

**PENGENALAN BIOINFORMATIKA:**  
***SEQUENCE ALIGNMENT***

Oleh:  
**Ika Rahayu, S.Si., M.Sc.**

**DEPARTEMEN BIODIKIMIA**  
**FAKULTAS KEDOKTERAN DAN ILMU KESEHATAN**  
**UNIVERSITAS KRISTEN KRIDA WACANA**  
**JANUARI**  
**2023**

## DAFTAR ISI

	halaman
Halaman judul.....	i
A Pendahuluan.....	1
B Sistem tiga domain.....	1
C Bangunan Dasar Kehidupan.....	4
D Ekspresi Gen.....	9
E <i>Polimerization Chain Reaction (PCR)</i> .....	11
F Sequencing.....	15
G Data Base .....	15
H <i>Sequence Alignment</i> .....	16
Daftar Pustaka.....	25

## **A. Pendahuluan**

Bioinformatika (Bio-informatika) merupakan suatu sistem manajemen informasi untuk biologi molekuler dan memiliki banyak aplikasi praktis. Konsep bioinformatika adalah adanya penerapan ilmu biologi dalam istilah “molekul” berkaitan dengan arti kimia fisika, dan juga penerapan "teknik informatika" yang berasal dari disiplin ilmu seperti matematika terapan, ilmu komputer dan statistik. Hal ini bertujuan untuk memahami dan mengatur informasi yang terkait dengan molekul-molekul dalam skala yang luas.

Ilmu bioinformatika dikembangkan dengan tujuan untuk mengatur data yang mempermudah peneliti dalam mengakses informasi yang ada dan mengirimkan entri baru saat diproduksi, misalnya struktur 3D protein di Protein Data Bank. Lebih dari itu, data yang sudah disimpan tidak akan bermanfaat jika tidak dilakukan analisis, maka bioinformatika berperan lebih dari sekedar penyimpanan *data base*. Oleh karena itu, bioinformatika juga membantu mengembangkan alat dan sumber daya yang dapat digunakan untuk analisis data. Analisis dapat dilakukan setelah mengurutkan protein tertentu, misalnya membandingkan sekuens protein tertentu dengan sekuens yang telah dikarakterisasi sebelumnya. Setelah alat dikembangkan, maka bioinformatika akan membantu menggunakan alat untuk melakukan analisis data sehingga data dapat diinterpretasikan hasilnya dan bermakna secara biologi.

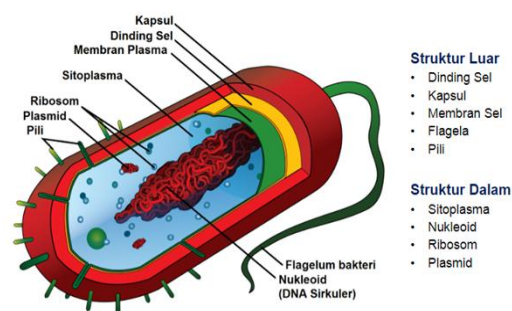
## **B. Sistem tiga domain**

Sebelum masuk ke ilmu bioinformatika lebih dalam, ada baiknya kita mengenal sistem tiga domain kehidupan, dimana hal ini terkait erat dengan perkembangan ilmu bioinformatika. Sistem tiga domain merupakan suatu sistem yang membagi kehidupan seluler menjadi domain bakteri, arkhea dan eukarya. Sebelum muncul sistem tiga domain ini, dahulu ada yang disebut sebagai sistem 5 Kingdom, yaitu Animalia, Plantae, Fungi, Protista, dan Monera. Klasifikasi 5 Kingdom ini merupakan suatu cara pengelompokan dan pengkategorian yang didasarkan pada ciri-ciri tertentu yang nampak pada organisme tersebut. Teori semakin berkembang dan sistem tiga domain kehidupan dicetuskan oleh Carl Woese pada 1990, yaitu bakteri, Eukariota, dan arkhea. Pada sistem ini, domain prokariota dipisahkan menjadi dua kelompok, yaitu *Eubacteria* (sekarang disebut *Bacteria*) dan *Archaeobacteria* (sekarang disebut *Archaea*). Woese menyatakan bahwa pengelompokan itu dibedakan atas dasar perbedaan gen 16S rRNA. Hal ini

menyebabkan dua kelompok tersebut dan eukariota muncul secara terpisah dari nenek moyang dengan genetika yang tidak berkembang dengan baik. Sistem tiga domain ini muncul oleh karena perkembangan ilmu bioinformatika.

## 1. Bakteri

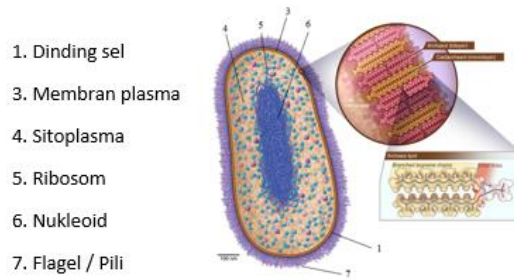
Bakteri merupakan suatu organisme yang biasanya bersel tunggal. Mayoritas bakteri memiliki dinding sel dengan peptidoglikan. Membran sel bakteri terbuat dari rantai asam lemak tidak bercabang yang bergabung dengan gliserol melalui ikatan ester. Sebagian besar bakteri tidak memiliki inti yang terikat pada membran. Bakteri terdapat di manapun tempat, dan beberapa bakteri dapat hidup di lingkungan yang ekstrim. Beberapa bakteri juga dapat melakukan proses fotosintesis (yaitu menghasilkan oksigen), misalnya *Cyanobacteria*.



Gambar 1. Struktur Bakteri

## 2. Arkhea

Arkhea merupakan satu domain atau kerajaan organisme bersel satu. Arkhea memiliki campuran ciri molekuler dari Bakteri dan Eukariota. Akan tetapi Arkhea dapat dibedakan dari bakteri dengan urutan rRNA yang unik. Sifat unik ini dapat memisahkan archaea dari dua domain kehidupan yang lain. Arkhea memiliki lipid membran yang unik - isoprenoid yang melekat pada gliserol melalui ikatan eter. Beberapa memiliki karakteristik metabolik yang tidak biasa dan banyak archaea yang hidup di lingkungan yang ekstrim.



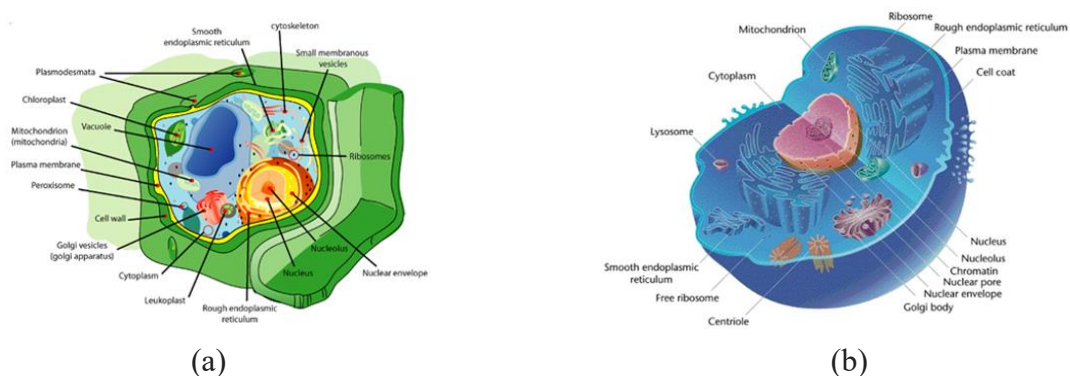
Gambar 2. Struktur Arkhea

### 3. Eukariota

Eukariota merupakan suatu domain yang anggotanya merupakan organisme uniseluler dan multiseluler yang memiliki membran inti (eukariotik). Merupakan domain yang memiliki anggota sangat banyak dimana anggota dari domain ini adalah kingdom protista, fungi, plantae, dan animalia. Berikut ini merupakan ciri-ciri domain Eukariota, yaitu:

- Memiliki dua tipe sel: Uniseluler (alga, protozoa, dan ragi) dan Multiseluler (jamur, tumbuhan, hewan)
- Beberapa (alga dan tumbuhan) melakukan fotosintesis
- Memiliki inti yang terikat membran
- Membran sel terbuat dari rantai asam lemak tidak bercabang yang bergabung dengan gliserol melalui ikatan ester
- Sel yang sangat terkompartemen
- Sel ini umumnya lebih besar dari Bakteri atau Archaea

Berikut ini merupakan stereotip sel tumbuhan dan hewan Eukariotik:



Gambar 3. Sel tumbuhan (a); Sel ukariotik (b)

## **C. Bangunan Dasar Kehidupan**

Terdapat empat bangunan dasar dalam kehidupan yang berperan penting dalam perkembangan kehidupan organisme dan memberikan ciri kehidupan tersebut. Empat bangunan dasar tersebut adalah asam nukleat, protein, karbohidrat dan lemak.

