IBT 432 Aplikasi Bioinformatika
Praktek: Anotasi sekuen ke genom

Riza Arief Putranto

www.esaunggul.ac.id

Pertemuan III – 22 September 2018

Prodi Bioteknologi

Fakultas Ilmu-Ilmu Kesehatan

Rencana Perkuliahan

1. ~~Kontrak belajar dan pengenalan bioinformatika aplikatif~~

2. ~~Database sekuen dan analisis genomika~~

3. Anotasi sekuen ke genom - Praktik

4. Analisis komparasi genomika I

5. Analisis komparasi genomika II

6. Analisis komparasi genomika III

7. Analisis komparasi genomika – Praktik

8. Protein modelling I

9. Protein modelling II

10. Protein modelling III

11. Protein modelling - Praktik

12. Visualisasi protein modelling

13. Visualisasi protein modelling - Praktik

14. Presentasi mahasiswa

Always remember this slide

Scanning....

Genomic DNA
(NC, NT, NW)

Model mRNA (XM) → Model protein (XP)

(XR)

Curated mRNA (NM) → Curated Protein (NP)

(NR)

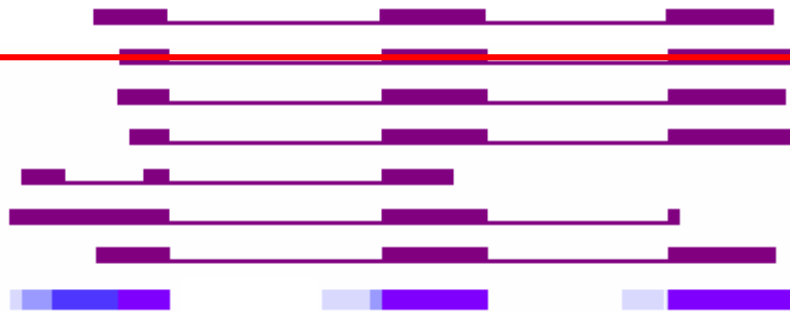RefSeq

GenBank
Sequences

R NM_002467
G BC000141
G BC000917
G K02276
G M13929
G M13930
G V00568
E ESTs

Remember this one also

Latar belakang praktikum

Kolaborasi INSDC merupakan basis data yang tergolong "umum", "besar" dan banyak digunakan oleh peneliti-peneliti di seluruh dunia. Saat ini, penambahan basis data sekuen terus terjadi dan akan terus bertambah dalam tahun-tahun mendatang, seiring dengan semakin turunnya biaya untuk melakukan sekuensing genomika, transkriptomika, epigenomika, serta proteomika. Perkembangan riset yang spesifik pada spesies tertentu juga memunculkan tipe basis data spesifik dimana informasi genomik yang dianotasi berasal dari 1 jenis kingdom, famili atau bahkan genus.

Keberadaan basis-basis data tersebut juga menunjukkan bahwa riset *in silico* atau riset laboratorium kering sedang menjadi tren serta mampu menjadi alat konfirmasi dan validasi untuk hasil riset "basah" yang dilakukan di laboratorium. Riset *in silico* dapat pula digunakan untuk mempercepat prediksi dan menajamkan riset "basah", seperti desain marka untuk seleksi pemuliaan tanaman hingga *drug design*. Hal tersebut diperkuat dengan fakta bahwa perkembangan teknologi komputasi yang pesat sudah memberikan kemudahan analisis bioinformatika skala kecil pada komputer desktop atau laptop.

Tujuan praktikum

1. Mahasiswa memahami cara pengambilan sekuen dari *database* nukleotida

2. Mahasiswa memahami runtutan penggunaan *database* nukleotida, BLAST dan bagaimana memanfaatkan informasi tersebut

3. Mahasiswa memahami cara anotasi manual dalam piranti lunak bioinformatika

4. Mahasiswa memahami makna exon (CDS) dan domain protein dalam anotasi

Sequence-to-Genome annotation

>seq
ATGGCCACCAACCAACTCCTCCCCCTCCTCCTCCTCGCCCTCGT
CGCCGGCTCCCATGCGGGCAGCATCGCTGTCTACTGGGGCCA
GAACGGCAATGAGGGCACCCTCGCCGACACCTGCTCCAGCG
GCCTCTACGCNTACGTCATCCTCGCCTTCCTCACCACATTCGG
CAACGGCCAGACCCCCGTCCTCAACCTCGCAGGCCACTGCGA
CCCCAACGCCGGCACCTGCACCAGCCTCTCCTCCGACATCCA
GTCCTGCCAGTCCCAGGGTGTCATCGTCCTTCTCTCCCTCGGC
GGTGGCTCCGGCAGCTACTCC

