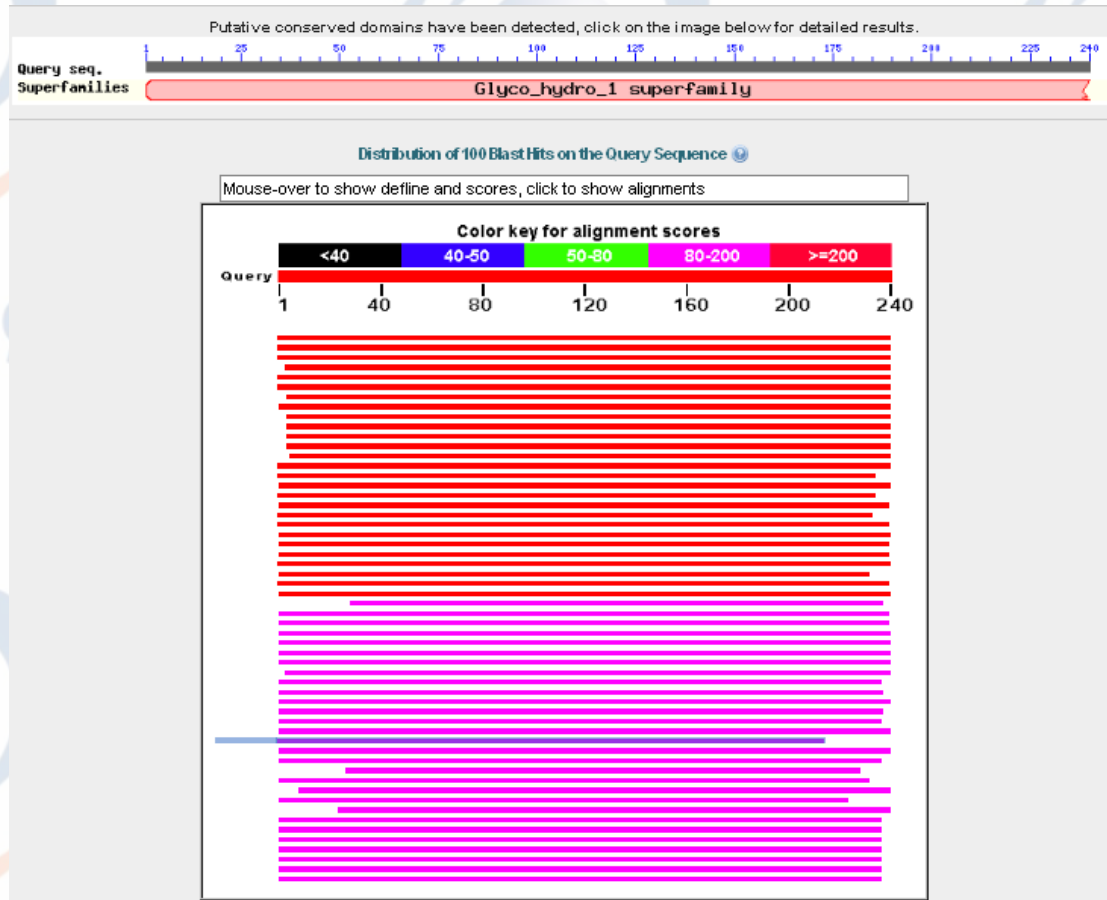
```
MFPEKFLWGVAQSGFQFEMGDKLRRNIDTNTDWWHWVRDKTNIEKGL
VSG DLPEEGINNYELYEKDHEIA RKLGLNAYRI GIEWSRIFPW
PTTFIDVDYS YNESYNLIED VKITKDTLEE LDEIANKREVAYYRSVINSL
RSKGFKVIIVNLNHFTLPYWLHDPIEARERALTNKRNGWVNPRTVIEFAKY
AAYIAYKFGDIVDMWSTFNEPMVVVELGYLAPYSGFPPGV LNPEAAKLAI
```

The screenshot shows the NCBI BLAST search interface. The query sequence is entered in the "Enter Query Sequence" field. The search settings are configured as follows:

- Database:** Non-redundant protein sequences (nr)
- Organism:** (Optional field)
- Exclude:**  Models (XM/XP)  Uncultured/environmental sample sequences
- Entrez Query:** (Optional field)
- Program Selection:**  blastp (protein-protein BLAST),  PSI-BLAST (Position-Specific Iterated BLAST),  PHI-BLAST (Pattern Hit Initiated BLAST)

The search button is labeled "BLAST". Below the search button, the search parameters are displayed: "Search database Non-redundant protein sequences (nr) using Blastp (protein-protein BLAST)". The checkbox "Show results in a new window" is checked.

4. Hasil *searching* akan didapat tampilan seperti berikut:



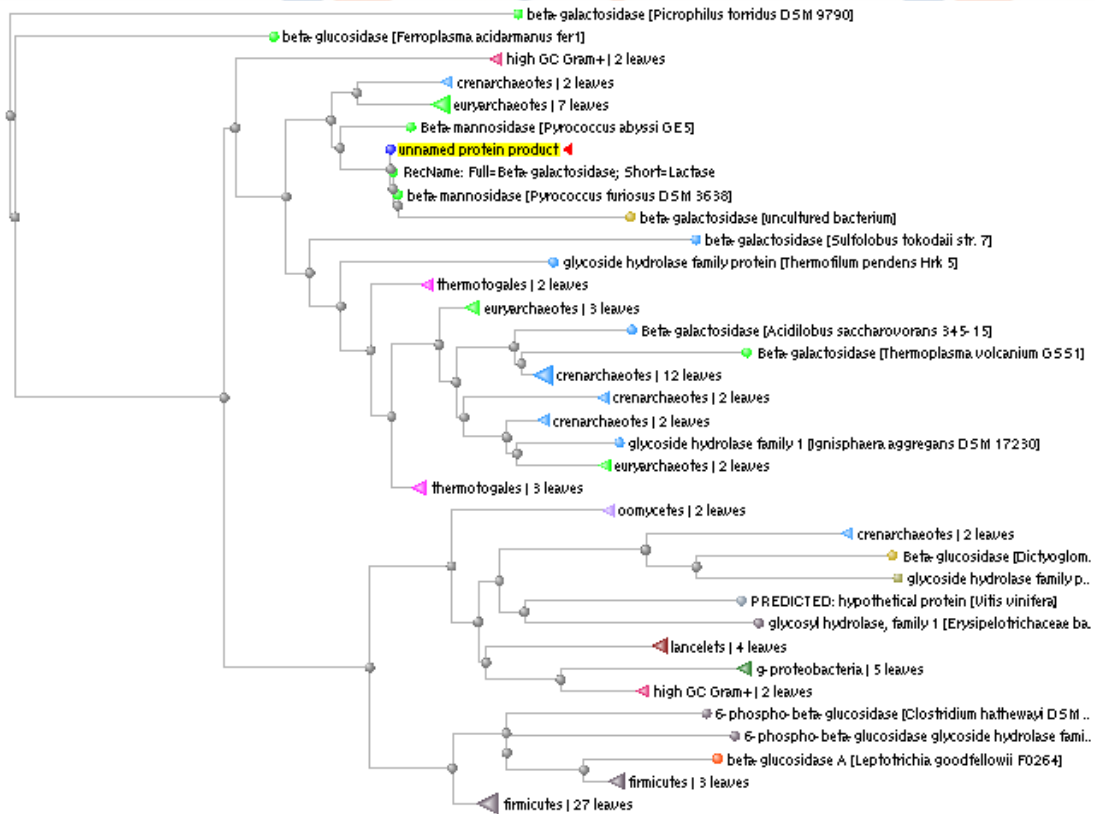
Sequences producing significant alignments:

Accession	Description	Max score	Total score	Query coverage	E value
<a href="#">O52629.1</a>	RecName: Full=Beta-galactosidase; Short=Lactase >gb AAB97862.1  beta-galactosidase	494	494	100%	2e-138
<a href="#">NP_578937.1</a>	beta-mannosidase [Pyrococcus furiosus DSM 3638] >gb AAC44387.1  beta-mannosidase	494	494	100%	4e-138
<a href="#">NP_126617.1</a>	Beta-mannosidase [Pyrococcus abyssi GES] >emb CAB49848.1  bgaL-1 beta-galactosidase	426	426	100%	1e-117
<a href="#">YP_002993739.1</a>	Beta-galactosidase [Thermococcus sibiricus MM 739] >gb ACS89390.1  Beta-galactosidase	337	337	98%	1e-90
<a href="#">ABY49715.1</a>	beta-galactosidase [uncultured bacterium]	328	328	100%	3e-88
<a href="#">YP_921073.1</a>	glycoside hydrolase family protein [Thermofilum pendens Hrk 5] >gb ABL79070.1  glycosidase	315	315	100%	3e-84
<a href="#">YP_184174.1</a>	beta-galactosidase [Thermococcus kodakarensis KOD1] >dbj BAD85950.1  beta-glycosidase	312	312	98%	3e-83
<a href="#">YP_003668816.1</a>	glycoside hydrolase family 1 [Staphylothermus hellenicus DSM 12710] >gb ADI31917.1	308	308	99%	4e-82
<a href="#">BAA78713.1</a>	beta-glycosidase [Thermococcus kodakarensis]	306	306	98%	1e-81
<a href="#">NP_142473.1</a>	beta-mannosidase [Pyrococcus horikoshii OT3] >dbj BAA29589.1  483aa long hypothetical	303	303	98%	1e-80
<a href="#">NP_127217.1</a>	beta-galactosidase [Pyrococcus abyssi GES] >emb CAB50447.1  bgaL-2 beta-galactosidase	299	299	98%	2e-79
<a href="#">NP_578085.1</a>	beta-galactosidase [Pyrococcus furiosus DSM 3638] >gb AAL80480.1  beta-galactosidase	299	299	98%	2e-79
<a href="#">ZP_04878911.1</a>	beta-galactosidase [Thermococcus sp. AM4] >gb EEB74957.1  beta-galactosidase [Thermococcus sp.]	297	297	97%	7e-79

5. Hasil blast akan menghasilkan lebih dari satu sekuen yang bersesuaian. Pilih hasil dengan skor paling tinggi. Dengan meng-klik referensi akan didapat keterangan lebih lanjut tentang protein tersebut.



- Klik “Distance tree of results” pada bagian akhir apabila ingin mengetahui phylogenetic tree antar protein yang didapatkan. Sebelum melakukan analisis ini, harus dipilih database protein yang akan dibandingkan.



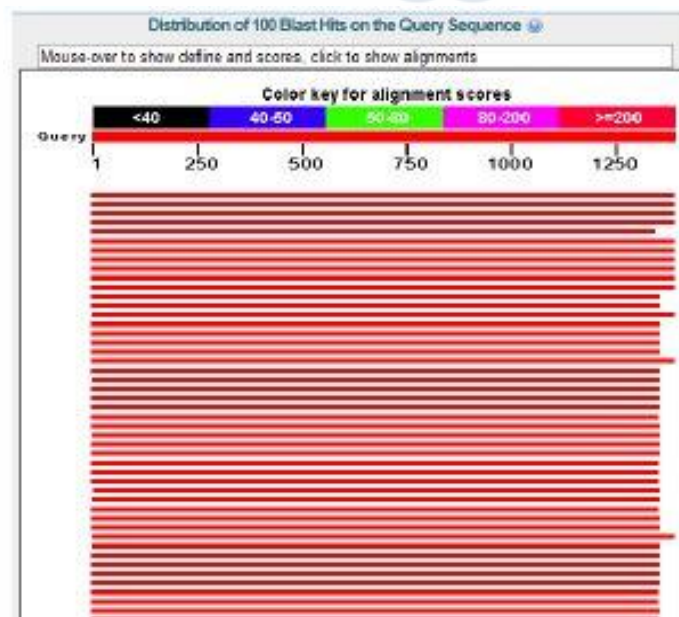
## 2. Analisis Hasil BLAST

BLAST tersebut memberikan informasi mengenai bakteri apa yang mempunyai kesamaan dengan urutan DNA sampel sehingga dapat digunakan untuk identifikasi bakteri. Informasi dari hasil BLAST tersebut berupa *Score*, *Query Coverage*, *E-value* dan *Maximum identity*.