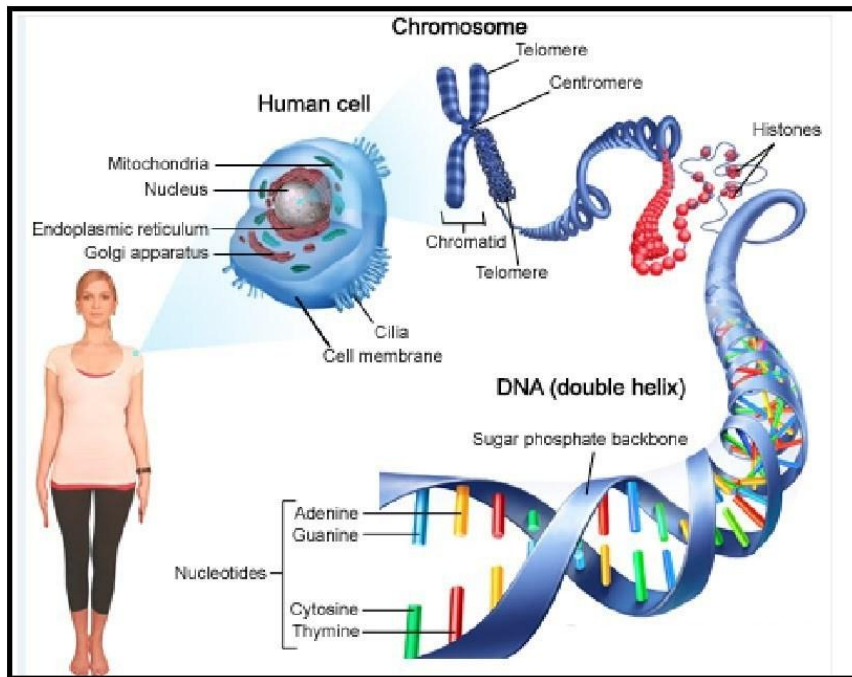
### **1. Asam nukleat**

Asam nukleat merupakan suatu molekul biokimia yang kompleks yang terdiri dari cincin heterosiklik nitrogen (1 atau 2), gula pentosa (ribosa atau deoksiribosa) dan gugus fosfat. Semua sel mengandung asam nukleat, bahkan virus yang secara teknis tidak hidup juga mengandung asam nukleat. Asam nukleat adalah makromolekul yang membawa informasi genetik di dalam sel. Dua asam nukleat yang paling umum ada dalam kehidupan adalah DNA (Asam Deoksiribonukleat) dan RNA (Asam Ribonukleat). DNA terdiri dari Adenin (A), Guanin (G), Sitosin (C), dan Timin (T). RNA terdiri dari A, G, C, dan Uracil (U). Berdasarkan strukturnya, A, G, C, T, dan U dapat dibagi menjadi dua kelompok, yaitu Purin dan Pirimidin

#### **a. DNA**

DNA atau Asam deoksiribonukleat merupakan suatu biomolekul yang menyimpan informasi genetik pada setiap organisme dan juga banyak virus. Informasi genetika ini berperan penting dalam pertumbuhan, perkembangan, dan fungsi suatu organisme dan juga virus. Informasi genetik ditangkap sebagai urutan nukleotida spesifik yang dikenal sebagai gen. Gen eukariotik terdiri dari ekson dan intron. Gen prokariotik biasanya hanya memiliki ekson. Ekson adalah urutan yang terdiri dari RNA dan sering kali merupakan kode untuk protein. Intron adalah suatu daerah berupa sekuens DNA yang ditemukan dalam gen dari sebagian besar organisme dan banyak virus. Sekuens ini merupakan sekuens yang dihapus sebelum RNA matang terbentuk.

Satu set DNA lengkap dari suatu organisme disebut Genome, gen termasuk di dalamnya. Jadi di dalam genom ini mengandung semua informasi yang penting menjalankan fungsi organisme. Ukuran genom berupa jumlah total DNA yang terkandung dalam satu salinan lengkap genom. Genom manusia terdiri dari 23 kromosom yang berpasangan dengan lebih dari 3 miliar base pair DNA (pasangan basa). Pada tubuh manusia, setidaknya ada 3 juta pasangan DNA, dan semuanya ini terdapat di dalam inti setiap sel.



Gambar 6. Struktur Kromosom

## b. RNA

Asam ribonukleat atau RNA adalah molekul polimer yang terlibat dalam berbagai peran biologis dalam mengkode, dekode, regulasi, dan ekspresi gen. RNA biasanya ditemukan dalam bentuk untai tunggal, akan tetapi RNA juga dapat beruntai ganda atau sebagian beruntai ganda ketika membentuk struktur seperti pada gambar 7.



Gambar 7. RNA

Gambar di sebelah kiri adalah bentuk molekul nyata tRNA dari sel ragi, khususnya tRNA yang membawa Fenilalanin. Struktur tRNA sering ditampilkan dalam bentuk

seperti gambar di sebelah kanan (di dalam kotak). RNA biasanya dibagi menjadi 2 kelompok dasar:

- Messenger RNA - RNA yang terbentuk dari ekson yang selanjutnya dibaca di ribosom dan ditranslasi menjadi protein
- RNA tanpa kode - semua RNA lain. Kelompok ini menyumbang sekitar 95% dari semua RNA yang diketahui

Dua dari jenis kelas RNA non-pengkode adalah RNA transfer (tRNA) dan RNA ribosom (rRNA). Ada banyak jenis RNA non-pengkode lainnya yang seringkali berkaitan dengan regulasi gen atau protein.

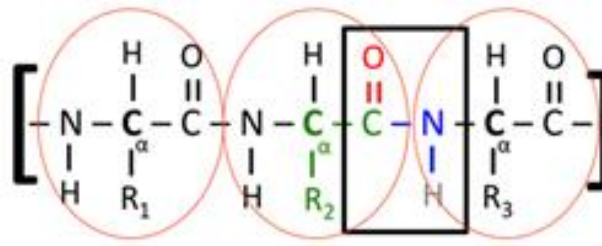
## 2. Protein

Protein merupakan suatu molekul yang terdiri dari asam amino. Asam amino-asam amino berikatan dengan ikatan polipeptida dan membentuk suatu protein. Kebanyakan organisme menggunakan 20 asam amino untuk membuat semua proteinnya, tetapi beberapa menggunakan 1 atau 2 lebih asam amino. Protein merupakan suatu biomolekul yang melakukan berbagai tugas di dalam sel. Semua asam amino memiliki struktur dasar yang serupa (N-C-C) dan memiliki sisi yang bervariasi.

	NONPOLAR, HYDROPHOBIC	R GROUPS	POLAR, UNCHARGED	
Alanine Ala A MW = 89	$\begin{array}{c} \text{OOC} \\   \\ \text{H}_3\text{N}^+ - \text{CH} - \text{CH}_3 \end{array}$		$\begin{array}{c} \text{H} - \text{CH} - \text{COO}^- \\   \\ \text{N} \text{H}_3 \end{array}$	Glycine Gly G MW = 75
Valine Val V MW = 117	$\begin{array}{c} \text{OOC} \\   \\ \text{H}_3\text{N}^+ - \text{CH} - \text{CH}(\text{CH}_3)_2 \end{array}$		$\begin{array}{c} \text{HO} - \text{CH}_2 - \text{CH} - \text{COO}^- \\   \\ \text{N} \text{H}_3 \end{array}$	Serine Ser S MW = 105
Leucine Leu L MW = 131	$\begin{array}{c} \text{OOC} \\   \\ \text{H}_3\text{N}^+ - \text{CH} - \text{CH}_2 - \text{CH}(\text{CH}_3)_2 \end{array}$		$\begin{array}{c} \text{OH} \\   \\ \text{CH}_2 - \text{CH} - \text{CH} - \text{COO}^- \\   \\ \text{N} \text{H}_3 \end{array}$	Threonine Thr T MW = 119
Isoleucine Ile I MW = 131	$\begin{array}{c} \text{OOC} \\   \\ \text{H}_3\text{N}^+ - \text{CH} - \text{CH}(\text{CH}_3) - \text{CH}_2 - \text{CH}_3 \end{array}$		$\begin{array}{c} \text{HS} - \text{CH}_2 - \text{CH} - \text{COO}^- \\   \\ \text{N} \text{H}_3 \end{array}$	Cysteine Cys C MW = 121
Phenylalanine Phe F MW = 131	$\begin{array}{c} \text{OOC} \\   \\ \text{H}_3\text{N}^+ - \text{CH} - \text{CH}_2 - \text{C}_6\text{H}_5 \end{array}$		$\begin{array}{c} \text{HO} - \text{C}_6\text{H}_4 - \text{CH}_2 - \text{CH} - \text{COO}^- \\   \\ \text{N} \text{H}_3 \end{array}$	Tyrosine Tyr Y MW = 181
Tryptophan Trp W MW = 204	$\begin{array}{c} \text{OOC} \\   \\ \text{H}_3\text{N}^+ - \text{CH} - \text{CH}_2 - \text{C}_8\text{H}_6\text{N}_2 \end{array}$		$\begin{array}{c} \text{NH}_2 \\   \\ \text{C} = \text{O} - \text{CH}_2 - \text{CH} - \text{COO}^- \\   \\ \text{N} \text{H}_3 \end{array}$	Asparagine Asn N MW = 132
Methionine Met M MW = 149	$\begin{array}{c} \text{OOC} \\   \\ \text{H}_3\text{N}^+ - \text{CH} - \text{CH}_2 - \text{CH}_2 - \text{S} - \text{CH}_3 \end{array}$		$\begin{array}{c} \text{NH}_2 \\   \\ \text{C} = \text{O} - \text{CH}_2 - \text{CH}_2 - \text{CH} - \text{COO}^- \\   \\ \text{N} \text{H}_3 \end{array}$	Glutamine Gln Q MW = 146
Proline Pro P MW = 115	$\begin{array}{c} \text{OOC} \\   \\ \text{H}_3\text{N}^+ - \text{CH} - \text{CH}_2 \\   \\ \text{HN} - \text{CH}_2 - \text{CH}_2 \end{array}$		<b>POLAR BASIC</b> $\begin{array}{c} \text{NH}_3^+ - \text{CH}_2 - (\text{CH}_2)_3 - \text{CH} - \text{COO}^- \\   \\ \text{N} \text{H}_3 \end{array}$	Lysine Lys K MW = 146
Aspartic acid Asp D MW = 133	<b>POLAR ACIDIC</b> $\begin{array}{c} \text{OOC} \\   \\ \text{H}_3\text{N}^+ - \text{CH} - \text{CH}_2 - \text{C}(=\text{O})\text{O}^- \end{array}$		$\begin{array}{c} \text{NH}_2 \\   \\ \text{N} \text{H}_2 - \text{C} = \text{NH} - (\text{CH}_2)_3 - \text{CH} - \text{COO}^- \\   \\ \text{N} \text{H}_3 \end{array}$	Arginine Arg R MW = 174
Glutamine acid Glu E MW = 147	$\begin{array}{c} \text{OOC} \\   \\ \text{H}_3\text{N}^+ - \text{CH} - \text{CH}_2 - \text{CH}_2 - \text{C}(=\text{O})\text{O}^- \end{array}$		$\begin{array}{c} \text{NH} \\   \\ \text{HN} = \text{N} - \text{CH}_2 - \text{CH} - \text{COO}^- \\   \\ \text{N} \text{H}_3 \end{array}$	Histidine His H MW = 155