Google Drive: goo.gl/BbeR9o

Praktikum A. Homology search

1. Buka situs BLAST NCBI: https://blast.ncbi.nlm.nih.gov/Blast.cgi
2. Lakukan Nucleotide BLAST

Praktikum A. Homology search

1. Input sekuen ke dalam Query
2. Lakukan Nucleotide BLAST dengan parameter: Database nucleotide collection, Optimize for Highly similar sequences (megablast),

Praktikum A. Homology search
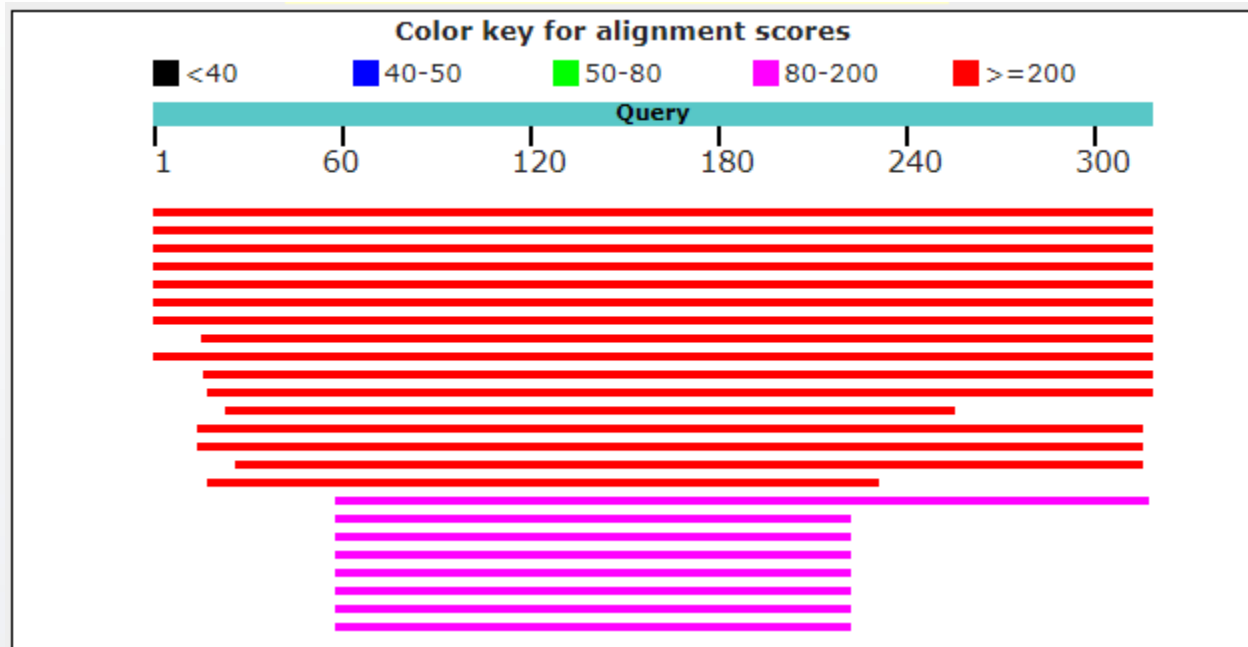
1. Apa hasil terbaik dari BLAST tersebut? Sekuen query teridentifikasi sebagai gen apa? Ambil hasil sekuen yang bukan *raw*
2. Sebutkan berapa E-value dan % homologi sekuen terbaik
3. Tuliskan no aksesi dari gen tersebut
4. Apakah no aksesi tersebut sudah tervalidasi oleh kurator? Pada tingkat mana? *Raw*, *model* atau *curated*?
5. Akses no aksesi tersebut dalam GenBank
6. Tuliskan no ID proteinnya dan copy paste sekuen nukleotida dan proteinnya ke dalam Notepad/Wordpad
7. Simpan sekuen dalam format .fasta

Praktikum A. Homology search

1. Lakukan anotasi manual di SnapGene untuk CDS (Exon) dan UTR
2. Berapa panjang basa CDS-nya? Berapa % GC content-nya?
3. Berapa panjang proteinnya? Ukuran molekul proteinnya?
4. Apakah ditemukan 5'-UTR?
5. Apakah ditemukan 3'-UTR?

Ingat dogma biologi molekuler!

Sekuen diatas adalah mRNA, tidak ada intron. Kita temukan sekuen DNA-nya melalui *sequence-to-genome annotation*

Praktikum B. Genome Data Viewer

1. Buka situs Genome Data Viewer: https://www.ncbi.nlm.nih.gov/genome/gdv/
2. Select organism: *Elaeis guineensis*
3. Ada berapa kromosom tanaman ini?

Praktikum B. Genome Data Viewer

1. Masukkan no aksesi gen yang di dapat sebelumnya ke **Search in genome GDV**
2. Akses gen tersebut dalam tampilan Genome Browser

Praktikum B. Genome Data Viewer

1. Terletak di kromosom berapa gen tersebut?
2. Catat lokasi gen tersebut di dalam kromosom dari basa berapa ke berapa?
3. Ada berapa jumlah exon dan intron?