*Score* adalah jumlah keselarasan semua segmen dari urutan database yang cocok dengan urutan nukleotida. Nilai skor menunjukkan keakuratan nilai penjajaran sekuens berupa nukleotida yang tidak diketahui dengan sekuens nukleotida yang terdapat di dalam genbank. Semakin tinggi nilai skor yang diperoleh maka semakin tinggi tingkat homologi kedua sekuens. *Query coverage* adalah persentasi dari panjang nukleotida yang selaras dengan database yang terdapat pada BLAST. *Max identity* adalah

nilai tertinggi dari persentasi identitas atau kecocokan antara sekuen *query* dengan sekuen database yang tersejajarkan.

Nilai *E-value* merupakan nilai dugaan yang memberikan ukuran statistik yang signifikan terhadap kedua sekuen. Nilai *E-value* yang semakin tinggi menunjukkan tingkat homologi antara sekuen semakin rendah, sedangkan nilai *E-value* yang semakin rendah menunjukkan tingkat homologi antar sekuen semakin tinggi. Nilai *E-value* bernilai 0 (nol) menunjukkan bahwa kedua sekuen tersebut identik



Gambar 20 Contoh Analisis BLAST Hasil Sekuensing

Tingkat homologi sekuen tersebut dapat ditunjukkan dengan nilai yang tertera pada warna grafik hasil BLAST. Nilai pada grafik yang berada di bawah angka 50 menunjukkan tingkat homologi kedua sekuen rendah yang dideskripsikan dengan warna hitam dan biru. Warna hijau, merah muda dan merah menunjukkan tingkat homologi yang semakin tinggi. Pada Gambar 20 Grafik hasil BLAST menunjukkan warna merah yang menandakan bahwa tingkat homologi sekuen yang tinggi. Homologi adalah kesimpulan bahwa dua sekuen tersebut sama dan memiliki hubungan evolusi.

Nilai *max identity* sebesar 99% mengindikasikan bahwa isolat dianggap sebagai spesies yang sama. Sedangkan homologi  $\geq 97\%$  dapat dinyatakan bahwa isolat yang dibandingkan berada pada genus yang sama dan homologi antara 89-93% menunjukkan famili yang berbeda. Tetapi hal ini



perlu ditelusuri lagi melalui analisis filogenetik dengan melihat percabangan yang dibentuk oleh isolat melalui pengamatan posisi yang ditempati diantara spesies-spesies yang lain atau spesies pembandingnya

### III.EVALUASI BELAJAR

#### a. Rangkuman

BLAST (Basic Local Alignment Search Tool) merupakan suatu program untuk pencarian kemiripan sekuen (sequence similarity) dan merupakan alat dalam identifikasi gen dan karakter genetik. Blast dapat melakukan pencarian sekuen melalui perbandingan dengan database DNA dalam waktu singkat analisisnya dapat berupa

- a) nucleotide blast (blastn) : membandingkan suatu sekuen nukleotida meragukan (query sequence) yang kita miliki dengan database sekuen nukleotida.
- b) protein blast (blastp) : membandingkan suatu sekuen asam amino yang kita miliki dengan database sekuen protein.
- c) blastx : membandingkan produk translasi konsep 6-frame sebuah sekuen nukleotida (translated nucleotide) yang kita miliki dengan database sekuen protein.
- d) tblastn : membandingkan suatu sekuen protein yang kita miliki dengan database sekuen nukleotida yang secara dinamis ditranslasi pada semua pembacaan 6 frame.
- e) tblastx : membandingkan suatu translasi 6 frame dari nukleotida

#### b. Latihan

1. Jelaskan 5 Metedo Analisis BLAST ?
2. Buat lah langkah – langkah analisis data menggunakan BLAST berdsarkan data sekuen DNA ?
3. Buat lah langkah – langkah analisis data menggunakan BLAST berdsarkan data sekuen protein!
4. Apa yang membedakan BLASTn dan BLASTp

#### c. Tugas

Identifikasi bakteri dengan metode 16S rRNA hasil hasil sekuens *partial gene*

#### d. Penilaian Tugas

- Tugas dibuat di blog mahasiswa
- Blog di link ke web hybrid learning.
- Blog tersebut harus mencantumkan logo dan nama Universitas Esa Unggul
- Diselesaikan sebelum batas akhir penyerahan tugas

#### VI. DAFTAR PUSTAKA

Brock, T. D. and M. T. Madigan. 1991. Biology of Microorganisms. Praticce Hall, Englewood Cliffs, New Jersey

Claveri, JM & Notredame, C. 2007. Bioinformatics for Dummies. 2nd Edition. Wiley Publishing. Indiana Canada.

Jonathan Pevsner. 2015. Bioinformatics and Functional Genomics

## BAB VI. MENDESAIN PRIMER DNA

### I. Pendahuluan

#### a. Pengantar

Pada saat sekarang ini desain primer merupakan salah satu faktor yang paling penting/langkah-langkah dalam sekuensing DNA sukses. Berbagai program bioinformatika yang tersedia untuk pemilihan pasangan primer dari urutan template. Program kebanyakan untuk desain primer PCR mencerminkan peran sentral PCR dalam biologi molekuler modern. Desain Primer merupakan salah satu bahan yang digunakan dalam proses perbanyakkan untai DNA dengan menggunakan metode *Polymerase Chain Reaction*

#### Kompetensi Dasar

Mahasiswa dapat memahami dan melakukan design primer dari gen tertentu untuk digunakan sebagai aplikasi gen yang ditargetkan.

#### Kemampuan Akhir yang Diharapkan

Mahasiswa diharapkan :

1. Mampu mendesign primer dari gen tertentu
2. Dapat mengkararakteristik primer yang baik
3. Dapat membedakan dalam mendesign degenerated primer dan primer spesifik

#### Kegiatan Pembelajaran

- Pembelajaran dilakukan dengan metoda contextual learning *Small grup discussion* dan Presentasi
- Mahasiswa mencari bahan pustaka, membuat bahan presentasi dan mempresentasikan hasil literasinya

## II. MATERI

### Karakteristik Primer Baik

Primer adalah oligonukleotida spesifik yang komplemen dengan daerah yang ditentukan pada DNA target sebagai tempat dimulainya sintesis

DNA baru dengan PCR (Glick and Pasternak, 2003). Untuk mendapatkan daerah tertentu pada DNA target diperlukan desain primer *forward* dan *reverse* dari daerah tersebut. Khusus untuk primer reverse diperoleh dari sekuen antisensenyanya. Jadi nukleotida yang didapat dari sekuen ujung 3' daerah DNA target tadi dicari komplementnya baru kemudian dibalik. Desain primer spesifik merupakan salah satu faktor yang mempengaruhi keberhasilan amplifikasi DNA dengan PCR. Komposisi, ukuran, dan homologi primer terhadap DNA target harus ditentukan dengan baik agar diperoleh produk amplifikasi yang diinginkan saat melakukan PCR.

Beberapa hal yang dipertimbangkan dalam mendesain primer adalah ukuran primer, temperatur *melting* ( $T_m$ ), dan komposisi primer. Saiki (1990) menyatakan bahwa primer yang digunakan bisa berukuran antara dua puluh sampai tiga puluh nukleotida. Sambrook dan Russel (2001) menerangkan bahwa sesuai dengan peraturan Wallace/"Wallace rule",  $T_m$  dapat ditentukan dengan rumus:

$T_m = 4^\circ\text{C} (G+C) + 2^\circ\text{C} (A+T)$ , dimana

G+C merupakan banyaknya basa G dan C dalam primer yang didesain, dan A+T merupakan banyaknya basa A dan T.

Namun penentuan  $T_m$  sangat bergantung pada konsentrasi  $\text{Na}^+$  pada buffer sehingga rumus  $T_m$

menjadi:

$$T_m = 81.50^\circ\text{C} + 16.6 (\log_{10}[\text{Na}^+]) + 0.41(\%[\text{G}+\text{C}]) - (500/n)$$

Dimana  $n$  merupakan jumlah nukleotida dan formula ini hanya bekerja pada  $[\text{Na}^+]$  1,0 M atau dibawahnya. Sambrook dan Russel (2001) lebih jauh menyatakan bahwa sebaiknya kandungan basa G+C adalah sekitar 50%. Apabila basa G dan C terdapat dalam jumlah banyak pada ujung 3' akan memungkinkan terjadinya kesalahan, misalnya terbentuk struktur "jepit rambut" (*hair pin*). Dalam penentuan primer, diusahakan agar tidak terjadi perpasangan sendiri (*self priming*) dan *dimer-duplex*. Primer yang rendah kandungan basa G dan C nya masih memungkinkan untuk dipilih, asal primernya berukuran lebih panjang untuk menghindari  $T_m$  yang terlalu rendah.  $T_m$  berhubungan dengan suhu *annealing* atau suhu dimana dimulainya hibridisasi/penempelan primer dengan komplementnya pada

*template*. Suhu *annealing* biasanya lebih rendah 5 °C dari T<sub>m</sub> (Sambrook dan Russel, 2001).

Beberapa hal yang perlu dipertimbangkan untuk mendesain primer sebagai berikut ;

### 1. Ukuran Primer dan Temperatur melting (T<sub>m</sub>)

Menurut Old dan Primrose (1994) bahwa primer yang digunakan sebaiknya berukuran 17 – 30 basa, sedangkan Saiki (1990) menyatakan 20 -30 basa dan ditentukan menurut sekuen DNA yang mengikat daerah yang akan di amplifikasi . Selanjutnya Guarvara-Garcia *et al.* (1997) menyatakan bahwa primer yang berukuran 20 basa atau lebih biasanya disukai sebab nilai T<sub>m</sub> nya dapat mencapai lebih dari 55°C. Perlu diingat bahwa primer yang berukuran pendek tidak dapat sempurna melakukan proses *annealing* disebabkan nilai T<sub>m</sub>-nya rendah. Oleh karena itu penentuan ukuran primer yang berkisar sekitar 20 basa yang paling banyak dipilih.

### 2. Komposisi Primer

Kandungan G + C pada primer sebaiknya sekitar 50 % (Old and Primrose 1994), sedangkan Guevara –Garcia (1997) menyatakan penentuan kandungan G+C tersebut sangat baik digunakan untuk menjamin stabilitas primer yang tinggi dengan penyebaran basanya yang merata ke arah ujung 5' pada primer. Apabila C dan G terdapat dalam jumlah yang banyak pada ujung 3' akan memungkinkan terjadinya banyak kesalahan, misalnya terbentuk struktur jepit rambut. Stabilitas primer biasanya akan menurun pada 5 basa terakhir daerah ujung 3'. Selanjutnya diusakan pula untuk mencegah terjadinya pemasangan sendiri (*self-priming*), dimer –duplex dan struktur jepit rambut (*hair pin*) yang sangat penting diperhatikan pada bagian akhir ujung 3' primer.

### 3. Desain Primer Menggunakan Program dari Internet

Pelaksanaan desain primer telah dimudahkan dengan adanya berbagai program yang dapat diakses dari internet atau paket program komputer seperti *oligo*, *primers*, *primer select* (DNA star) dan *Cybergene* yang dapat membantu untuk mendesain primer (Guevara-Garcia *et al.* 1997)

Program *Primer3plus* dapat digunakan untuk menentukan posisi primer *forward* dan *reverse* dalam satu sekuen DNA yang panjang sesuai dengan kriteria yang di berikan misalnya beberapa jumlah basa, kandungan G+C

damn Tm/Ta. Program tersebut secara langsung menentukan calon primer dan letaknya tanpa mempertimbangkan apakah primer yang dianalisis berada pada sekuen yang diinginkan sehingga bila digunakan untuk mendesain primer spesifik akan menimbulkan kesulitan.

Pada program Cybergenese hanya memasukan sejumlah sekuen basa yang akan dipilih sebagai primer, kemudian dianalisis apakah ada atau tidaknya sruktur sekunder seperti polindrom dan jepitan rambut, dan juga dapat menghitung beberapa kandungan G+C serta Tm/Ta. Sekuen primer dianalisis secara manual terlebih dahulu dan dipilih yang dianggap layak, kemudian dimasukan dalam program Cybergene.se, primer yang baik apabila hasil analisisnya tidak dityemukan sruktur sekunder.