Gambar 8. Jenis-jenis asam amino



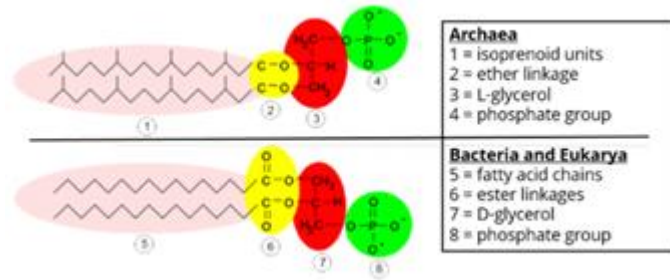


Gambar 9. Struktur asam amino

Perhatikan gambar diatas, masing-masing memiliki pola N-C-C, C pertama dihubungkan ke rantai samping (direpresentasikan sebagai R), C kedua terikat pada N asam amino berikutnya (ditunjukkan oleh persegi panjang hitam). Ikatan C- N dikenal sebagai ikatan peptida. Semua asam amino dibuat melalui jalur yang sudah mapan. Anda juga akan melihat bagaimana tiga asam amino (lingkaran merah) saling terkait. Semua asam amino berasal dari zat antara di jalur metabolisme primer sel: glikolisis, siklus asam sitrat, dan jalur pentosa fosfat. Protein melakukan begitu banyak pekerjaan di dalam sel sehingga penting untuk mengetahui berapa banyak dari setiap jenis yang ada. Penghitungan protein adalah salah satu hal yang diperhatikan oleh proteomik. Hal ini berkaitan dengan diagnosis banyak jenis penyakit. Sifat protein cukup tidak stabil. Semua protein memiliki umur simpan setelah itu dihancurkan oleh sel dan konten yang dihasilkan didaur ulang. Protein dapat bertahan selama beberapa detik atau beberapa bulan tergantung pada sifat protein. Umur rata-rata protein manusia sedikit di atas 1 hari.

### 3. Lemak

Lemak terdiri dari gliserol yang melekat pada asam lemak atau unit isoprenoid. Asam lemak terikat pada gliserol menggunakan ikatan ester sedangkan unit isoprenoid terikat pada gliserol menggunakan ikatan eter. Lipid dibuat dengan menggabungkan gliserol dan asam lemak (Bakteri dan Eukariota) atau unit isoprenoid (Archaea). Sekarang mari kita tinjau struktur lipid:



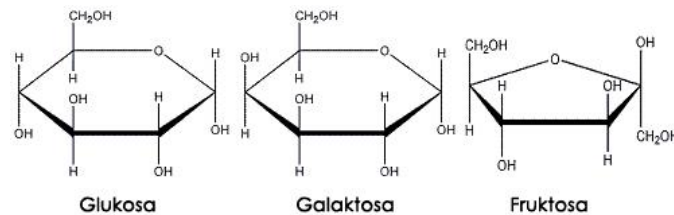
Gambar 10. Lemak dalam Arkhea, Bakteri dan Eukrya

#### 4. Karbohidrat

Karbohidrat terdiri dari Karbon, Hidrogen, dan Oksigen, sering disebut sakarida. Karbohidrat dapat dibedakan menjadi :

##### a. Monosakarida

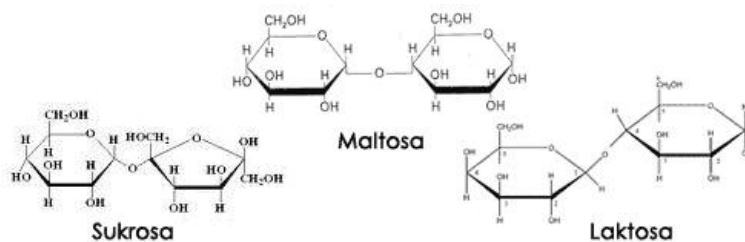
- Dibuat dengan proses yang disebut glukoneogenesis



Gambar 11. Monosakarida

##### b. Disakarida

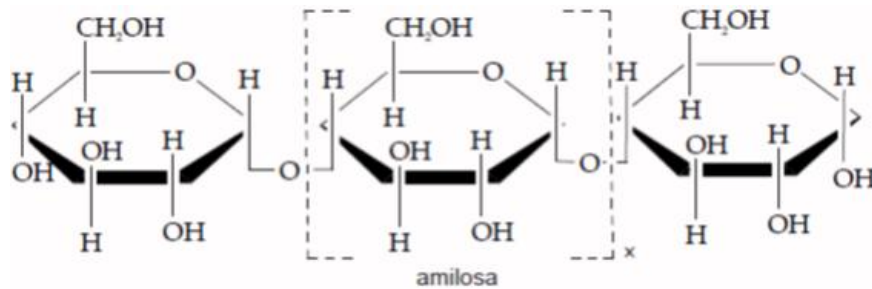
- Dibuat dengan menggabungkan dua monosakarida
- Misalnya: Sukrosa (glukosa + fruktosa) dan laktosa (glukosa + galaktosa)



Gambar 12. Disakarida

### c. Polisakarida

- Dibuat dengan menggabungkan tiga atau lebih monosakarida
- Misalnya: amilosa (banyak molekul glukosa dalam satu rantai)



Gambar 13. Polisakarida

Empat bangunan dasar ini sangat penting dalam membangun kehidupan manusia. Materi genetik dan protein menjadi bahan yang penting untuk dikaji dalam Bioinformatika, baik dalam pengkajian perubahan dalam urutan DNA, RNA atau protein sampai pada desain obat baru menggunakan informasi protein target. Untuk itu kita perlu memahami bagaimana proses sintesis protein terjadi didalam sel.

## D. Ekspresi Gen

Ekspresi gen merupakan proses penerjemahan informasi genetik yang dikode di dalam gen menjadi urutan asam amino pada sintesis protein. Dogma sentral biologi molekuler menjelaskan mengenai proses perubahan gen dari DNA menjadi RNA, dan RNA menjadi protein. Dogma ini menjelaskan bagaimana proses pembacaan materi genetik menjadi protein yang berperan di setiap tahap metabolisme di dalam tubuh suatu organisme. Dogma sentral biologi molekuler terbagi atas 3 tahapan besar, yaitu replikasi, transkripsi, dan translasi. Ketiga tahap ini memungkinkan penyalinan materi genetik menjadi protein.

### 1. Replikasi

Replikasi DNA merupakan proses penggandaan rantai ganda DNA. Proses ini terjadi sebelum pembelahan sel. Proses penggandaan tersebut memanfaatkan enzim DNA polimerase yang membantu pembentukan ikatan pada nukleotida-nukleotida penyusun polimer DNA. Proses replikasi DNA dapat dilakukan secara *in vitro* dengan

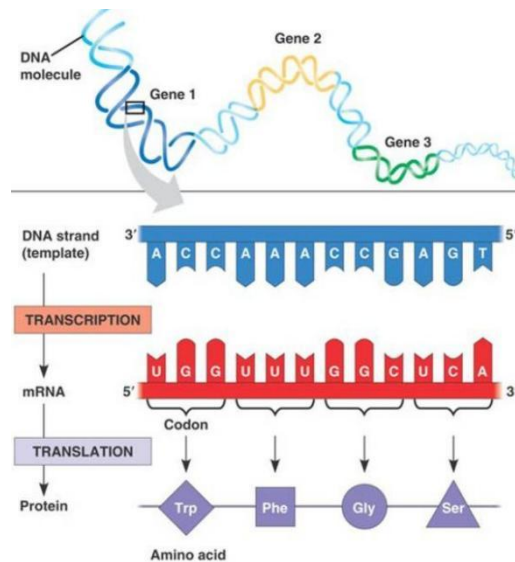
teknik disebut sebagai *Polymerase Chain Reaction*. Teknik inilah yang dikembangkan untuk melakukan penggandaan DNA dengan berbagai tujuan, seperti studi polimorfisme, deteksi suatu penyakit, deteksi suatu organisme baru. Dan tentu saja prosesnya dilakukan dengan melibatkan ilmu bioinformatika. Di Bab selanjutnya akan kita bahas mengenai proses PCR dan sequencing untuk mendapatkan hasil sekuens DNA, RNA atau protein.

## **2. Transkripsi**

Fungsi dasar lain yang harus dijalankan oleh DNA sebagai materi genetik adalah fungsi fenotipik, yaitu kemampuan DNA dalam mengatur pertumbuhan dan diferensiasi individu organisme. Kemampuan ini akan menghasilkan suatu fenotipe tertentu. Fungsi fenotipik dilakukan melalui ekspresi gen. Tahap pertama yang dilalui adalah proses transkripsi. Pada proses ini terjadi perubahan urutan basa molekul DNA menjadi urutan basa molekul RNA. Transkripsi juga dapat diartikan sebagai proses sintesis RNA dengan menggunakan salah satu untai molekul DNA sebagai cetaknya. Transkripsi merupakan sintesis RNA dari salah satu rantai DNA, yaitu rantai cetakan yang disebut sense, sedangkan pasangan rantai DNanya disebut rantai antisense. Proses ini terjadi di dalam inti sel.