Praktikum B. Genome Data Viewer

1. Klik kanan di sekuen XM tersebut
2. Download FASTA (Visible Range) – file akan tersimpan dalam format FASTA

Praktikum B. Genome Data Viewer

1. Buka sekuen tersimpan tersebut dalam SnapGene
2. Berapa panjang basa sekuen DNA dari gen tersebut?

Praktikum B. Genome Data Viewer

1. Akses view to sequence untuk tiap exon
2. Anotasi kembali sekuen DNA dari gen tersebut: CDS (exon 1, exon 2), intron, 5'UTR dan 3'UTR. **Perhatikan langkah-langkahnya dan catat!**

Praktikum B. Genome Data Viewer

1. Anotasi kembali sekuen DNA dari gen tersebut: CDS (exon 1, exon 2), intron, 5'UTR dan 3'UTR. **Perhatikan langkah-langkahnya dan catat!**

Praktikum B. Genome Data Viewer

Anotasi selesai jika seluruh bagian sekuen telah teridentifikasi

Tugas Homology Search

>ajusi

AAAATATATTAAAATAATAATATTTTGTGTAATAGCCCCCAAATTCAAATATCTAATCCGTTCTATATA
TTTCCCAAACTTAGTGCCACTTAGACACAATAAGAAAACCGTTAACAAGAAGAAAAAAAAAG
ATCGAAAATGGAATATCAAACTAACTTCTTAAGTGGAGAGTTTTCCCCGGAGAACTCTTCTTCAA
GCTCATGGAGCTCACAAGAATCATTCTTGTGGGAAGAGAGTTTCTTACATCAATCATTTGACCAA
TCCTTCCTTTTATCTAGCCCTACTGATAACTACTGTGATGACTTCTTTGCATTTGAATCATCAATCAT
AAAAGAAGAAGGAAAGAAGCCACCGTGGCGGCCGAGGAGGAGGAGAAGTCATACAGAGG
AGTGAGGAAACGGCCGTGGGGGAAATTCGCGGCCGAGATAAGAGACTCAACGAGGAAAGG
GATAAGAGTGTGGCTTGGGACATTCGACACCGCGGAGGCGGCGGCTCTCGCTTATGATCAGGC
GGCTTTCGCTTTGAAAGGCAGCCTCGCAGTACTCAATTTCCCCGCGGATGTCGTTGAAGAATCT
CTCCGGAAGATGGAGAATGTGAATCTCAATGATGGAGAGTCTCCGGTGATAGCCTTGAAGAGA
AAACACTCCATGAGAAACCGTCCTAGAGGAAAGAAGAAATCTTCTTCTTCTTCGACGTTGACAT
CTTCTCCTTCTTCCTCCTCCTCCTATTCATCTTCTTCGTCTTCTTCTTCTTTGTCGTCAAGAAGTAGA
AAACAGAGTGTTGTTATGACGCAAGAAAGTAATACAACACTTGTGGTTCTTGAGGATTTAGGTG
CTGAATACTTAGAAGAGCTTATGAGATCATGTTCTTGATAATCTCTGCTTCTACAATTTTTATGTAA
TTTGAGTGAGCTCTGTTTTTTTAAGAAAAAATCCGTTGAAAAGTTTTGGAAGAGATGTGATGGC
CGTGTATAGATGTCAACTATTTGCGTTACTGTTGGAAACATGTATAGATGTGTAGTTCTTTATCTTG
GTGCTTAATTTTAATATTTATTTTTGCCATAATTCTACAAAACCGATCCGAG

Tugas Homology Search

1. Apa hasil terbaik dari BLAST tersebut?
2. Sekuen query teridentifikasi sebagai gen apa? Ambil hasil sekuen yang bukan *raw*
3. Sebutkan berapa E-value dan % homologi sekuen terbaik
4. Tuliskan no aksesi dari gen tersebut
5. Apakah no aksesi tersebut sudah tervalidasi oleh kurator? Pada tingkat mana? *Raw*, *model* atau *curated*?
6. Akses no aksesi tersebut dalam GenBank
7. Anotasi sekuen tersebut dalam SnapGene: CDS, 5'UTR dan 3'UTR

Bukti yang harus dilampirkan – print pdf/image dari
1. Halaman hasil BLAST nucleotide
2. Halaman GenBank
3. Halaman Map dari gen hasil anotasi di SnapGene

Laporan dan Tugas disubmit oleh Kormat
Google Drive: goo.gl/BbeR9o
Folder: 3b-Laporan dan Tugas Praktikum
Maksimal tanggal Jum'at, 5 Oktober 2018

It was still the third course, don't get dizzy yet