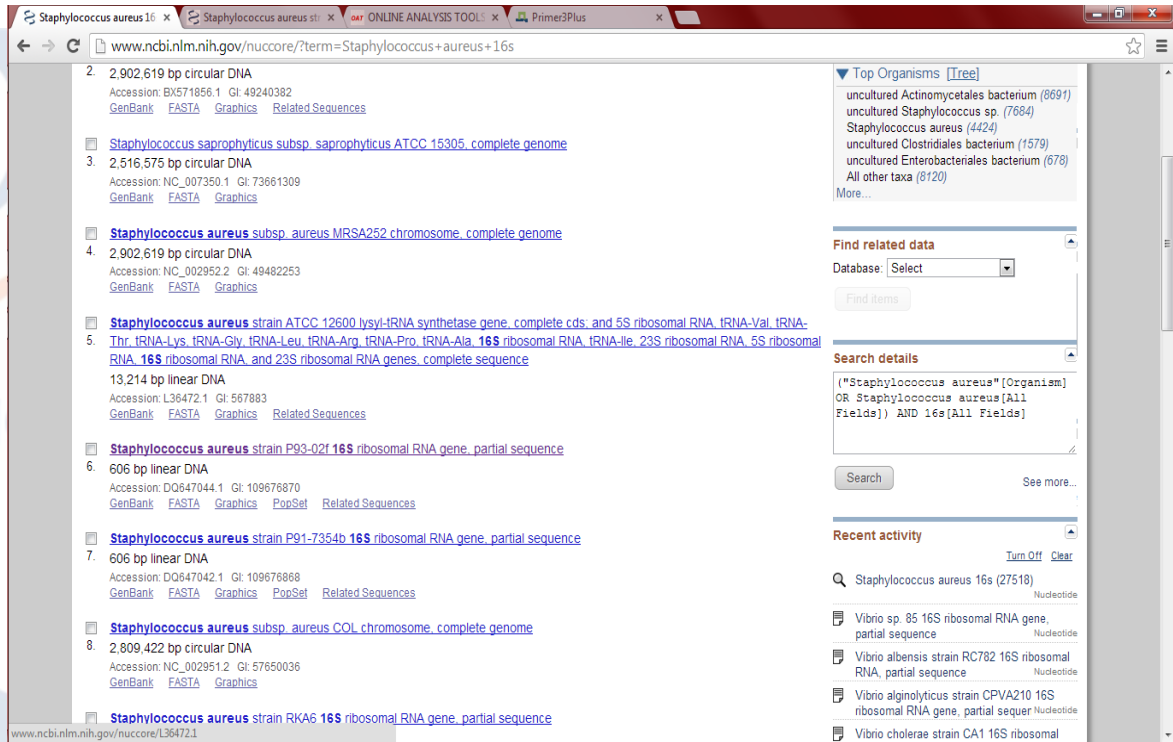
Pada prinsipnya program untuk mendesai primer dari internet mempunyai kemiripan dalam proses serta hasilnya. Pemilihan salah satu program yang akan digunakan mendesain primer bergantung pada kemudahan mengaksesnya, kemudahan prosedurnya serta kelengkapan informasi yang dapat diberikan oleh program tersebut. Namun demikian analisis yang dilakukan tidak menjamin bahwa primer yang dipilih secara langsung berhasil dengan baik sebab masih perlu dilakukan optimasi.

### **Design Primer spesifik dengan Primer3Plus**

Desain primer pada daerah 16S rDNA digunakan untuk mendeteksi pada kelompok prokariot, salah satunya bakteri yang terdapat dalam produk herbal. Indikasi kontaminan pada produk herbal seperti obat – obatan lebih banyak ditemukan mikroorganisme dari kelompok bakteri dan jamur. Salah satu bakteri yang kontaminan pada produk herbal adalah *Staphylococcus aureus* berdasarkan publikasi jurnal “ *Bacterial profiles and consumer preference of some indigenous orally consumed herbal medications in Nigeria*”.

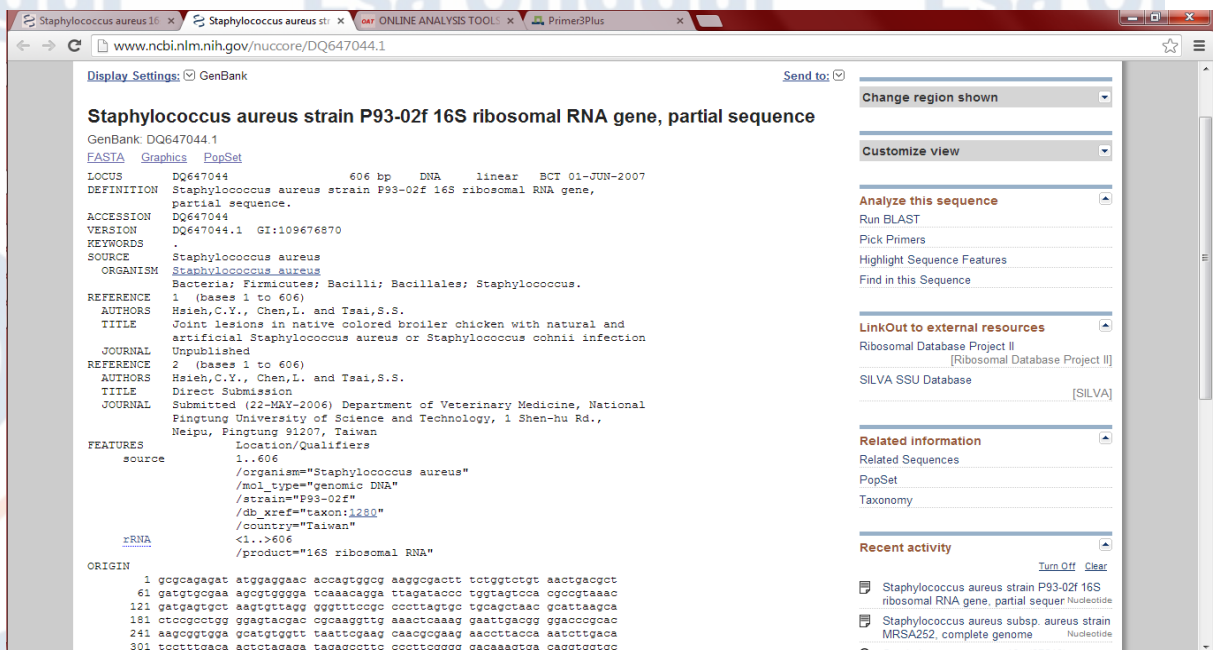
Desain primer yang digunakan dari program dari internet yaitu **Primer3plus** pada laman web **Online Analysis Tools**. Berikut langkah langkah dalam mendesain primer spesifik pada *Staphylococcus aureus* .

1. Mencari sekuen daerah 16S bakteri *Staphylococcus aureus* pada website [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov) pada kotak search dan memilih nukleotida pada kotak database



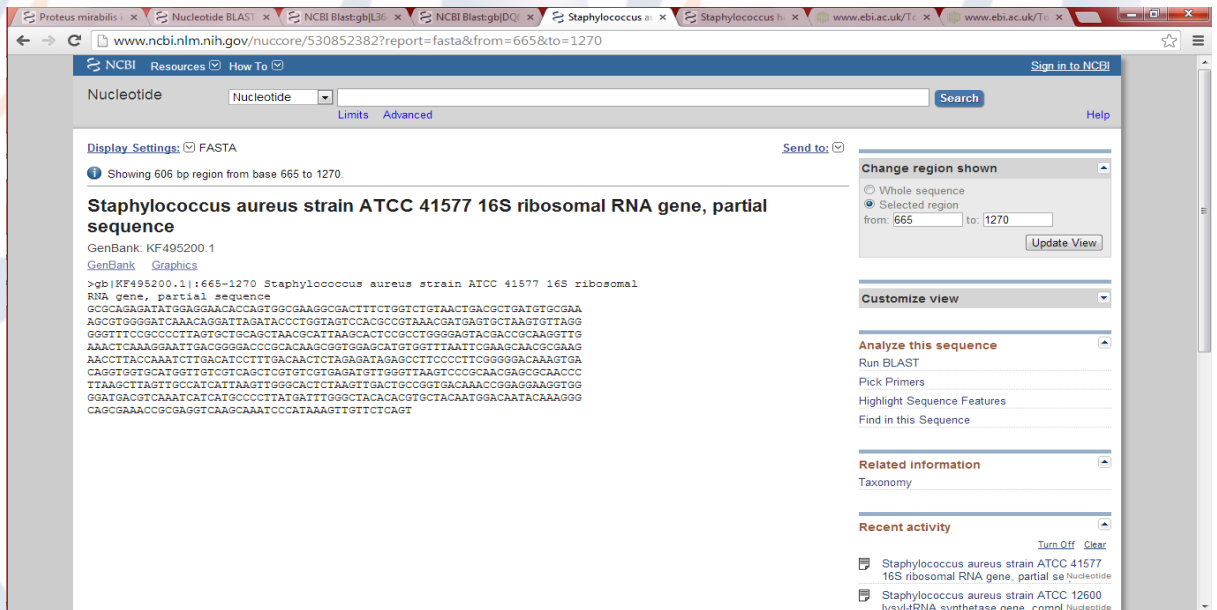
Gambar 1. Website www.ncbi.nlm.nih.gov

2. Kemudian akan keluar beberapa Spesies bakteri *Staphylococcus aureus* dengan strain berbeda yang memiliki kode sekuen 16s RNA gene, lalu dipilih yang memiliki 'partial sequence' yang berarti bahwa daerah yang kita inginkan telah memiliki sekuen yang lengkap, lalu klik sehingga diperoleh informasi untuk strain



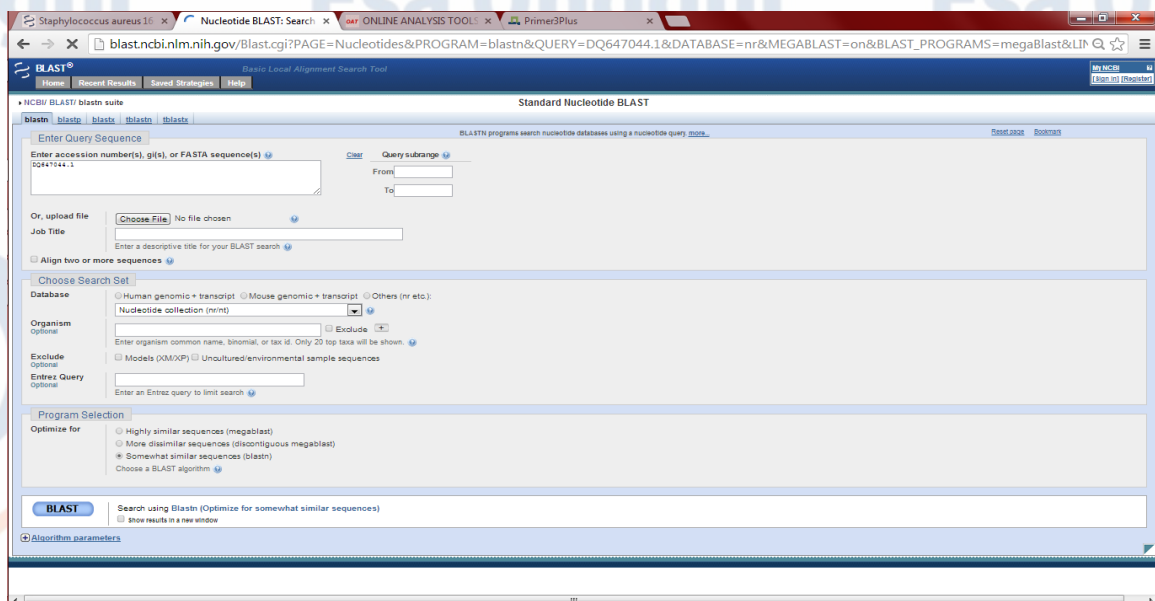
Gambar 2. Partial sequence pada *Staphylococcus aureus*