## **3. Translasi**

Translasi adalah proses penerjemahan urutan nukleotida yang ada pada molekul mRNA menjadi rangkaian asam-asam amino yang menyusun suatu polipeptida atau protein. Pada translasi hanya molekul mRNA yang ditranslasi, sedangkan rRNA dan tRNA tidak ditranslasi. Translasi berlangsung di dalam sitoplasma dan ribosom. Translasi merupakan proses penerjemahan satu kode genetik menjadi protein yang sesuai. Disini terjadi pembacaan kodon mRNA menjadi asam amino. Dimulai dengan kodon start dan diakhiri dengan kodon stop pada terminasi. Asam amino yang dibawa oleh tRNA awal adalah metionin. Metionin adalah asam amino yang disandi oleh AUG. Setiap asam amino bisa dikode oleh satu/lebih kodon. Sejumlah 64 kodon yang mungkin diekspresikan menjadi asam amino disebut kode genetik. Kode genetik menghasilkan 20 asam amino yang berbeda. Proses translasi berhenti bila tempat A bertemu kodon akhir yaitu UAA, UAG, UGA.



Gambar 15. Transkripsi dan Translasi

### ***E. Polymerization Chain Reaction (PCR)***

*Polymerase Chain Reaction* (PCR) merupakan suatu teknik sintesis dan amplifikasi DNA secara *in vitro*. Teknik ini pertama kali dikembangkan oleh Karry Mullis pada tahun 1985. Teknik PCR dapat digunakan untuk menggandakan atau mengamplifikasi segmen DNA target secara *in vitro*. Dapat menghasilkan DNA dalam jumlah jutaan dalam waktu singkat (beberapa jam) (Handoyo dan rudiretna, 2000). Proses ini menggunakan menggunakan alat *thermalcycler*. *Thermalcycler* merupakan sebuah mesin yang memiliki kemampuan untuk memanaskan sekaligus mendinginkan tabung reaksi dan mengatur temperatur untuk tiap tahapan reaksi. Berikut ini merupakan komponen yang diperlukan untuk PCR yang digunakan untuk melakukan PCR:

#### 1. Template DNA

DNA template (cetakan) yaitu fragmen DNA yang akan dilipatgandakan dan berasal dari patogen yang terdapat dalam spesimen klinik. Template DNA merupakan urutan DNA yang akan disalin, disebut juga DNA target. Template ini dapat memperkuat (menyalin) sepotong DNA 50 ~ 4000 pasangan basa (mungkin

lebih). DNA dapat diisolasi dari suatu organisme sebelum dapat disalin (materi isolasi DNA sudah ada di materi sebelumnya).

## 2. Primer oligonukleotida

Primer ini merupakan untai sintetik pendek dari untai tunggal DNA yang sama persis dengan awal dan akhir fragmen DNA yang akan diamplifikasi. Primer berupa primer *reverse* dan primer *forward*. Amplifikasi DNA pada PCR dapat dicapai jika menggunakan primer oligonukleotida. Primer DNA dalam suatu sekuens oligonukleotida biasanya pendek. Primer ini berfungsi untuk mengawali sintesis rantai DNA. Umumnya primer yang digunakan pada PCR terdiri dari 20-30 nukleotida.

## 3. Enzim DNA polimerase

Enzim ini merupakan enzim termostabil Taq dari bakteri termofilik *Thermus aquaticus*. Polimerase membangun untai DNA baru dari arah 5' ke 3'. Molekul yang baru dipolimerisasi adalah pelengkap untai cetakan dan identik dengan untai mitra cetakan. DNA polimerase harus *Thermostable (Heat-stable)* karena suhu yang digunakan dalam PCR tinggi. Enzim ini disebut Taq polimerase karena enzim ini diisolasi dari bakteri *Thermus aquaticus* yang hidup di suhu tinggi/ air panas.

## 4. Deoksiribonukleotida trifosfat (dNTP)

Deoksiribonukleotida trifosfat (dNTPs) adalah blok bangunan dalam reaksi PCR. Mereka merupakan monomer yang digunakan DNA polimerase untuk membentuk DNA. Tersedia dalam bentuk larutan yang berisi A, T, C, dan G yang akan membangun untai DNA baru akan melekat pada ujung 3' primer ketika proses pemanjangan dan ion magnesium menstimulasi aktivasi polimerase.

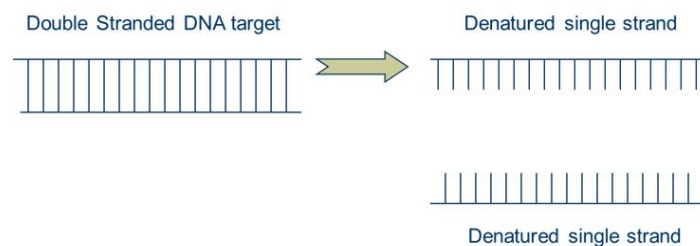
## 5. Buffer

Buffer merupakan komponen pendukung lain. Buffer dapat berfungsi mempertahankan pH. Larutan buffer PCR umumnya mengandung 10 – 50mM Tris-HCl (pH 8,3-8,8), 50 mM KCl, 0,1% gelatin atau BSA (Bovine Serum Albumin), Tween 20 sebanyak 0,01% atau dapat diganti dengan Triton X-100 sebanyak 0,1%, disamping itu perlu ditambahkan 1,5 mM MgCl<sub>2</sub>. Untuk bekerja dengan baik, Taq

membutuhkan  $Mg^{2+}$ . Konsentrasi ion magnesium perlu dioptimalkan dengan masing-masing target dan kombinasi primer. Terlalu sedikit magnesium bisa mempengaruhi jumlah produk PCR, bisa sedikit atau tidak ada produk PCR. Jika jumlah magnesium terlalu banyak maka akan dapat menimbulkan produk yang tidak diinginkan. Ada 3 langkah PCR, setiap langkah dilakukan pada suhu yang berbeda-beda dan dilakukan dengan jumlah siklus tertentu.

### 1. Denaturasi

Di dalam proses PCR, Pemanasan diatas suhu  $90^{\circ}C$  dapat merusak ikatan hidrogen DNA dan dapat memisahkan DNA double strand menjadi single strand. Denaturasi DNA merupakan proses yang bertujuan untuk melakukan pembukaan DNA untai ganda menjadi DNA untai tunggal. Ini biasanya berlangsung sekitar 3 menit (waktu disesuaikan ketika melakukan optimasi), untuk meyakinkan bahwa molekul DNA terdenaturasi menjadi DNA untai tunggal. Denaturasi yang tidak lengkap mengakibatkan DNA mengalami renaturasi (membentuk DNA untai ganda lagi) secara cepat, dan ini mengakibatkan gagalnya proses PCR.

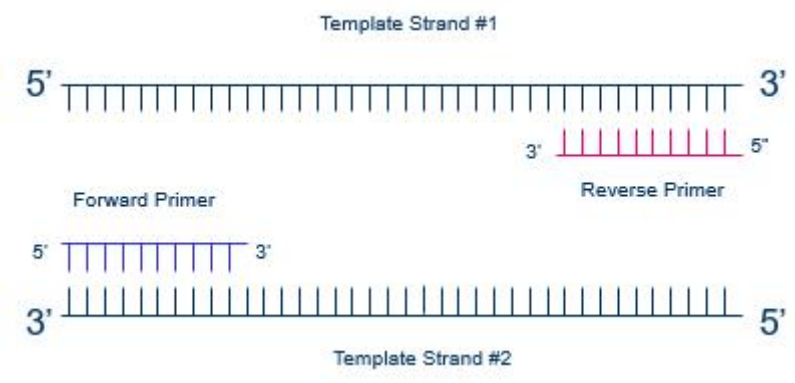


Gambar 16. Denaturasi

### 2. Annealing (penempelan primer)

Penempelan primer ke Target, disebut Hibridisasi. Suhu yang digunakan untuk penempelan primer tergantung pada panjang primer dan jumlah G-C. Kriteria umum yang digunakan untuk merancang primer yang baik adalah panjang basa berukuran 18 – 25 basa dan mengandung 50 – 60 % G-C. Kedua primer tersebut sebaiknya sama. Sekuens DNA dalam masing-masing primer itu sendiri juga sebaiknya tidak saling berkomplemen, karena hal ini akan mengakibatkan terbentuknya struktur sekunder pada primer tersebut dan mengurangi efisiensi PCR. Waktu annealing yang biasa digunakan dalam PCR adalah 30 – 45 detik. Semakin panjang ukuran primer, semakin

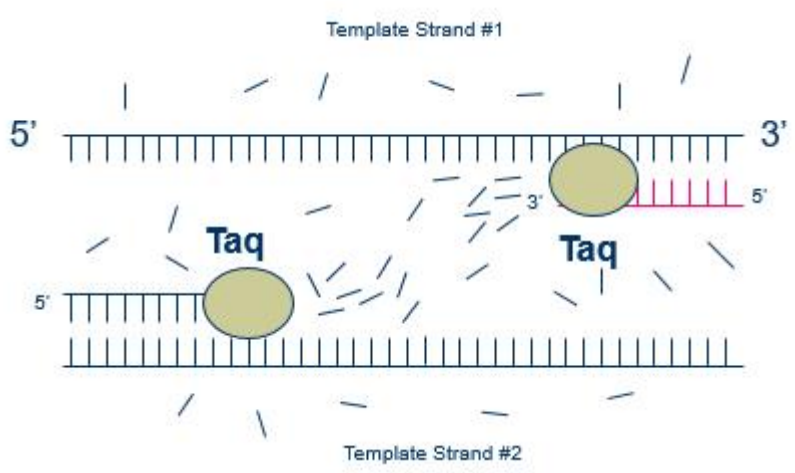
tinggi temperaturnya. Kisaran temperatur penempelan yang digunakan adalah antara 36°C sampai dengan 72°C, namun suhu yang biasa dilakukan itu adalah antara 50 – 60°C.



Gambar 17. Annealing

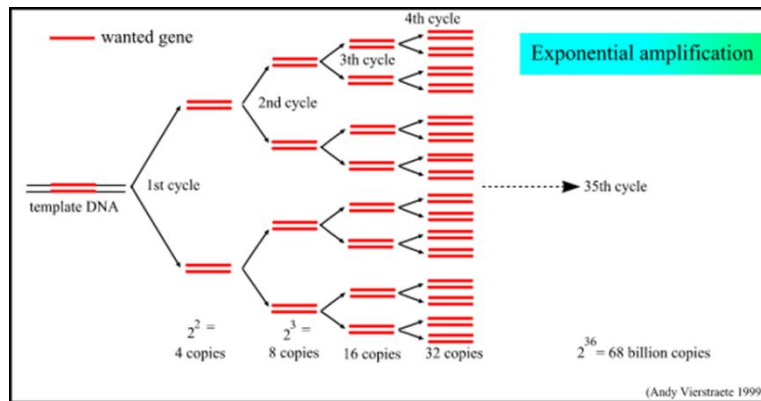
### 3.Pemanjangan (Extention)

Selama tahap ini Taq polimerase memulai aktivitasnya memperpanjang DNA primer dari ujung 3', dari arah 5' ke 3'. Kecepatan penyusunan nukleotida oleh enzim tersebut pada suhu 72°C diperkirakan 35 – 100 nukleotida/detik, bergantung pada buffer, pH, konsentrasi garam dan molekul DNA target. Pada akhir siklus PCR pertama, ada dua untai DNA baru identik terbentuk. Jumlah DNA yang terbentuk dapat diperkirakan dengan menghitung  $2^n$ , dengan  $n$  adalah jumlah siklus.



Gambar 18. Extention





Gambar 19. Hasil amplifikasi DNA

Hasil amplifikasi PCR (amplikon) ini yang akan digunakan untuk analisa selanjutnya. Setelah amplifikasi PCR dapat dilanjutkan dengan analisa menggunakan gel elektroforesis dan juga sequencing.

## F. Sequencing

Informasi genetik yang terdapat dalam DNA dapat diketahui dengan menggunakan teknik DNA Sequencing, metode yang digunakan untuk menentukan urutan basa nukleotida (Adenine, Guanine, Cytosine dan Thymine) pada molekul DNA. Prinsipnya adalah menggunakan metode PCR. DNA akan dianalisa urutan basanya dijadikan sebagai cetakan (template). Cetakan tersebut diamplifikasi menggunakan enzim dan bahan – bahan yang mirip dengan reaksi PCR dan ada penambahan reaksi tertentu. Dari hasil ini akan didapatkan urutan basa nukleotida yang sangat bermanfaat untuk pemeriksaan lebih lanjut, misalnya mengetahui terjadinya polimorfisme, mutasi, dan berbagai perubahan yang terjadi pada urutan basa nukleotida suatu organisme.

## G. Data Base

Serangkaian data base mengenai DNA, Genomic, cDNA, dan banyak lagi, dapat ditemukan (dan disimpan) di banyak data base. Tiga dari yang paling umum / populer adalah: National Center for Biotechnology Information (NCBI), European Molecular Biology Laboratory (EMBL), DNA Databank of Japan (DDBJ).

*National Center for Biotechnology Information* (NCBI) merupakan suatu bagian dari *National Library of Medicine* (NLM) Amerika Serikat, cabang dari *National Institutes*

*of Health* (NIH). Organisasi ini didanai oleh pemerintah Amerika Serikat. Serangkaian database mengenai bioteknologi dan biomedis disimpan di NCBI. Ini merupakan sumber penting untuk alat dan layanan bioinformatika. Genbank merupakan database utama termasuk untuk sekuens DNA, dan PubMed merupakan database bibliografi untuk literatur biomedis. Database lain termasuk database NCBI Epigenomics.

NCBI bertanggung jawab untuk menyediakan basis data sekuens DNA GenBank sejak tahun 1992. GenBank berkoordinasi dengan laboratorium individu dan database sekuens lainnya, seperti *European Molecular Biology Laboratory* (EMBL) dan *DNA Data Bank of Japan* (DDBJ). Selain GenBank, NCBI juga menyediakan *Gene*, *Online Mendelian Inheritance in Man*, *Molecular Modeling Database* (struktur protein 3D), *dbSNP* (database polimorfisme nukleotida tunggal), *Reference Sequence Collection*, peta genom manusia, dan browser taksonomi, dan koordinat dengan Institut Kanker Nasional untuk menyediakan Proyek Anatomi Genom Kanker. NCBI memiliki perangkat lunak yang tersedia melalui browser internet, misalnya, BLAST. BLAST merupakan program yang digunakan untuk mencari kesamaan urutan sekuens DNA, RNA atau protein. BLAST juga dapat digunakan untuk perbandingan urutan terhadap basis data DNA GenBank dalam waktu kurang dari 15 detik. BLAST adalah alat yang ampuh untuk menemukan urutan yang mirip dengan urutan kueri dalam organisme yang sama atau organisme yang berbeda. Itu mencari urutan kueri pada database dan server NCBI dan memposting hasilnya kembali ke browser orang tersebut dalam format yang dipilih. Urutan input ke BLAST sebagian besar dalam format FASTA atau GenBank sementara output dapat dikirimkan dalam berbagai format seperti format HTML, XML, dan teks biasa.

## **H. Sequence Alignment**

*Sequence alignment* digunakan untuk menemukan kesamaan antara dua (atau lebih) *sequence* DNA, RNA atau protein. *Sequence alignment* merupakan suatu susunan dari dua atau lebih *sequence* DNA, RNA atau protein yang fokus pada *similarity*. Hasil *alignment* dari urutan antara dua *sequence* yang diuji akan diisi dengan gaps (dashes/tanda hubung) sehingga jika terdapat kesamaan maka kolom terisi karakter yang identik dari urutan yang terlibat.



Gambar 20. *Alignment* dua *sequence* DNA

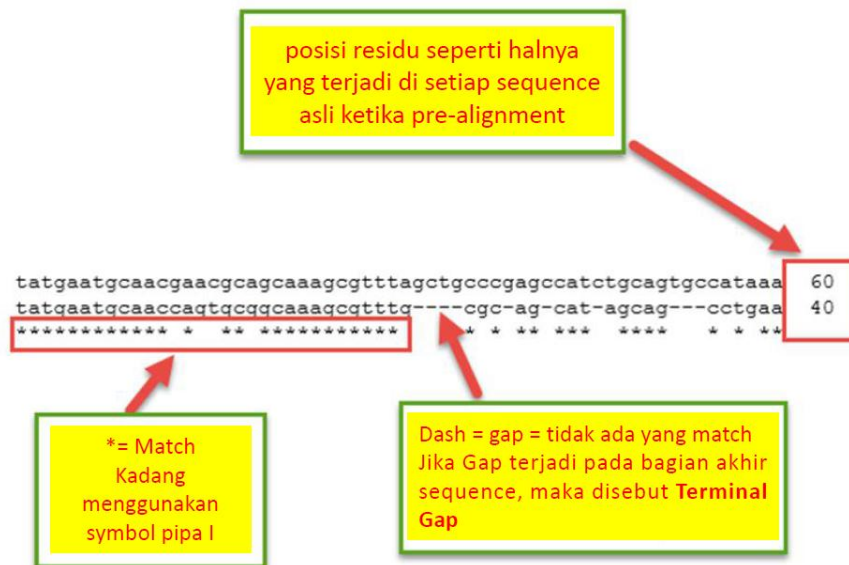
*Pairwise sequence alignment* merupakan suatu metode yang digunakan untuk menemukan kesejajaran lokal atau global yang paling cocok dari sekuens protein (asam amino) atau DNA (asam nukleat). Tujuannya adalah untuk menemukan homolog (kerabat) dari suatu gen atau produk-gen dalam *database* yang telah diketahui. Jadi, *sequence* yang didapatkan dari hasil penelitian dibandingkan dengan *database library* sehingga akan didapatkan informasi yang tepat. Informasi yang didapatkan disini sangat berguna untuk menjawab berbagai pertanyaan biologi, seperti:

- Identifikasi *sequence* yang tidak diketahui struktur atau fungsinya.
- Studi tentang evolusi molekuler.

*Global alignment* antara dua *sequence* adalah *alignment* dimana semua karakter di kedua *sequence* berpartisipasi dalam *alignment*. *Global alignment* berguna sebagian besar untuk menemukan urutan yang terkait erat. Metode *local alignment* menemukan daerah terkait dalam *sequence* yang dapat terdiri dari subset dari karakter tiap *sequence*. Misalnya, posisi 20-40 dari urutan A mungkin disejajarkan dengan posisi 50-70 dari urutan B. Ini merupakan suatu teknik yang lebih fleksibel daripada *global alignment* dan memiliki keuntungan bahwa wilayah terkait yang muncul dalam urutan berbeda dalam dua protein (yang dikenal sebagai domain *shuffling*) dapat diidentifikasi sebagai hal yang terkait.

*Pair-wise Sequence Alignment* merupakan prosedur menyusun dua *sequence* DNA, RNA, atau protein untuk mengidentifikasi daerah kesamaan yang mungkin merupakan konsekuensi dari hubungan fungsional, struktural, atau evolusi. *Alignment* dapat dibuat dengan berbagai metode. Hasil *alignment* biasanya memiliki residu nukleotida atau asam amino yang disejajarkan dalam kolom yang berurutan dengan celah yang disisipkan di antara residu atau tanpa celah; tergantung pada metode *alignment* yang digunakan. *Sequence alignment* biasanya digunakan untuk mempelajari evolusi *sequence* dari *common ancestor* seperti *sequence* protein atau *sequence* DNA. *Alignment* ini juga biasanya digunakan untuk mengetahui adanya

*mismatches* dalam *alignment* sesuai dengan mutasi yang terjadi, dan juga *gaps* terkait dengan adanya mutasi insersi atau delesi. *Sequence alignment* juga mengacu pada proses membangun *alignment* yang signifikan dalam *database* dari *sequence* yang mungkin tidak berhubungan.