3. Selanjutnya untuk memperoleh sekuen nukleotida yang dicari klik FASTA dengan tampilan seperti gambar dibawah ini



Gambar 3. Fasta Nukleotida pada GenBank

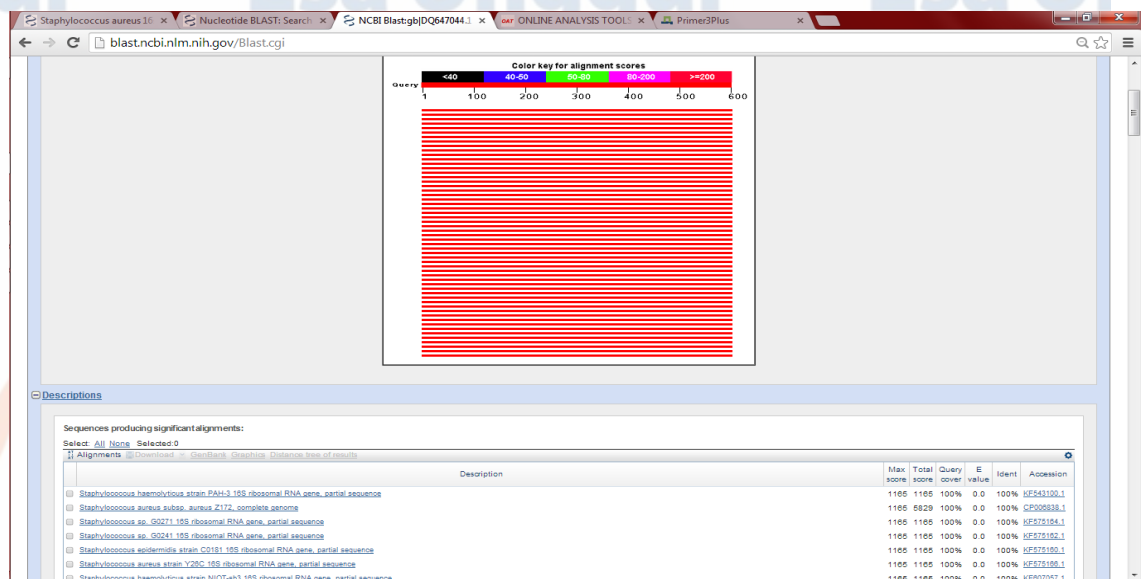
4. Setelah kita mendapatkan sekuen gen yang dicari. Kemudian dicari gen lainnya yang mirip dengan menggunakan program BLAST, Run BLAST spesies yang diambil langsung pada laman tersebut. Hasil terlihat pada gambar dibawah ini.



Gambar 4. Tampilan "nucleotide blast" (bagian bawah).

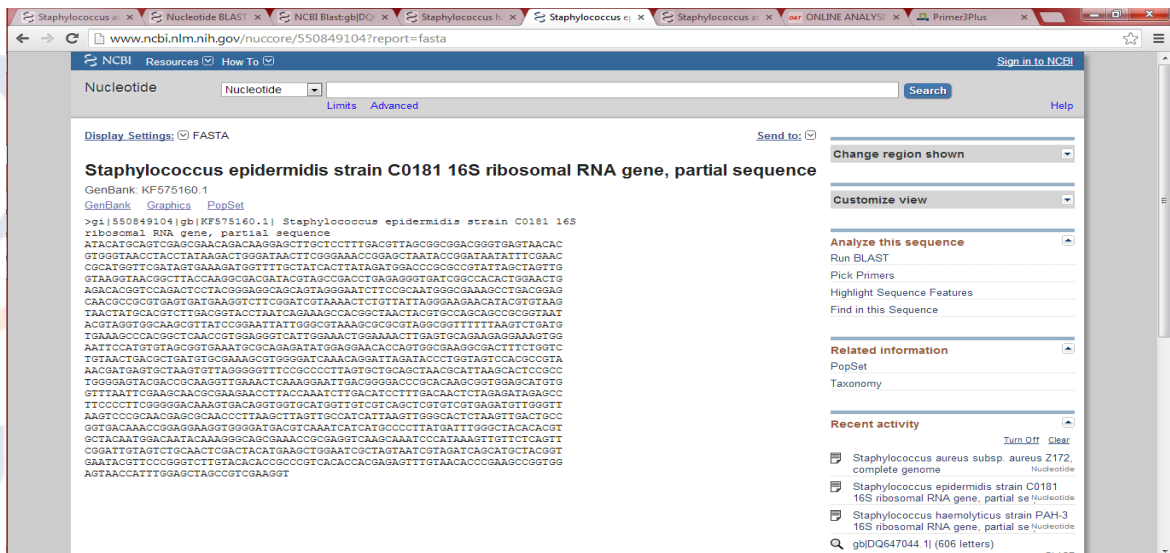


5. Setelah di BLAST akan keluar beberapa sekuen gen yang memiliki kemiripan dengan sekuen yang kita miliki. Dari gambar terlihat garis-garis berwarna merah yang berarti sekuen-sekuen yang ada memiliki skor lebih dari 200. Kemudian dipilih organisme lain yang akan dibandingkan. Dalam memilihnya, dipilih yang berwarna merah dan juga sekuen yang *complete cds*. Untuk mengoleksinya dipilih "Description" sehingga akan keluar "result" yaitu hasil koleksi organisme pilihan kita. Terlihat pada gambar di bawah



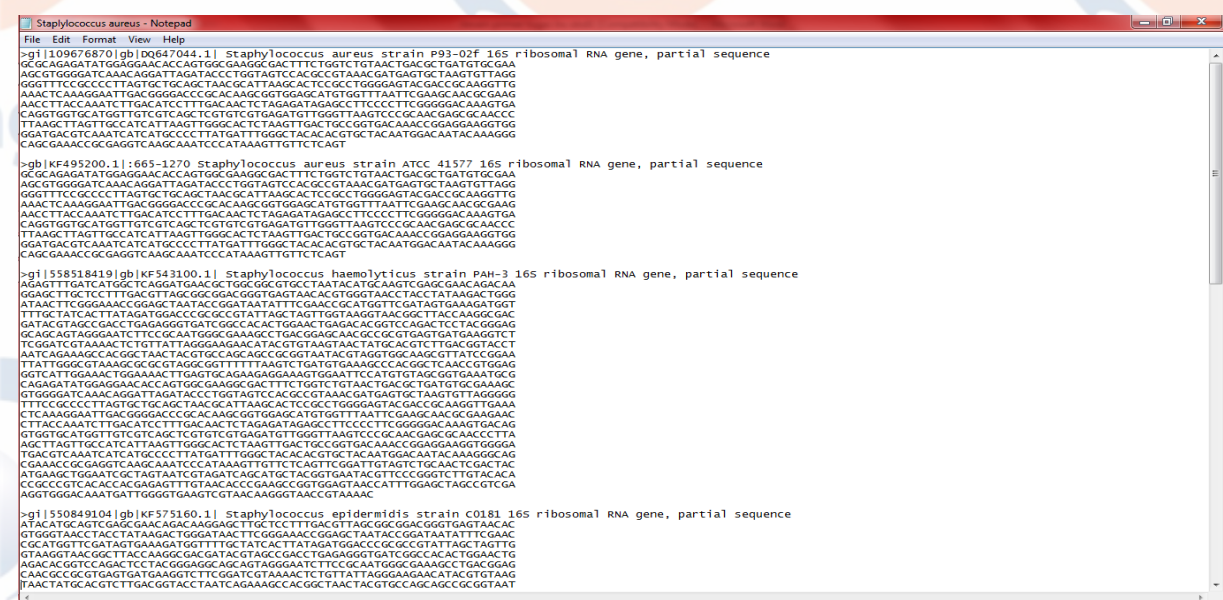
Gambar 5. Hasil Blast- N dengan beberapa sekuen GenBank

6. Setelah mengklik kode organisme lain pada kotak accession, kemudian kita akan memperoleh informasi tentang organisme yang telah dipilih tersebut yang memiliki kemiripan dengan spesies yang kita punya, lalu untuk melihat sekuen nukleotidanya kita klik FASTA. Setelah itu semua sekuen nukleotida organisme yang kita pilih tadi dikopi ke notepad



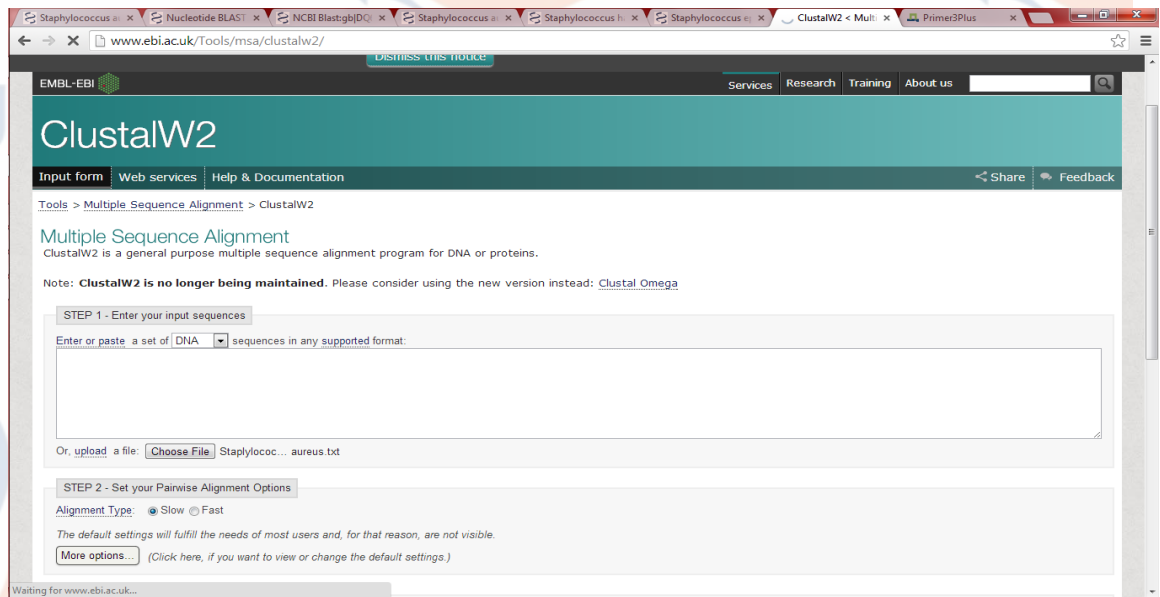
Gambar 6. Sekuen DNA spesies lain dari hasil BLAST

7. Kemudian dipilih “Get Selected Sequences” lalu selanjutnya dipilih “FASTA” dankemudian di “APPLY” sehingga didapat sekuen- sekuen gen dari sejumlah organisme yang kita pilih. Setiap sekuen kita *copy* dan dipindahkan ke *file* notepad yang sebelumnya (Gambar 3) lalu *file* tersebut disimpan. Terlihat pada gambar di bawah



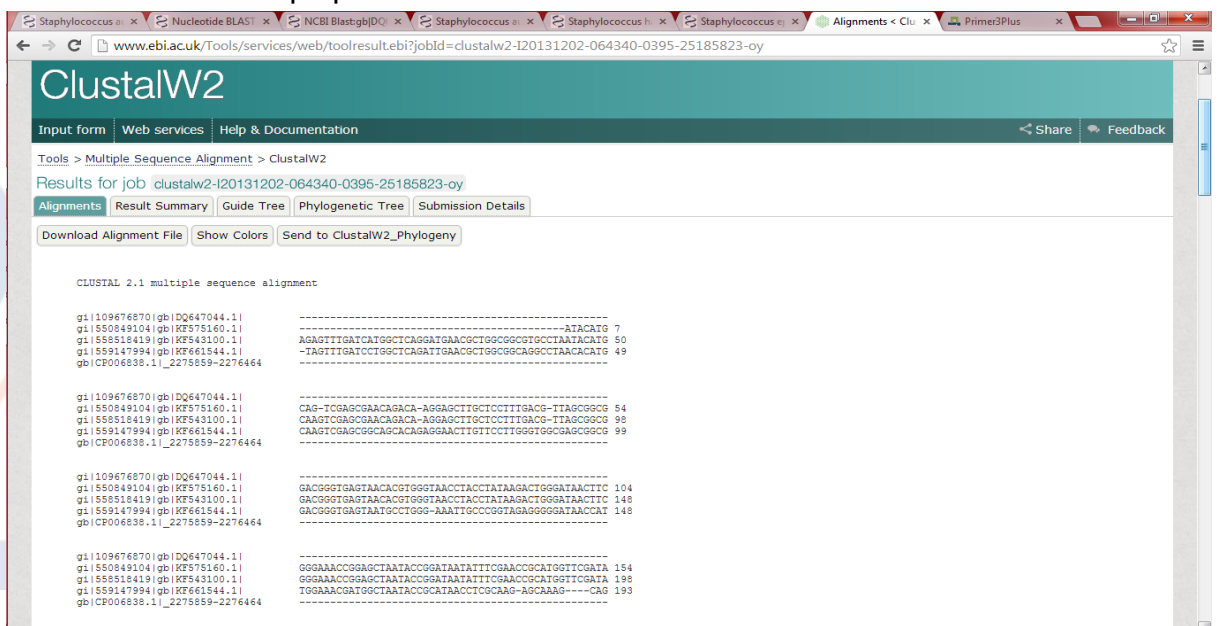
Gambar 7. Sekuen-sekuen gen bakteri 16S hasil BLAST dalam file notepad.