Gambar 21. Hasil *alignment*

Jika hasil dari *alignment* ini memiliki urutan yang homolog, kita dapat menyimpulkan bahwa sequence tersebut memiliki *ancestor* bersama dan mempunyai fungsi yang sama dengan homolog. Jika hasil alignment tidak menunjukkan urutan yang homolog, maka nampak terjadinya perubahan pada gen, meski seringkali tidak mengubah sifat fisiokimia protein yang dikodekan. Jadi semakin tinggi tingkat kemiripan dalam urutan berpasangan, maka semakin besar kemungkinan kedua urutan kode memiliki fungsi serupa.

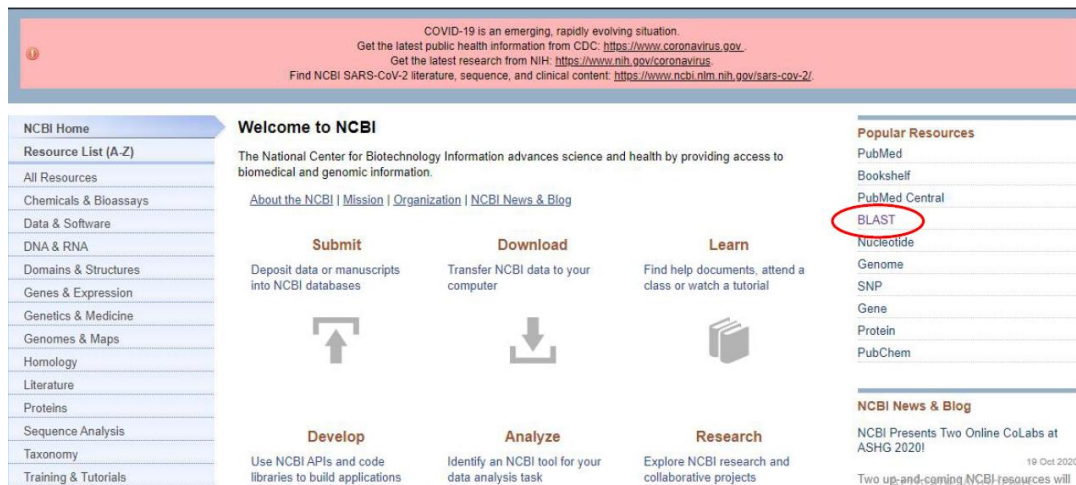
### ***Alignment* menggunakan NCBI**

Alignment menggunakan NCBI dapat dilakukan dengan dua cara, yaitu membandingkan urutan DNA, RNA atau protein dengan database, dan menggunakan dua atau lebih urutan DNA, RNA atau protein untuk dibandingkan. Metode ini dapat dilakukan menggunakan software BLAST yang terdapat di NCBI.

#### **1. Membandingkan urutan asam nukleat atau protein dengan database**

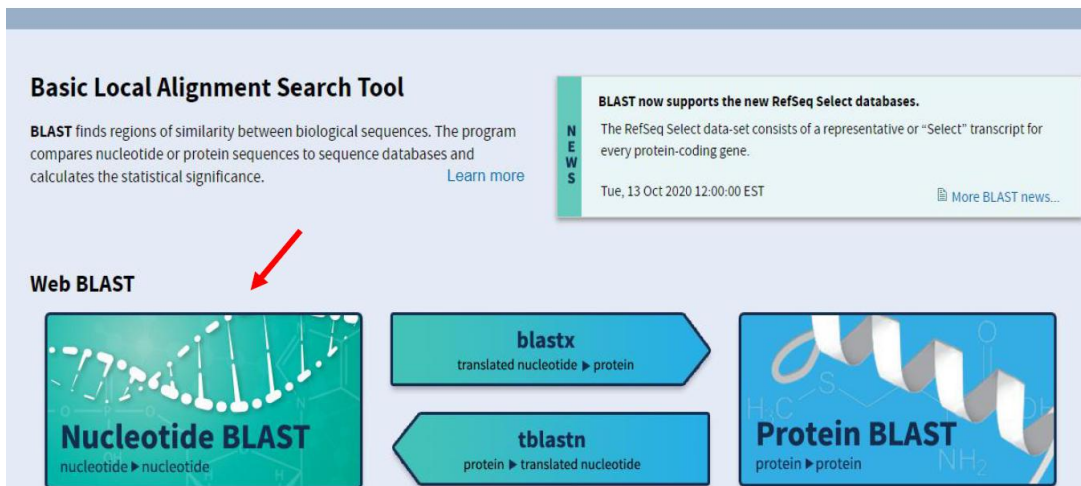
Berikut ini adalah langkah langkah yang dilakukan untuk membandingkan urutan asam nukleat yang kita punyai dengan database. Misalnya kita melakukan studi dalam pencarian organisme hasil *sequencing* kita.

- a. Buka laman NCBI (<https://www.ncbi.nlm.nih.gov/>), akan muncul tampilan seperti dibawah ini, kemudian pilih BLAST.



Gambar 22. Halaman depan NCBI

- b. Menu akan menunjukkan lama BLAST yang menampilkan beberapa pilihan, seperti berikut:



Gambar 23. BLAST tools

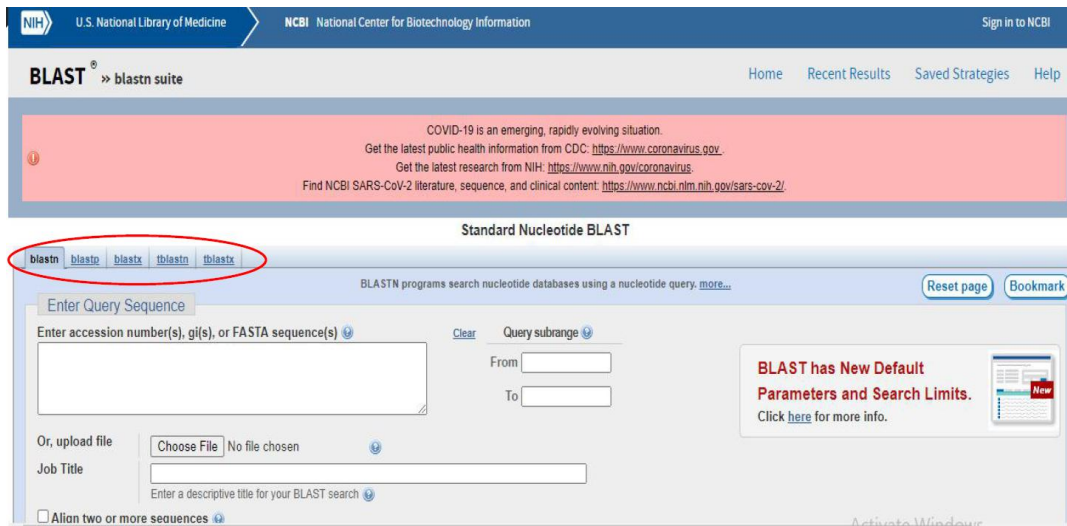
Berikut merupakan keterangan dari web BLAST yang ditunjukkan.

- i. *Nucleotide blast* (blastn) : membandingkan suatu sekuen nukleotida (*query sequence*) yang kita miliki dengan database sekuen nukleotida.
- ii. *Protein blast* (blastp) : membandingkan suatu sekuen asam amino yang kita miliki dengan database sekuen protein.
- iii. *blastx* : membandingkan produk translasi konsep 6-frame sebuah sekuen nukleotida (*translated nucleotide*) yang kita miliki dengan database sekuen protein.

- iv. *tblastn* : membandingkan suatu sekuen protein yang kita miliki dengan database sekuen nukleotida yang secara dinamis ditranslasi pada semua pembacaan 6 frame.
- v. *tblastx* : membandingkan suatu translasi 6 frame dari nukleotida.

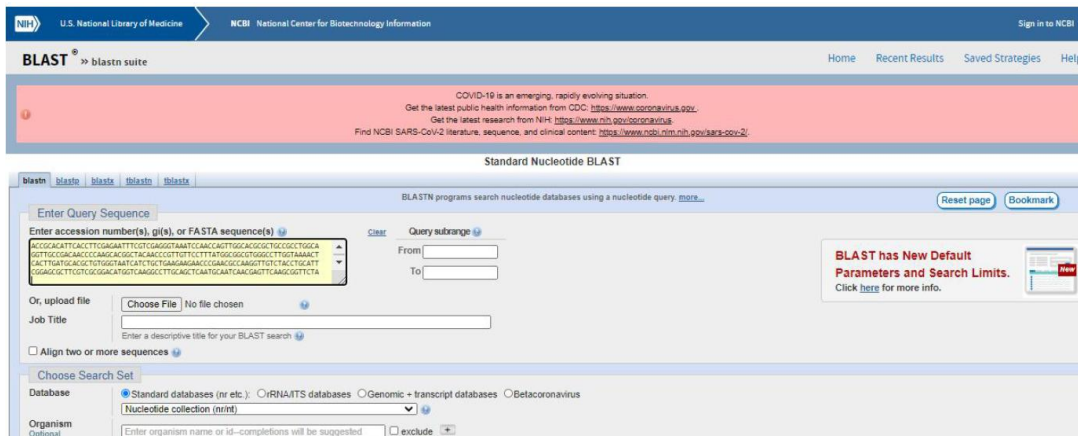
Web BLAST dapat dipilih sesuai dengan tujuan kita dan *sequence* yang kita miliki

- c. Jika kita ingin sampel yang kita miliki adalah *sequence* DNA, maka pilih *blastn*



Gambar 24. Standard nucleotide BLAST

- d. Masukkan *sequence* ke dalam kotak *query*



Gambar 25. Memasukkan *sequence*

- e. Pilih jenis organisme, *highly similar sequence*, dan klik BLAST. Setelah ini tunggu beberapa saat karena software akan memproses data kita.