- Setelah itu sekuen nukleotida yang telah terdapat pada file notepad tadi akan digunakan untuk mendesain primer dengan menggunakan program ClustalW multiple alignment pada website [www.ebi.ac.uk/tools/msa/clustalw2](http://www.ebi.ac.uk/tools/msa/clustalw2) untuk melakukan pensejajaran sekuen nukleotida yang kita miliki dengan nukleotida dengan mengupload file notepad tadi dan kemudian diklik submit

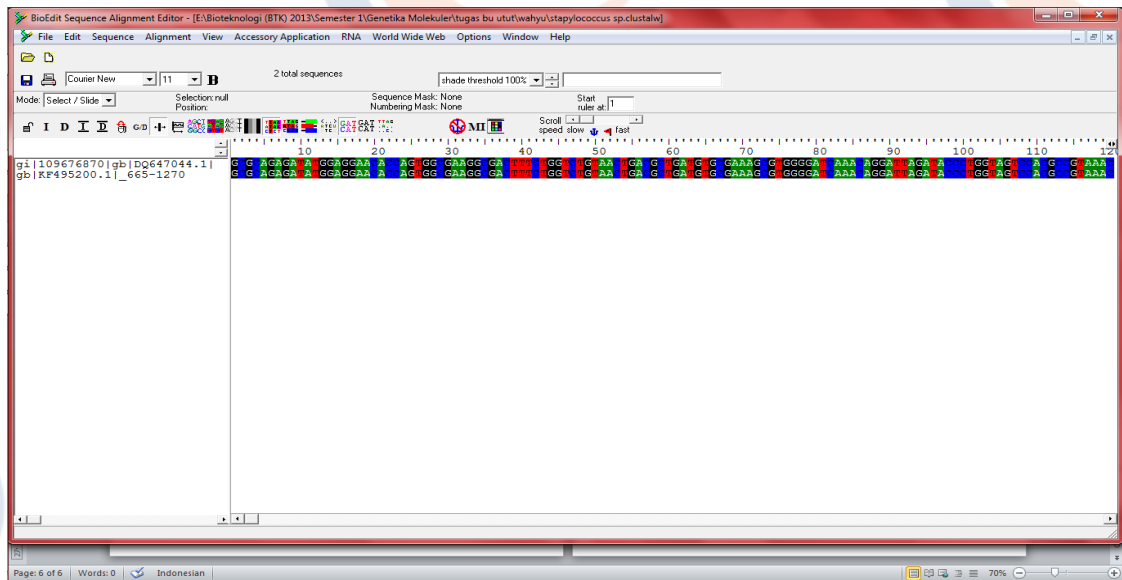


Gambar 8. Alignment beberapa spesies *Staphylococcus aureus*

- Kemudian akan muncul hasil pensejajaran yang memiliki tanda bintang merupakan daerah yang conserve dan digunakan sebagai desain primer yang spesifik terhadap spesies *Staphylococcus aureus* namun variabel terhadap spesies lain

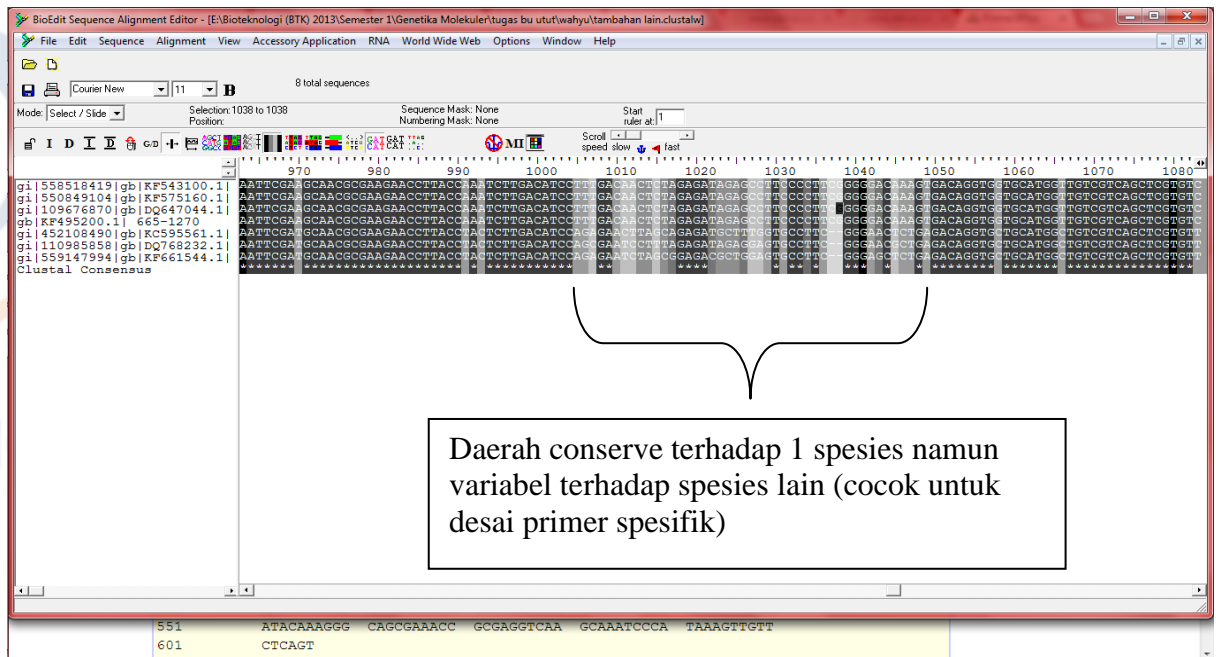
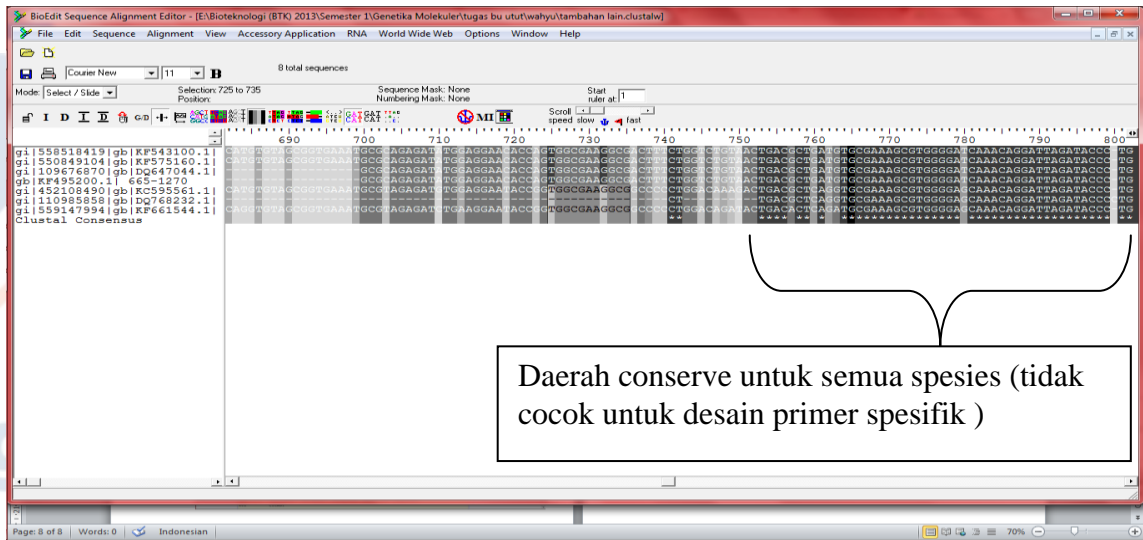


10. Sebelum mengalignment dengan beberapa spesies yang berbeda terlebih dahulu dialignment dengan spesies yang sama antara *Staphylococcus aureus* strain P93-02f dengan *Staphylococcus aureus* strain ATCC 41577 untuk mengetahui daerah conserve yang spesifik terhadap spesies tersebut dengan ditandai daerah yang berwarna.



Gambar 10. Daerah conserve pada *Staphylococcus aureus*

11. Kemudian File di Download dan disimpan dalam program ClustalW *Multiple alignment* yang dapat langsung dibuka dalam BIOEDIT. Kemudian akan muncul sekuen-sekuen yang telah diClustalW. Untuk mempermudah kita untuk mencari sekuen yang *conserve*, sekuen-sekuen gen yang *conserve* bisa ditandai dengan memilih “*Shade Identities and similarities in alignment window*”. Pada daerah conserve yang berwarna hitam. Untuk desain primer diambil daerah yang conserve terhadap satu spesies yang sama tetapi tidak conserve terhadap spesies lain (daerah variabel) pada gambar dibawah ini ditandai daerah yang tidak berwarna hitam.



Gambar 11. Daerah Conserve untuk desain primer spesifik

12. Setelah itu untuk mendesain primer yang spesifik untuk *Staphylococcus aureus* digunakan website [www.primer3plus.com](http://www.primer3plus.com) dengan cara mengupload atau mengkopi sekuen nukleotida *Staphylococcus aureus* yang menjadi target kita ke kotak paste sequence or upload file, pada kotak task kita pilih cloning, dan pada bagian include region, dituliskan posisi nukleotida kita untuk bagian forward dan untuk reversenya dengan menuliskan selisih posisi nukleotida antara posisi reverse dikurang forward. Lalu klik pick primers. Berikut hasil desain primer dengan menggunakan *Primer3plus* pada spesies bakteri *Staphylococcus aureus*

**Primer3Plus**  
pick primers from a DNA sequence

Pair 1:  
 Left Primer 1: gii109676870(gb|DQ647044.1) Staphylococcus  
 Sequence: CTGGTCTGTAAGTACTGACGCTGA  
 Start: 42 Length: 21 bp Tm: 59.1 °C GC: 52.4% ANY: 3.0 SELF: 1.0  
 Right Primer 1: gii109676870(gb|DQ647044.1) Staphylococcus  
 Sequence: GAAGGGGAAGGCTCTATCTCT  
 Start: 336 Length: 21 bp Tm: 58.1 °C GC: 52.4% ANY: 3.0 SELF: 2.0  
 Product Size: 295 bp

Send to Primer3Manager | Reset Form

```

1      GCGCAGAGAT  ATGGAGGAAC  ACCAGTGGCG  AAGCGCACTT  TCTGGTCTGT
51     AACTGACGCT  GATGTGCGAA  AGCGTGGGGA  TCAACACAGGA  TTAGATACCC
101    TGGTAGTCCA  CGCGTA AAC  GATGAGTGTCT  AAGTGTAGG  GGGTTCCGCG
151    CCCTTAGTGC  TGCAGCTAAC  GCATTAAGCA  CTCCGCTGG  GGAGTACGAC
201    GGC AAGGTTG  AAATC A AAG  GAATGACGG  GGACCCGCAC  AAGCGGTGGG
251    GCATGTGGTT  TAATTCGAAG  CAACGGGAAG  AACCTTACCA  AATCTTGACA
301    TCCTTGACA  ACTTAGAGA  TAAGCCCTTC  CCCTTCGGGG  GACAAAGTGA
351    CAGGTGGTGC  ATGGTGTGCG  TCAGCTCGTG  TCGTGAGAGG  TTGGSTTAAG
401    TCCCGCACAG  AGCGCAACCC  TTAAGTGTAG  TTGCCATCAT  TAAGTTGGGG
451    ACTCTAAGTT  GACTGCCGGT  GACAAACCGG  AGGAAGTGG  GGATGACGTC
501    AAATCATCAT  GCCCCTTATG  ATTTGGGCTA  CACAAGTGTCT  ACAATGGACA
551    ATACAAAGGG  CAGCGAAACC  GCGAGGTCAA  GCAAAATCCCA  TAAAGTTGTT
601    CTCAGT
  