Job Title

Enter a descriptive title for your BLAST search

Align two or more sequences

Choose Search Set

Database  Standard databases (nr etc.):  rRNA/ITS databases  Genomic + transcript databases  Betacoronavirus

Nucleotide collection (nr/nt)

Organism   exclude

Enter organism name or id—completions will be suggested

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown

Exclude  Models (XM/XP)  Uncultured/environmental sample sequences

Limit to  Sequences from type material

Entrez Query  [YouTube](#) [Create custom database](#)

Enter an Entrez query to limit search

Program Selection

Optimize for  Highly similar sequences (megablast)

More dissimilar sequences (discontiguous megablast)

Somewhat similar sequences (blastn)

Choose a BLAST algorithm

**BLAST** Search database Nucleotide collection (nr/nt) using Megablast (Optimize for highly similar sequences)

Show results in a new window

Gambar 26. Alignment sequence

f. Berikut ini adalah hasil yang akan kita dapatkan dengan acuan *library* pada *database*

BLAST » blastn suite » results for RID-SZBX17DA014

Home Recent Results Saved Strategies Help

[< Edit Search](#) Save Search Search Summary

How to read this report? BLAST Help Videos Back to Traditional Results Page

Your search is limited to records that include: bacteria (taxid:2)

Job Title **Nucleotide Sequence**

RID [SZBX17DA014](#) Search expires on 10-22 11:48 am [Download All](#)

Program BLASTN [Citation](#)

Database nt [See details](#)

Query ID lcl|Query\_34225

Description None

Molecule type dna

Query Length 3080

Other reports [Distance tree of results](#) [MSA viewer](#)

Filter Results

Organism  exclude

only top 20 will appear

Type common name, binomial, taxid or group name

[+ Add organism](#)

Percent Identity  to

E value  to

Query Coverage  to

[Filter](#) [Reset](#)

Descriptions Graphic Summary Alignments Taxonomy

Sequences producing significant alignments

Download Manage columns Show 100

select all 100 sequences selected

	Description	Max Score	Total Score	Query Cover	E value	Per Ident	Accession
<input checked="" type="checkbox"/>	<a href="#">Pseudomonas syringae pv. actinidiae str. Shaanxi M228 chromosome, complete genome</a>	5688	5688	100%	0.0	100.00%	<a href="#">CP032631.1</a>
<input checked="" type="checkbox"/>	<a href="#">Pseudomonas syringae pv. actinidiae strain CRAFRU 14.08, complete genome</a>	5688	5688	100%	0.0	100.00%	<a href="#">CP019732.1</a>
<input checked="" type="checkbox"/>	<a href="#">Pseudomonas syringae pv. actinidiae strain CRAFRU 12.29, complete genome</a>	5688	5688	100%	0.0	100.00%	<a href="#">CP019730.1</a>
<input checked="" type="checkbox"/>	<a href="#">Pseudomonas syringae pv. actinidiae strain NZ-47, complete genome</a>	5688	5688	100%	0.0	100.00%	<a href="#">CP017009.1</a>
<input checked="" type="checkbox"/>	<a href="#">Pseudomonas syringae pv. actinidiae strain NZ-45, complete genome</a>	5688	5688	100%	0.0	100.00%	<a href="#">CP017007.1</a>
<input checked="" type="checkbox"/>	<a href="#">Pseudomonas syringae pv. actinidiae ICMP 18884, complete genome</a>	5688	5688	100%	0.0	100.00%	<a href="#">CP011972.2</a>
<input checked="" type="checkbox"/>	<a href="#">Pseudomonas syringae pv. actinidiae ICMP 18708, complete genome</a>	5688	5688	100%	0.0	100.00%	<a href="#">CP012179.1</a>
<input checked="" type="checkbox"/>	<a href="#">Pseudomonas syringae pv. actinidiae ICMP 9853, complete genome</a>	5666	5666	100%	0.0	99.87%	<a href="#">CP018202.1</a>
<input checked="" type="checkbox"/>	<a href="#">Pseudomonas syringae pv. actinidiae strain MAFF212063 chromosome, complete genome</a>	5661	5661	100%	0.0	99.84%	<a href="#">CP024712.1</a>
<input checked="" type="checkbox"/>	<a href="#">Pseudomonas syringae pv. actinidiae strain P155 chromosome, complete genome</a>	5502	5689	100%	0.0	100.00%	<a href="#">CP032871.1</a>
<input checked="" type="checkbox"/>	<a href="#">Pseudomonas syringae group genomosp. 3 isolate CFBP6411 genome assembly, chromosome I</a>	5330	5514	99%	0.0	98.96%	<a href="#">LT963408.1</a>
<input checked="" type="checkbox"/>	<a href="#">Pseudomonas syringae strain Ps25 chromosome</a>	5162	5162	100%	0.0	96.92%	<a href="#">CP034558.1</a>
<input checked="" type="checkbox"/>	<a href="#">Pseudomonas syringae pv. tomato strain B13-200 chromosome, complete genome</a>	5162	5162	99%	0.0	97.03%	<a href="#">CP019871.1</a>
<input checked="" type="checkbox"/>	<a href="#">Pseudomonas syringae pv. tomato strain delta IV IX chromosome</a>	5140	5140	100%	0.0	96.79%	<a href="#">CP047072.1</a>
<input checked="" type="checkbox"/>	<a href="#">Pseudomonas syringae pv. tomato strain delta VI chromosome</a>	5140	5140	100%	0.0	96.79%	<a href="#">CP047071.1</a>
<input checked="" type="checkbox"/>	<a href="#">Pseudomonas syringae pv. tomato strain delta X chromosome, complete genome</a>	5140	5140	100%	0.0	96.79%	<a href="#">CP047073.1</a>

Gambar 27. Hasil Alignment

Hasil ini menunjukkan kesamaan *sequence* dengan homolog nya, yaitu beberapa organisme yang memiliki kesamaan.

- g. Berikut ini gambaran lebih jelas kesamaan antara sequence dengan *library*. Jika teridentifikasi 100% maka tidak ada gap diantara dua *sequence* tersebut.

Score	Expect	Identities	Gaps	Strand
5688 bits(3080)	0.0	3080/3080(100%)	0/3080(0%)	Plus/Plus

Features: [chromosomal replication initiator protein DnaA](#)  
[DNA polymerase III subunit beta](#)

```

Query 1   TTATAGCTCTCTGTGAGCAACCTACATGTGGATAAGTCGTCCTCGGATCGTTACAATGGC 60
Sbjct 1   TTATAGCTCTCTGTGAGCAACCTACATGTGGATAAGTCGTCCTCGGATCGTTACAATGGC 60

Query 61   GGCTGTTTTTGCCTCACC GGCTTTCAACTCAGGGGATATCCGTGTCAGTGGAACTTTGGC 120
          |||
Sbjct 61   GGCTGTTTTTGCCTCACC GGCTTTCAACTCAGGGGATATCCGTGTCAGTGGAACTTTGGC 120

Query 121  AGCAGTGCCTGGAGCTTTTGC GCGACGAGCTGCCTGCCAGCAATTCAACACTTGGATCC 180
          |||
Sbjct 121  AGCAGTGCCTGGAGCTTTTGC GCGACGAGCTGCCTGCCAGCAATTCAACACTTGGATCC 180

Query 181  GTCCGCTACAGGTCGAAGCCGAAGGCGACGAGTTGCGTGTGTACGCACCCAACCGCTTTG 240
          |||
Sbjct 181  GTCCGCTACAGGTCGAAGCCGAAGGCGACGAGTTGCGTGTGTACGCACCCAACCGCTTTG 240

Query 241  TTCTCGATTGGGTCAATGAAAAATACCTCGGCCGTTTGCTTGAGCTTCTCGGCGAACCTG 300
          |||
Sbjct 241  TTCTCGATTGGGTCAATGAAAAATACCTCGGCCGTTTGCTTGAGCTTCTCGGCGAACCTG 300

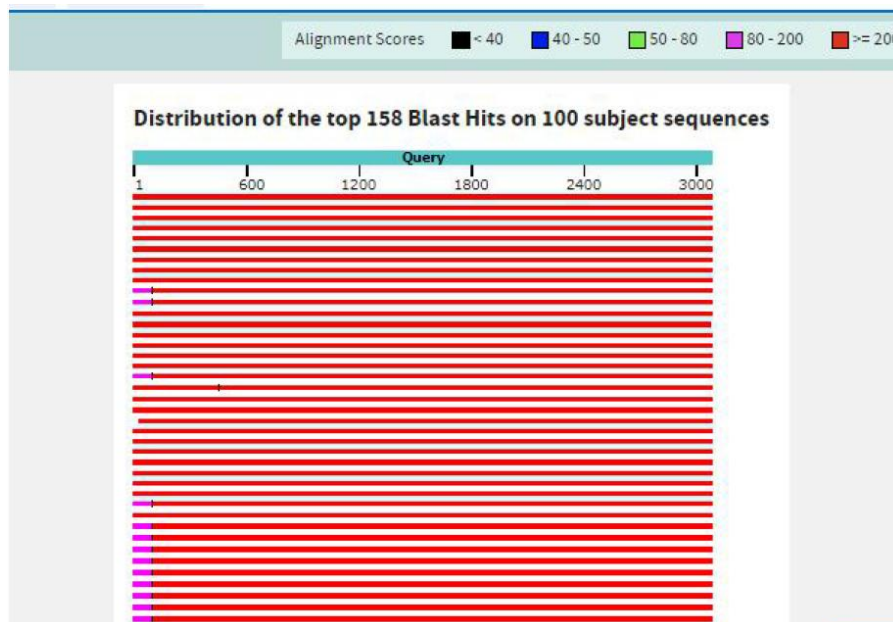
Query 301  GTCAAGGCATGGCGCCCGCTTTTCTTATTAATAGGAAGCAAGCGCAGCTCCGCGCCGC 360
          |||
Sbjct 301  GTCAAGGCATGGCGCCCGCTTTTCTTATTAATAGGAAGCAAGCGCAGCTCCGCGCCGC 360

Query 361  GTGCCGCACCGAATGCGCCACTGGCTGCTGCAGCGTCCAGGCGTTGTCTGCCAATTCGG 420
          |||
Sbjct 361  GTGCCGCACCGAATGCGCCACTGGCTGCTGCAGCGTCCAGGCGTTGTCTGCCAATTCGG 420

Query 421  TCAGCAGCGTTTCAGTGTCTGCCCCCGTCATGGCCGCTCCTGCACCGATGGTTGCAGCGC 480
          |||
Sbjct 421  TCAGCAGCGTTTCAGTGTCTGCCCCCGTCATGGCCGCTCCTGCACCGATGGTTGCAGCGC 480

Query 481  CAGTGCCTGTGCATAACGTTGCCACTCACGACGAGCCGTCGCGCGACAGCTTCGATCCGA 540
          |||
Sbjct 481  CAGTGCCTGTGCATAACGTTGCCACTCACGACGAGCCGTCGCGCGACAGCTTCGATCCGA 540

```



Gambar 28. Tingkat kesamaan dua *sequence*

Warna merah ini menunjukkan bahwa ada lebih dari 200 basa nuklotida yang sama diantara dua *sequence* homolog.