```

Untuk desain Primer *forward* terdapat pada urutan basa nukleotida ke 42. Sedangkan untuk desain primer *reverse* diambil pada urutan basa nukleotida ke 336. Nukleotida yang didapat dari sekuen ujung 3 tadi kemudian dicari komplementnya lalu kemudian dibalik. Atau dapat menggunakan program “Reverse Complement” yang hal ini langsung terprogram dari *Primer3plus* beserta dengan nilai Tm dan persentase G+C nya. Hasil desain primer spesifik yang diperoleh untuk *Staphylococcus aureus* adalah untuk Sebagai berikut

	Start	BP	(Tm) <sup>o</sup> C	GC	Self	Nukleotida	Product Size
Forward	42	21	59.1	52.4%	1.00	CTGGTCTGTAAGTACTGACGCTGA	606 bp
Reverse	336	21	58.1	52.4%	2.00	GAAGGGGAAGGCTCTATCTCT	

Hasil desain primer ini dapat digunakan untuk mengidentifikasi bakteri *Staphylococcus aureus* pada suatu produk herbal yang terkontaminasi oleh ada nya bakteri tanpa perlu waktu lama dengan identifikasi secara 16S rDNA. Apabila bakteri yang mengkontaminasi suatu produk herbal kemudian kita isolasi DNA bakteri langsung dari sampel produk herbal tersebut dan di amplifikasi dengan primer spesifik terhadap *Staphylococcus aureus* yang telah di desain, maka apabila hasil amplifikasi DNA bakteri hasil PCR setelah

di visualisasi dengan gel agarose dapat dilihat ukuran pita sekitar 600 bp sehingga dapat diambil kesimpulan bahwa produk tersebut terkontaminasi spesies *Staphylococcus aureus*.

Primer spesifik bagi *Staphylococcus aureus* dapat didisain melalui bioinformatika. Dengan Primernya yaitu:

- *Forward* : 5' CTGGTCTGTAAGTACGCTGA'3
- *Reverse* : 3' GAAGGGGAAGGCTCTATCTCT'5

### III. EVALUASI BELAJAR

#### a. Rangkuman

Desain primer merupakan bagian dari bioinformatika yang menjadi faktor yang paling penting/langkah-langkah menentukan sekuen DNA yang belum diketahui. Berbagai program bioinformatika yang tersedia untuk pemilihan pasangan primer dari urutan template. Program kebanyakan untuk desain primer PCR mencerminkan peran sentral PCR dalam biologi molekuler modern. Desain Primer merupakan salah satu bahan yang digunakan dalam proses perbanyakan untai DNA dengan menggunakan metode Polymerase Chain Reaction. Primer adalah oligonukleotida spesifik yang komplemen dengan daerah yang ditentukan pada DNA target sebagai tempat dimulainya sintesis DNA baru dengan metode PCR

#### b. Latihan

1. Jelaskan Karakteristik Primer yang baik ?
2. Apa yang menjadi pertimbangan yang penting dalam mendesain primer?
3. Jelaskan tentang primer Forward dan primer Reverse !
4. Apa perbedaan degenerated primer dengan primer spesifik ?

#### c. Tugas

Buatlah desain primer dari gen spesifik dan spesies bakteri spesifik

#### Penilaian Tugas

1. Tugas dibuat di blog mahasiswa
2. Blog di link ke web hybrid learning.

3. Blog tersebut harus mencantumkan logo dan nama Universitas Esa Unggul
4. Diselesaikan sebelum batas akhir penyerahan tugas

#### **IV. DAFTAR PUSTAKA**

Claverie, J.M. & C. Notredame. 2003. Bioinformatics For Dummies. 2nd ed. Wiley Publishing, Inc. New York. p10-68. p215-338.

Glick, B.R. & J.J. Pasternak. 2003. Molecular Biology: Principles and Application of Recombinant DNA. ASM Press. Washington, D.C.

Saiki, R.K. 1990. Amplification of genomic DNA. In: Innis, M.A, Gelfand, D. H, Sninsky, J.J White (eds) PCR Protocol: A Guide to Methode and Applications. Academic Press, San Diego, CA. p13-20

Sambrook, J. & D.W. Russel. 2001. Molecular Cloning A Laboratory Manual Vol. 2. 3rd ed. Spring Harbor Laboratory Press. New York. p10.1-10.49



## BAB VII. ANALISIS POHON FILOGENETIK

### I. Pendahuluan

#### a. Pengantar

Filogenetik dikenal sebagai bidang ilmu yang tidak terlepas dengan analisis bioinformatika. Filogenetika menyediakan fasilitas dalam bidang epidemiologi manusia, ekologi, dan evolusi biologi. Ketertarikan peneliti menggunakan analisis filogenetika tidak jarang membuat sedikit membingungkan dikarenakan ketidakmertian dalam menggunakan beberapa metode dalam analisis filogenetik. Analisis filogenetik tidak terlepas dari evolusi biologis. Evolusi adalah proses gradual, suatu organisme yang memungkinkan spesies sederhana menjadi lebih kompleks melalui akumulasi perubahan dari beberapa generasi. Keturunan akan mempunyai beberapa perbedaan dari nenek moyangnya sebab sedang berubah dalam sebuah evolusi. Dalam mempelajari variasi dan diferensiasi genetik antar populasi, jarak genetik dapat dihitung dari jumlah perbedaan basa polimorfik suatu lokus gen masing-masing populasi berdasarkan urutan DNA

#### b. Kompetensi Dasar

Mahasiswa dapat memahami dan melakukan analisis filogenetik dari data sekuens DNA sebagai dasar kekerabatan Organisme.

#### c. Kemampuan Akhir yang Diharapkan

Mahasiswa diharapkan :

- a. Mampu mendefinisikan apa itu filogenetik
- b. Mampu membuat pohon filogenetik dari data sekuens DNA
- c. Mampu menjelaskan dan membaca analisis percabangan pada pohon filogenetik

#### d. Kegiatan Pembelajaran

1. Pembelajaran dilakukan dengan metoda contextual learning *Small grup discussion* dan Presentasi

2. Mahasiswa mencari bahan pustaka, membuat bahan presentasi dan mempresentasikan hasil literasinya

## **II. MATERI**

### **A. Analisis Filogenetika Molekuler**

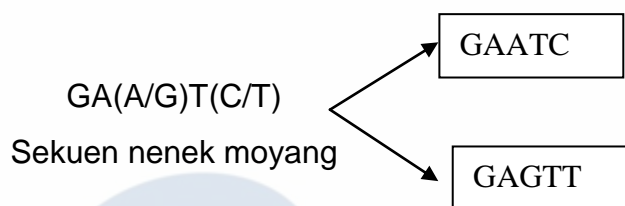
Konstruksi pohon filogenetika adalah hal yang terpenting dan menarik dalam studi evolusi. Terdapat beberapa metode untuk mengkonstruksi pohon filogenetika dari data molekuler (nukleotida atau asam amino) Analisis filogenetika dari keluarga sekuen nukleotida atau asam amino adalah analisis untuk menentukan bagaimana keluarga tersebut diturunkan selama proses evolusi. Hubungan evolusi diantara sekuen digambarkan dengan menempatkan sekuen sebagai cabang luar dari sebuah pohon. Hubungan cabang pada bagian dalam pohon merefleksikan tingkat dimana sekuen yang berbeda saling berhubungan. Dua sekuen yang sangat mirip akan terletak sebagai neighboring outside dari cabang-cabang dan berhubungan dalam cabang umum (Common branch)

Filogenetika digambarkan sebagai klasifikasi secara taksonomi dari organisme berdasarkan pada sejarah evolusi mereka, yaitu filogeni mereka dan merupakan bagian integral dari ilmu pengetahuan yang sistematis dan mempunyai tujuan untuk menentukan filogeni dari organisme berdasarkan pada karakteristik mereka. Lebih lanjut filogenetika adalah pusat dari evolusi biologi seperti penyingkatan keseluruhan paradigma dari bagaimana organisme hidup dan berkembang di alam

Analisis filogenetika sekuen asam amino dan protein biasanya akan menjadi wilayah yang penting dalam analisis sekuen. Selain itu, dalam filogenetika dapat menganalisis perubahan yang terjadi dalam evolusi organisme yang berbeda. Berdasarkan analisis, sekuen yang mempunyai kedekatan dapat diidentifikasi dengan menempati cabang yang bertetangga pada pohon. Ketika keluarga gen ditemukan dalam organisme atau kelompok organisme, hubungan filogenetika diantara gen dapat memprediksikan kemungkinan yang satu mempunyai fungsi yang ekuivalen. Prediksi fungsi ini dapat diuji dengan eksperimen genetik. Analisis filogenetika juga digunakan untuk mengikuti perubahan yang terjadi secara cepat yang mampu mengubah suatu spesies, seperti virus

## B. Hubungan Filogenetik dengan penjejeran (Alignment)

Hubungan analisis filogenetika dengan alignment/penjejeran sekuen ketika sekuen nukleotida atau protein dari dua organisme yang berbeda memiliki kemiripan, maka mereka diduga diturunkan dari sekuen common ancestor. Sekuen penjejeran akan menunjukkan dimana posisi sekuen adalah tidak berubah/conserved dan dimana merupakan divergent/atau berkembang menjadi berbeda dari common ancestor seperti diilustrasikan pada ilustrasi di bawah ini. Sekuen 1 dan 2 diasumsikan berasal dari nenek moyang yang sama (common ancestor). Total terdapat dua sekuen yang berubah.



Studi sekuen biologi selalu tidak dapat dihindarkan dari penjejeran sekuen/alignment. Tujuan dari proses pensejajaran adalah mencocokkan karakterkarakter yang homolog, yaitu karakter yang mempunyai nenek moyang yang sama. Ketika menghomologikan sekuen, kolom dari penjejeran dapat digunakan untuk berbagai macam aplikasi seperti mengidentifikasi residu dengan struktur yang homolog atau yang mempunyai fungsi yang sama.