## 2. Membandingkan dua atau lebih urutan asam nukleat atau protein

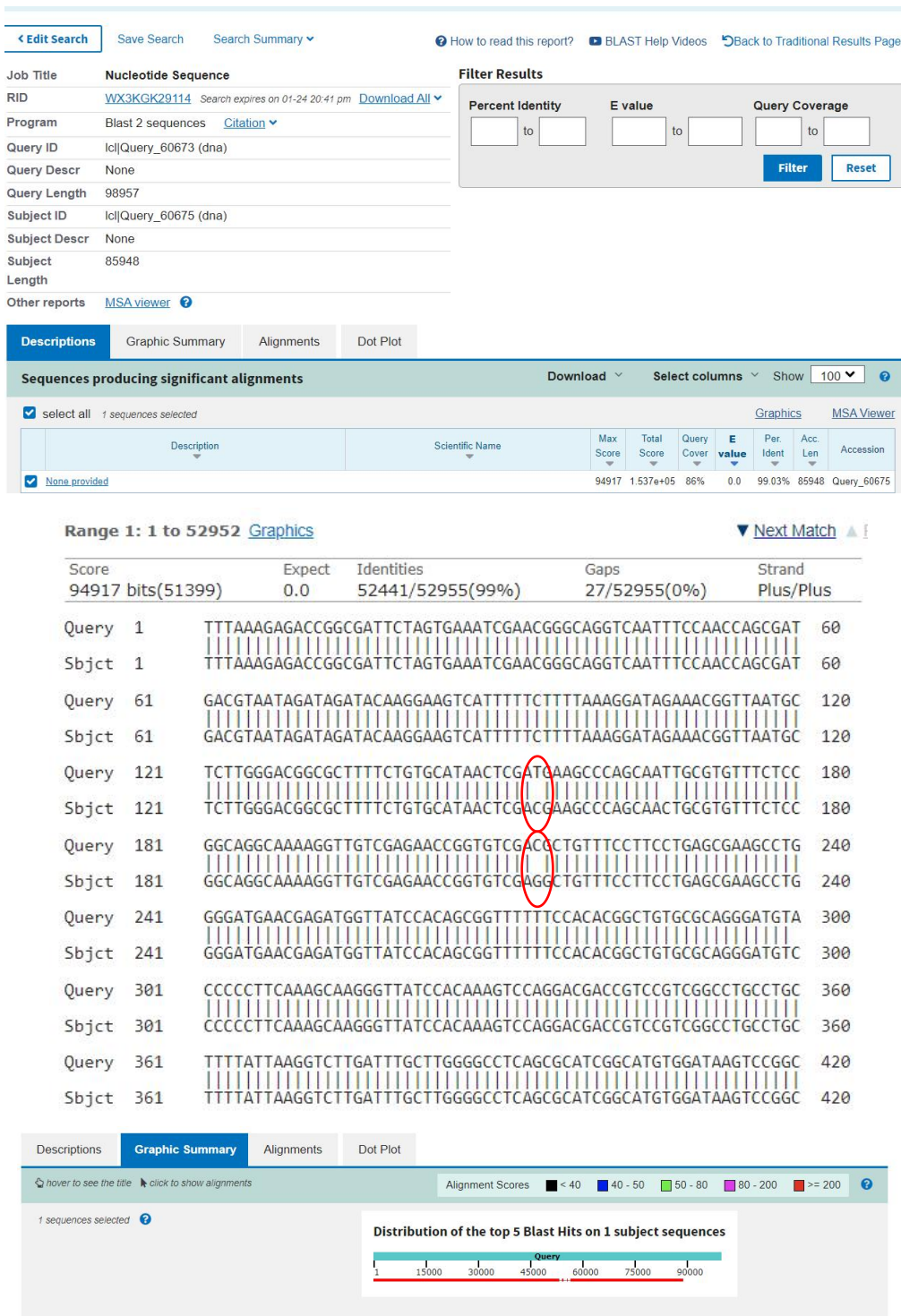
Berikut ini adalah metode yang digunakan untuk membandingkan dua atau lebih sequence. Disini akan nampak perbedaan antara kedua atau lebih *sequence* yang dapat dilihat dari hasil *gap*-nya. Berikut adalah langkah-langkah yang akan kita lakukan:

- Menyiapkan dua sequence yang akan kita bandingkan
- Buka BLAST dan pilih *align two or more sequence*, kemudian masukkan *sequence* yang akan kita bandingkan. Pilih *highly similar sequence* kemudian klik BLAST.

The screenshot shows the NCBI BLAST web interface. The top navigation bar includes 'blastn', 'blastp', 'blastx', 'tblastn', and 'tblastx'. The main heading is 'Align Sequences Nucleotide BLAST'. Below this, there are three main sections: 'Enter Query Sequence', 'Enter Subject Sequence', and 'Program Selection'. The 'Enter Query Sequence' section has a text area containing the sequence: CAGAGGACCTGCAGGCTGGAGAA, TACTGCAACTAGACGAGCCCGCAGGCAGCCCAACCCGCGCCT, CCTGCACCGAGAGATGGAATA, and AAGCCCTTGAACCAGC. There is also a 'Job Title' field and a checked checkbox for 'Align two or more sequences'. The 'Enter Subject Sequence' section has a text area containing the sequence: GTGGAGCAGTGCACCCAGCATC, TGTTCCTTACCAGCTGGAGAACTACTGCAACTAGCCCGCCCTGAG, GCGCGCTGCTGCTCCCGCACC, and CCNAACCCATAAA. The 'Program Selection' section has radio buttons for 'Highly similar sequences (megablast)', 'More dissimilar sequences (discontiguous megablast)', and 'Somewhat similar sequences (blastn)', with 'Highly similar sequences (megablast)' selected. At the bottom, there is a 'BLAST' button and a checkbox for 'Show results in a new window'.

Gambar 29. Alignment dua sequence

- Berikut merupakan hasil dari *alignment* dua *sequence*. Hasil menunjukkan bahwa tingkat kesamaan adalah 99% (52441bp yang sama). Dimana terdapat 27 gap dari 52955 pasangan basa, ditandai dengan lingkaran merah. Hasil grafik menunjukkan bahwa terdapat lebih dari 200 pasangan basa yang memiliki kesamaan (warna garis merah). Dapat disimpulkan bahwa kedua sequence menunjukkan kesamaan karakter dan homolog.



Gambar 30. Hasil *alignment* dua *sequence*

Berikut adalah contoh penerapan ilmu bioinformatika, dimana kita dapat menggunakannya untuk melakukan identifikasi *sequence* yang kita tidak ketahui dengan membandingkannya dengan *database library*. Pada buku seri berikutnya kita akan membahas penggunaan ilmu bioinformatika untuk melakukan uji filogenetik.

## DAFTAR PUSTAKA

- Bernstein, F.C., Koetzle, T.F., Williams, G.J., Meyer, E.F., Brice Jr., M.D., Rodgers, J.R., *et al.* 1977. The Protein Data Bank. A computer-based archival file for macromolecular structures. *Eur J Biochem*, 80(2):319-324.
- Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., *et al.* 2000. The Protein Data Bank. *Nucleic Acids Res*, 28(1):235-242.
- Campbell, N.A., Reece, J.B., Mitchell, L.G. 2000. *Biologi*. Jakarta: Erlangga.
- Campbell, N.A., Jane, B.R. 2005. *Biology 7th ed, International Edition*. San Fransico: Pearson Educational Inc.
- Clancy, S. 2008. DNA transcription. *Nature Education*, 1(1):41.
- Clancy, S., Brown, W. 2008. Translation: DNA to mRNA to Protein. *Nature Education*, 1(1):101.
- Hartwell, L.H., Leroy, H., *et al.* 2000. *Genetics: From Genes to Genomes*. New York: McGraw-Hill Companies, Inc.
- Handoyo, D., Rudiretna, A. 2000. General Principles and Implementation of Polymerase Chain Reaction. *Unitas*, 9 (1): 17-29.
- Luscombe, N.M., Greenbaum, D., Gerstein, M. 2001. What is bioinformatics? A proposed definition and overview of the field. *Methods Inf Med*, 40(4): 346-358.
- National Center for Biotechnology Information. 2022. <https://www.ncbi.nlm.nih.gov>.
- Pearson, W.R., Lipman, D.J. 1988. Improved tools for biological sequence comparison. *Proc Natl Acad Sci USA*, 85(8):2444-2448.
- Phillips, T. 2008. Regulation of transcription and gene expression in eukaryotes. *Nature Edu*, 1(1):199.
- Stansfield, W.D., Colome, J.S., Cano, R.J. 2006. *Biologi Molekuler dan Sel*. Jakarta: Erlangga.
- van der Gulik, P.T.S., Hoff, W.D., Speijer, D. 2017. In defence of the three-domains of life paradigm. *BMC Evol Biol*, 17(1):218.
- Woese, C., Kandler, O., Wheelis, M. 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci USA*, 87 (12): 4576–4579.
- Yuwono, Triwibowo. 2010. *Biologi Molekuler*. Jakarta: Erlangga.