## C. Kontruksi Pohon Filogenetik

Filogenetik merupakan kajian mengenai hubungan evolusi diantara organisme atau gen dari segi taksonomi, dianalisis dengan menggunakan kombinasi antara biologi, molekuler dan teknik statistik. Dasar klasifikasitersebut adalah persamaan dan perbedaan sifat morfologi, anatomi dan molekuler. Sistem tersebut mencerminkan urutan perkembangan serta jauh dekatnya kekerabatan antartakson, serta menggambarkan persamaan dan perbedaan sifat berupa morfologi dan anatomi. Taksonomi filogenetik merupakan pengelompokan spesies atau jenis baru dengan cara analisis molekuler dan morfologi. Klasifikasi sistem filogenetik disusun berdasarkan persamaan fenotip yang mengacu pada sifat-sifat bentuk luar, faal, tingkah laku yang dapat diamati, dan pewarisan keturunan yang

mengacu pada hubungan evolusioner jenis nenek moyang hingga cabang-cabang keturunannya

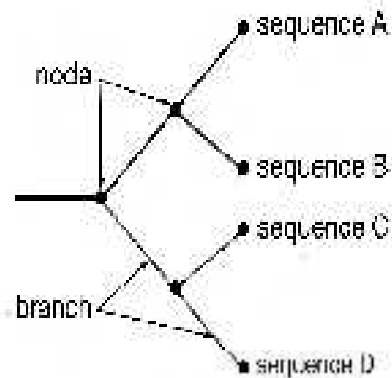
Filogenetik berdasarkan perbandingan genom dengan pemisahan ke dalam hubungan evolusioner (*clades*), kemungkinan akan menggantikan phenotypica (*phenetic*) taksonomi. Pengelompokan organisme terdiri dari phenetic sistem yaitu pengelompokan organisme berdasarkan kesamaan sifat karakteristik fenotipik (fisik dan kimia). Pengelompokan Phenetic, mungkin atau tidak mungkin berkorelasi dengan hubungan evolusi. Filogenetik sistem merupakan pengelompokan organisme berdasarkan pada kesamaan warisan evolusi. Teknik sekuensing DNA dan RNA dianggap memberikan filogeni paling berarti untuk menentukan nenek moyang dan evolusi yang terjadi. Oleh karena itu tujuan dari sistematika ini adalah untuk menciptakan suatu klasifikasi yang mencerminkan sejarah evolusi organisme. Maka, diperlukan pengelompokan spesies ke dalam taksa :

1. Monofiletik yaitu jika nenek moyang tunggalnya menghasilkan semua spesies turunan yang berasal dari takson tersebut, bukan spesies pada takson lain.
2. Polifiletik yaitu jika anggotanya diturunkan dari dua atau lebih bentuk nenek moyang yang tidak sama bagi semua anggotanya.
3. Parafiletik yaitu jika takson itu tidak meliputi spesies yang memiliki nenek moyang yang sama yang menurunkan spesies yang termasuk dalam takson tersebut

#### **D. Konsep Pohon Evolusi**

Pohon evolusi adalah sebuah grafik dua dimensi yang menunjukkan hubungan diantara organisme atau lebih spesifik lagi adalah sekuen gen dari organisme. Pemisahan sekuen disebut taxa (atau taxon jika tunggal) yang didefinisikan sebagai jarak filogenetika unit pada sebuah pohon. Pohon terdiri dari cabangcabang luar (*outer branches*) atau daun-daun (*leaves*) yang merepresentasikan taxa dan titik-titik (*nodes*) dan cabang merepresentasikan hubungan diantara taxa, yang diilustrasikan sebagai A-D pada Gambar 2.

### A. Rooted tree



### B. Unrooted tree



Figure 6.5. Structure of evolutionary trees.

### Gambar 2. Struktur pohon evolusi

Oleh karena itu, sekuen A dan B dipisahkan dari sekuen common ancestor yang direpresentasikan dengan titik-titik di bawahnya; C dan D adalah mempunyai kemiripan. Pada Gambar 2 menunjukkan bahwa sekuen A/B dan C/D memiliki common ancestor yang sama yang ditunjukkan dengan sebuah titik pada bagian paling rendah dari pohon. Hal ini sangat penting untuk mengenali bahwa masing-masing titik dalam pohon direpresentasikan sebuah pemisahan garis evolusi gen ke dalam dua spesies yang berbeda. Panjang masing-masing cabang pada titik berikutnya menunjukkan jumlah sekuen yang berubah yang terjadi sebelum level pemisahannya. Contohnya, panjang cabang antara titik A/B dan B menunjukkan spesies mempunyai rata-rata evolusi yang sama

Dalam mengkonstruksi pohon filogenetika dapat diklasifikasikan menjadi 2 kategori yang digunakan sebagai strategi untuk menghasilkan pohon filogenetika terbaik. Kategori pertama adalah memeriksa semua atau sejumlah besar kemungkinan pohon filogenetika dan memilih satu yang terbaik dengan kriteria-kriteria tertentu. Biasanya disebut dengan metode exhaustivesearch. Metode maximum parsimony, Fitch Margoliash dan maximum likelihood termasuk dalam kategori ini. Kategori yang kedua adalah

memeriksa hubungan topologi lokal dari pohon dan mengkonstruksi pohon terbaik dengan langkah demi langkah. Metode *Neighbor-joining* dan beberapa metode Distance lainnya adalah termasuk dalam kategori yang kedua ini. Dalam makalah ini akan dibahas tiga metode saja yaitu: (1) *Maximum parsimony*, (2) *Distance* dan (3) *Maximum likelihood* yang secara umum digunakan untuk membentuk pohon evolusi atau pohon terbaik untuk mengamati variasi sekuen dalam kelompok. Masing-masing metode ini digunakan untuk tipe analisis yang berbeda.

### **Metode *maximum parsimony***

*Parsimony* atau metode minimum *evolution* pertama kali digunakan dalam filogenetik oleh Camin and Sokal pada tahun 1965 (FELSENSTEIN, 1978). Metode ini memprediksikan pohon evolusi/ *evolutionary tree* yang meminimalkan jumlah langkah yang dibutuhkan untuk menghasilkan variasi yang diamati dalam sekuen. Untuk alasan ini, metode ini juga sering disebut sebagai metode evolusi minimum/*minimum evolution method*. Sebuah *multiple sequence alignment* dibutuhkan untuk memprediksi posisi sekuen yang sepertinya berhubungan. Posisi ini akan menampilkan kolom vertikal dalam *multiple sequence alignment*. Untuk masing-masing posisi yang disejajarkan, pohon filogenetika membutuhkan perubahan evolusi dalam jumlah terkecil untuk menghasilkan pengamatan perubahan sekuen yang diidentifikasi. Analisis ini terus menerus dilakukan terhadap masing-masing posisi dalam penjejeran sekuen. Akhirnya, pohon yang menghasilkan jumlah perubahan terkecil secara keseluruhan dihasilkan untuk semua posisi sekuen yang diidentifikasi. Metode ini berguna untuk sekuen yang mirip dan dalam jumlah yang sedikit. Algoritma yang digunakan tidak rumit tetapi dijamin untuk dapat menemukan pohon yang terbaik, sebab semua kemungkinan pohon yang dibentuk berhubungan dengan kelompok sekuen yang diperiksa. Untuk alasan ini, metode ini cukup membutuhkan banyak waktu dan tidak berguna untuk data sekuen dalam jumlah besar dan asumsi lain harus dibuat untuk root pohon yang diprediksikan.

## **Metode jarak/distance method**

Metode jarak bekerja pada jumlah perubahan diantara masing-masing pasangan dalam kelompok untuk mengkonstruksi pohon filogenetika dalam kelompok. Pasangan sekuen yang mempunyai jumlah perubahan terkecil diantara mereka disebut neighbors. Pada pohon, sekuen-sekuen ini menggunakan secara bersama-sama satu titik atau posisi common ancestor dan masing-masing dihubungkan titik oleh sebuah cabang. Tujuan dari metode jarak adalah metode untuk mengidentifikasi pohon pada posisi neighbors dengan benar, dan juga mempunyai cabang yang menghasilkan data orisinil sedekat mungkin. Penemuan neighbors terdekat diantara kelompok sekuen dengan metode jarak biasanya langkah pertama dalam memproduksi sebuah multiple sequence alignment. Metode jarak pertama kali ditemukan oleh Feng dan Doolittle; pengelompokan program oleh penulis tersebut menghasilkan sebuah penjejeran dan pohon dari set sekuen protein (FENG dan DOOLITTLE, 1996).

Program CLUSTALW, digunakan untuk neighborjoining distance method sebagai panduan untuk multiple sequence alignment. Program PAUP versi 4 merupakan pilihan untuk membentuk sebuah analisis filogenetika dengan distance method. Program PHYLIP package yang membentuk analisis distance termasuk program yang secara otomatis dibaca dalam sekuen dalam PHYLIP infile format dan secara otomatis menghasilkan file yang disebut dengan tabel distance. Dalam pengukuran jarak genetik menggunakan model substitusi nukleotida, suatu sekuen DNA akan dibandingkan satu nukleotida dengan nukleotida lainnya. Jarak ini dapat mengukur suatu sekuen nukleotida baik yang menyandi protein maupun tidak. Pada jarak matrik (distance matrices) yang dihasilkan, mereka mungkin digunakan sebagai input yang mengikuti program analisis jarak dalam PHYLIP. Program PHYLIP semua secara otomatis membaca input file yang disebut infile dan menghasilkan sebuah outfile. Jadi, nama file harus diedit ketika menggunakan program ini. Sebagai contoh, distance outfile harus diedit untuk memasukkan hanya tabel distance dan jumlah taxa, dan ketika file disimpan dengan nama sekuen infile. Analisis distance dalam program PHYLIP adalah sebagai berikut:

1. FITCH mengestimasi sebuah pohon filogenetika yang mengasumsikan penambahan panjang cabang menggunakan metode Fitch-Margoliash dan tidak mengasumsikan sebuah molecular clock (mengikuti rata-rata evolusi sepanjang cabang yang bervariasi).
2. KITCH mengestimasi sebuah pohon filogenetika tetapi dengan mengasumsikan molecular clock.
3. NEIGHBOR mengestimasi pohon filogenetika menggunakan neighbor joining atau metode unweighted pair group dengan rata-rata aritmatika (UPGMA). Metode neighbor joining tidak mengasumsikan molecular clock dan menghasilkan unrooted tree.

Metode UPGMA mengasumsikan sebuah molecular clock dan rooted tree. Metode ini secara normal menghitung skor similaritas yang didefinisikan sebagai jumlah total dari jumlah sekuen yang identik dan jumlah substitusi konservatif dalam penjejeran dua sekuen dengan gap yang diabaikan. Skor identitas antara sekuen menunjukkan hanya identitas yang mungkin ditemukan dalam penjejeran. Untuk analisis filogenetik digunakan skor jarak antara dua sekuen. Skor diantara dua sekuen adalah jumlah posisi yang tidak cocok/mismatch dalam penjejeran atau jumlah posisi sekuen yang harus diubah untuk menghasilkan sekuen yang lain. Gap mungkin diabaikan dalam kalkulasi atau diberi perlakuan seperti substitusi. Ketika sebuah skoring atau matrik substitusi digunakan, kalkulasi menjadi lebih kompleks tetapi secara prinsip tetap sama.

### **Metode neighbor - joining (NJ)**

Metode neighbor-joining sangat mirip dengan metode Fitch dan Margoliash kecuali tentang pemilihan sekuen untuk berpasangan ditentukan oleh perbedaan algoritma. Metode neighbor-joining sangat cocok ketika rata-rata evolusi dari pemisahan lineage adalah di bawah pertimbangan yang berbeda-beda. Ketika panjang cabang dari pohon yang diketahui topologinya berubah dengan cara menstimulasi tingkat yang bervariasi dari perubahan evolusi, metode neighbor-joining adalah yang paling cocok untuk memprediksi pohon dengan benar. Neighbor-joining memilih sekuen yang jika digabungkan akan memberikan estimasi terbaik dari panjang cabang yang paling dekat merefleksikan jarak yang nyata diantara sekuen. Pada Gambar 4



menunjukkan pohon filogenetika yang dikonstruksi dengan metode Neighbor-joining.

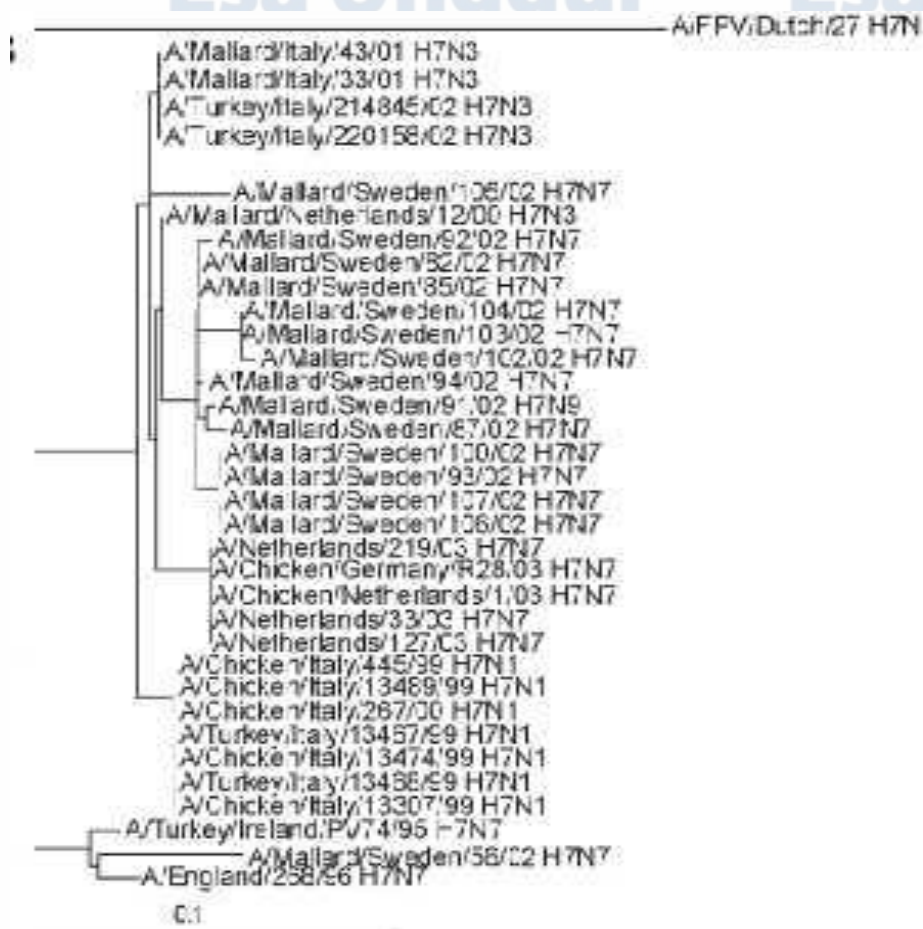
### **Metode unweighted pair group dengan rata-rata aritmetika (UPGMA)**

Metode jarak yang telah diuraikan di atas memberikan sebuah estimasi yang baik dari sebuah pohon evolusi dan tidak terpengaruh oleh variasi dalam rata-rata perubahan sepanjang cabang dari pohon. Metode UPGMA adalah metode sederhana untuk konstruksi pohon yang mengasumsikan rata-rata perubahan sepanjang pohon adalah konstan dan jaraknya kira-kira ultrametric (ultrametric biasanya diekspresikan sebagai molecular clock tree). Metode UPGMA dimulai dengan kalkulasi panjang cabang diantara sekuen paling dekat yang saling berhubungan, kemudian rata-rata jarak antara sekuen ini atau kelompok sekuen dan sekuen berikutnya atau kelompok sekuen dan berlanjut sampai semua sekuen yang termasuk dalam pohon. Akhirnya metode ini memprediksi posisi root dari pohon.

### **Pemilihan Outgroup**

Jika kita ingin secara independen mendapatkan informasi yang meyakinkan dari sekuen lebih berhubungan, sebuah prosedur dapat diikuti dengan menambahkan sekuen pada pohon dan yang paling dekat dengan root. Modifikasi dapat meningkatkan prediksi dari pohon dengan metode di atas yaitu dengan menambahkan outgroup pada langkah akhir dari prosedur. Satu atau lebih sekuen jenis ini disebut sebagai outgroup. Sebagai contoh, sekuen A dan B berasal dari spesies yang telah diketahui terpisah satu dengan yang lain pada awal evolusi berdasarkan catatan fosil. A dan B kemudian diperlakukan sebagai outgroup. Pemilihan satu atau lebih outgroup dengan distance method dapat juga membantu dengan lokalisasi root dari pohon. Root akan ditempatkan diantara outgroup dan titik yang menghubungkan sekuen. Sekuen dari outgroup semestinya berkorelasi dekat dengan sekuen-sekuen yang dianalisa, tetapi juga mempunyai perbedaan yang signifikan antara outgroup dengan sekuen yang lain daripada diantara sekuen itu sendiri. Pemilihan sekuen outgroup yang terlalu jauh kemungkinan akan berperan terhadap prediksi pohon menjadi salah akibat terdapat perbedaan yang secara random yang lebih banyak diantara sekuen outgroup

dengan sekuen lainnya. Perubahan multiple sequence pada masing-masing situs menjadi lebih mungkin dan akan lebih kompleks untuk genetic rearrangements yang kompleks. Untuk alasan yang sama, menggunakan sekuen yang terlalu berbeda dalam metode jarak dari prediksi filogenetik dapat berperan terhadap kesalahan yang terjadi (SWOFFORD et al., 1996). Jumlah perbedaan yang meningkat, perubahan histori sekuen pada masing-masing situs menjadi lebih kompleks dan menjadi sulit untuk memprediksi.



Gambar 3. Pohon filogenetika DNA Maximum Likelihood sekuen hemagglutinin H7 virus Avian Influenza H7 low pathogenic

### Analisis Bootstrap

Analisis filogenetika set sekuen yang menjejerkan dengan baik adalah jelas sebab posisi yang bertanggung jawab dalam sekuen dapat diidentifikasi dalam multiple sequence alignment dari sekuen. Tipe perubahan dalam penjejeran posisi atau jumlah yang berubah dalam penjejeran antara

pasangan sekuen menyediakan dasar untuk menentukan hubungan filogenetika antara sekuen berdasarkan metode analisis filogenetika. Penentuan perubahan sekuen yang telah terjadi menjadi sulit sebab multiple sequence alignment mungkin tidak optimal dan sebab perubahan yang banyak terjadi pada penjejeran posisi sekuen. Pilihan metode multiple sequence alignment tergantung pada tingkat variasi diantara sekuen. Jika penjejeran yang cocok telah ditemukan, pertanyaannya adalah bagaimana prediksi filogenetika didukung oleh data dalam multiple sequence alignment.

Dalam metode bootstrap, data dilakukan resampled, dengan secara random memilih kolom vertikal dari sekuen yang dijejerkan untuk menghasilkan penjejeran, dan dalam pengaruh sebuah penjejeran baru dengan panjang yang sama. Masingmasing kolom digunakan lebih dari satu kali dan beberapa kolom mungkin tidak digunakan pada semua penjejeran yang baru. Pohon-pohon kemudian diprediksi dari beberapa penjejeran ini dari resampled sekuen. Untuk cabang-cabang dalam topologi filogenetika yang diprediksi menjadi signifikan jika set data resampled seharusnya berulang kali (sebagai contoh  $> 70\%$ ) memprediksi cabang-cabang yang sama.

Analisis bootstrap adalah metode yang menguji seberapa baik set data model. Sebagai contoh validitas penyusunan cabang dalam prediksi pohon filogenetik dapat diuji dengan resampled dari kolom dalam multiple sequence alignment untuk membentuk beberapa penjejeran baru. Penampakan cabang dalam pohon dari sekuen resampled ini dapat diukur. Alternatifnya, sekuen kemungkinan harus dikeluarkan dari analisis untuk menentukan berapa banyak sekuen yang mempengaruhi hasil dari analisis. Bootstrap analysis didukung oleh sebagian besar paket software menguji cabang-cabang yang dapat dipercaya.

Jarak genetik berdasarkan metode algoritma pembentukan pohon akan menampilkan data berupa pohon filogenetika. Pohon filogenetika memberi informasi tentang pengklasifikasian populasi berdasarkan hubungan evolusionernya. Dalam rekonstruksi pohon filogenetika, data molekul lebih banyak dipakai karena dianggap lebih stabil dalam proses evolusi dibandingkan dengan data morfologi. Pohon filogenetika dapat berakar (rooted) atau tidak berakar (unrooted), tergantung metode analisis yang

dipergunakan. Akar pada pohon menggambarkan titik percabangan pertama atau asal masing-masing populasi dengan asumsi bahwa laju evolusi berjalan konstan. Pola percabangan pohon dibentuk berdasarkan jarak matrik antar pasangan populasi yang dapat menggambarkan fusi genetik yang terjadi pada kelompok tersebut. Panjang cabang menggambarkan jumlah substitusi basa yang dapat berupa polimorfisme DNA atau haplotipe. Metode pengolahan data yang digunakan harus sesuai dengan set data yang ada, agar dapat menghasilkan pola percabangan (topologi) serta panjang cabang yang benar.

### **Konstruksi Pohon Filogenetik dengan MEGA 6**

program MEGA, (Molecular Evolutionary Genetics Analysis) merupakan program untuk membantu menganalisis gen, DNA, kromosom, ataupun protein pada berbagai organisme, sehingga dapat disusun kesamaan dan perbedaan materi tersebut pada organisme, serta kedekatan kekerabatan antar spesies dengan penyusunan pohon phylogenetik. Penggunaan Gen Bank NCBI dan Program MEGA akan sangat membantu dalam pembuatan karya ilmiah seperti skripsi dan lainnya.

Tujuan dari perangkat lunak MEGA adalah untuk memberikan alat untuk mengeksplorasi, menemukan, dan menganalisis urutan DNA dan protein dari evolusi perspektif. Versi pertama dikembangkan untuk komputasi terbatas sumber daya yang tersedia pada komputer pribadi rata-rata di awal 1990-an. MEGA juga termasuk penjelajah matriks dan filogenik jarak dan sebagai modul grafis canggih untuk representasi visual data masukan dan output hasil. Fitur-fitur ini, yang kita bahas dibawah ini, selain mengatur MEGA lain urutan program analisis komparatif. Seperti versi sebelumnya, MEGA dirancang khusus untuk mengurangi waktu dibutuhkan untuk tugas-tugas biasa dalam analisis data dan untuk menyediakan metode statistik analisis genetik molekuler evolusi dalam komputasi mudah digunakan meja kerja.

Pohon filogenetik dibuat untuk memvisualisasikan hubungan evolusi diantara berbagai spesies atau benda-benda lain. Pohon filogenetik yang berupa diagram bercabang-cabang ini dapat dikonstruksi berdasarkan kesamaan atau perbedaan sifat fisik atau genetik seperti sekuen DNA,

sekuen asam amino (protein), pola pemotongan enzim restriksi, ukuran allele pada analisa microsatellite, dan lain-lain. Dalam artikel ini kita akan membuat pohon filogenetik berdasarkan sekuen DNA gen 16S ribosomal RNA dari bakteri-bakteri.

### **III. EVALUASI BELAJAR**

#### **a. Rangkuman**

Konstruksi pohon filogenetika adalah hal yang terpenting dan menarik dalam studi evolusi. Terdapat beberapa metode untuk mengkonstruksi pohon filogenetika dari data molekuler (nukleotida atau asam amino) Analisis filogenetika dari keluarga sekuen nukleotida atau asam amino adalah analisis untuk menentukan bagaimana keluarga tersebut diturunkan selama proses evolusi. Hubungan evolusi diantara sekuen digambarkan dengan menempatkan sekuen sebagai cabang luar dari sebuah pohon. Hubungan cabang pada bagian dalam pohon merefleksikan tingkat dimana sekuen yang berbeda saling berhubungan. Dua sekuen yang sangat mirip akan terletak sebagai neighboring outside dari cabang-cabang dan berhubungan dalam cabang umum (Common branch)

#### **b. Latihan**

1. Jelaskan tentang Filogenetik secara sistematis?
2. Tuliskan Metode yang terdapat dalam analisis filogenetik
3. Apa yang dimaksud dengan Nilai Bootstrap dan UPGMA?
4. Bagaimana sistem kekerabatan yang terbentuk dalam satu percabangan

#### **c. Tugas**

Buatlah konstruksi pohon filogenetik dari bakteri kelompok actinobacteria dan analisis nilai bootstrap yang terbentuk (sertakan bakteri dari kelompok lain sebagai outgroup).

#### **Penilaian Tugas**

1. Tugas dibuat di blog mahasiswa
2. Blog di link ke web hybrid learning.

3. Blog tersebut harus mencantumkan logo dan nama Universitas Esa Unggul
4. Diselesaikan sebelum batas akhir penyerahan tugas

#### **IV. DAFTAR PUSTAKA**

Claverie, J.M. & C. Notredame. 2003. Bioinformatics For Dummies. 2nd ed. Wiley Publishing, Inc. New York. p10-68. p215-338.

Glick, B.R. & J.J. Pasternak. 2003. Molecular Biology: Principles and Application of Recombinant DNA. ASM Press. Washington, D.C.

Saiki, R.K. 1990. Amplification of genomic DNA. In: Innis, M.A, Gelfand, D. H, Sninsky, J.J White (eds) PCR Protocol: A Guide to Methode and Applications. Academic Press, San Diego, CA. p13-20

Sambrook, J. & D.W. Russel. 2001. Molecular Cloning A Laboratory Manual Vol. 2. 3rd ed. Spring Harbor Laboratory Press. New York. p10.1-10.49